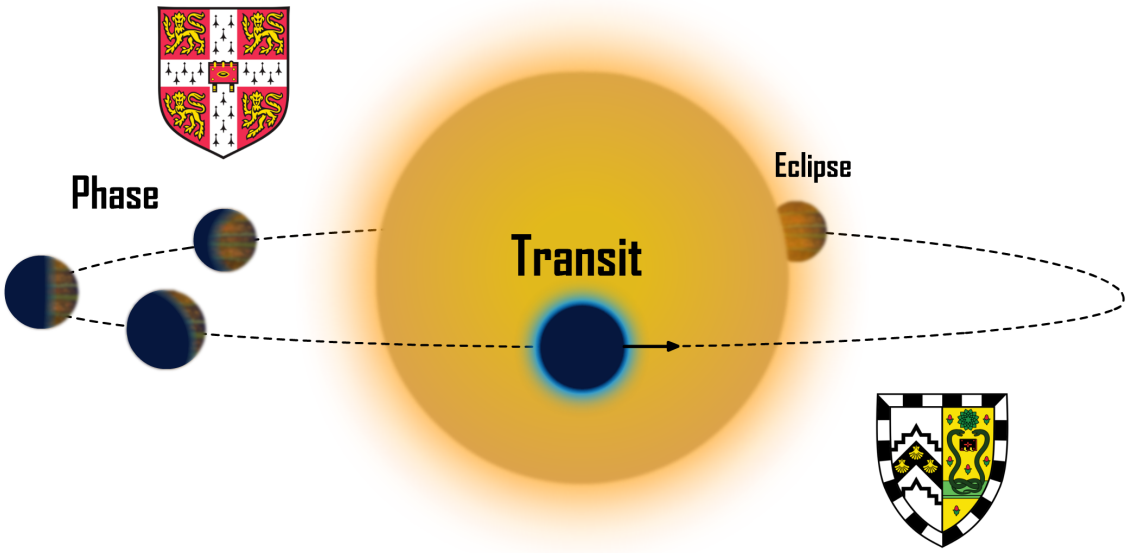
Revealing the Nature of Exoplanetary Atmospheres



Ryan John MacDonald

Supervisor: Dr Nikku Madhusudhan

Institute of Astronomy
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

Gonville and Caius College June 2019

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Chapters 2 & 3 draw upon material published as 'HD 209458b in new light: evidence of nitrogen chemistry, patchy clouds and sub-solar water' in *Monthly Notices of the Royal Astronomical Society* (MacDonald & Madhusudhan, 2017a). This original research was conducted by myself, published in co-authorship with Nikku Madhusudhan.

Chapter 4 draws upon material from two published works: (i) 'Signatures of Nitrogen Chemistry in Hot Jupiter Atmospheres' in *The Astrophysical Journal Letters* (MacDonald & Madhusudhan, 2017b), and (ii) 'Detection of titanium oxide in the atmosphere of a hot Jupiter' in *Nature* (Sedaghati et al., 2017). The former piece of original research was conducted by myself, published in co-authorship with Nikku Madhusudhan. Material from the latter work draws only from my own original contribution to the analysis, which was published in co-authorship with Elyar Sedaghati, Henri Boffin, Siddharth Gandhi, Nikku Madhusudhan, Neale Gibson, Mahmoudreza Oshagh, Antonio Claret, and Heike Rauer.

Chapters 6 & 7 draw upon material published as 'The metal-rich atmosphere of the exo-Neptune HAT-P-26b' in *Monthly Notices of the Royal Astronomical Society* (MacDonald & Madhusudhan, 2019). This original research was conducted by myself, published in co-authorship with Nikku Madhusudhan.

Ryan John MacDonald
June 2019

Revealing the Nature of Exoplanetary Atmospheres

Ryan John MacDonald

The study of planets orbiting other stars, exoplanets, has entered the era of characterisation. When an exoplanet passes in front, or transits, its parent star, absorption features are imprinted into starlight passing through the planetary atmosphere. By analysing the resultant transmission spectrum, one can reveal the chemical composition of these distant worlds. Already, this technique has yielded detections of multiple chemical species in exoplanetary atmospheres. However, measuring the abundances of these molecules and atoms, as required to infer exoplanet formation mechanisms, has been challenged by the prospect of atmospheric clouds.

In this thesis, I introduce a new approach to retrieve properties of exoplanetary atmospheres. I demonstrate that a generalisation of common 1D exoplanet atmosphere retrieval models to include 2D properties can break cloud-chemistry degeneracies. This algorithm, implemented by a new atmospheric retrieval code, POSEIDON, enables precise constraints on the chemistry, cloud properties, and temperature structure of exoplanetary atmospheres. By applying POSEIDON to observed transmission spectra of giant exoplanets, new insights into exoplanet atmospheres have been obtained. I present evidence of inhomogeneous clouds and disequilibrium nitrogen chemistry (NH_3 and HCN) in hot Jupiter atmospheres, the first detection of titanium oxide (TiO) in an exoplanetary atmosphere, and evidence of metal hydrides, possibly from a recent impact, in the atmosphere of a Neptune-mass exoplanet. Finally, looking to the future, I examine the ability of the upcoming *James Webb Space Telescope* to probe the atmosphere of a Neptune-mass exoplanet in unprecedented detail.

These results offer extraordinary promise for the retrieval of atmospheric properties from exoplanet spectra. Planets possessing cloudy skies can still be precisely characterised, expanding the potential for new discoveries in the years to come.

To all those who dream of distant worlds.

We will cross that ocean, we will voyage to the stars.

Acknowledgements

No journey exists in isolation. A PhD is both a humbling and rewarding adventure, for it is only when you reach the edge of the known that you truly appreciate how far we have yet to travel. First, I would like to express gratitude towards my PhD supervisor, Dr Nikku Madhusudhan, for guiding me on this epic journey of discovery. Those late nights discussing the philosophy of science, and the mysteries of distant worlds, have had a profound impact on my development as a scientist.

I would like to thank all those teachers over the years who encouraged and supported me on my journey to become a Physicist. In particular, Hannah Weller, Graham Tavener, Colin Shipway, Patrick Baird, Robin Nicholas, and Neil Bowles. Thanks for putting up with all the pedantic questions. A special shout-out must also go to my Chemistry teacher Torben Smith, who, to this very day, epitomises my image of the mad scientist (*vita aeterna*!).

To all those working in science communication, I can never express how important you have been in igniting the passion for science in myself and countless others. In particular, the late Stephen Hawking was the shining light that drew me first to University College, Oxford, before crossing over to Gonville & Cauis College in Cambridge. Sincere thanks also go to untold science fiction authors through the years, amongst them Robert Heinlein, Kevin J. Anderson, Kim Stanley Robinson, and James S. A. Corey (Daniel Abraham + Ty Franck), whose dreams of other worlds we are now beginning to realise.

My time in Cambridge would not have been the same without the people I have had the pleasure to meet at the Institute of Astronomy. Many thanks go to all my office mates over the years, in particular Douglas Boubert, Dilon Dong, Sebastian Marino, Anjali Piette, Namrah Habib, Jess Rigley, Chris Bambic, and Jeff Jennings. I wish you all the best on your own voyages of discovery.

To Nick Henden, Douglas Boubert, and Adam Jermyn: thank you for the wonderful 2 years in the Astrohouse. Who would have thought that four astrophysicists (especially us pesky theorists!) could thrive in the real world? Live long and prosper.

Evenings in Cambridge would not have been the same without the boardgame crew here at the IoA. Thanks to Jasleen Matharu, Nick Henden, Douglas Buisson, Matteo Sbroscia, Lukas Hergt, Rob Whittaker, and Sophie Koudmani for all the fun times.

It has been my pleasure to discuss our future in space with Deyan Mihaylov and Sergei Dyda over many a cup of tea. I hope we meet again on Mars someday.

Life is never linear, but through the twists and turns I have been highly fortunate to call some incredible people my friends. To Alexander Macquisten: thanks for always being the one person I could talk to in school (especially when I went off on a tangent about three year old Samurais, or some equally imaginative topic). To Elliot Reynolds: my time at Oxford wouldn't have been the same without our heated late-night discussions (especially when I refused to yield ground just to keep the debates flowing!). To Rachael Martin: thanks for all the fascinating discussions over the years, I wish you all the best with your entry into public life.

I could not have asked for a more wonderful family. Without you, Mum, I could never have gotten to where I am today. You gave up so much to support Brönte and myself through difficult times, always putting us before yourself. You are the best Mother anyone could hope for; I will always love you. To Brönte: there are so many things that only we can understand, so many precious moments we have shared together. You'll always be my 'bredbin'; our streets are paved with gold. And of course, to my grandparents: thank you for always encouraging me to pursue my dreams, and showing me that anything is possible. When you said the sky is the limit, I'm not sure alien skies is what you had in mind!

Finally, to Jennifer:

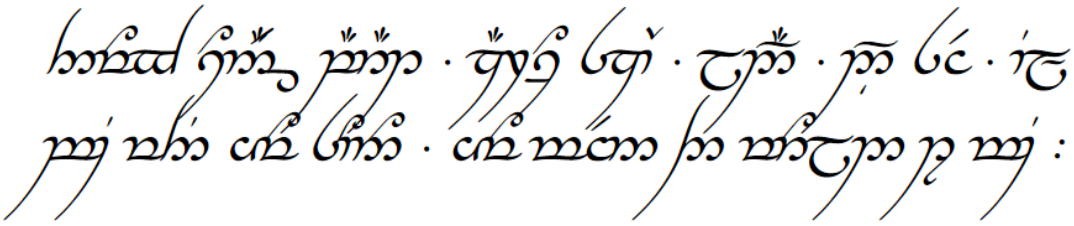


Table of contents

List of figures

List of tables

Chapter 1

Distant Worlds

1.1 The search for extrasolar planets

Contrast the world around us, vibrant and full of wonders, with the immense darkness of the night-time sky. Punctuated by but a few points of light, standing guard far above us, the stars have been our constant companion across the reaches of time. But all is not still in the void. When our ancestors first observed certain stars to wander across the sky, they could scarcely imagine that our own world is but a wandering star.

Speculation that the Earth does not exist in isolation in the cosmos extends back thousands of years. The Greek philosopher Epicurus asserted in his letter to Herodotus (c. 300 BCE) that "there is an infinite number of worlds, some like this world, others unlike it." Even before the invention of the telescope in 1608 permitted the scientific study of the planets in our own solar system, the Italian friar Giordano Bruno imagined the possibility of planets orbiting other stars. In 'On the infinite universe and worlds' (Bruno, 1584) he wrote that "there are countless suns and countless earths all rotating round their suns in exactly the same way as the seven planets of our system. We see only the suns because they are the largest bodies and are luminous, but their planets remain invisible to us because they are smaller and non-luminous."

Bruno's remarkable insight already suggests the immense difficulty involved in directly observing planets orbiting other stars, *extrasolar planets* (or exoplanets). If one were to search for a planet like the Earth around a nearby sun-like star in the same manner as we observe the planets in our solar system (i.e. reflected visible light), it would be less than a billionth as bright and separated by only a fraction of an arcsecond from its parent star. Such a feat remains infeasible even today. Rather than attempting direct observation, the key early insight was *indirect* inference.

Shortly after the discovery of Neptune (Galle, 1846), based on indirect predictions from residuals in the orbit of Uranus (Le Verrier, 1846; Adams, 1846), the idea that such methods could be applied to infer the existence of extrasolar planets began to take shape. As early as 1851, it was noted, in the context of variable stars, that "periodical obscuration or total disappearance of the star may arise from transits of the star by its attendant planets" (Lardner, 1851). The first published claim of an exoplanet came in 1855, when Captain W.S. Jacob at the Madras Observatory reported deviations in the orbits of the binary star system 70 Ophiuchi (Jacob, 1855). Jacob suggested the gravitational influence of an unseen companion of planetary nature to be the culprit, a view that was enhanced by subsequent observations (See, 1895), though later shown to be unlikely from orbital stability considerations (Moulton, 1899).

Further claims of exoplanets based on perturbations in the positions of stars (astrometry) emerged in the 20th century. Suggestions of a 16 M_J planet around 61 Cygni (Strand, 1943) and a 10 M_J planet around 70 Ophiuchi (Reuyl & Holmberg, 1943) were eventually ruled out (Heintz, 1978). Similar claims surrounding planetary mass companions to the nearby stars Lalande 21185 (van de Kamp & Lippincott, 1951) and Barnard's star (van de Kamp, 1963) ignited decades of controversy, before these claimed planets were also excluded (e.g. Gatewood, 1974; Choi et al., 2013).[1] This troubled history of astrometric exoplanet searches did little to endear the fledgling field as an area of serious astronomical research.

As is so often the case in science, success can come from unexpected directions. Whilst astrometry garnered the most attention, Belorizky (1938) and Struve (1952) deduced that Jupiter-mass planets orbiting a star could induce potentially detectable Doppler shifts in the stellar spectrum and, should the planetary orbit lie nearly edge on, transits. However, the expectation that other solar systems would possess similar architectures to our own[2] cast such searches in dubious light. An observer hoping to detect a Jupiter-mass exoplanet at a similar orbital separation would have to achieve sensitivity to stellar radial velocities around 10 ms^{-1}, stable over many years, or search a vast array of stars for low probability transits. Nevertheless, such radial velocity precisions were achieved by the late 1970s (Campbell & Walker, 1979) and the theory of exoplanet transits saw further refinement (Rosenblatt, 1971; Borucki & Summers, 1984).

[1] Ironically, a planet candidate around Barnard's star was recently identified by Ribas et al. (2018), albeit with notably different properties to the claims of the 1960s.

[2] Struve (1952) explicitly mentioned that Jupiter-mass planets in orbits far smaller than Mercury, as we now know to exist, could have been detectable even with the spectrographs of the 1950s.

Despite the challenges involved, the first indications (later confirmed) of wide-orbit substellar objects of, or near[3], planetary mass arose from radial velocity studies in the late 1980s (Campbell et al., 1988; Latham et al., 1989).

Ultimately, the 1990s became the decade which saw the first unambiguous exoplanet detections. By precisely measuring variations in the arrival time of regular radio pulses from the pulsar PSR B1257+12, Wolszczan & Frail (1992) detected the presence of two exoplanets close to Earth in mass ($2.8\,M_\oplus$ and $3.4\,M_\oplus$) with orbital periods of 98.2 and 66.6 days. These worlds became the first confirmed planets orbiting another star. Shortly thereafter, the first detection of an exoplanet orbiting a sun-like star followed from radial velocity measurements of the star 51 Pegasi (Mayor & Queloz, 1995). This planet, 51 Pegasi b, upended classical notions of planetary system architecture. With a minimum mass around $0.5\,M_J$, orbiting 7 times closer than Mercury, and a period of just 4.2 days, 51 Pegasi b represented the first of an entirely new class of planets that came to be known as *hot Jupiters*. Though some scepticism as to the existence of hot Jupiters persisted, all doubt was lifted when a known radial velocity exoplanet, HD 209458 b, was observed to transit in 1999 (Henry et al., 2000; Charbonneau et al., 2000). In the space of a decade, three powerful exoplanet detection techniques had emerged. After millenia of wondering, and nearly 150 years of unsuccessful searches, at the dawn of the 21st century the era of exoplanets had finally arrived.

1.2 Exoplanet detection methods

Since the discovery of the first exoplanets, many methods have been applied, with varying degrees of success, to their detection and study. I shall now elaborate on the key physical principles at the heart of the main detection methods currently in use.

1.2.1 Radial velocities

The radial velocity method was the first widely applied exoplanet detection technique, responsible for the discovery of the majority of the first 100 exoplanets. The essential idea behind this technique is that the gravitational influence of planetary companions cause a star to execute an orbit about the mutual centre of mass of the system. As the star orbits, its projected line-of-sight velocity relative to a distant observer, or

[3]The companion to HD 114762 discovered by Latham et al. (1989) has a minimum mass of $11\,M_J$, so may eventually prove to be a brown dwarf rather than an exoplanet.

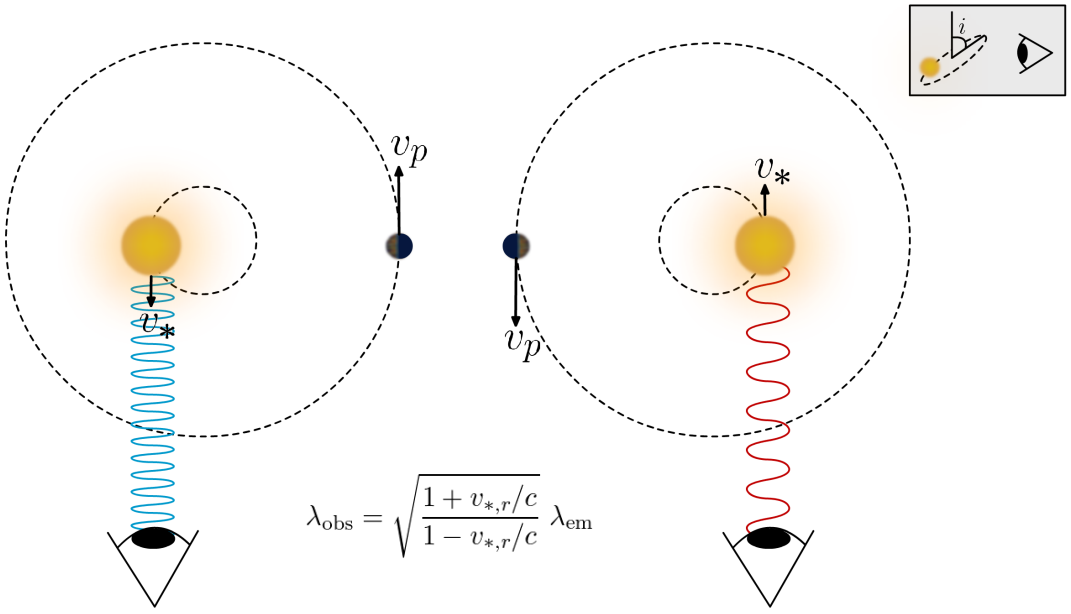


Fig. 1.1 Illustration of the radial velocity method for detecting exoplanets. The stellar spectrum seen by an observer is progressively blueshifted (left) and redshifted (right) as the star orbits the system barycentre. The relation between the observed and emitted wavelength, as a function of the line-of-sight stellar radial velocity, $v_{*,r}$, is given by the relativistic Doppler equation. Inset: the *a priori* unknown inclination between the orbital plane and that of the sky.

radial velocity, changes over time, causing absorption lines in the star's spectrum to be Doppler shifted in a periodic fashion. This process is illustrated in Figure 1.1.

For a single planet system[4], it can be shown from Kepler's laws that the characteristic scale of stellar radial velocity variations is given by (e.g. Murray & Correia, 2010)

$$K_* = \frac{M_p \sin i}{(M_* + M_p)^{2/3}} \left(\frac{2\pi G}{P}\right)^{1/3} \frac{1}{\sqrt{1-e^2}} \tag{1.1}$$

where K_* is the semi-amplitude of the stellar radial velocity, M_p and M_* are the planet and star masses, respectively, i is the inclination angle between the orbital plane and the plane of the sky, G is the universal gravitational constant, P is the orbital period, and e is the orbital eccentricity. In the case of a circular orbit, the radial velocity variations are sinusoidal, with a maximum value of K_*.[5] This equation

[4]Multiplanet systems, to first order, can be treated as a linear superposition of single planet signals.

[5]Eccentric orbits have a maximum of $K_*(1 + e\cos\omega)$, where ω is the argument of pericentre.

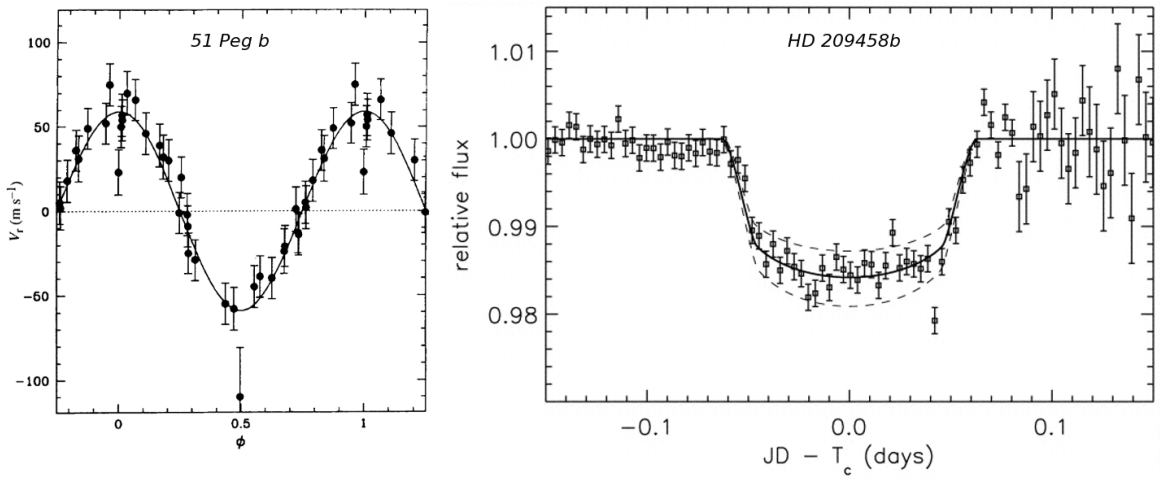


Fig. 1.2 Early observational evidence for the existence of exoplanets. Left: stellar radial velocity curve induced by 51 Pegasi b (Mayor & Queloz, 1995) - the first exoplanet detected around a main-sequence star. Right: transit of the hot Jupiter HD 209458b (Charbonneau et al., 2000) - the first known transiting exoplanet.

immediately offers two key insights: (i) for a given star, radial velocity semi-amplitudes are maximised for high mass, low period, planets, and (ii) as $\sin i$ is unknown *a priori*, by measuring K_* one obtains a constraint on $M_p \sin i$ rather than M_p (i.e. a *minimum* planetary mass).[6] In particular, point (i) illustrates why detections of hot Jupiters were readily achieved by early radial velocity surveys.

Were a distant observer to examine our own solar system, the Sun's radial velocity would be dominated by that of Jupiter: $K_{\odot,J} = 12.5$ ms^{-1}. The Earth, by comparison, would only offer a radial velocity of $K_{\odot,\oplus} = 0.09$ ms^{-1}. Hot Jupiters, on the other hand, induce radial velocities many times in excess of that of Jupiter, with the first detection (Figure 1.2, left) yielding $K_* = 59$ ms^{-1} (Mayor & Queloz, 1995). Radial velocity surveys today have reached precisions of $K_{*,\mathrm{min}} \approx 0.3$ ms^{-1} (e.g. Figure 1.3, left), which precludes detection of Earth-mass planets in the habitable zones of sun-like stars. However, the lower mass and closer-in habitable zones of M dwarf stars leads to characteristically larger radial velocity signals, such that terrestrial mass planets around nearby stars, such as Proxima Centauri b (Anglada-Escudé et al., 2016) and Ross 128b (Bonfils et al., 2018) are now being discovered.

[6] If the radial velocity of the *planet*, K_p, can be measured (e.g. observing Doppler shifted thermal emission from the atmosphere - section 1.4.4) then $K_p/K_* = M_p/M_*$ gives the absolute planet mass.

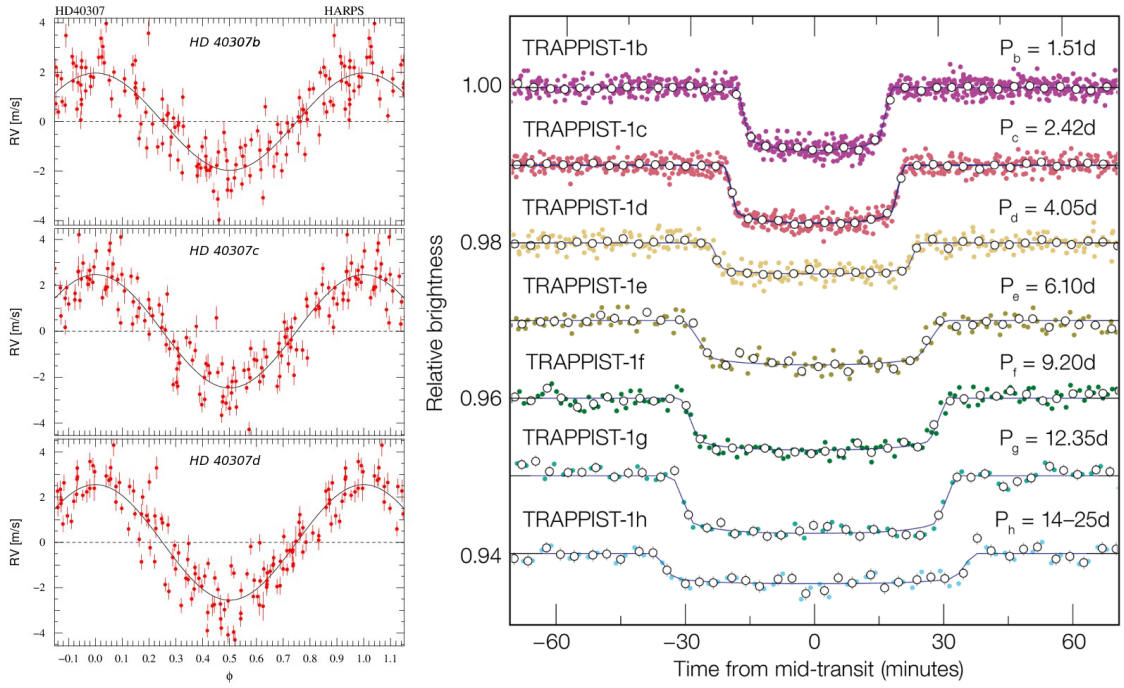


Fig. 1.3 Modern observations of multiplanet solar systems.
Left: HARPS radial velocity curves due to three low-mass planets ($M_p \sin i$ = 4.2, 6.9, and 9.2 $M_\oplus$) in the HD 40307 system (Mayor et al., 2009). A further three planets were inferred in 2013, raising the multiplicity to six (Tuomi et al., 2013). Note that these semi-amplitudes are $\sim 20\times$ smaller than that from 51 Pegasi b (Figure 1.2, left). Right: Spitzer transit light curves for the seven terrestrial planets in the TRAPPIST-1 system (Gillon et al., 2017).

1.2.2 Exoplanet transits

If the orbital plane of a planetary system appears nearly edge on ($i \approx 90°$) with respect to an observer, any planets present in the system will periodically pass in front, or *transit*, their parent star. An illustration of the geometry of an exoplanet transit is shown in Figure 1.4. As transits across other stars are far too distant to be spatially resolved, an observer will instead note a decrease in the stellar flux during a transit due to obscuration by the planet. To first approximation, the fractional flux decrement, or *transit depth*, is given by the ratio of the projected planetary and stellar areas

$$\delta = \frac{F_{\rm out} - F_{\rm in}}{F_{\rm out}} \approx \left(\frac{R_p}{R_*}\right)^2 \tag{1.2}$$

where $F_{\rm out}$ and $F_{\rm in}$ are the fluxes outside and during transit, and R_p and R_* are the planetary and stellar radii. This expression assumes the stellar surface to have a

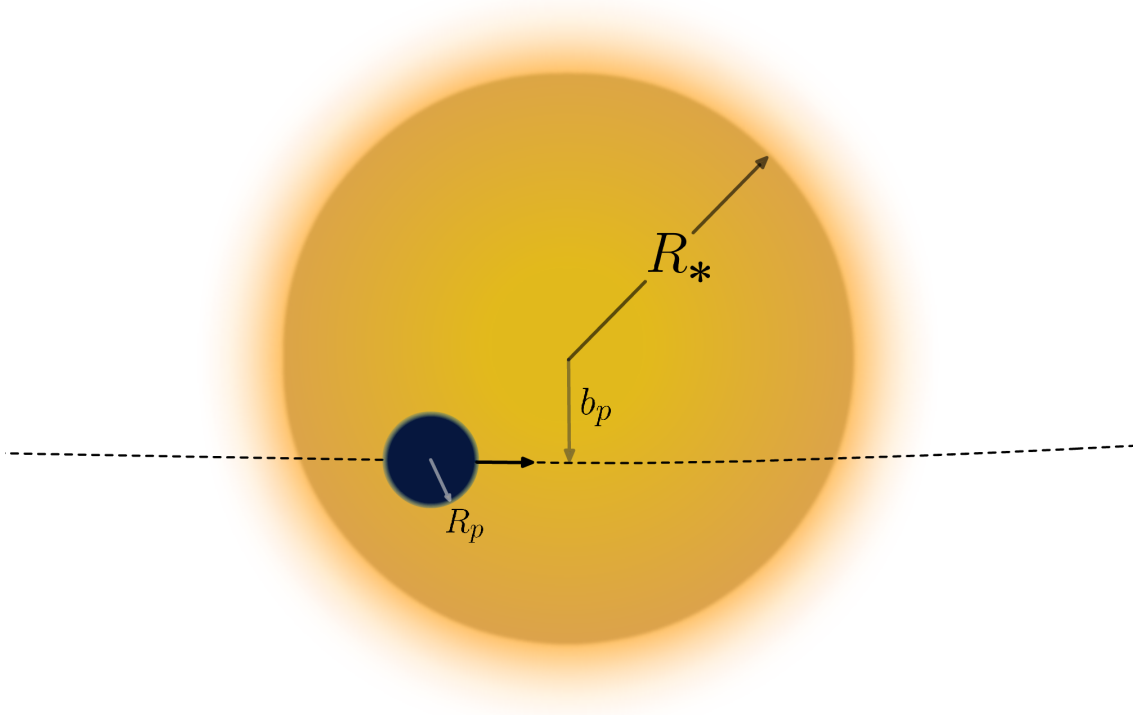


Fig. 1.4 Illustration of the transit method for detecting exoplanets.
If the orbital plane of a system appears nearly edge-on, planets will cross the disk of the star. For systems with $i \neq 90°$, the projected distance between the centre of the planet and centre of the star at mid-transit is given by the planetary impact parameter, $b_p \approx a \cos i$. The condition for a transit to occur is then $b_p \leqslant R_* + R_p$.

uniform brightness (i.e. neglects limb darkening), that the planet fully overlaps the star, and the planet is perfectly opaque. As we will see in section 1.4.1, the relaxation of the final assumption provides a powerful mechanism to probe exoplanet atmospheres.

For a Jupiter-radius planet, Equation 1.2 yields $\delta_J \approx 1\%$, such that photometric transit detections are well-within the range of ground-based telescopes (e.g. Figure 1.2, right).[7] Earth-size planets around sun-like stars produce transits some 100× smaller, $\delta_\oplus \approx 8 \times 10^{-5}$, necessitating space-based observations. For example, the Kepler mission has achieved photometric precisions around bright stars as low as 10 ppm (Borucki, 2018). However, Equation 1.2 demonstrates that smaller stars, such as M dwarfs ($R_* \approx 0.1 R_\odot$), will possess enhanced transit depths. This can be seen in the right panel of Figure 1.3, where the transits of the terrestrial ($R_p \approx R_\oplus$) planets orbiting the M dwarf TRAPPIST-1 (Gillon et al., 2017) have transit depths of the order 1%.

[7] Remarkably, the transit shown in Figure 1.2 was observed with a 10 cm telescope aperture.

The radius of an exoplanet, R_p, can be inferred from a transit (should R_* be otherwise known from stellar evolution models, spectral type, etc.). The transit duration and light curve shape additionally yield the planetary impact parameter, $b_p = a\cos i$ (for a circular orbit), where $a = a_* + a_p$ is the relative star-planet semi-major axis (deduced from the orbital period), and hence the inclination. If a radial velocity observation of the system is also available, then the $\sin i$ mass degeneracy can be lifted and the absolute planet mass, M_p, obtained. Even without radial velocity measurements, multiplanet systems can permit mass determinations through variations in the time of each transit due to planet-planet gravitational interactions (e.g. Agol & Fabrycky, 2018). With R_p and M_p in hand, one can then calculate a bulk density - the first step towards characterising the nature of the planet.

However, one notable limitation to this technique is the narrow range of orbital inclinations which permit a transit to be observed. From the condition $b_p \leqslant R_* + R_p$, integration over a uniform distribution of $\cos i$ shows that the probability of an observable transit occurring for a planet in a circular orbit[8] is $P_{\rm trans} = (R_* + R_p)/a$ (Winn, 2010). This expression demonstrates that close-in planets around large stars are the most likely to exhibit transits. For example, distant observers would have to contend with $P_{\rm trans} = 0.5\%$ for Earth, whilst close-in hot Jupiters ($a \approx 0.05$ au) present a more favourable $P_{\rm trans} = 10\%$. Transit surveys must thus observe large numbers of stars to ensure a sufficient sample of transits, especially for wide-orbit exoplanets.

Nevertheless, the transit method has proven extraordinarily successful, yielding the most exoplanet detections of any method to date. The first new exoplanet discovered by the transit method was confirmed in 2003 (Konacki et al., 2003), demonstrating the efficacy of ground-based instruments to detect transiting exoplanets. Large ground-based surveys, such as WASP (Pollacco et al., 2006), have yielded many exoplanets around nearby bright stars ideal for atmospheric characterisation (section 1.4.1). The era of space-based transit detections began with the launch of the CoRoT mission (Baglin et al., 2006), reporting its first exoplanet detection in 2008 (Barge et al., 2008) and 29 detections over the mission lifetime. The power of the transit method was demonstrated by the Kepler mission (Borucki et al., 2010), which reported over 2600 exoplanet discoveries from 2009-2018. The successor to Kepler, TESS (Ricker et al., 2014), launched in 2018, is now undertaking a 2 year survey of 500,000 nearby bright stars ($V \leqslant 12$), with the first detections already reported (e.g. Vanderspek et al., 2018; Jones et al., 2018). Clearly the transit method has a bright future.

[8]For eccentric orbits with the same a, $P_{\rm trans}$ is enhanced by a multiplicative factor of $1/(1-e^2)$.

1.2.3 Gravitational microlensing

If one wishes to reliably infer properties of the entire population of exoplanets in our galaxy, radial velocity and transit surveys do not provide the entire picture. As we have seen in the last two sections, radial velocity surveys are most sensitive to high-mass, low period planets, whilst transit surveys similarly favour close-in planets. Planets with wide orbital separations are more difficult to detect via these methods. Furthermore, signal to noise considerations render radial velocity and transit detections challenging for all but relatively nearby ($D \lesssim 2$ kpc), bright stars (Sahu et al., 2006). Gravitational microlensing is a technique which can overcome some of these limitations.

First proposed as an exoplanet detection method by Mao & Paczynski (1991) and Gould & Loeb (1992), gravitational microlensing exploits the general relativistic deflection of light in curved space-time to infer the presence of distant exoplanets. When a distant star passes close behind a nearer star from the perspective of an observer, light rays from the background ('source') star traversing the gravitational field of the foreground ('lensing') star experience a change in trajectory. In this manner, light rays on straight line paths away from the observer are bent into the line of sight, resulting in extended, intensified, images of the background star. Should the background star lie precisely behind the foreground star, axial symmetry produces an *Einstein ring*, with angular radius θ_E, as illustrated in Figure 1.5. In the more general case of a slightly off-axis alignment, two images are produced, one arc beyond θ_E and one within. Due to the small angles involved ($\theta_E \sim 1$ mas), the images are unresolved and only the magnified intensity of the background star is observed. Near alignment, the magnification of the background star can be expressed as (Bennett, 2008)[9]

$$A_S \approx \frac{\theta_E}{\theta_S} = \frac{1}{\theta_S}\sqrt{\frac{4GM_L}{c^2}\left(\frac{D_S - D_L}{D_S D_L}\right)} \tag{1.3}$$

where θ_S is the angle between the source star and the lens, θ_E is the angular Einstein radius, M_L is the mass of the lensing star, c is the speed of light, and D_S and D_L are the distances from the observer to the source and lens, respectively. The approximation holds for $\theta_S \ll \theta_E$. Due to the relative velocities of the source and lens, θ_S is a function a time and hence the magnification will increase when approaching closest alignment and decrease thereafter. Should the foreground star possess a planet in an orbit close to the Einstein radius, $r_E = \theta_E D_L$ (typically around 4 au), the gravitational field of the

[9] As $\theta_S \to 0$, higher-order finite source size effects allow large, but not infinite, magnification.

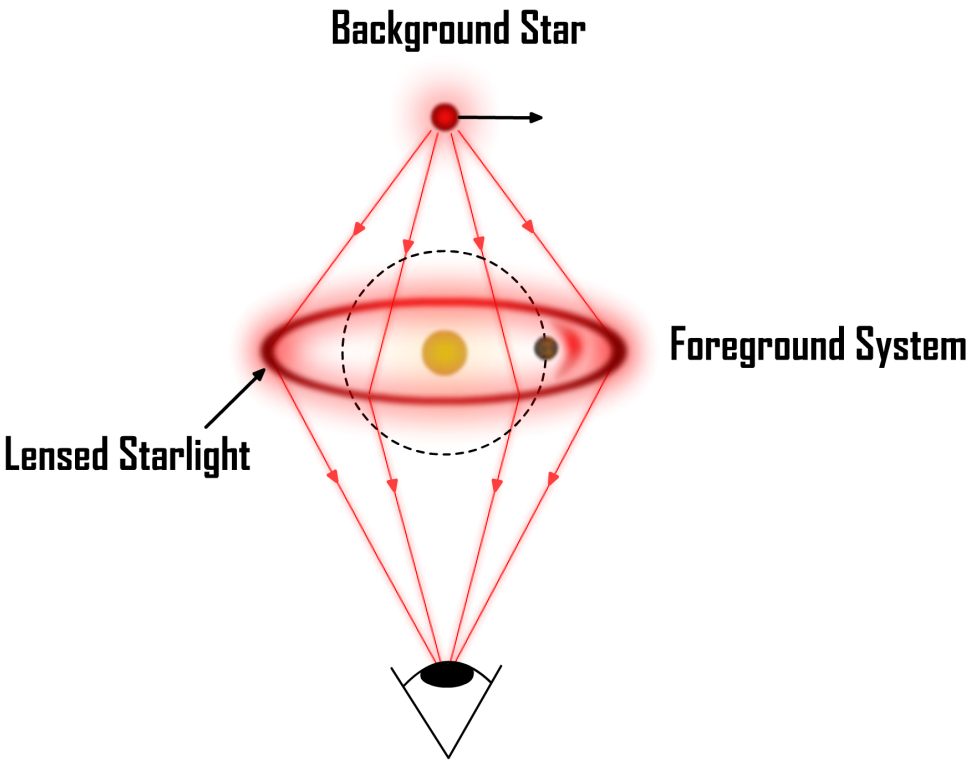


Fig. 1.5 Illustration of gravitational microlensing for exoplanet detection. Light from a background star experiences a change in trajectory after passing through the curved space-time in the gravitational field of a foreground star. Rays that would otherwise have missed an observer are deflected into the line of sight, resulting in an extended, magnified image. A planet orbiting the foreground star acts as an additional gravitational lens, perturbing, and further magnifying, the lensed background starlight.

planet further distorts the image of the background star, resulting in a magnification 'spike' (Figure 1.6, left). Importantly, a microlensing alignment is a one-time event.

The first exoplanet discovered by microlensing was presented by Bond et al. (2004). As the probability of any given pair of stars aligning to produce a microlensing event is small ($\sim 10^{-6}$), microlensing surveys, such as OGLE (Udalski et al., 1992), tend to observe densely-packed stellar regions towards the galactic bulge ($D_S \sim 8$ kpc), such that detected exoplanets are around $D_L \approx D_S/2 \sim 4$ kpc from Earth (Gaudi, 2010). Most microlensing events tend to have $A_S \lesssim 100$, though magnifications as high as $A_S \sim 3000$ have been recorded (Dong et al., 2006). Notably, microlensing is sensitive to exoplanets in wide orbits beyond the snow line (~ 2.7 au in the solar system, see section 1.3.3) and even free-floating exoplanets (e.g. Mróz et al., 2018). Gravitational microlensing thus complements radial velocity and transit detections, providing crucial insights into the exoplanet population and planetary formation mechanisms.

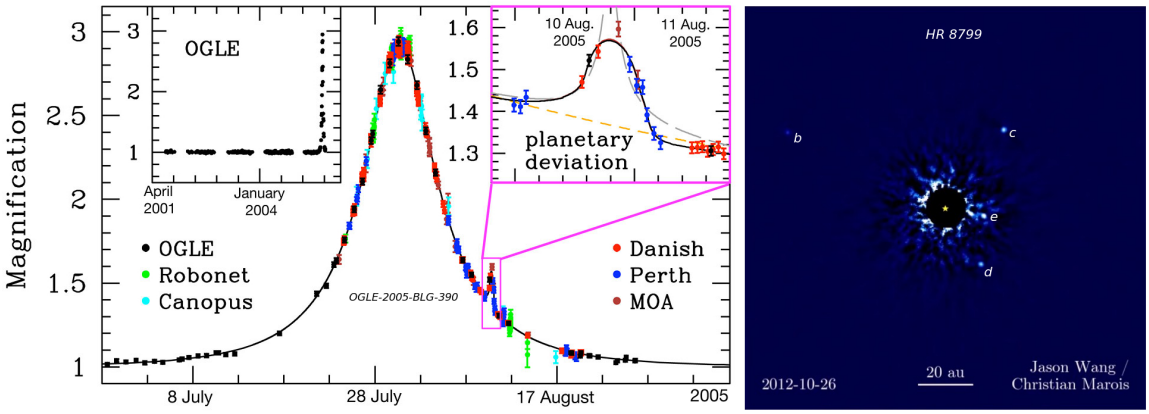


Fig. 1.6 Observations of wide-orbit exoplanet systems.
Left: microlensing detection of an exoplanet at 2.6 au (Beaulieu et al., 2006) (credit: ESO). Right: direct image of the four-planet HR 8799 system (e.g. Marois et al., 2010) with a = 14.5, 24, 38, and 68 au (credit: Jason Wang / Christian Marois).

1.2.4 Direct imaging

The previous three exoplanet detection methods, despite their different physical principles, share one common aspect: the existence of a planet is *indirectly* inferred through its influence on starlight. Whether via a Doppler shifted host star spectrum, a stellar flux decrease during transit, or gravitational magnification of a background star, none of these techniques measure photons from the planet itself. In contrast, the technique of direct imaging aims to observe either reflected light or thermal emission from an exoplanet *directly* by spatially resolving the planet and star as separate point sources.

As alluded to in the introduction, direct imaging of exoplanets presents extraordinary technical challenges. In essence, this boils down to two considerations: (i) host star stellar fluxes far exceed planetary fluxes, and (ii) planet-star angular separations are small, even for the nearest stars. For example, a Jupiter analogue orbiting a star 10 pc away has a planet-star reflected light flux ratio of $F_J/F_* \approx 0.1(R_p/a)^2 \sim 10^{-9}$ (Brown & Burrows, 1990), separated by a spatial scale of $\delta\theta = a/d \approx 0.5''$. For an Earth analogue at the same distance, the reflected light flux ratio falls further to $\sim 10^{-10}$ at $\delta\theta \approx 0.1''$. Suppressing starlight to such low flux ratios falls far below the capability of current instruments, prohibiting the detection of solar system analogues via reflected light. For these reasons, direct imaging of exoplanets, despite being considered since the 1970s (Bonneau et al., 1975; Bracewell, 1978), has instead achieved measured success only since the 2000s by turning to thermal emission from young exoplanets.

The thermal emission flux ratio of a planet-star pair can be roughly estimated as

$$\frac{F_p}{F_*} \approx \frac{B_\lambda\left(T_p\right)}{B_\lambda\left(T_*\right)}\left(\frac{R_p}{R_*}\right)^2 \tag{1.4}$$

where $B_\lambda\left(T\right)$ is the spectral radiance of a black body at temperature T, and T_p and T_* are representative temperatures of the planet and star, respectively. Due to the planetary thermal emission peaking in the infrared, whilst that of the star peaks in the optical, the black body ratio becomes more favourable at longer wavelengths (asymptotically approaching T_p/T_*). Considering a young (< 100 Myr) Jovian planet around a solar analogue, and taking an effective temperature from residual formation heat of $T_p \sim 1000\,\mathrm{K}$ (Biller & Bonnefoy, 2018), Equation 1.4 predicts a flux ratio of $\sim 10^{-4}$ at $3\,\mu\mathrm{m}$. We thus see that thermal emission from young, warm, giant planets observed at infrared wavelengths present especially favourable flux ratios. If such planets additionally reside at wide-orbital separations ($a > 10\,\mathrm{au}$), then direct detection lies within the capabilities of current instruments (Bowler & Nielsen, 2018).

Direct detection of wide-orbit exoplanets from ground-based telescopes are enabled by two key innovations: (i) a coronagraph to block the light of the star within the central observing region of a telescope, thereby improving the planet-star contrast ratio; and (ii) adaptive optics techniques to counter the influence of atmospheric turbulence which otherwise washes out small angular separations (Pueyo, 2018). An illustration of the direct imaging of an exoplanetary system is shown in Figure 1.7. In this manner, the first directly imaged exoplanet, 2M J1207b, a $\sim 5M_J$, $\sim 8\,\mathrm{Myr}$ old planet with $T_{\mathrm{eff}} \approx 1200\,\mathrm{K}$ was detected at $\delta\theta = 0.8''$ (corresponding to $a = 40\,\mathrm{au}$, roughly equivalent to Pluto's orbital separation) by VLT-NACO (Chauvin et al., 2004) and later by Hubble (Song et al., 2006). This was shortly followed by the first direct detection of a multiplanet system, HR 8799 (Marois et al., 2008, 2010), shown on the right of Figure 1.6. The current generation of instruments capable of directly imaging exoplanets, SPHERE (Beuzit et al., 2008) and GPI (Macintosh et al., 2008), both exploit 8 m telescopes with adaptive optics and coronagraphs to achieve sensitivities down to $F_p/F_* \sim 10^{-6}$ and $\delta\theta \gtrsim 0.2\,''$. This has enabled the discovery of exoplanets approaching Jupiter-mass with $a \approx 10\,\mathrm{au}$, such as the $2\,M_J$ planet 51 Eridani b (Macintosh et al., 2015). Though the direct detection of true solar system analogues will have to await the development of future instruments, direct imaging is already offering constraints on the occurrence rate of wide-orbit exoplanets (Bowler & Nielsen, 2018) and yielding direct spectra of their atmospheres (see section 1.4.5).

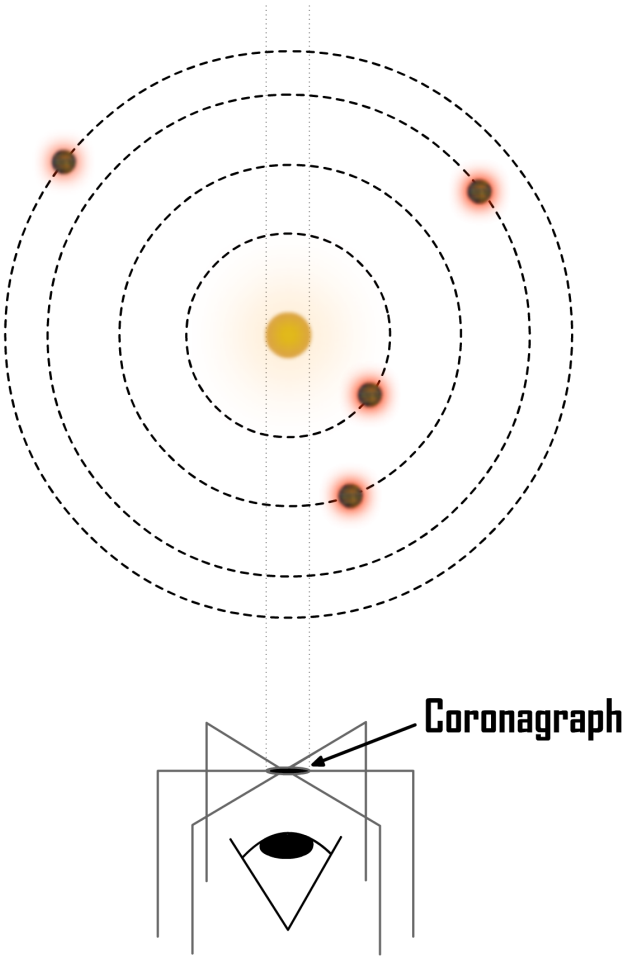


Fig. 1.7 Illustration of the direct imaging method for detecting exoplanets. A coronagraph suppresses central starlight by either physical obstruction or cancellation via interference. Sufficiently bright planets may then be observed at sub-arcsecond separations either by space-based diffraction-limited imaging or from the ground using adaptive optics to negate the influence of atmospheric turbulence.

1.2.5 Other methods

The methods already discussed have yielded the vast majority of exoplanet discoveries to date. For completeness, I also briefly mention here other methods which have led to the discovery of at least one exoplanet[10] and provide references to recent reviews for the interested reader. *Pulsar timing* (Kramer, 2018), though responsible for the first exoplanet detections (Wolszczan & Frail, 1992), have yielded only 6 detections over the last 27 years. More recently, a similar method, *stellar pulsation timing* (Hermes, 2018),

[10] All discovery counts listed are from the NASA Exoplanet Archive, accessed on 14 June 2019.

has been applied to detect 2 exoplanets around pulsating variable stars (Silvotti et al., 2007; Murphy et al., 2016). Variations in the timing or duration of a planetary transit or eclipse due to planet-planet gravitational interactions, *transit timing / duration variations* (Agol & Fabrycky, 2018), first demonstrated by Holman et al. (2010), have resulted in 26 detections. Non-transiting planets with high reflectivity or thermal emission can be detected by *orbital brightness modulations* (see section 1.4.3), as first achieved by Charpinet et al. (2011) and resulting in 6 detections to date. Finally, the oldest method used to search for exoplanets, *astrometry* (Malbet & Sozzetti, 2018), the measurement of the position and motion of stars, has only recently yielded a convincing detection of a sub-stellar object (Sahlmann et al., 2013) - though it is proving useful to further characterise known exoplanets (e.g Snellen & Brown, 2018). In what follows, exoplanets discovered by these methods will be denoted by 'other'. Having reviewed the principles underlying exoplanet detection methods, I now turn towards the current understanding of the exoplanet population enabled by these discoveries.

1.3 The diversity of worlds

We are the first generation to *know* about the existence of planets orbiting other stars. Over the last 30 years, over 4,000 exoplanets have been detected. These discoveries are now providing answers to fundamental questions about planetary systems which lay beyond reach when the solar system was the only known example. Here, I review the current understanding of the broad characteristics of exoplanets as of early 2019. I first summarise properties of the known exoplanet population, before exploring how the physical properties of exoplanets connect to their interior composition. Finally, the current paradigms of planetary formation, and the ways in which the diversity of worlds has influenced these formation theories, will be examined.

1.3.1 The exoplanet population

Extrasolar planets exhibit remarkable variety. In Figure 1.8, I compare the known exoplanet population, in terms of mass vs. semi-major axis and radius vs. orbital period, to the solar system planets. All substellar objects included in the exoplanet.eu database (Schneider et al., 2011) are displayed, though I note that the objects with $M_p \gtrsim 10 - 20\,M_J$ are better thought of as brown dwarfs (Chabrier et al., 2014). A passing glance reveals that exoplanets naturally separate into three broad populations:

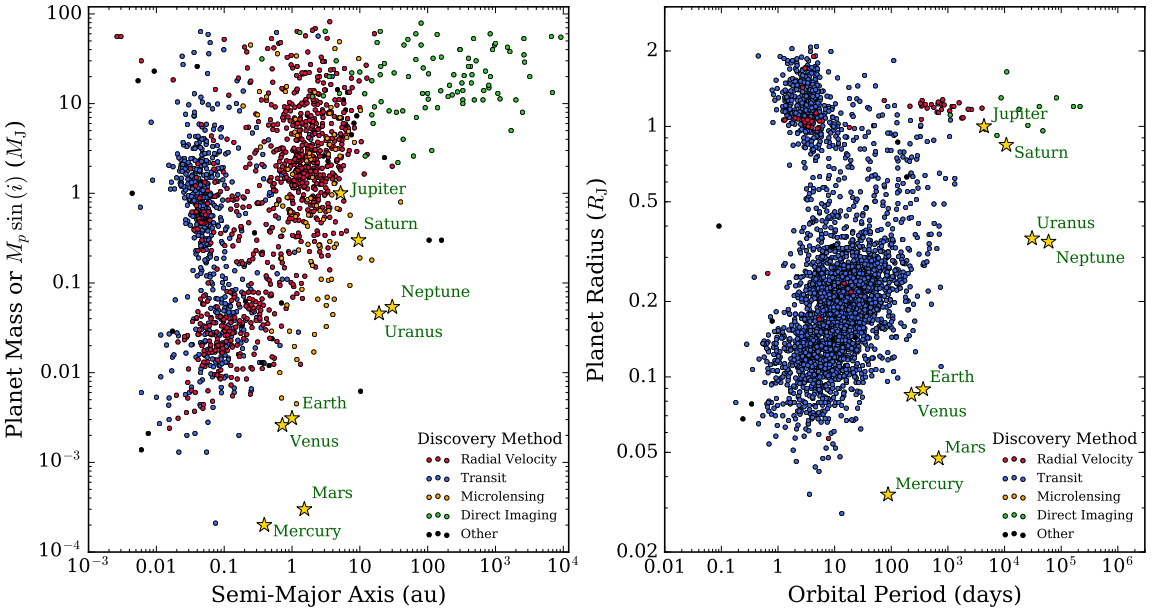


Fig. 1.8 The population of discovered exoplanets.
Left: exoplanets with known M_p (or $M_p \sin i$, for many radial velocity planets) and semi-major axis, coloured according to detection method. Right: exoplanets with known R_p and orbital period, coloured according to detection method. The solar system planets (gold stars) are overlaid for comparison. The data used is from the exoplanet.eu database (accessed on 14 June 2019), with 1865 exoplanets shown on the left and 3025 exoplanets on the right. The total number of exoplanets is 4039.

1. **Hot Jupiters**. These giant planets are similar to Jupiter in mass ($M_p \sim 0.3 - 3\,M_J$), though possess a larger average radius ($R_p \sim 0.8 - 2\,R_J$). They typically reside in close-in orbits ($a < 0.1\,\text{au}$ / $P < 10\,\text{days}$), resulting in high temperatures ranging from $\sim 800 - 4000\,\text{K}$. Their densities are consistent with a H_2-He composition (see section 1.3.2).

2. **Cool giants**. This population of giant planets, ranging from $M_p \sim 0.5 - 10\,M_J$, reside at much larger orbital separations ($a > 0.4\,\text{au}$ / $P > 100\,\text{days}$) than the hot Jupiters. At the highest semi-major axes explored by current radial-velocity observations, this population encompasses Jupiter.

3. **Sub-giants**. This particularly extensive population, with $M_p \lesssim 0.1\,M_J$ ($\lesssim 30\,M_\oplus$) and $R_p \lesssim 0.5\,R_J$ ($\lesssim 6\,R_\oplus$), encompasses both Neptune-like gaseous planets and smaller planets consistent with predominantly rocky compositions. The known sub-giant population resides in shorter period orbits ($P < 100\,\text{days}$) than the terrestrial planets of our own solar system.

The separation of Figure 1.8 according to mass, radius, and detection method illustrates an important consideration: all exoplanet detection methods only illuminate *specific* regions of the possible discovery space. This is particularly striking in two areas: (i) the cool giant population is almost entirely absent in the R_p - P diagram (these planets are detected preferentially by their radial-velocities, yielding $M_p \sin i$ but no information on R_p); (ii) the sub-giant population is more prevalent in R_p - P space vs. M_p - a space (transits yields R_p but little information on M_p, whilst radial-velocity detections are more challenging for lower-mass planets).[11] It can also be seen that planets discovered by microlensing cover a range of masses across ~ 4 orders of magnitude, from $\sim M_\oplus$ to the brown dwarf limit and beyond, though only a limited range of semi-major axes centred around $a \approx 2\,\mathrm{au}$ is covered. Microlensing detections are particularly important, however, as they can include planets currently inaccessible to other methods, such as Earth-mass planets near the snow line ($\gtrsim 2\,\mathrm{au}$). Finally, direct imaging, though yet to reach sub-Jovian masses, has played a crucial role in revealing most of the very wide-orbit ($> 100\,\mathrm{au}$) planets found to date.

These above considerations are examples of *detection biases*. Indeed, if one were to take Figure 1.8 at face value, it could be construed that all the solar system planets (bar Jupiter, and perhaps Saturn) are extreme outliers from the exoplanet population. In reality, vast regions of parameter space (e.g. wide-orbit ice-giants like Uranus and Neptune) remain *terra incognita* until improvements in detection method sensitivity and / or extended observational baselines open these new worlds to scientific inquiry. It will be fascinating to see whether additional populations of planets emerge in these uncharted regions in the future.

Nevertheless, detection biases are sufficiently well understood that they can be corrected to reveal underlying properties of the exoplanet population. In particular, one set of quantities of interest are *occurrence rates* - the average number of planets per star satisfying a given range of properties. With estimates of occurrence rates, especially from the large number statistics provided by Kepler, we can finally obtain quantitative answers to age-old questions about the intrinsic abundance, or rarity, of different types of planets in our galaxy. Extensive reviews on occurrence rates can be found in the literature (e.g. Winn & Fabrycky, 2015; Winn, 2018; Bowler & Nielsen, 2018), so here I will merely summarise the main traits of the exoplanet population.

[11] Indeed, far more exoplanets have measured radii than masses at present, which largely reflects the success of the Kepler space telescope at detecting sub-giants.

I. **Planets are common in our galaxy**
Sun-like stars have ~ 0.9 planets per star with $P < 400\,\mathrm{day}$ and $R_p > R_\oplus$ (Zhu et al., 2018). Planets on wider orbits yield consistent results: $1.6^{+0.7}_{-0.9}$ planets per star ranging from $0.5 - 10$ au and $5\,M_\oplus - 10\,M_J$ (Cassan et al., 2012). It therefore appears that planets are at least as common as stars in our galaxy.

II. **Free-floating ('rogue') planets may be more common than stars**
Microlensing surveys have inferred a background of $\sim 1.2 - 1.8$ unbound planets per main sequence star in our galaxy (Sumi et al., 2011; Clanton & Gaudi, 2017). Recent studies suggest this 'sea' of unbound planets has masses predominantly similar to that of the Earth (Mróz et al., 2017; OGLE Collaboration et al., 2019).

III. **Hot Jupiters are rare**
Despite representing many of the first exoplanet discoveries, their intrinsic occurrence rate around main sequence stars is only $\sim 0.5 - 1\%$ (e.g. Wright et al., 2012; Fressin et al., 2013; Deleuil et al., 2018). Hot Jupiters simply represent the 'low-hanging fruit' preferentially discovered by early exoplanet discoveries.

IV. **Massive planets on very wide orbits are rare**
The dearth of planets found by direct imaging surveys indicates an occurrence rate of $\sim 1\%$ for planets with $\sim 5 - 13\,M_J$ at projected separations from $\sim 5 - 500\,\mathrm{au}$ (Bowler, 2016; Galicher et al., 2016; Vigan et al., 2017). Giant planets in the widest known orbits thus appear to be as rare as the close-in hot Jupiters.

V. **Cool giants are more common than hot Jupiters**
Around solar-type (FGK) main sequence stars, the occurrence rate of cool giants is $\sim 5 - 10\%$ (Cumming et al., 2008; Mayor et al., 2011; Santerne et al., 2016). In particular, Jupiter analogues appear to be relatively rare, with long-term radial velocity monitoring yielding frequencies of $\sim 3 - 6\%$ (Rowan et al., 2016; Wittenmyer et al., 2016).

VI. **Giant planets are less common around metal-poor and low-mass stars**
Stars with metallicities $0.5\times$ that of the Sun are $\sim 5 - 10\times$ less likely to contain a giant planet than stars with $2\times$ solar metallicity (e.g. Fischer & Valenti, 2005; Sousa et al., 2011). Similarly, M dwarf stars ($0.1 - 0.5\,M_\odot$) are $\sim 3\times$ less likely to host giant planets with $M_p \gtrsim 0.3\,M_J$ than Sun-like FGK stars (Cumming et al., 2008; Bonfils et al., 2013).

VII. **Sub-giants are the most common type of planet**
Planets with radii ranging from $1-4\,R_\oplus$ (i.e. sizes between Earth and Neptune) with $P \lesssim 1\,\mathrm{yr}$ occur around $\sim 50\%$ of Sun-like stars (Winn & Fabrycky, 2015). Many of these planets reside in compact multiplanet systems, with as many as 8 planets orbiting within $P < 1\,\mathrm{yr}$ (Shallue & Vanderburg, 2018).

VIII. **Small planets are more common around low-mass stars**
Planets ranging from $1-4\,R_\oplus$ are $\sim 2-3\times$ more common around M dwarfs than around Sun-like stars (Howard et al., 2012; Mulders et al., 2015). Gaidos et al. (2016) report the average number of such planets with $P < 180\,\mathrm{day}$ to be 2.2 ± 0.3 per M dwarf. The occurrence of planets with $R_p \sim R_\oplus$ is relatively insensitive to stellar metallicity, as opposed to the clear trend seen for giant planets (Sousa et al., 2008; Buchhave et al., 2012). However, short-period ($10-100\,\mathrm{days}$) sub-Neptunes ($1.7-4.0\,R_\oplus$) have recently been found to increase in frequency for metal-rich stars (Petigura et al., 2018).

IX. **Earth-like ('habitable') planets are common**
The frequency of planets with $R_p \sim R_\oplus$ in orbits receiving an insolation capable of supporting liquid water (the 'habitable zone'), often called $\eta_\oplus$, has been estimated at $\sim 25\%$ (e.g. Petigura et al., 2013; Dressing & Charbonneau, 2015). Such estimates of course depend on the precise definition of 'habitable', and are currently limited by small sample sizes. A recent illustrative example is provided by Hsu et al. (2019), who derive an occurrence rate of $24^{+11}_{-10}\%$ for planets with $1-1.75\,R_\oplus$ in $237-500\,\mathrm{day}$ orbits around Sun-like stars.

These findings have provided a tantalising glimpse of the wider picture of planetary systems in our galaxy. In particular, the abundance of small planets, many of which reside within their star's 'habitable zone', has attracted intense attention due to the potential to remotely search for signs of life on such planets in the near future (e.g. see Kaltenegger, 2017). The road to this ambitious goal is long, but one which naturally aligns with a broader trend: the shift from exoplanet detection to exoplanet *characterisation*. The first step in this direction comes from relating the mass and radius of a planet to its overall composition.

1.3.2 The composition of exoplanets

In our own solar system, there is a clear difference between the masses and radii of the terrestrial and giant planets. There is a discontinuous jump between the properties of the largest rocky planet, Earth, and the smallest gas giant, Neptune, from $1 \rightarrow 3.88\,R_\oplus$ and $1 \rightarrow 17.15\,M_\oplus$. As we saw in the last section, this distinction is less clear-cut for exoplanets, with the most common planet population spanning a continuum of masses and radii between those of Earth and Neptune (see Figure 1.8). This raises an interesting question: where do we draw the line between 'rocky' and 'gaseous' planets? To answer this, and related questions about the physical nature of exoplanets, I will now briefly examine how one can infer the overall interior composition of exoplanets.

The subset of exoplanets with M_p and R_p known to within 30% precision is shown in Figure 1.9. When the physical properties of exoplanets are expressed in this manner, trends not apparent on $M_p{-}a$ or $R_p{-}P$ diagrams emerge. A population of small ($\sim R_\oplus$), low mass ($\sim 5M_\oplus$), planets (including Earth and Venus) undergo a transition around $10\,M_\oplus$ into a population of large ($\sim 1.4\,R_J$), high mass ($\sim M_J$), planets (including Jupiter and Saturn). The low-mass solar system giants, Uranus and Neptune, reside near the transition. This transition is far from a general rule, however, as even the limited number of planets with well-constrained physical properties (≈ 400) exhibit significant scatter about this overall trend. For each planet, I additionally estimate its equilibrium temperature as $T_{\rm eq} \approx T_* \sqrt{R_*/2a}$ (where T_* is the stellar effective temperature, and I assume planetary energy balance, uniform heat distribution, a circular orbit, and zero albedo). This allows one to see that most of the hottest known planets are in the high mass population (i.e. the hot Jupiters), though there are some known small planets with $T_{\rm eq} \gtrsim 2000\,{\rm K}$ (sometimes considered 'lava planets').

To further interpret this diagram, theoretical models of planetary interiors are required. Taking a spherical planet in hydrostatic equilibrium, supplemented by an *equation of state* ($P = f(\rho, T)$, with P pressure, ρ density, and f a function unique to a given material), one can numerically integrate progressive spherical shells to obtain predicted values of M_p and R_p (e.g. Seager et al., 2007; Zeng & Jacobsen, 2017). By repeating this exercise for different potential materials, or combinations of materials, each with known equations of state, one can construct *mass-radius relations* for a range of interior compositions. Figure 1.9 shows four such mass-radius relations from Seager et al. (2007): (i) a planet made of a mixture of cold ($T = 0\,{\rm K}$) H_2 and He in solar proportions, (ii) a 'water world' (100% H_2O), (iii) a 'rocky world' (100% $MgSiO_3$), and (iv) an 'iron world' (100% Fe). The three lower curves, corresponding to higher

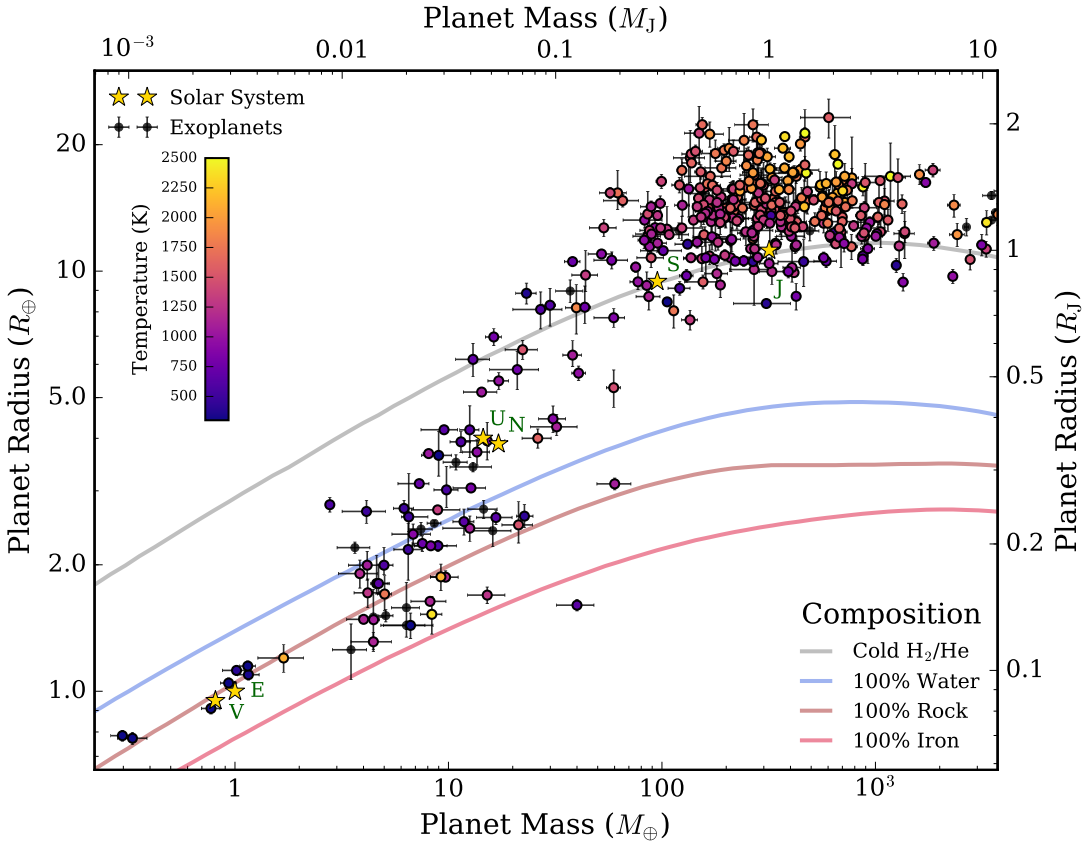


Fig. 1.9 Mass-radius diagram for well-constrained exoplanets.
All exoplanets with M_p and R_p determined to $< 30\%$ precision are shown - 435 planet currently fulfil this criteria. Each planet is colour-coded according to an estimate of its equilibrium temperature. The solar system planets are indicated by gold stars. Theoretical mass-radius relations for various different interior compositions, ranging from pure Fe to a H_2-He mixture, are overlaid for comparison (Seager et al., 2007). Exoplanet properties are from the exoplanet.eu database (accessed 14 June 2019), with updated system parameters for the TRAPPIST-1 planets from Dorn et al. (2018).

densities, show the typical range of M_p and R_p values for planets considered 'rocky'. For example, Earth and Venus contain $\approx 30\%$ Fe and $\approx 70\%$ $MgSiO_3$ by mass, shifting them towards the Fe curve. For R_p above the 100% H_2O curve, a significant component of a planet's size comes from a gaseous envelope. Lying between the H_2O and H_2/He curves, Uranus and Neptune are referred to as *ice-giants* due to a large fraction of their masses likely deriving from H_2O. Finally, note that the majority of hot Jupiters have R_p above the maximum predicted for H_2/He. This feature, suggesting hot Jupiters possess an additional source of heating beyond that provided by their star, is known as the 'hot Jupiter radius inflation problem' (for a recent review, see Laughlin, 2018).

Classification of exoplanets

Given the apparent transition in physical properties of the exoplanet population due to different planets possessing different interior compositions, it is only natural that attempts have been made to categorise exoplanets. In Table 1.1, I show two such classification schemes, one according to the mass of an exoplanet (Stevens & Gaudi, 2013) and one according to its radius (Borucki et al., 2011). If only M_p or R_p is known, such classification schemes can establish a zeroth-order estimate for the nature of a planet. Whilst useful in a broad sense, Figure 1.9 demonstrates that the true picture can be more complex. Indeed, there are planets with $M_p > 10\,M_\oplus$ which have radii consistent with a rocky interior, alongside planets with $M_p < 10\,M_\oplus$ that have radii indicating a gaseous envelope. All this is to say that one should be cautious assigning exoplanets to categories if only one of the mass or radius is accurately determined. This caveat in mind, all known exoplanets with $R_p \lesssim 2\,R_\oplus$ appear consistent with a rocky interior, whilst those with $R_p \gtrsim 3.5\,R_\oplus$ are consistent with a gaseous envelope.

For planets with extremely precise measurements of R_p and M_P, it is now possible to go beyond simple classification to derive statistical constraints on potential interior compositions. A notable example is provided by the seven terrestrial planets in the TRAPPIST-1 system (Gillon et al., 2017) (seen in Figure 1.9 as seven blue circles in the lower left near Earth and Venus). Combining precise radii from transits (Delrez et al., 2018) with masses derived from transit timing variations has enabled the densities of these planets to be established at the 5-12% level (Grimm et al., 2018), permitting detailed statistical studies of their interiors (Dorn et al., 2018). This promising avenue

Table 1.1 Exoplanet classification schemes via mass or radius

Category	$M_p^\dagger$	$R_p^\ddagger$
Sub-Earths	$10^{-8}\,M_\oplus - 0.1\,M_\oplus$	–
Terrestrial / Earths	$0.1\,M_\oplus - 2\,M_\oplus$	$\leqslant 1.25\,R_\oplus$
Super-Earths	$2\,M_\oplus - 10\,M_\oplus$	$1.25\,R_\oplus - 2\,R_\oplus$
Neptunes	$10\,M_\oplus - 100\,M_\oplus$	$2\,R_\oplus - 6\,R_\oplus$
Jupiters	$100\,M_\oplus - 1000\,M_\oplus$	$6\,R_\oplus - 15\,R_\oplus$
“ ”	$0.31\,M_J - 3.15\,M_J$	$0.54\,R_J - 1.34\,R_J$
Super-Jupiters	$3.15\,M_J - 13\,M_J$	$\geqslant 1.34\,R_J$

† Based on Stevens & Gaudi (2013). ‡ Based on Borucki et al. (2011).
Conversion factors: $M_J = 317.82\,M_\oplus$, $R_J = 11.21\,R_\oplus$.

is, however, limited by the availability of just two data-points for each planet (R_p and M_p), resulting in a wide range of degenerate solutions for more complicated multilayer interior models (Suissa et al., 2018). As we shall see in section 1.4, spectra of exoplanetary atmospheres provide a complementary and rich window into the nature of exoplanets. Before turning to atmospheric characterisation, the focus of this thesis, I will conclude this section with a brief overview of how the observed properties and composition of the exoplanet population connects to the process of planetary formation.

1.3.3 Exoplanet formation paradigms

The present day orbits, physical characteristics, and compositions of the planets in any given system are intrinsically linked to the manner in which they formed. To understand how the past shaped the present, one requires knowledge of the initial conditions in protoplanetary disks, physical processes and disk chemistry, mechanisms via which planets form, how young planets interact with the disk, and the subsequent coupled planet-disk evolution (to name but a few considerations). The discovery of thousands of exoplanets has provided a treasure trove of information to constrain aspects of planetary system formation theories beyond the limited picture provided by the solar system. Here I will broadly outline the mechanisms by which planets are currently understood to form, and how this links to their composition.

Following the formation of a star, conservation of angular momentum causes collapsing material from the initial molecular cloud to flatten into a rotating *protoplanetary disk*. Observed disks have radii $> 30\,\mathrm{au}$ (e.g. Tobin et al., 2015) and typically contain gas for 3-10 million years (Myr) (Hillenbrand, 2008; Roberge & Kamp, 2010). This sets a natural timescale during which gaseous planets must form. Two principle mechanisms have been proposed for giant planet formation: (i) 'bottom-up' formation via *core-accretion* (e.g. Pollack et al., 1996) or (ii) 'top-down' formation via *gravitational instability* (e.g. Boss, 1997). Smaller, rocky, planets can only form via bottom-up processes (Izidoro & Raymond, 2018). A schematic illustration of these two formation pathways is shown in Figure 1.10. I will now summarise each theory in turn.

Formation by core-accretion

The core-accretion theory of planet formation (Safronov, 1972; Lissauer, 1993; Pollack et al., 1996) is the currently accepted paradigm under which most planets form (D'Angelo & Lissauer, 2018; Izidoro & Raymond, 2018). The process begins when

Fig. 1.10 Illustration of the two main pathways for planetary formation. Left: stages of planet formation according to the core-accretion theory (e.g. Pollack et al., 1996). Right: stages of planet formation according to the gravitational instability theory (e.g. Boss, 1997). Credit: NASA / ESA & A. Feild (STScI).

interstellar dust (∼µm in size) grains present in the protoplanetary disk gradually grow from adhesive two-body collisions, eventually reaching ∼ cm-sized objects (Dominik et al., 2007). Further growth up to ∼ km-sized *planetesimals* is thought to be accom-

plished by a subtle effect called the streaming instability (e.g. Youdin & Goodman, 2005; Schäfer et al., 2017). For sizes $\gtrsim$ km, the effects of gravity become increasingly more important than the planetesimal material properties. Low-velocity collisions of 1-100 km-sized planetesimals in nearby orbits build planetary embryos (or *cores*) composed of heavy-element materials. Embryos with greater initial masses accrete material more rapidly (Greenberg et al., 1978), leading to the formation of protoplanets over ~ 0.1 Myr (Benvenuto et al., 2009) until a local 'feeding zone' has been depleted. It is at this point that terrestrial and giant planet formation diverge.

Small, predominately rocky, planets are thought to form in the hotter inner regions of protoplanetary disks, whilst giant planets form in the cooler outer regions. This distinction can be understood in terms of the location of the H_2O *snow line* - the radial distance where the disk temperature becomes cool enough for water to condense into ice ($\approx 150\,\mathrm{K}$, or $2.7\,\mathrm{au}$ for the early solar system Martin & Livio (2012)). Past the snow line, the abundance of ice enhances the disk solid surface density by a factor of ~ 4 (Lecar et al., 2006), resulting in the rapid growth of cores with $M_{\mathrm{core}} > M_{\oplus}$ (Pollack et al., 1996). These massive cores slowly accrete a gaseous envelope over the next ~ 2 Myr (Lissauer et al., 2009). Once (or if) the condition $M_{\mathrm{envelope}} > M_{\mathrm{core}}$ is reached, the envelope collapses and a period of rapid runaway gas accretion begins (Pollack et al., 1996; Hubickyj et al., 2005), lasting until all gas in the vicinity has been dispersed. Giant planets with $M_{\mathrm{core}} > 5 - 20\,M_{\oplus}$, such as Jupiter and Saturn, thereby undergo runaway accretion to obtain large H_2-He envelopes. Planets with less-massive cores, such as Uranus and Neptune, can fail to reach the runaway stage during the disk lifetime and hence attain considerably lower final masses (cf. $M_{\mathrm{Neptune}} \approx 0.05\,M_J$). Whilst gaseous planets form in this manner over $\sim$ Myr timescales (Movshovitz et al., 2010), rocky planet cores accrete little gas and instead grow by collisions and mergers over ~ 100 Myr timescales (Chambers & Wetherill, 1998; Raymond et al., 2009). The layout of the solar system, with rocky planets inside the location of the primordial snow line and giant planets outside, matches well with the core-accretion paradigm.

The discovery of exoplanets presented challenges for this 'classical' formation sequence. In particular, the population of hot Jupiters and short period exoplanets underline that the present day orbits of planets need not lie where they initially formed. The movement of planets during and post-formation is referred to as *migration* (Ward, 1997). Planetary cores which reach $M_{\mathrm{core}} \gtrsim 0.1 - 1\,M_{\oplus}$ whilst gas remains can exert gravitational torques on the protoplanetary disk, resulting in an exchange of orbital angular momentum between the planet and the disk (Goldreich & Tremaine, 1979).

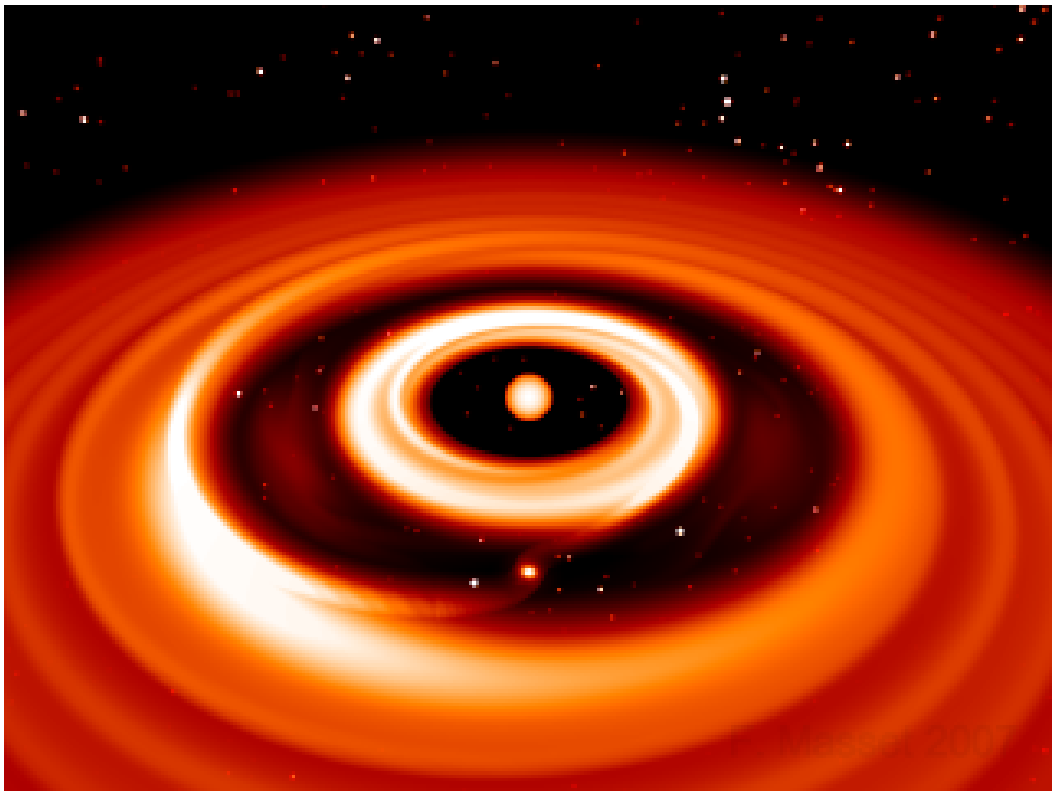


Fig. 1.11 Simulation of a giant planet undergoing Type II migration. An $\sim M_J$ planet generates tidal toques which exchange angular momentum with gas in the protoplanetary disk, resulting in a gap opening around the planet's orbit. The planet then follows the evolution of the gas in the disk, typically spiralling inwards as gas accretes onto the star. Credit: Frédéric Masset.

Spiral density waves thereby excited in the gas tend to cause the planet to rapidly migrate towards the star over $\sim 0.1\,\mathrm{Myr}$ timescale (Lubow & Ida, 2010) in a process called *Type I migration*. The solar system terrestrial planets grew too slowly to be affected by Type I migration, though this is not necessarily the case for exoplanet systems. More massive planets ($M_p \sim 30\,M_\oplus$) exert stronger tidal torques, which can result in the formation of a gap in a protoplanetary disk where gas is depleted (Lin & Papaloizou, 1986). As the gas viscously evolves, typically accreting onto the star over Myr timescales, the gap and planet follow the gas and spiral inwards. This process, known as *Type II migration*, is visually depicted in Figure 1.11. Type II migration has been offered as an explanation for how hot Jupiters reached their present-day close-in orbits (e.g. Lin et al., 1996). *In situ* formation of hot Jupiters is considered unlikely (e.g. Bodenheimer et al., 2000), though planet-planet scattering is also a viable mechanism (Weidenschilling & Marzari, 1996).

Formation by gravitational instability

An alternative theory of planet formation is the disk instability (or gravitational instability) model (Kuiper, 1951; Cameron, 1978; Boss, 1997). This formation pathway proceeds when a random overdensity of gas in a protoplanetary disk begins to collapse under its own gravity, as shown in Figure 1.10. Sufficiently massive, low temperature, disks can then fragment as the overdensity becomes self-gravitating and rapidly accrues material (in a similar manner to how stars themselves form). Gravitational instabilities tend to form high-mass ($\sim 10\,M_J$) planets at large orbital separations ($\sim 10 - 100\,\mathrm{au}$) (Rafikov, 2005). Though gravitational instability can form planets more rapidly than core-accretion ($\sim 10^3\,\mathrm{yr}$), it is thought to be an unlikely origin for most exoplanets observed to date (Kratter & Lodato, 2016), especially hot Jupiters (Forgan & Rice, 2013). However, some of the very wide-orbit exoplanets now being discovered by direct imaging surveys, such as the HR 8799 system (Marois et al., 2010) (see Figure 1.6), have been suggested as evidence that gravitational instability offers a viable second avenue to form giant planets (Boley, 2009; Dodson-Robinson et al., 2009).

Testing planet formation theories

Observations of exoplanet properties offers the ability to test predictions of these two formation theories. In particular, the higher occurrence rates for giant planets around metal-rich and more massive stars (section 1.3.1) is readily explained by core-accretion. Higher metallicity or mass stars possess disks containing a greater inventory of material to build planetesimals, matching the observed trends for giant planets (Mulders, 2018). Gravitational instability, however, is inconsistent with the host-star metallicity-planet occurrence trend (Cai et al., 2006; Durisen et al., 2007). Another line of evidence comes from the JUNO mission (Bolton et al., 2017) inference of a $6 - 25\,M_\oplus$ core for Jupiter (Wahl et al., 2017), matching the prediction of $\sim 10\,M_\oplus$ under core-accretion scenarios (Mizuno et al., 1978). Although the formation of giant planet cores is not a requirement for planets formed by gravitational instability, a $\sim 6\,M_\oplus$ core could be created for planets like Jupiter after their formation through sedimentation of rock and ice into their gaseous atmosphere (Boss, 1998). Furthermore, the evidence of a large population of free-floating planets (section 1.3.1) supports the notion that planet-planet scattering may have ejected large numbers of planets during their formation and early evolution (e.g. during migration). On balance, the weight of evidence currently favours core-accretion as the dominant mechanism of planet formation.

Linking formation to composition

The present day composition of planets is fundamentally linked to their formation. The initial location in the protoplanetary disk where they formed and any subsequent migration history determines the elemental composition of the materials accreted by the planet. We have already seen that accretion in the outer disk beyond the H_2O snow line provides an inventory of ices to build a planetary core, hence resulting in a greater H_2O content in their interior (Alessi et al., 2017). The suggestion of a compositional transition in Figure 1.9 above $\sim 10\,M_\oplus$ provides support for runaway accretion of H_2/He predicted by the core-accretion theory (Pudritz et al., 2018). Indeed, the fact that there are no well-characterised exoplanets with $M_p \gtrsim 80\,M_\oplus$ consistent with a rocky interior (Figure 1.9) suggests that planet formation by core-accretion occurs frequently in our galaxy.

However, the bulk composition of a planet, accessed via M_p and R_p, is far from the only way to probe its formation. It has been increasingly recognised in recent years that the *atmospheric composition* of exoplanets provide crucial clues to their formation and evolution. A well-known example is that the carbon-to-oxygen elemental ratio, C/O, in a planetary atmosphere is less sensitive to chemical processes than the abundance of individual molecules, and therefore can reflect the C/O of the material from which the planet formed (e.g. Madhusudhan, 2012; Madhusudhan et al., 2016b). Crucially, the C/O ratio changes throughout a disk as snow lines are crossed due to the condensation of different volatile species (e.g. H_2O, CO_2, CO) at different temperatures (Öberg et al., 2011; Piso et al., 2016). The C/O ratio of an atmosphere today therefore provides a powerful diagnostic of the planet's formation location, its migration and accretion history, and the chemical evolution and physical structure of the disk from which it formed (e.g. Thiabaud et al., 2015; Madhusudhan et al., 2014c). Further considerations indicate that the time at which a planet formed can influence its final atmospheric composition (e.g. Cridland et al. (2017) found that early formation near the snow line resulted in more NH_3 than N_2 vs. late formation at the same location), and that solid planetesimals ablated in an atmosphere can enrich it in heavy-metals (Pinhas et al., 2016). In short, revealing the nature of exoplanetary atmospheres opens an unprecedented window into the formation and evolution of planetary systems. With this extraordinary promise in mind, I shall now elaborate on how one can uncover the composition of exoplanet atmospheres.

1.4 Observing exoplanetary atmospheres

In the decades since exoplanets were first detected, increasing efforts have been devoted towards the characterisation of these distant worlds. As we saw in section 1.3.2, the measured mass and radius of a planet inform our knowledge of its overall composition. However, from mass and radius alone a far away observer could infer that the Earth and Venus are similar planets, when in reality the surface of Venus is $\approx 450\,\mathrm{K}$ hotter than Earth, the pressure is $\approx 90\times$ greater, and suffers the minor inconvenience of H_2SO_4 on a rainy day. These stark differences largely arise from the contrasting nature of the Terran and Venusian atmospheres, highlighting that understanding the conditions on exoplanets is intrinsically linked to revealing the nature of their atmospheres.

The study of exoplanet atmospheres is a rapidly expanding frontier. The key additional dimension permitting the observational study of exoplanet atmospheres is variation of properties with *wavelength*. Spectroscopic observations of exoplanets can reveal signatures of their chemical composition, temperature structure, circulation patterns, clouds, geological activity, and, potentially, the presence of life. In this section, I review five techniques which have successfully revealed properties of exoplanet atmospheres. The first three techniques are mostly applicable to the atmospheres of transiting exoplanets, whilst the later two can be more generally applied. For each method, I briefly outline its underlying principles, the properties of exoplanetary atmospheres to which it is sensitive, and offer a few discovery highlights.

1.4.1 Transmission spectroscopy

Transmission spectroscopy (Seager & Sasselov, 2000; Brown, 2001) is the most widely applied method of studying exoplanet atmospheres. Consider a system with a transiting exoplanet, as illustrated in Figure 1.12. When the planet transits its host star, a small fraction of the light from its star will pass through the atmosphere of the planet (shown by the blue annulus). As photons transmit through the atmosphere, their interactions with atoms and molecules can cause them to change direction (scatter) or be absorbed. Where photons fail to reach an observer, the planet will appear to possess an enhanced radius. The fundamental insight here is that the processes governing the interaction of light with atoms and molecules are *wavelength dependent*, and hence the 'effective radius' of the planet will also vary with wavelength. By comparing wavelengths where the planet appears larger with known absorption features of various chemical species, the composition of the atmosphere can thereby be revealed.

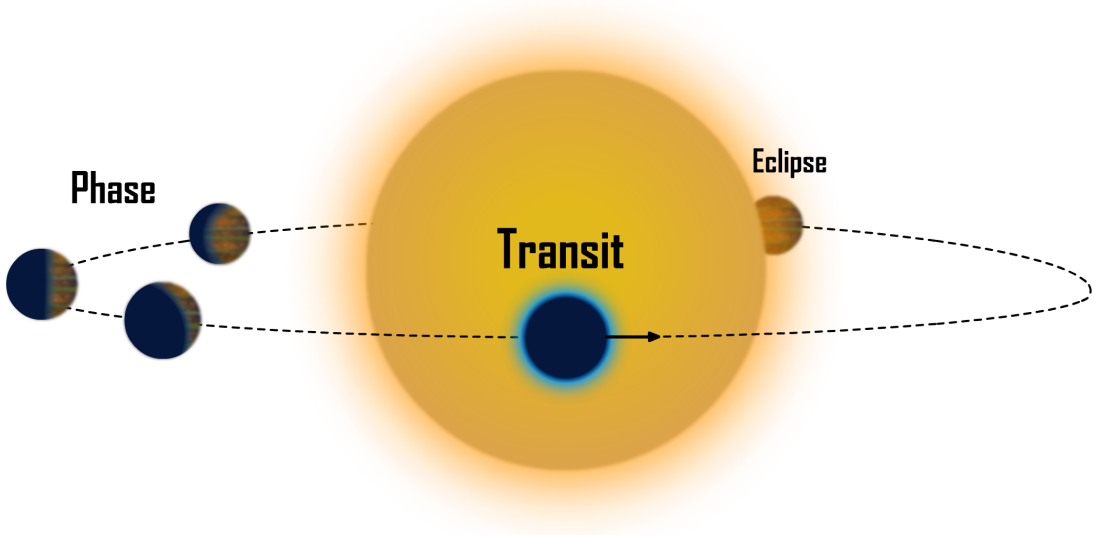


Fig. 1.12 Three methods to characterise exoplanet atmospheres.
An exoplanet orbit with $i \approx 90°$ exhibits transits when the planet passes in front of its host star and (secondary) eclipses when it passes behind. During transit, starlight passing through the planet's atmosphere accrue absorption features, producing a *transmission spectrum* diagnostic of conditions at the day-night terminator. Subtracting light observed during the eclipse (star only) from the total observed flux just before the eclipse (star and planet combined) reveals reflected light and thermal emission from the planet's dayside, encoded into an *occultation spectrum*. During other times, or phases, different hemispheres of the planet reflect and emit radiation. The star and planet combined flux at these different orbital phases constitute a *phase curve*.

To gain a rough indication of the size of this effect, consider a planet with radius[12] R_p and an atmosphere of height H_A. Now assume that at a particular wavelength the entire atmosphere becomes opaque, such that the effective projected area of the planet is $\pi(R_p + H_A)^2$. The transit depth in this absorption feature is then

$$\delta_{p+\mathrm{atm}} = \frac{F_{\mathrm{out}} - F_{\mathrm{in}}}{F_{\mathrm{out}}} = \frac{(R_p + H_A)^2}{R_*^2} \approx \left(\frac{R_p}{R_*}\right)^2 + \frac{2R_p\,H_A}{R_*^2} \tag{1.5}$$

The first term is simply the standard transit depth (Equation 1.2) that would be seen at wavelengths where the atmosphere is transparent. The second term encodes additional absorption due to the atmosphere (neglecting the small term H_A^2/R_*^2). To estimate

[12] For a gaseous planet without a solid surface, R_p is set sufficiently deep such that the planet is opaque at all wavelengths (typically at a pressure $\gtrsim 1$ bar).

the magnitude of this absorption, we need an estimate of H_A. For an isothermal atmosphere in hydrostatic equilibrium obeying the ideal gas law (see section 2.2.2), the density falls off as $\rho(r) \propto e^{-(r-R_p)/H}$, where H is the *scale height*, given by

$$H = \frac{k_B\, T}{\mu\, g} \tag{1.6}$$

Here, k_B is the Boltzmann constant, T the atmospheric temperature, μ the atmospheric mean molecular mass, and g the planet's surface gravity. Taking $H_A \approx 5\, H$ therefore reasonably approximates an opaque atmosphere (Seager, 2010). Choosing our planet to be representative of the prototypical hot Jupiter HD 209458b ($R_p = 1.36\, R_J$, $R_* = 1.16\, R_\odot$, $T \approx 1400\,\mathrm{K}$, $\mu = 2.3$ atomic mass units, and $g = 9.8\,\mathrm{ms}^{-2}$ - example values from Torres et al. (2008) and Guillot (2010)) yields $H \approx 550\,\mathrm{km}$ and hence $\delta_{p\,+\,\mathrm{atm}} = (1.45\% + 0.08\%)$ for the first and second terms in Equation 1.5, respectively. We thus see that even in this idealised case the contribution of an atmosphere to the transit depth is $< 0.1\%$.

In reality, the entire atmosphere will rarely be completely opaque at a given wavelength. It is also worth noting that the orbital geometry of an exoplanet transit dictates that the observed rays pass through the boundary between the dayside and nightside of the planet (see Figure 1.12). For close-in tidally locked planets, such as hot Jupiters, the region of the atmosphere being probed is therefore the *terminator*. These considerations will be addressed in much greater detail in Chapter 2, where the calculation of theoretical spectra for realistic atmospheres will be explained. For now, one can conceptually note that, as different atoms and molecules have different intrinsic absorption strengths, we can imagine replacing $H_A \rightarrow N_\lambda\, H$, where N_λ is the number of scale heights above R_p for which the planet is effectively opaque at wavelength λ (determined by the composition of the atmosphere). Substituting this into Equation 1.5, we thereby obtain a function δ_λ, referred to as a *transmission spectrum*.

The appearance of a transmission spectrum is schematically illustrated in Figure 1.13 for a typical hot Jupiter. At short optical wavelengths, the transit depth exhibits a 'slope' due to scattering out of the observer's line of sight becoming more efficient at shorter wavelengths. This slope arises from H_2 Rayleigh scattering in clear hydrogen dominated atmospheres, though small particle 'hazes' can result in a steeper slope if they scatter more efficiently (e.g. Sing et al., 2016; Pinhas & Madhusudhan, 2017). At wavelengths longer than $\sim 0.5\,\mu\mathrm{m}$, strong absorption from the resonance lines of atomic Na and K is anticipated for planets with $T \gtrsim 800$ K (Burrows et al., 2000; Sudarsky et al.,

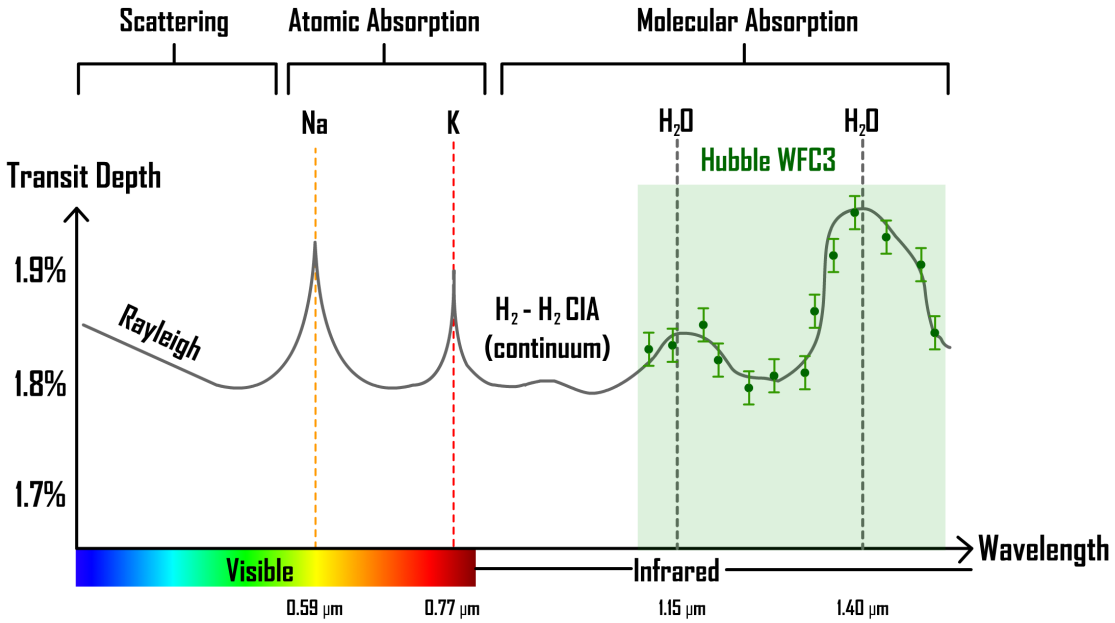


Fig. 1.13 Schematic illustration of a hot Jupiter transmission spectrum. Variation of the transit depth with wavelength for a typical hot Jupiter atmosphere is sketched by the grey curve. From left to right, sources of opacity are provided by: H_2 Rayleigh scattering, alkali atomic absorption, H_2-H_2 collision-induced absorption (CIA), and molecular absorption. Wavelengths with prominent absorption features are indicated. The green shaded region and data points illustrate the wavelength range probed by the Hubble Space Telescope's Wide Field Camera 3 (WFC3).

2000). Absorption features due to molecules, such as H_2O, CH_4, and CO_2, tend to be more prominent in the infrared. In relatively transparent wavelength regions with little scattering or atomic / molecular absorption, collision-induced absorption effectively provides the floor (or continuum) to the spectrum, limiting the deepest altitude which can be probed (Brown, 2001). Relatively flat floors, cutting off spectral features, can also be provided by high-altitude clouds (e.g. Ehrenreich et al., 2014; Kreidberg et al., 2014a) or refractive surfaces (Bétrémieux & Kaltenegger, 2014; Robinson et al., 2017). Finally, I note that the amplitude of each absorption feature scales with both the abundance of the responsible chemical species and the atmospheric scale height. From Equation 1.6, we thus see that the optimum atmospheres for transmission spectroscopy have high-temperatures, low gravity, and a low mean molecular weight (i.e. typically H_2 dominated). This explains why hot Jupiters, with $T \gtrsim 1000\,\mathrm{K}$, have proven especially amendable to transmission spectroscopy.

Highlights from transmission spectroscopy

Transmission spectroscopy was the first method to successfully detect an exoplanet atmosphere. Charbonneau et al. (2002) reported a detection of Na absorption at 589 nm during a transit of HD 209458b observed by the Hubble Space Telescope (HST) (later independently verified by Snellen et al., 2008; Sing et al., 2008). Shortly thereafter, UV transmission spectra of the same planet with HST yielded detections of atomic H, O, and C (Vidal-Madjar et al., 2003a,b), indicating an extended escaping atmosphere. In a similar manner, escaping He has recently been detected (Spake et al., 2018). Ground-based transmission spectra are also now yielding detections of Na and K (Redfield et al., 2008; Sing et al., 2011; Wilson et al., 2015). Molecular absorption due to the 1.4 µm H_2O feature has been robustly detected by HST Wide Field Camera 3 (WFC3) spectroscopy (illustrated in Figure 1.13) for $>$ 10 hot Jupiters (e.g. Deming et al., 2013; Huitson et al., 2013; Mandell et al., 2013; Kreidberg et al., 2014b; Sing et al., 2016; Tsiaras et al., 2018) and two Neptune-mass exoplanets (Fraine et al., 2014; Wakeford et al., 2017). Finally, signatures of scattering and flat transmission spectra have indicated that clouds and hazes are a common feature in many exoplanet atmospheres (Lecavelier Des Etangs et al., 2008; Pont et al., 2013; Kreidberg et al., 2014a; Knutson et al., 2014a; Ehrenreich et al., 2014; Sing et al., 2016).

1.4.2 Occultation spectroscopy

Occultation spectroscopy (e.g. Seager & Sasselov, 2000; Sudarsky et al., 2003) allows one to probe the conditions in the dayside atmosphere of an exoplanet. Similar to transmission spectroscopy, occultation spectroscopy involves a comparison of the total flux received from a system at two different points in time: (i) shortly before the planet passes behind its star, and (ii) during the (secondary) eclipse (see Figure 1.12). Prior to the eclipse, the total flux is a combination of the stellar flux along with reflected light and thermal emission from the planetary dayside. During the eclipse, the contributions from the planet vanish, leaving only that from the star.

The fractional flux decrement, or *eclipse depth*, can be expressed as

$$\delta_{\mathrm{ec.},\lambda} = \frac{F_{\mathrm{out}} - F_{\mathrm{in}}}{F_{\mathrm{out}}} = \frac{(F_p + F_*) - F_*}{F_p + F_*} \approx \frac{F_p}{F_*} \tag{1.7}$$

where it is assumed that $F_* \gg F_p$ and the flux wavelength dependence is implicit. Considering first optical wavelengths ($\lambda \lesssim 1\,\mu\mathrm{m}$), the dominant contribution to F_p is

reflected or scattered light from the planetary dayside. In this limit, the flux ratio is

$$\frac{F_p}{F_*} \approx A_{g,\lambda} \left(\frac{R_p}{a} \right)^2 \tag{1.8}$$

where $A_{g,\lambda}$ is the planetary geometric albedo[13] at wavelength λ and it has been assumed that the planet is fully illuminated prior to the eclipse. The variable reflective ability of the atmosphere is encoded by $A_{g,\lambda}$ (also known as a *reflection spectrum*), which contains a wealth of information about the composition, temperature, and cloud properties of the atmosphere. For a hot Jupiter eclipse, taking $A_{g,\lambda} \approx 0.3$ (Brown & Burrows, 1990), $R_p = 1.4\,R_J$, and $a = 0.05\,\mathrm{au}$ yields $\delta_{\mathrm{ec.,\,optical}} \approx 5 \times 10^{-5}$. Though small, such optical eclipse depths have been successfully measured by Kepler (e.g. Esteves et al., 2013; Huber et al., 2017). Moving now to longer wavelengths ($\lambda \gtrsim 1\,\mu\mathrm{m}$), the flux ratio becomes increasingly dominated by thermal emission, such that

$$\frac{F_p}{F_*} \approx \frac{B_\lambda\left(T_p\right)}{B_\lambda\left(T_*\right)} \left(\frac{R_p}{R_*} \right)^2 \tag{1.9}$$

where $B_\lambda\left(T\right)$ is the spectral radiance of a black body at temperature T. Indeed, we have already encountered this equation in the context of directly imaged planets in section 1.2.4 (Similarly, Equation 1.8, when multiplied by a phase function to account for partial illumination, also holds for directly imaged planets). At the elevated temperatures of hot Jupiters, quite favourable infrared flux ratios can result. Taking $R_p = 1.4\,R_J$, $R_* = R_\odot$, $T_p = 1400\,\mathrm{K}$, and $T_* = T_\odot$ yields $\delta_{\mathrm{ec.,\,infrared}} \approx 10^{-3}$ at $3.6\,\mu\mathrm{m}$ (the wavelength of one of the Spitzer IRAC bands). This ratio only improves for longer wavelengths, due to the stellar flux (peaking in the optical) declining more rapidly relative to that from the planet (peaking in the infrared). We thus see that infrared eclipse depths are $\sim 10\times$ smaller than transit depths ($\approx 1\%$). Nevertheless, eclipse depths contain highly complementary information about exoplanetary atmospheres.

Infrared eclipse depths are more commonly known as *emission spectra*. Unlike transmission spectra, where attenuated stellar photons are observed, emission spectra directly sample thermal photons emitted from the planet. Equation 1.9 is a rough approximation to emission spectra (see e.g. Gandhi & Madhusudhan, 2017, for an extensive treatment), but the way these spectra encode atmospheric properties can

[13] Historically defined as the ratio of the scattered flux from the fully-illuminated planet (as seen from the star) to that of an isotropically scattering (Lambertian) disk with the same projected area placed at the same distance (Seager, 2010).

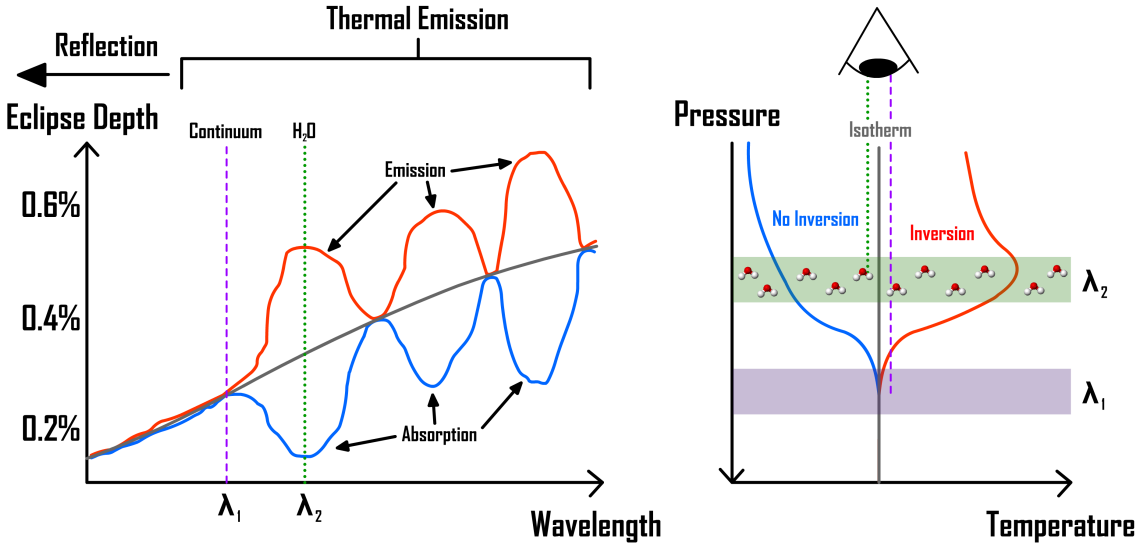


Fig. 1.14 Schematic illustration of a hot Jupiter emission spectrum. Left: eclipse depth wavelength-variation for a hot Jupiter with three possible temperature structures: isothermal (grey), thermally inverted (red), and no inversion (blue). Thermal emission dominates over reflected light across the shown wavelength range ($\lambda \gtrsim 1\,\mu$m). Right: the three corresponding pressure-temperature profiles. Formation of absorption or emission features is illustrated via two reference wavelengths. At λ_1 (purple dashed line), where H_2O has low opacity, sight lines penetrate into the deep atmosphere where all the models have identical temperatures (and hence fluxes). At λ_2 (green dotted line), where H_2O has high opacity, sight lines only reach upper layers, where the colder profile (blue) emits less flux, hence forming an absorption feature, whilst the hotter profile (red) emits more flux, thereby forming an emission feature.

be conceptually understood by a careful interpretation of T_p. In eclipse geometry, the atmosphere is observed face-on. At wavelengths where the atmosphere is relatively transparent, the emergent flux arises from the deep atmosphere at temperature $T_{\rm deep}$. However, at wavelengths where a molecule in the atmosphere strongly absorbs, only flux from the upper atmosphere, with temperature $T_{\rm up}$, reaches the observer. If $T_{\rm up} < T_{\rm deep}$, the resulting drop in flux creates an absorption feature, whilst if $T_{\rm up} > T_{\rm deep}$ (a *thermal inversion*), the increased flux creates an *emission* feature. In the case where $T_{\rm up} = T_{\rm deep}$ (an isotherm), all layers radiate with the same flux and no molecular features are imprinted into the spectrum. T_p can thus be interpreted as the (wavelength-dependent) temperature of the planetary photosphere. The link between temperature structure and molecular features in emission spectra is illustrated in Figure 1.14. The key takeaway is that emission spectra of exoplanets are sensitive to *both* the composition and the shape of the temperature profile in their dayside atmosphere.

Highlights from emission spectroscopy

The first exoplanet secondary eclipses were detected with Spitzer for the hot Jupiters HD 209458b (Deming et al., 2005) and TrES-1b (Charbonneau et al., 2005). Spitzer is capable of observing secondary eclipses at 3.6 µm and 4.5 µm[14], providing favourable infrared contrasts. Spitzer has also detected eclipses for exo-Neptunes (Deming et al., 2007) and super-Earths (Demory et al., 2012). Ground-based eclipse detections (such as in the K-band at ∼2 µm) soon followed (e.g. Sing & López-Morales, 2009; Croll et al., 2010). The first multi-wavelength eclipse depth measurements (i.e. emission spectra) were obtained from Spitzer IRAC photometry / IRS spectroscopy for the hot Jupiters HD 209458b (Knutson et al., 2008) and HD 189733b (Charbonneau et al., 2008; Grillmair et al., 2008). HST WFC3 observations now offer complementary near-infrared emission spectra displaying significant H_2O absorption features (e.g. Kreidberg et al., 2014b; Line et al., 2016). Evidence for thermal inversions, despite early controversies (Knutson et al., 2008; Diamond-Lowe et al., 2014; Line et al., 2016), rest on more solid ground following a recent 5σ detection of H_2O emission from the hot Jupiter WASP-121b (Evans et al., 2017). Thermal inversions in hot Jupiter atmospheres are potential indicators of strong optically absorbing molecules, such as TiO or VO (Hubeny et al., 2003; Fortney et al., 2008). Most recently, emission spectra of 'ultra-hot' Jupiters ($T_p \gtrsim 2500$ K) are providing evidence for H- opacity and H_2O dissociation (Arcangeli et al., 2018; Kreidberg et al., 2018; Parmentier et al., 2018).

1.4.3 Phase curves

Exoplanet atmospheres are inherently three-dimensional entities. Close-in planets are subject to significant tidal forces from their star, readily capable of synchronising their rotational and orbital periods in much less than the age of a system. The expectation is that exoplanets with $P \lesssim 10$ day will be tidally locked (Guillot et al., 1996), such that, in the planet's frame of reference, there exists a permanent dayside and nightside with a terminator serving as their boundary. The differing stellar insolation received by different regions of the atmosphere leads to the expectation of large temperature gradients around the planet, in turn driving non-uniform chemistry, clouds, and circulation patterns (e.g. Kataria et al., 2016). From the perspective of a distant observer, an exoplanet presents different hemispheres as it rotates about its axis, with the nightside seen during transit and the dayside seen just before and after eclipse (see

[14]Prior to depletion of its liquid He coolant in 2009, Spitzer could reach up to 24 µm.

Figure 1.12). From this geometry, we have seen in the last two sections two methods to probe specific regions of the atmosphere: (i) the terminator (via transmission spectra), and (ii) the dayside (via occultation spectra). Here, I will summarise the powerful method of exoplanet *phase curves*, which can longitudinally map properties across the entire surface of an exoplanet atmosphere.

Phase curves refer to variations in the observed brightness of a system as a function of the orbital phase of an exoplanet. These variations are a result of different hemispheres coming into view as the planet rotates. For tidally locked planets in a circular orbit, there is a one-to-one correspondence between orbital phase and the hemisphere centred on a particular planetary longitude. Different hemispheres modulate the total flux received from the system due to their differing albedo (at optical wavelengths) and thermal emission (at infrared wavelengths). These flux modulations enable longitudinal properties of an exoplanet atmosphere to be measured. At optical wavelengths, phase curves can show brightness variations due to reflective clouds, enabling one to identify longitudes with active cloud formation (e.g. Demory et al., 2013; Esteves et al., 2013). Infrared phase curves exhibit variable brightness due to higher thermal emission when hotter hemispheres rotate into view, enabling the construction of coarse longitudinal temperature maps (Knutson et al., 2007; Cowan & Agol, 2008). By combining cloud and temperature maps (from both optical and infrared phase curves) with knowledge of the condensation temperatures of different chemical species (e.g. Lee et al., 2016), it is also possible to constrain the chemical composition of exoplanet clouds (Parmentier et al., 2016; Oreshenko et al., 2016).

An infrared phase curve of the hot Jupiter HD 189733b is shown in Figure 1.15 (left). Phase curves have three observable features (after Parmentier & Crossfield, 2018). First, the eclipse depth provides a reference flux (as only the star is visible during eclipse), with flux variations above this level attributable to the planet. Secondly, the flux difference between the minimum and maximum of the curve traces the temperature difference between the coolest and hottest atmospheric regions. Thirdly, any phase offset between the flux maximum and eclipse (phase = 0.5) provides a measure of the longitude of highest temperature. Similar phenomenology holds for optical phase curves in terms of reflected flux maxima, minima, and offsets. I further note that phase curves can also be observed for non-transiting planets (e.g. Crossfield et al., 2010), though the lack of a baseline eclipse prohibits measurements of a planet's temperature contrast. However, such phase curves can serve as orbital brightness variations to enable the detection of non-transiting exoplanets (Millholland & Laughlin, 2017).

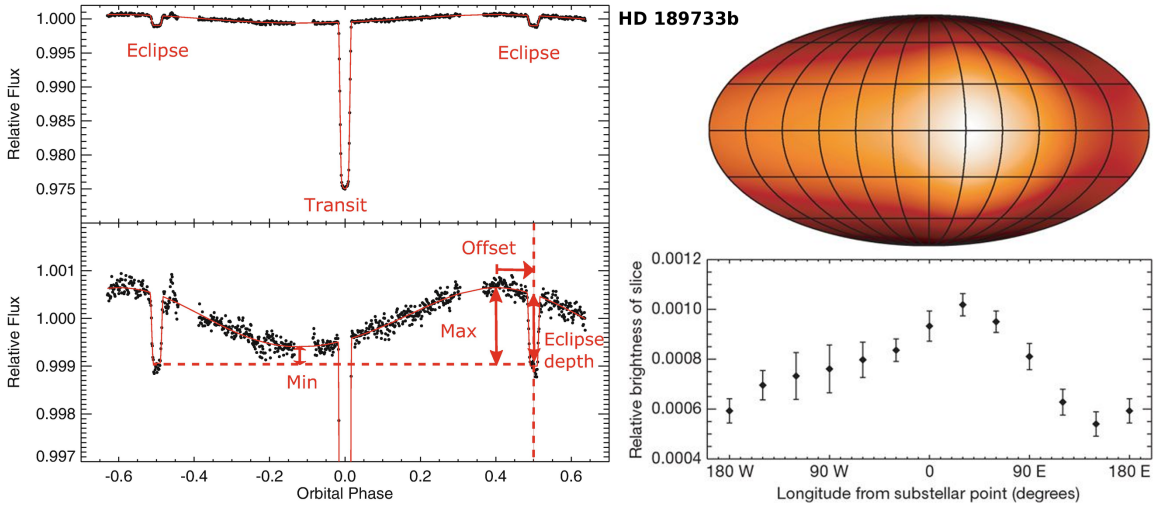


Fig. 1.15 Thermally mapping an exoplanet atmosphere using phase curves. Left: infrared phase curve of the hot Jupiter HD 189733b observed by Spitzer at 3.6 µm (Knutson et al., 2012). The top panel shows the full light curve, with the bottom panel zoomed in to render clear thermal flux variations due to the changing face of the planet. Key observable features of the phase curve, annotated by Parmentier & Crossfield (2018), are: the minimum and maximum planetary flux, the eclipse depth, and the phase offset between the maximum and secondary eclipse. Right: longitudinal temperature map and relative brightness of HD 189733b obtained by inversion of a 8 µm Spitzer phase curve (Knutson et al., 2007). A clear eastwards offset is seen, indicating atmospheric energy transport away from the substellar point.

Observations of phase curve amplitudes and offsets offer valuable insights into exoplanet atmospheric dynamics. In the absence of winds, the hottest location should be at the substellar point (centre of the dayside), implying no offset between the infrared flux maximum and secondary eclipse. However, atmospheric circulation models predict fluid motions caused by a super-rotating equatorial jet advect energy to eastern longitudes (e.g. Showman & Guillot, 2002; Cho et al., 2003). Transported air eventually cools via radiative heat loss, such that competition between advection and radiative cooling sets the longitude of the hottest point in the atmosphere (Showman & Polvani, 2011; Komacek & Showman, 2016). The further east the hotspot lies, the earlier before secondary eclipse the flux maximum occurs, causing an eastwards phase offset. This situation is seen in Figure 1.15 (right), where a temperature map of HD 189733b displays an eastwards offset of $\sim 30°$ (Knutson et al., 2007). Similarly, cooler air transported from the nightside can promote the formation of reflective clouds near the western edge of the dayside, leading to optical phase curves with (westwards) offsets observed after the eclipse (e.g. Demory et al., 2013; Shporer & Hu, 2015).

If one can observe phase curves at multiple wavelengths, as opposed to broadband measurements thus far discussed, the technique of *phase curve spectroscopy* can be harnessed. A spectroscopic phase curve can yield atmospheric emission or reflection spectra from *multiple* hemispheres across the planet, producing spectra as a function of longitude (Knutson et al., 2009, 2012). As we saw in section 1.4.2, emission spectra probe different altitudes in the dayside atmosphere at different wavelengths (higher in the atmosphere / lower pressures inside absorption features, deeper / higher pressures outside). This implies that spectroscopic phase curves can be inverted to yield 2-dimensional maps (in longitude and pressure) of properties in an exoplanet atmosphere (Showman et al., 2009). Wavelength-dependent phase curves can thus serve as a probe of different circulation patterns at different pressures, as revealed by wavelength-dependent phase offsets (Knutson et al., 2009; Stevenson et al., 2014).

Before moving on, I briefly note the main challenges hindering the wider applicability of this method. First, the time required to measure a full-orbit phase curve necessarily exceeds that of a transit or eclipse ($\sim$ days vs. hours for a typical hot Jupiter). Spectroscopic phase curves observed with Hubble can also suffer from coverage gaps due to obscuration of the target by the Earth. Furthermore, multiple phase curves may be required, in practice, to assess the reliability of the observations (Stevenson et al., 2014, 2017). One must also contend with the need to disentangle the contributions of thermal emission and reflected radiation at wavelengths where they overlap, along with higher-order effects such as relativistic beaming and stellar ellipsoidal variations (see Parmentier & Crossfield, 2018, for a comprehensive discussion). Nevertheless, despite these complications, phase curves have yielded some fascinating and unique results.

Highlights from phase curves

The first coarsely sampled phase curve was obtained by Harrington et al. (2006) using 24 µm Spitzer photometry. This was shortly followed by an 8 µm Spitzer phase curve continuously sampled between transit and eclipse (Knutson et al., 2007). Inversion of this phase curve provided the first temperature map of a hot Jupiter (HD 189733b), and revealed an eastwards hotspot (see Figure 1.15). A multi-wavelength infrared phase curve (in four Spitzer bands) was reported by Knutson et al. (2012), again for HD 189733b, revealing an $\approx 200-500$ K day-night temperature contrast and suggestions of disequilibrium carbon chemistry. Optical phase curves were first obtained with the CoRoT space telescope (Snellen et al., 2009). Kepler has since provided many precise optical phase curves, offering detections of westwards phase offsets attributable to

inhomogeneous ('patchy') cloud coverage (Demory et al., 2013; Shporer & Hu, 2015). The first full-orbit spectroscopic phase curve was presented by Stevenson et al. (2014, 2017), who used HST WFC3, and later Spitzer, to suggest high-altitude nightside clouds for the hot Jupiter WASP-43b. Phase curves have recently been extended to lower mass exoplanets, with the first temperature map of a super-Earth, 55 Cancri e (Demory et al., 2016), revealing a $\sim 1300\,\mathrm{K}$ day-night temperature contrast - possibly due to magma flows. Phase curves have also been obtained for ultra-hot Jupiters (Armstrong et al., 2016; Kreidberg et al., 2018), with long-term Kepler monitoring of HAT-P-7b's optical phase curve revealing *temporal variability* in the form of a phase offset oscillating between west and east over ~ 50 orbital periods - suggestive of cloud motion (i.e. 'exoweather', see Armstrong et al., 2016) or magnetic effects (Rogers, 2017). Finally, despite most infrared phase curves displaying eastwards hotspots (Parmentier & Crossfield, 2018), a westwards infrared phase offset has recently emerged (Dang et al., 2018) - perhaps indicating non-synchronous rotation, magnetic effects, or eastern clouds. The wide variety of phenomena inferred, with only ≈ 20 phase curves observed to date, suggests many more surprises await.

1.4.4 High-resolution Doppler spectroscopy

The three techniques for observing exoplanet atmospheres discussed so far typically yield low-resolution ($R = \lambda/d\lambda \lesssim 100$) spectra. Furthermore, with the exception of phase curves, their applicability is restricted to transiting planets. However, as we saw in section 1.2.2, the probability of a transit tends to be $\lesssim$ 1-10%. Statistically speaking, any method relying on transits to observe exoplanet atmospheres is therefore necessarily limited to a small subset of the exoplanet population. Here, and in the following subsection, I will cover two methods for observing the atmospheres of *non-transiting* exoplanets. I begin with ground-based high-resolution spectroscopy.

High-resolution Doppler spectroscopy (Snellen et al., 2010; Birkby, 2018) is an extension of the radial velocity method (see section 1.2.1). The central concept is that radiation absorbed by, reflected off, or thermally emitted from a planet's atmosphere is Doppler shifted over the course of its orbit. For a single-planet system, the spectrum of the planet follows a radial velocity curve with a semi-amplitude given by

$$K_p = \frac{M_* \sin i}{(M_p + M_*)^{2/3}} \left(\frac{2\pi G}{P}\right)^{1/3} \frac{1}{\sqrt{1-e^2}} \tag{1.10}$$

Comparing this expression with the corresponding one for K_* (Equation 1.1), the exoplanet radial velocity is higher than its star by a factor of M_*/M_p. Radiation from exoplanets therefore experience Doppler shifts $\gtrsim 100\times$ that of their star. Substituting typical hot Jupiter values ($M_* = M_\odot$, $M_p = M_J$, $P = 4\,\mathrm{day}$, and $\sin i \approx 1$) into Equation 1.10 yields $K_p \approx 130\,\mathrm{km\,s^{-1}}$, corresponding to a maximum redshift of $z = K_p/c \approx 5 \times 10^{-4}$. An absorption line in the spectrum of a hot Jupiter at rest-wavelength $\lambda_{\mathrm{em}} = 2.310\,\mu\mathrm{m}$ then varies from $\lambda_{\mathrm{obs}} \approx 2.309$–$2.311\,\mu\mathrm{m}$ over the course of its orbit. Such shifts are too small to register at the low-resolutions of spectroscopic phase curves ($R \approx 40$ for HST WFC3 spectra, e.g. Stevenson et al., 2014). However, large ground-based telescopes can utilise high-resolution ($R \gtrsim 25{,}000$) spectrographs to resolve molecular bands into individual transitions. As we shall see, the key to this technique is the time-varying nature of molecular lines in exoplanet spectra.

Figure 1.16 (upper left) shows a simulation of thermal emission from an exoplanet atmosphere as a function of orbital phase at high-spectral resolution. Absorption lines undergo sinusoidal Doppler variations tracing the planet's radial velocity curve. The maximum redshift occurs at $\phi = 0.25$, when the planet's velocity is maximally-aligned away from the observer, whilst the maximum blueshift occurs half an orbit later at $\phi = 0.75$. The greatest rate of change of redshift occurs at superior ($\phi = 0.5$) and inferior ($\phi = 0.0, 1.0$) conjunction - for transiting exoplanet systems, $\phi = 0.5$ corresponds to the eclipse and $\phi = 0.0, 1.0$ to the transit. The line intensity varies throughout the orbit, according to the brightness temperature of the hemisphere facing the observer. For tidally locked planets, the line intensity is greatest around superior conjunction (see section 1.4.3). Absorption lines due to Earth's atmosphere ('tellurics') occur at constant wavelengths as a function of the exoplanet's orbital phase, and stellar lines appear relatively stationary ($K_* \ll K_p$), such that the time-varying nature of planetary absorption lines enables them to be disentangled from the dominant flux of the star. Indeed, one would never observe a clean signal like shown in Figure 1.16 (upper left), even after correcting for Earth's atmosphere, as small planet-star contrasts ($F_p/F_* < 10^{-3}$) leave signatures of the planet hidden in the noise.

High-resolution spectroscopy relies on model spectra to extract the signal of an exoplanet atmosphere from the overwhelming stellar background. This is unlike methods like transmission spectroscopy, where a spectrum is first observed and later compared with models (see section 1.5). Figure 1.16 (lower left) shows a high-resolution model emission spectrum in the rest-frame of a hot Jupiter (de Kok et al., 2013). At this spectral resolution, individual CO absorption lines are clearly resolved. This

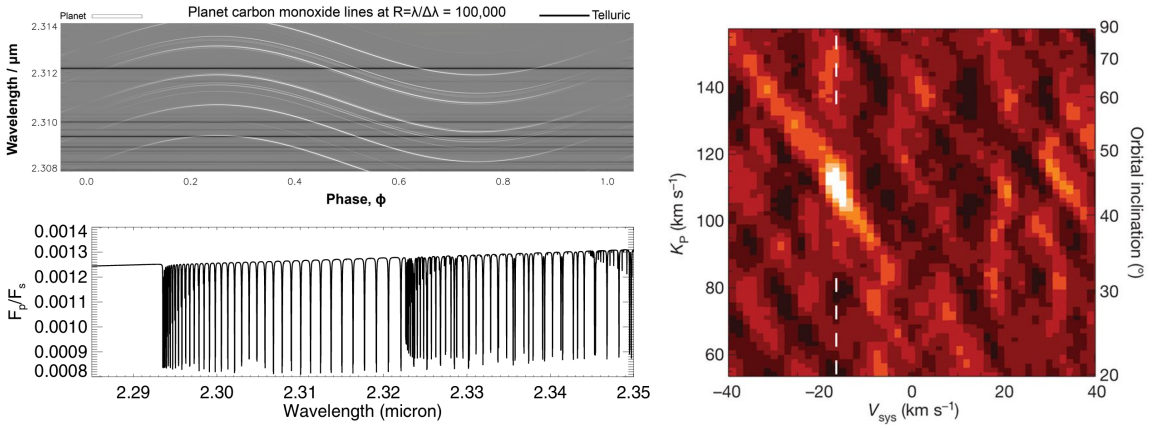


Fig. 1.16 Exoplanet atmospheres via high-resolution spectroscopy. Upper left: simulated Doppler shifting CO lines at high spectral resolution from a hot Jupiter atmosphere (Birkby, 2018). The thermal line intensity (white) is greatest around superior conjunction ($\phi = 0.5$), when the planet is maximally-illuminated, and minimal at inferior conjunction ($\phi = 0.0, 1.0$). Horizontal black bands are absorption from Earth's atmosphere ('tellurics'). Lower left: 'template' emission spectrum of a hot Jupiter atmosphere showing rest-frame CO absorption lines (de Kok et al., 2013). Right: atmospheric detection of the non-transiting exoplanet τ Boötis b, accomplished via cross-correlation of a CO template model with high-resolution CRIRES spectra (Brogi et al., 2012). A $> 6\sigma$ detection of CO occurs at the known line-of-sight velocity of the system (white dashed line) and a K_p corresponding to $i \approx 45°$.

template, and other competing models, can then be *cross-correlated* (i.e. convolved) with the observed spectrum of the system, resulting in a cross correlation function that peaks for velocities where a model line matches a line in the observations[15]. A recent overview of the cross correlation method is provided by Birkby (2018), but here I will summarise the salient points. The degree of alignment between a template and observations increases with the number of matching lines as $\sqrt{N_{\rm lines}}$ (Snellen et al., 2015), such that cross correlation functions combine many individual lines to extract the planet signal from the background. By summing the cross correlation function over all orbital phases for a grid of K_p values, a significance plot, such as that shown in Figure 1.16 (right), is produced. Where a high-signal peak is observed at the known system velocity, the motion of the planet (i.e. a determination of K_p) has been detected. Furthermore, a detection of the planet relying on a model containing a particular molecule (here, CO) serves as a detection of that molecule in the atmosphere.

[15]This can be visualised as follows: rotate the model in the lower left of Figure 1.16 such that λ is on the y-axis and F_p/F_* is into the page. Then slide the 'comb' of lines vertically through the 'observed' spectrum in the upper left of Figure 1.16 until a velocity is found where the lines coincide.

Atmospheric detections via high-resolution spectroscopy can further constrain exoplanet properties. First, recall from section 1.2.1 that the $\sin i$ ambiguity for non-transiting planets only yields a minimum M_p from the stellar radial velocity. However, with a determination of K_p from the Doppler shifted planet spectrum, $\sin i$ cancels in the ratio of Equations 1.10 and 1.1, such that $K_p/K_* = M_*/M_p$ yields the absolute planet mass (typically at 5% precision, see de Kok et al., 2013). Substitution of M_p and K_p into Equation 1.10 then determines i. This is illustrated for the non-transiting exoplanet τ Boötis b in Figure 1.16 (right). Secondly, the planetary rotation speed can be inferred. Due to the differing line-of-sight velocity across a rotating visible hemisphere, the cross correlation function is broadened (Snellen et al., 2014; Brogi et al., 2016). High-resolution spectroscopy can thereby directly measure rotation speeds without the assumption of synchronous rotation (Rauscher & Kempton, 2014).

High-resolution spectroscopy also offers some advantages over low-resolution spectroscopy. First, every molecule has a unique sequence of absorption lines at high-resolution, enabling their unambiguous detection (typically at $\gtrsim 5\sigma$). Shifts in line positions due to mass-differences between isotopes can even be detected (Mollière & Snellen, 2019). Secondly, additional Doppler shifts (beyond those from the system velocity, the barycentric motion of Earth, and the planet's orbit and rotation) caused by $\sim \mathrm{km\,s^{-1}}$ atmospheric winds can be detected (Snellen et al., 2010; Miller-Ricci Kempton & Rauscher, 2012; Zhang et al., 2017). Such wind speed measurements provide information on atmospheric circulation complementary to phase curve offsets (section 1.4.3). Thirdly, lines appear sharper at high-resolution, potentially opening the atmospheres of cloudy planets with flat (low-resolution) transmission spectra to study (de Kok et al., 2013; Pino et al., 2018). Finally, lines appearing in emission or absorption serve as a sensitive test for thermal inversions (Schwarz et al., 2015).

Nevertheless, this technique does face some drawbacks. Ground-based observations operate at wavelengths where Earth's atmosphere is relatively transparent, but must still correct for telluric absorption (Brogi et al., 2012; de Kok et al., 2013). Data reduction procedures also tend to lose the continuum level of the planetary spectrum (e.g. Birkby et al., 2013; Piskorz et al., 2016). Without a continuum, abundance constraints for detected molecules tend to be weak (Birkby, 2018). Attempts to alleviate this difficulty by combining high-resolution and low-resolution spectroscopy are ongoing (e.g. de Kok et al., 2014; Brogi et al., 2017; Pino et al., 2018; Brogi & Line, 2019). Finally, the success of this method crucially hinges on accurate theoretical models, especially line positions (Hoeijmakers et al., 2015; Schwartz & Cowan, 2015).

Highlights from high-resolution spectroscopy

The first detection of an exoplanet atmosphere using high-resolution Doppler spectroscopy came from a Very Large Telescope (VLT) CRIRES detection of CO absorption during a transit of the hot Jupiter HD 209458b (Snellen et al., 2010). This detection served simultaneously as the first robust ($> 5\sigma$) detection[16] of CO and measurement of wind speeds ($2 \pm 1\,\mathrm{km\,s^{-1}}$) in an exoplanet atmosphere. The first detection of an atmosphere around a non-transiting hot Jupiter was reported by Brogi et al. (2012), who detected CO absorption near the eclipse of τ Boötis b (see Figure 1.16, right). Detections in hot Jupiter atmospheres soon expanded to include H_2O (e.g. Birkby et al., 2013; Lockwood et al., 2014; Piskorz et al., 2017). An exoplanet's rotation period was first determined by Snellen et al. (2014), who observed broadened CO lines in the emission spectrum of the directly imaged planet β Pictoris b (Snellen et al., 2014)[17]. The assumption of tidal locking for hot Jupiters can be directly tested via their rotation period (e.g. Brogi et al., 2016, measured the rotation period of HD 189733b to be consistent with its orbital period). Recently, high-resolution spectroscopy has detected several chemical species beyond CO and H_2O, namely: TiO (Nugroho et al., 2017) and HCN (Hawker et al., 2018; Cabot et al., 2019) in hot Jupiters, along with Fe, Fe+, and Ti+ in an ultra-hot Jupiter (Hoeijmakers et al., 2018a).

1.4.5 Direct imaging spectroscopy

The four techniques for observing exoplanet atmospheres discussed so far focus primarily on close-in exoplanets, in particular hot Jupiters. However, highly irradiated giant planets represent only part of the exoplanet population (see section 1.3.1). Direct imaging, on the other hand, is uniquely sensitive to wide-orbit ($a \gtrsim 10\,\mathrm{au}$), young ($\lesssim 100\,\mathrm{Myr} \rightarrow T_p \sim 1000\,\mathrm{K}$), giant exoplanets (see section 1.2.4). For directly imaged planets, the spatial separation between the planetary and stellar fluxes allows one to obtain a direct spectrum of the planet. This technique, *direct imaging spectroscopy*, allows atmospheric characterisation for an entirely distinct class of planets.

Modern $\gtrsim 8\,\mathrm{m}$ ground-based telescopes (e.g. Keck, VLT, Gemini South) are capable of measuring direct thermal emission spectra of exoplanets with wide-orbits around nearby ($< 100\,\mathrm{pc}$) stars. These facilities use integral field spectographs (e.g. OSIRIS, SPHERE, GPI), often in tandem with adaptive optics and a coronagraph, to image

[16] Space-based low-R spectra do not currently cover wavelengths with strong CO absorption bands.

[17] With $a \approx 10\,\mathrm{au}$, β Pictoris b is not tidally locked. Its rotation period was thus *a priori* unknown.

exoplanets at multiple wavelengths. This tends to yield low-resolution ($R \sim 30$) spectra (e.g. Macintosh et al., 2015; Zurlo et al., 2016), though moderate resolution spectra ($R \sim 4000$) have been reported for especially bright planets sufficiently far from their star (Konopacky et al., 2013; Barman et al., 2015). Example direct spectra of the four planets in the well-studied HR 8799 system are shown in Figure 1.17.

Relating observed spectra of a directly imaged exoplanet to their underlying atmospheric properties can be challenging. Unlike transiting planets, for which R_p and often M_p are known, directly imaged exoplanets have *a priori* unknown physical properties. Atmospheric characterisation must then proceed with their spectra as the only observational constraint, with (model-dependent) masses and radii inferred during modelling. Traditionally, grids of model atmospheres spanning different properties (e.g. $T_{\rm eff}$, R_p, M_p, compositions, clouds, etc.) are compared to a spectrum to assess the combination(s) which best fit the observations (e.g. Konopacky et al., 2013; Macintosh et al., 2015). These grid (or 'forward model') approaches often draw analogies between the spectral types of brown dwarfs. More recently, inverse (or 'retrieval') techniques have been developed to explore the vast parameter space of possible atmospheric models (Lee et al., 2013; Line et al., 2015; Lavie et al., 2017). Retrieval methods utilise Bayesian frameworks to automatically fold the inherent uncertainty from the unknown values of R_p and M_p into reported constraints on atmospheric properties. I will discuss these two modelling approaches in greater detail in section 1.5.

Despite the inherent challenges, spectra of directly imaged exoplanets are rich in information. Presently-accessible near-infrared observations cover spectral regions containing H_2O, CO, and CH_4 absorption features, allowing their unique identification with moderate resolution spectra. Detections of multiple carbon and oxygen-bearing molecules allows estimation of abundance ratios, such as C/O and C/H, offering insights into aspects of planet formation (see section 1.3.3). As with eclipse spectra, constraints can also be obtained on atmospheric temperature structures, disequilibrium chemistry, and cloud properties (Lee et al., 2013; Lavie et al., 2017). As a planet rotates, inhomogeneous features, such as patchy clouds, can translate into flux variability over time (Apai et al., 2013). By adding time as a dimension to directly imaged planet spectra, it will soon be possible to construct surface feature maps - as has already been accomplished for brown dwarfs showing such rotational modulations (Apai et al., 2013; Crossfield et al., 2014; Radigan, 2014; Karalidi et al., 2016). Future telescopes, such as WFIRST (Spergel et al., 2015), will extend direct imaging spectroscopy to optical wavelengths, yielding the first reflection spectra (Marley et al., 1999; Sudarsky

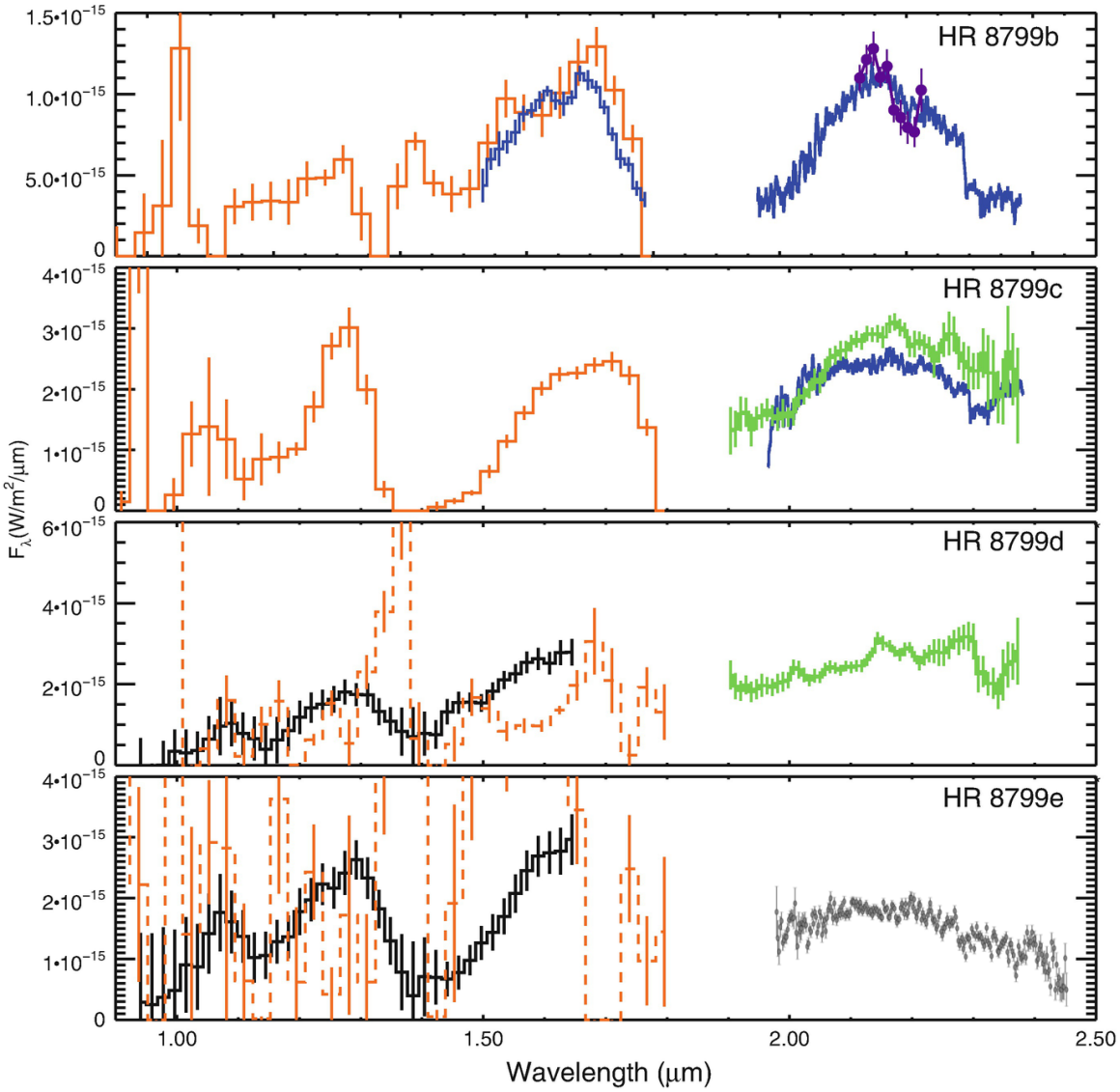


Fig. 1.17 Direct imaging spectroscopy of a multiplanet system. The four panels show near-infrared thermal emission spectra of the exoplanets HR 8799b,c,d,e (based on the compilation by Konopacky & Barman, 2018) observed by a variety of instruments: OSIRIS (blue / purple) (Bowler et al., 2010; Barman et al., 2011, 2015; Konopacky et al., 2013), P1640 (orange) (Oppenheimer et al., 2013), GPI (green) (Ingraham et al., 2014), SPHERE (black) (Bonnefoy et al., 2016), and GRAVITY (grey) (Gravity Collaboration et al., 2019). These spectra have enabled detections of subsets of H_2O, CO, and CH_4 in the atmospheres of this system.

et al., 2000) of exoplanet atmospheres. For a recent look at the potential of reflection spectra to characterise cool giant exoplanets (more analogous to Jupiter and Saturn than young, hot, wide-orbit imaged exoplanets), see MacDonald et al. (2018).

Finally, it is possible to combine direct imaging with high-resolution spectroscopy to detect exoplanet atmospheres at lower contrasts and angular separations than possible with imaging alone (Sparks & Ford, 2002; Riaud & Schneider, 2007). In section 1.4.4, we saw that cross correlation of high-resolution spectra with template models can utilise the many molecular lines present in a planet spectrum to extract atmospheric signatures otherwise hidden in noise. Whilst this technique does not extract the planet spectrum itself (as would be obtained from direct imaging alone), it can robustly detect molecular species present in the model templates. To combine these techniques, adaptive optics and a coronagraph are first applied to a directly imaged system to improve the contrast ratio of the planet-star flux by $\sim 10^5$ (Snellen et al., 2015). Due to the star and planet being *spatially* resolved on a detector, it is then possible to subtract the isolated spectrum of the star (and residual tellurics) from the total flux at the location where the planet is known to reside (e.g. from low-spectral resolution imaging). At this point, the high-resolution spectrum of the planet remains buried in noise. Cross correlation with high-resolution model templates can then extract the planet signal, effectively lowering the minimum contrast for which an atmosphere can be detected to $F_p/F_* \sim 10^{-7}$ (Snellen et al., 2015). Such low contrast ratios raise the possibility of detecting molecules in the atmospheres of terrestrial planets, including the prospective biosignature O_2, around the nearest stars with the next generation of ground-based telescopes in the late-2020s (Lovis et al., 2017; Hawker & Parry, 2019). However, detailed atmospheric studies of true Earth analogues ($F_p/F_* \sim 10^{-10}$), as required to statistically assess the frequency of life in nearby systems, will likely have to await dedicated space-based direct imaging spectroscopy (Fujii et al., 2018).

Highlights from direct imaging spectroscopy

Early reconnaissance of directly imaged exoplanet atmospheres took the form of near-infrared broadband photometry (Marois et al., 2008; Hinz et al., 2010, e.g.). These observations suggested directly imaged planets are redder than brown dwarfs at similar effective temperatures, indicating that their lower gravity influences their thermal emission (Barman et al., 2011). Photometry also revealed a relative absence of CH_4 absorption, compared to CO, taken as evidence of disequilibrium carbon chemistry (Hinz et al., 2010; Galicher et al., 2011; Skemer et al., 2012). The first spectra of directly imaged planets were obtained from HR 8799c (Janson et al., 2010) and HR 8799b (Bowler et al., 2010). Low-resolution ($R \sim 60$) spectra of both planets have been interpreted as evidence of patchy clouds (Barman et al., 2011). HR 8799 has now

become the first system (besides our solar system) with direct spectra for multiple planets (Oppenheimer et al., 2013). Notably, the spectra from all four planets are distinct, as seen in Figure 1.17. The first moderate resolution ($R \sim 4000$) direct exoplanet spectrum provided clear detections of H_2O and CO on HR 8799c (Konopacky et al., 2013). Their inferred abundances suggested a super-solar C/O ratio, providing evidence supporting formation by core-accretion. The first unambigious detections of CH_4 in an exoplanet atmosphere were reported by Barman et al. (2015) (for HR 8799b) and Macintosh et al. (2015) (for 51 Eridani b). Direct imaging was first combined with high-resolution spectroscopy by Snellen et al. (2014) for β Pictoris b, and has recently been combined with moderate resolution spectra to produce 'molecule maps' of the same system (Hoeijmakers et al., 2018b). Most recently, the first application of optical interferometry to direct imaging, combining the four telescopes at the VLT into an effective $\sim 100\,\mathrm{m}$ telescope, has yielded an $R \sim 500$ spectrum of HR 8799e (Gravity Collaboration et al., 2019). Though direct spectroscopy is the youngest method of probing exoplanet atmospheres, it is already showing remarkable promise.

1.5 Characterisation of exoplanet atmospheres

Relation of exoplanet spectra to underlying atmospheric properties requires theoretical models. In the previous section, I outlined the atmospheric features to which various observations are sensitive, and provided illustrative examples of model spectra (e.g. the schematic diagrams in Figures 1.13 and 1.14 for transmission and emission spectra). The technical details behind constructing model spectra will be addressed in subsequent chapters. Here, I outline how theoretical models of exoplanet spectra are utilised to characterise exoplanet atmospheres. The two broad approaches, self-consistent modelling and atmospheric retrieval, will be considered in turn.

1.5.1 Self-consistent modelling

Theoretical models of exoplanet atmospheres must contend with a paucity of spectral observations. This was especially true in the early 2000s, when observations of exoplanet atmospheres were nearly non-existent. It is against this backdrop that the first self-consistent models of exoplanet atmospheres were developed (Seager & Sasselov, 1998; Goukenleuque et al., 2000; Barman et al., 2001; Sudarsky et al., 2003). Self-consistent models circumvent the lack of knowledge about any given planetary atmosphere,

appealing to first-principles physics and chemistry to predict theoretical spectra. This approach starts from (assumed) atmospheric conditions and works forward to compute the corresponding planetary spectrum. It is therefore also referred to as *forward modelling*. By repeating the calculation for different atmospheric properties (e.g. composition, clouds), a grid of models can be created. Such grids are instrumental to predict how various atmospheric properties influence theoretical spectra (of great utility for planning observations), and can also be useful to fit observed spectra.

Early self-consistent models were adapted from stellar atmospheres codes (e.g. Seager & Sasselov, 1998; Sudarsky et al., 2003) or solar system planet codes (Marley et al., 1999). 1D models such as these take as input system properties (e.g. star-planet separation, stellar spectrum, planet surface gravity) and an elemental composition for the atmosphere. The atmospheric temperature structure, subject to the stellar flux above and internal heating below, is updated to satisfy the constraints of radiative-convective equilibrium and hydrostatic equilibrium. As the temperature in each atmospheric layer updates, chemical equilibrium is often assumed to specify how the input elemental abundances partition into different molecules and atoms. Chemical species then perturb the temperature in a given layer by absorbing, emitting, or scattering photons - encoded by their respective cross sections, assumed known from quantum mechanics or laboratory spectra (Tennyson & Yurchenko, 2018). Cloud formation is sometimes included, which can strongly influence the propagation of radiation. An iterative procedure accounts for these coupled processes, eventually converging on a *self-consistent* solution. The equation of radiative transfer is then solved throughout the plane-parallel atmosphere to compute spectra for the desired viewing geometry. An example of self-consistent transmission spectra is shown in Figure 1.18, demonstrating that observations of the super-Earth GJ 1214b indicate the existence of high-altitude clouds (Morley et al., 2013). For further details on the practicalities of self-consistent modelling, I refer the reader to the recent reviews of Marley & Robinson (2015), Hubeny (2017), and Heng & Marley (2018).

Whilst traditional 1D self-consistent models are still widely used (e.g. Mollière et al., 2015; Gandhi & Madhusudhan, 2017), further complexities can be considered in the modelling process. One refinement is to relax some underlying assumptions. An important example is models exploring departures from chemical equilibrium (e.g. Moses et al., 2011; Venot et al., 2012), such as from vertical mixing and photochemistry. An entirely alternative approach is to construct a 3D Global Circulation Model (GCM) of an atmosphere, which numerically solves the Navier-Stokes equations for a rotating

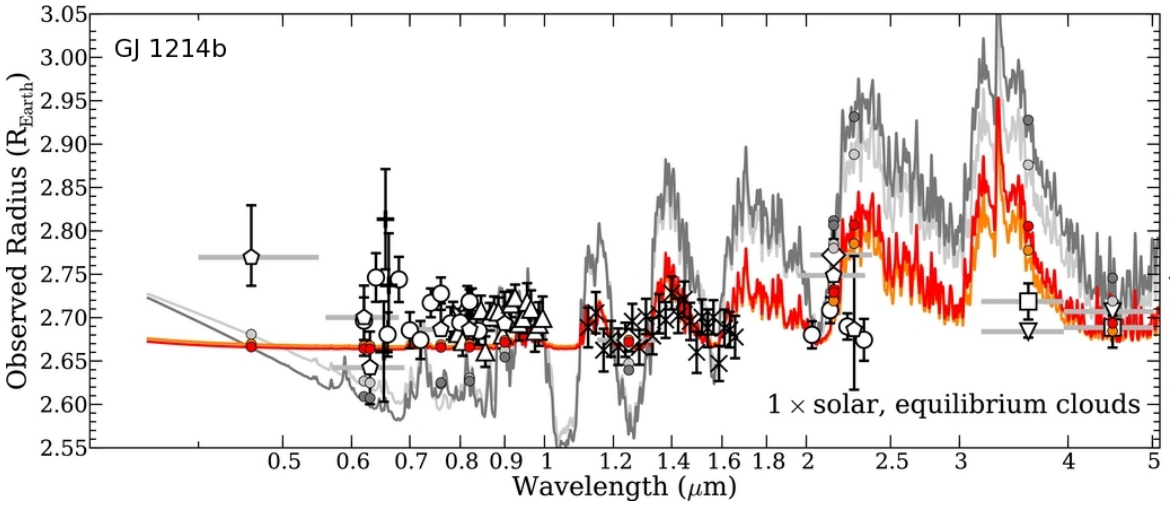


Fig. 1.18 Self-consistent models of an exoplanet atmosphere. Transmission spectra of the super-Earth GJ 1214b are compared against four self-consistent models (adapted from Morley et al., 2013). Cloud-free models (grey) provide a poor fit to the observations compared to models with clouds (red and orange). The red and dark grey models assume localisation of heat on the permanent dayside, whilst the orange and light grey models assume uniform redistribution of heat around the planet (e.g. by winds). By comparing models with the data, one can confidently conclude the atmosphere has clouds (but infer little about heat redistribution).

fluid to predict the composition, temperature structure, and cloud formation as a function of longitude, latitude, and depth. GCMs have been constructed for planets ranging from highly-irradiated hot Jupiters (e.g. Showman et al., 2008; Helling et al., 2016; Lines et al., 2018) to habitable exoplanets (e.g. Joshi, 2003; Yang et al., 2013; Boutle et al., 2017). In particular, GCMs of hot Jupiters have achieved great success in predicting and understanding phase curve offsets (see section 1.4.3).

Self-consistent modelling offers a direct link between the physics and chemistry of an atmosphere to spectral predictions. However, our understanding of the physicochemical processes in exoplanet atmospheres is likely incomplete, especially for conditions far removed from solar system experiences. Perhaps the greatest limitation of this approach is the computational burden required to iterate until a self-consistent solution is found. Consequently, modern self-consistent grids (e.g. Goyal et al., 2019) typically contain no more than $\sim 10^4$ models. This renders a thorough exploration of possible atmospheres infeasible (even for 1D models), such that regions of parameter space expressing alternative solutions (e.g. disequilibrium chemical abundances), can be missed entirely. This *degeneracy problem*, along with the computational bottleneck, were key motivators behind the development of an alternative approach: atmospheric retrieval.

1.5.2 Atmospheric retrieval

Over the last decade, the increasing quality of exoplanet spectra has rapidly given birth to a fresh approach to characterise exoplanet atmospheres. Rather than starting with an assumed atmosphere, computing a spectrum, and comparing with observations (cf. self-consistent models), one can instead start with observations and seek the range of atmospheric properties consistent with them. In this approach, also called *inverse* modelling, it is the observations, not assumptions about physics and chemistry, that drive atmospheric interpretations. As one is extracting, or *retrieving*, the state of an atmosphere from observed spectra, this approach is known as atmospheric retrieval.

Atmospheric retrieval algorithms couple a model atmosphere and radiative transfer prescription, i.e. a forward model, with a statistical sampling algorithm (e.g. Madhusudhan & Seager, 2009; Line et al., 2013), as illustrated in Figure 1.19. Retrieval codes describe many atmospheric characteristics by parametrised functions (e.g. the temperature profile), generate simulated spectra for many sampled parameter values, and compare the resultant spectra with observations to identify parameter ranges consistent with the data. As the chemistry, temperature structure, and cloud properties are described by free parameters, time-intensive iterations to enforce self-consistency with assumed principles (e.g. radiative-convective equilibrium) are no longer required. By mapping the entire parameter space, atmospheric retrieval techniques enable statistically rigorous constraints on atmospheric properties. In this manner, one can go beyond detection of atoms and molecules to measure their atmospheric abundances. If the obtained abundances of key molecules, such as H_2O, are sufficiently precise, they can constrain plausible planetary formation and migration pathways (see section 1.3.3 and, e.g. Öberg et al., 2011; Mordasini et al., 2016).

The forward models employed by retrieval codes must be both simple and flexible. Simplicity enables the evaluation of many millions ($\gtrsim 10^7$) of models, as required to thoroughly map a high-dimensional parameter space, in a timely fashion. Flexibility is built into the minimal assumption approach of retrieval methods, exploring both models that conform to expectations (e.g. chemical equilibrium) or present serendipitous results (e.g. extreme disequilibrium chemistry), during the fitting procedure. This being said, some assumptions remain in most retrieval forward models, namely: hydrostatic equilibrium, local thermodynamic equilibrium, and 1D, plane-parallel, geometries. Free parameters constituting the atmospheric state then control: chemical species abundances (e.g. H_2O, CH_4, CO, CO_2, Na, K, etc.), the pressure-temperature (P-T) profile, and clouds / hazes. Typically, this results in $\gtrsim 15$ free parameters.

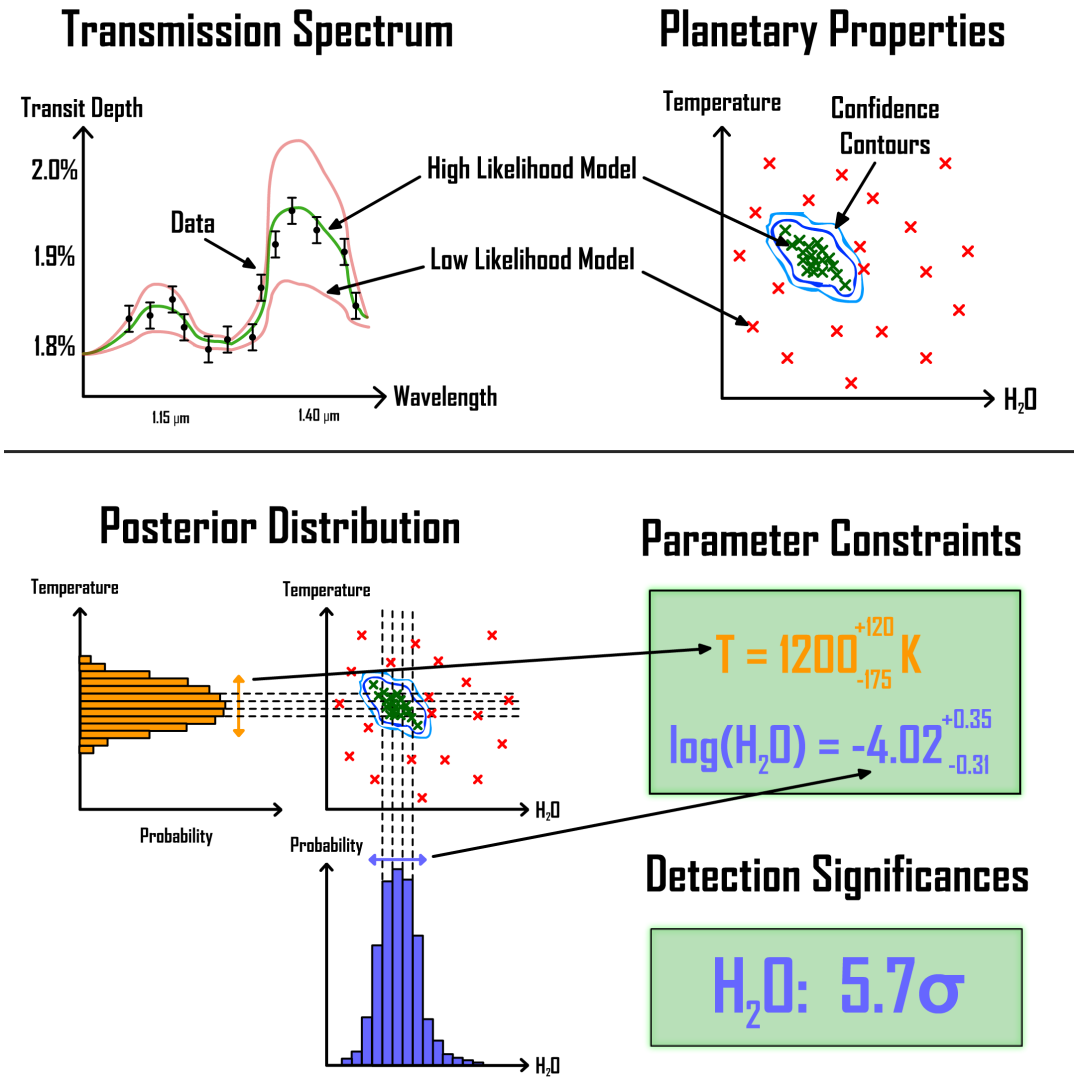


Fig. 1.19 Schematic illustration of exoplanet atmospheric retrieval. Top: transmission spectrum data is compared to a range of possible model spectra. A simplified toy model is shown (top-left), controlled by two parameters: the atmospheric temperature and H_2O abundance. Taking a parameter combination which fits the data (green), increasing the H_2O abundance for fixed temperature, or *vice versa*, strengthens absorption features, such that a low-likelihood fit (red) is obtained. The same holds if either parameter is reduced while the other is fixed. A statistical algorithm then maps the parameter space (top-right) to identify parameter combinations consistent with the data. Bottom: the resulting posterior distribution is integrated vertically and horizontally, resulting in (marginalised) probability distributions for each model parameter. The 1σ histogram ranges are reported as parameter constraints. Running an additional, nested, retrieval without a given molecule included (here, H_2O) yields a detection significance via Bayesian model comparison (see Chapter 2).

Given the large parameter space requiring exploration, efficient and robust statistical sampling algorithms are central to all retrieval codes. Early exoplanet retrieval studies employed grid-based exploration (Madhusudhan & Seager, 2009) or optimal estimation[18] (Lee et al., 2012). However, the limited *a priori* ground-truth of exoplanet atmospheres, coupled with the highly non-linear nature of the potential parameter space, necessitated the introduction of sophisticated Bayesian techniques (see Trotta, 2017, for an excellent introduction) capable of deriving full posterior probability distributions. In particular, Markov Chain Monte Carlo (MCMC) (e.g. Foreman-Mackey et al., 2013) and Nested Sampling (Skilling, 2004) have emerged as the techniques of choice, with various implementations employed in the atmospheric retrieval literature (Line et al., 2013; Benneke & Seager, 2013; Waldmann et al., 2015a; Lavie et al., 2017; Wakeford et al., 2017). Some authors have also explored the application of machine learning techniques to atmospheric retrieval (Waldmann, 2016; Márquez-Neila et al., 2018; Zingales & Waldmann, 2018; Soboczenski et al., 2018). A detailed discussion of how Bayesian statistics is used in the context of atmospheric retrieval will be provided in Chapter 2.

Atmospheric retrieval codes have seen wide application to the characterisation of exoplanet atmospheres. They span the full range of atmospheric observations I outlined in section 1.4, namely: transmission spectra (Madhusudhan & Seager, 2009; Benneke & Seager, 2012; Waldmann et al., 2015a; MacDonald & Madhusudhan, 2017a), emission spectra (Lee et al., 2012; Barstow et al., 2013; Line et al., 2013; Waldmann et al., 2015b; Gandhi & Madhusudhan, 2018), phase-resolved spectra (Stevenson et al., 2014; Kreidberg et al., 2018), direct imaging spectra (Lee et al., 2013; Todorov et al., 2016; Lavie et al., 2017), and, most recently, high-resolution Doppler spectroscopy (Brogi et al., 2017; Brogi & Line, 2019). Approaches have also been developed for (simulated) reflection spectra, in anticipation of future telescopes (Lupu et al., 2016; Feng et al., 2018). Retrievals have provided substantial insights into exoplanet atmospheres, with some recent highlights including: precise measurements of molecular abundances (Madhusudhan et al., 2014b; Kreidberg et al., 2014b; Wakeford et al., 2017), detections of new chemical species (Sedaghati et al., 2017; Chen et al., 2018), and robust detections of temperature inversions (Evans et al., 2017; Sheppard et al., 2017).

However, atmospheric retrieval is far from a straightforward endeavour. Without a dataset of sufficient precision, spectral resolution, and wavelength coverage, any atmospheric constraints will be little more than artefacts of the particular noise

[18]Optimal estimation has a rich heritage in planetary science, such as for Earth remote sensing (Rodgers, 2000) and atmospheric studies of the solar system giants (e.g. Irwin et al., 2008, 2018).

instance in the data. Even with reliable observations, degeneracies present in the forward model used by a retrieval code can lead to surfaces of solutions that hinder parameter inferences (e.g. Heng & Kitzmann, 2017). In particular, transmission spectra models for planets with clouds suffer from cloud-chemistry degeneracies which can limit chemical abundance determinations to the order-of-magnitude level (Benneke, 2015). As most transmission spectra show evidence for some degree of cloudiness (Sing et al., 2016; Barstow et al., 2017), the problem of precisely characterising the atmospheres of cloudy planets has been considered a fundamental challenge (Madhusudhan et al., 2016b). In the years to come, transmission spectroscopy will be at the vanguard of efforts to characterise lower-mass exoplanets, including temperate habitable worlds. It is therefore vital to develop new atmospheric retrieval algorithms capable of robustly inferring the underlying properties of exoplanets, especially those that possess clouds. The present work represents a step in this direction.

1.6 Thesis outline

In this thesis, I present a new approach to the atmospheric retrieval of exoplanet transmission spectra. A new algorithm is created, which offers a solution to retrieving precise atmospheric properties of cloudy exoplanets. Armed with this technique, I conduct detailed atmospheric retrieval studies for several exoplanets, revealing many tantalising insights, and a few surprises, into the nature of these distant worlds. The layout of this thesis is as follows:

- In Chapter 2, I introduce POSEIDON – a custom-built atmospheric retrieval code capable of *simultaneously* obtaining precise constraints on an exoplanet's composition and cloud properties from transmission spectra.
- In Chapter 3, this new retrieval code is applied to reveal the atmospheric nature of the canonical hot Jupiter HD 209458b – the exoplanet with the most precise transmission spectrum currently available.
- In Chapter 4, I present evidence of new molecules in hot Jupiter atmospheres, including the first detection of a heavy molecule in an exoplanet atmosphere.
- In Chapter 5, I extend POSEIDON's opacity database to include chemical species important across a wide range of atmospheres, and introduce a new method for rapidly calculating molecular absorption cross sections.

- In Chapter 6, POSEIDON is applied to characterise the atmosphere of a Neptune-mass exoplanet – representing one of the first precision retrieval studies of a low-mass exoplanet atmosphere.
- In Chapter 7, I investigate the ability of the upcoming James Webb Space Telescope (JWST) to characterise the atmosphere of a Neptune-mass exoplanet.
- Finally, in Chapter 8, I offer concluding remarks about the impact of this body of work, before discussing promising directions for future research.

Chapter 2

POSEIDON: Atmospheric Retrieval for Cloudy Exoplanets

2.1 Revenge of the exoclouds

We stand at the precipice of a new age – one where the vision of characterising exoplanets in exquisite detail is rapidly being realised. Using state-of-the-art atmospheric retrieval techniques, it is now possible to use spectroscopic observations to infer properties of exoplanet atmospheres. However, the dream of extracting precise chemical abundances from exoplanet transmission spectra, a crucial ingredient to constrain the formation of exoplanets, has been frustrated by the influence of high-altitude clouds. Despite launching telescopes into space, astronomy's oldest nemesis strikes again[1].

Studies of transiting exoplanets have been revolutionised over the last 6 years by the Hubble Space Telescope's (HST) Wide Field Camera 3 (WFC3). WFC3 offers precise, low-resolution ($R = \Delta\lambda/\lambda \approx 50$), infrared transmission and emission spectra. Crucially, the spectral range of WFC3 observations covers multiple strong H_2O absorption features (illustrated in Figure 1.13). As of early 2019, WFC3 spectra have enabled detections of H_2O in > 20 exoplanetary atmospheres (Madhusudhan, 2019). However, in almost all cases, the amplitudes of H_2O absorption features in transmission spectra are significantly lower than those expected of a cloud-free solar-composition atmosphere – instead of $\sim$ 5-10 scale heights (Seager, 2010; Madhusudhan et al., 2014a), typically $\sim$ 2 scale heights are observed (Deming et al., 2013; Kreidberg et al., 2015; Sing et al.,

[1] A popular account of how exoplanet astronomers came to terms with, and learned to embrace, the discovery of exoclouds is provided in MacDonald (2018).

2016; Wakeford et al., 2019). Taken at face value, this could imply a plethora of atmospheres inherently depleted in oxygen (Madhusudhan et al., 2014b). Alternatively, they may be explained by invoking high altitude (P <1 mbar) opaque cloud decks (Deming et al., 2013) or uniform-in-altitude grey opacity (Pont et al., 2013). Given the increasing number of low-amplitude, or even flat, observed spectra (e.g. Kreidberg et al., 2014a; Knutson et al., 2014a,b; Ehrenreich et al., 2014; Sing et al., 2016), the consideration of clouds has been elevated to the forefront of transmission spectroscopy.

The fundamental issue with deriving chemical abundances for cloudy exoplanetary atmospheres lies in innate degeneracies between clouds and chemistry. A wide range of solutions exist, spanning high-altitude clouds with concealed solar abundances to low-altitude, or non-existent, clouds with sub-solar abundances. This naturally leads to extremely loose constraints consistent with the full range from sub-solar through super-solar abundances (e.g. Benneke, 2015). It is thus clear that clouds pose an existential challenge to robustly estimating chemical abundances. However, most efforts to retrieve atmospheric properties of cloudy atmospheres have employed one-dimensional cloud models – i.e. homogeneous terminator clouds. This is despite predictions from Global Circulation Models (GCMs) that temperature contrasts of many hundreds of degrees may fuel a prominence of *partially cloudy* terminators on tidally locked hot Jupiters (Parmentier et al., 2016). The effect of these 'patchy' clouds on transmission spectra retrievals has recently been examined by Line & Parmentier (2016). They demonstrated, for a simplified model applied to near-infrared (WFC3) observations, that patchy clouds may introduce an additional degeneracy, beyond that of uniform clouds, by mimicking high mean molecular weight atmospheres. However, the implications of non-uniform cloud coverage within a full retrieval framework, considering spectral observations from the optical through the infrared, has yet to be explored.

In this chapter, I offer a potential solution to the problem of interpreting transmission spectra of cloudy exoplanets. I introduce POSEIDON, a new atmospheric retrieval algorithm that includes generalised two-dimensional inhomogeneous cloud distributions. By not assuming global cloud coverage across the terminator, regions without clouds are sampled during transit – effectively allowing one to 'peer below' the clouds and break many of the degeneracies between clouds and chemical abundances. The method I propose enables the *simultaneous* retrieval of cloud/haze properties and precise molecular abundance constraints. POSEIDON represents a step beyond traditional 1D atmospheric retrieval approaches and, in so doing, enables the nature of cloudy exoplanet atmospheres to be revealed.

In what follows, I describe the calculation of theoretical transmission spectra, i.e. the forward model of POSEIDON, in section 2.2. The process of retrieving atmospheric parameters from an observed spectrum is outlined in section 2.3. Finally, in section 2.4, the forward model and retrieval approach are validated.

2.2 Transmission spectra with 2D clouds

Extracting atmospheric properties from a spectrum involves two components: (i) a parametric forward model; and (ii) a statistical retrieval algorithm to sample the model parameter space. Typical forward models assume a 1D, plane-parallel, geometry. For transmission spectra, this amounts to assuming homogeneous properties (e.g. chemistry and clouds) across an exoplanet's terminator. Here, I will describe a 2D forward model formulation capable of accounting for azimuthally inhomogeneous clouds.

The forward model behind POSEIDON computes the transmission spectrum of an exoplanet as it transits its host star. The day-night terminator is modelled assuming hydrostatic equilibrium and a terminator-averaged temperature structure and composition. The model allows for inhomogeneous azimuthal cloud and haze distributions. The transmission of radiation through the planetary atmosphere is evaluated for a grid of rays incident on the terminator.

2.2.1 Radiative transfer

Consider the geometry of an exoplanet transit depicted in Figure 2.1. A planet of radius R_p with an atmosphere of height H_A transits a star of radius R_*. Given the large distances to extrasolar systems ($D \gg R_*$), all rays can be assumed to be parallel upon reaching the observer. Each ray is characterised by its *impact parameter*, b (not to be confused with the orbital impact parameter of the planet, b_p), the lowest altitude in the atmosphere experienced by the ray[2]. In general, the terminator may have different properties at different polar angles ϕ (illustrated in Figure 2.1, right).

Our goal is to compute the transmission spectrum of this planet. Recall from section 1.2.2 the definition of the transit depth

$$\Delta_\lambda = \frac{F_{\lambda,\,\mathrm{out}} - F_{\lambda,\,\mathrm{in}}}{F_{\lambda,\,\mathrm{out}}} \tag{2.1}$$

[2]Deflection of rays due to refraction does allow rays to reach altitudes $r_{\mathrm{min}} < b$ (Bétrémieux & Kaltenegger, 2014), but this effect is weak for H_2 dominated atmospheres (Robinson, 2017).

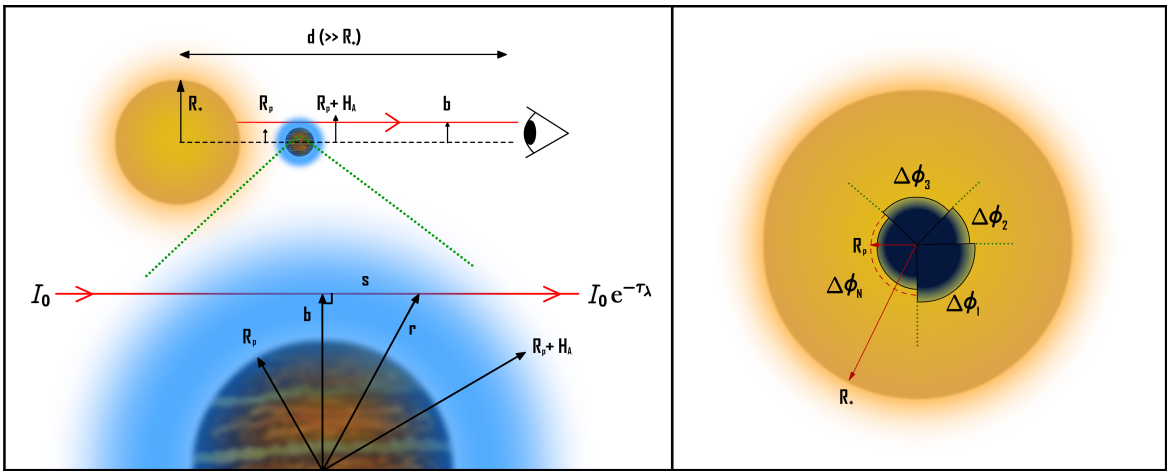


Fig. 2.1 Radiative transfer geometry for an exoplanet transit.
Left: stellar rays of an initial intensity I_0 at impact parameter b are attenuated due to passage through an atmosphere of height H_A. Right: the transit as viewed from the observer's perspective. For 2D models, the terminator is divided into N regions of polar angular extent $\Delta\phi_n$, each of which may possess different cloud opacity as a function of b (radial extent exaggerated). Note that the observed transit radius R_p represents an average radius at which the atmosphere becomes opaque.

where $F_{\lambda,\,\mathrm{out}}$ is the (spectral) flux observed outside of transit and $F_{\lambda,\,\mathrm{in}}$ is the flux observed during transit. Here I have redefined $\delta_\lambda \rightarrow \Delta_\lambda$ to emphasise that this quantity, the *2D transit depth*, is different from the transit depth of a 1D atmosphere (as we shall see). The fluxes are defined (Seager, 2010) as integrals of the spectral intensity over solid angle, projected into the observer's line of sight

$$F_\lambda = \int_\Omega I_\lambda \, \hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{k}} \, d\Omega \tag{2.2}$$

where I_λ is the spectral intensity, $\hat{\boldsymbol{n}}$ is a unit vector in the direction of beam propagation, $\hat{\boldsymbol{k}}$ is a unit vector in the direction of the observer, and Ω is the solid angle subtended by the source at the observer. I_λ is formally defined as the amount of energy, originating from a differential solid angle $d\Omega$, passing through unit area per unit wavelength per unit time (Seager, 2010). From Figure 2.1, we see that $\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{k}} = 1$. In this plane-parallel geometry, differential solid angles are given by $d\Omega = dA/D^2$, where dA is a (projected) differential area element on the source and D is the distance to the system.

The flux just before transit, $F_{\lambda,\,\mathrm{out}}$, has contributions from the stellar surface and thermal flux from the planetary nightside, such that

$$F_{\lambda,\,\mathrm{out}} = \int_{\Omega_*} I_{\lambda,\,*}\, d\Omega + \int_{\Omega_p} I_{\lambda,\,p,\,(\mathrm{night})}\, d\Omega \tag{2.3}$$

where $I_{\lambda,\,*}$ is the intensity of the stellar photosphere and $I_{\lambda,\,p\,(\mathrm{night})}$ is the thermal intensity from the planetary nightside. The first integral, in principle, can be evaluated over a limb darkening profile. However, standard transit data reduction algorithms (e.g. Mandel & Agol, 2002) account for limb darkening when deriving $R_{p,\,\mathrm{eff}}(\lambda)/R_*$ (this 'effective' radius includes atmospheric absorption). But the transit depth from a planet with radius $R_{p,\,\mathrm{eff}}$ across a *uniform* stellar disk is simply $(R_{p,\,\mathrm{eff}}/R_*)^2$. From a modelling perspective, we can therefore compute spectra for a uniform stellar intensity and directly compare to the observed $(R_{p,\,\mathrm{eff}}(\lambda)/R_*)^2$. The second term ('nightside pollution') can be neglected if (Kipping & Tinetti, 2010): (i) $I_{\lambda,\,p\,(\mathrm{night})} \ll I_{\lambda,\,*}$ (only violated for ultra-hot planets with high day-night heat transport at long wavelengths); or (ii) the observed transit depths have been corrected for the nightside flux. I assume that either of these hold, such that $I_{\lambda,\,p\,(\mathrm{night})} \to 0$. With these considerations, we have

$$F_{\lambda,\,\mathrm{out}} = I_{\lambda,\,*}\left(\frac{\pi R_*^2}{D^2}\right) \tag{2.4}$$

The flux observed during transit, $F_{\lambda,\,\mathrm{in}}$, has three components: (i) flux from the unobscured portion of the stellar surface, (ii) stellar flux passing through the planetary atmosphere, and (iii) planetary nightside thermal emission. Together, these read

$$F_{\lambda,\,\mathrm{in}} = \int_{\Omega_*-\Omega_p} I_{\lambda,\,*}\, d\Omega + \int_{\Omega_p} \tilde{I}_{\lambda,\,*}\, d\Omega + \int_{\Omega_p} I_{\lambda,\,p,\,(\mathrm{night})}\, d\Omega \tag{2.5}$$

where $\Omega_* - \Omega_p$ denotes the solid angle of the unobscured portion of the stellar surface and $\tilde{I}_{\lambda,\,*}$ is the (attenuated) intensity of stellar rays after passing through the planetary atmosphere. As before, the nightside thermal emission (third term) is taken to be negligible. Before evaluating these integrals, we must specify how the atmosphere interacts with the incident stellar intensity to produce $\tilde{I}_{\lambda,\,*}$.

When a ray passes through any medium, photons can be added or subtracted from the beam during its propagation. The evolution of the beam intensity as it travels a distance ds is given by the *radiative transfer equation* (Chandrasekhar, 1960)

$$\frac{dI_\lambda}{ds} = -\kappa_\lambda I_\lambda + \epsilon_\lambda \tag{2.6}$$

where κ_λ is the *absorption coefficient*, encoding absorption of photons by the medium and scattering out of the beam, and ϵ_λ is the *emission coefficient*, which encodes emission of photons and scattering into the beam. This equation is understood intuitively as accounting for losses ($-\kappa_\lambda I_\lambda$) and additions ($\epsilon_\lambda$) to the beam intensity. It is conventional to recast Equation 2.6 in terms of the *optical depth*, defined as the integral of the absorption coefficient along a distance s within a medium

$$\tau_\lambda = \int_{s=0}^{s} \kappa_\lambda(s')\, ds' \tag{2.7}$$

The optical depth is the central property governing the appearance of transmission spectra. Using this definition, substitution of $d\tau_\lambda = \kappa_\lambda ds$ into Equation 2.6 gives

$$\frac{dI_\lambda}{d\tau_\lambda} = -I_\lambda + S_\lambda \tag{2.8}$$

where the source function, $S_\lambda \equiv \epsilon_\lambda/\kappa_\lambda$, has been defined. In the case of transmission of stellar radiation, there is no emission and negligible scattering into the beam. The latter is a statement that it is unlikely for a photon scattered in a random direction to still reach the observer[3], so any scattering can effectively be considered as a loss term (i.e. equivalent to absorption). For these reasons, in transit geometry $\epsilon_\lambda = 0$ and hence $S_\lambda = 0$. As κ_λ is the sole factor encoding intensity losses (due to absorption and scattering combined), it is often termed the *extinction coefficient*. Equation 2.8 can now be trivially integrated along the path of a stellar ray to yield

$$\tilde{I}_{\lambda,*}(\tau_\lambda) = I_{\lambda,*}\, e^{-\tau_\lambda} \tag{2.9}$$

This exponential attenuation of intensity in the atmosphere is known as Beer's law. For the slant geometry relevant to 2D transmission spectra, τ_λ is a function of both the impact parameter and the terminator sector the ray traverses. The former can be seen by noting that lower b results in larger path lengths through the atmosphere (see Figure 2.1, left). The latter arises from κ_λ, in general, taking different values around the terminator (e.g. regions with strongly absorbing clouds have higher κ_λ than clear regions) - as visualised on the right of Figure 2.1. From Equation 2.7 and the geometry

[3]Strongly *forward-scattering* hazes can provide an interesting exception, see Robinson (2017).

in Figure 2.1, the slant optical depth can be written as

$$\tau_\lambda(b,\phi) = 2\int_{s=0}^{s_{\rm end}} \kappa_\lambda(s',\phi)ds' = 2\int_{b}^{R_p+H_A} \kappa_\lambda(r,\phi)\left(\frac{r}{\sqrt{r^2-b^2}}\right)dr \tag{2.10}$$

where ϕ is a polar angle specifying the region of the terminator the ray traverses, r is the radial distance from the centre of the planet to the current position of the ray, and the factors of 2 arise from the two symmetric components of the ray path. The relation $s^2 = r^2 - b^2$ was used in the last step to change variables. The dependence of κ_λ on altitude arises from the absorption strengths of atoms and molecules depending on the pressure and temperature in a given layer (as we shall see in section 2.2.3). With an expression for $\tilde{I}_{\lambda,*}$ in hand, we can now evaluate the flux during transit.

To evaluate the integrals for $F_{\lambda,\rm in}$ in Equation 2.5, I first note an important symmetry. Strictly speaking, the integral over the planetary atmosphere (the second term) should be carried out over a polar coordinate system centred on the planet. However, assuming the planet is fulling overlapping the star, the assumption of uniform stellar brightness means the integral over the planet is identical regardless of where it is placed on the stellar disk[4]. Without loss of generality, the planet can therefore be placed along the axis of the star-observer line-of-sight (i.e. in the centre of the stellar disk). This symmetry implies that transmission spectra are invariant as the planet crosses the stellar disk (e.g. Figure 1.4) and need only be calculated once[5].

Having aligned the planet with the star-observer axis, a mutual polar coordinate system (b, ϕ) can be defined. All solid angle integrals can now be expressed with a common area element, $d\Omega = dA/D^2 = b\,db\,d\phi/D^2$. The in-transit flux (Equation 2.5) can therefore be written as

$$F_{\lambda,\rm in} = \frac{1}{2\pi}\int_0^{2\pi}\int_{R_p+H_A}^{R_*} I_{\lambda,*}\left(\frac{2\pi b}{D^2}\right)db\,d\phi + \frac{1}{2\pi}\int_0^{2\pi}\int_0^{R_p+H_A}\tilde{I}_{\lambda,*}(b,\phi)\left(\frac{2\pi b}{D^2}\right)db\,d\phi \tag{2.11}$$

where the lower b limit of the second integral is set to zero to account for the possibility of rays passing through the planet below the (observed) R_p at wavelengths where the atmosphere is especially transparent. Where this may occur is governed by the extinction coefficient at different azimuthal angles and altitudes, as regions with low

[4]I implicitly assume here that all rays follow straight trajectories mapping to the stellar surface behind the planet. In cases with strong refraction and / or scattering the symmetry is only approximate.

[5]This symmetry also explains why the planetary impact parameter, b_p, does not influence the transmission spectrum for $b_p < R_* - R_p - H_A$ (i.e. for non-grazing transits).

opacity possess deeper radii where the atmosphere becomes opaque. A common lower value for all wavelengths is therefore not assumed. Substituting Equation 2.9 into the second term, and integrating over b in the first term, produces

$$F_{\lambda,\,\mathrm{in}} = \frac{1}{2\pi}\int_0^{2\pi} \frac{\pi I_{\lambda,\,*}}{D^2}\left(R_*^2 - R_p^2 - 2R_pH_A - H_A^2 + 2\int_0^{R_p+H_A} b\, e^{-\tau_\lambda(b,\,\phi)} db\right) d\phi \quad (2.12)$$

Using $2\int_{R_p}^{R_p+H_A} b\, db = 2R_pH_A + H_A^2$, and splitting the radial integral at $b = R_p$ gives

$$F_{\lambda,\,\mathrm{in}} = \frac{1}{2\pi}\int_0^{2\pi} \frac{\pi I_{\lambda,\,*}}{D^2}\left(R_*^2 - R_p^2 + 2\int_0^{R_p} b\, e^{-\tau_\lambda(b,\,\phi)} db + 2\int_{R_p}^{R_p+H_A} b\left(e^{-\tau_\lambda(b,\,\phi)} - 1\right) db\right) d\phi \quad (2.13)$$

Finally, recalling that $F_{\lambda,\,\mathrm{out}} = I_{\lambda,\,*}\,(\pi R_*^2/D^2)$ (Equation 2.4), these two fluxes can be substituted into the Equation 2.1 to obtain[6]

$$\Delta_\lambda = \left(\frac{R_p}{R_*}\right)^2 + \frac{1}{2\pi}\int_0^{2\pi} \frac{2}{R_*^2}\left(\int_{R_p}^{R_p+H_A} b\left(1 - e^{-\tau_\lambda(b,\,\phi)}\right) db - \int_0^{R_p} b\, e^{-\tau_\lambda(b,\,\phi)} db\right) d\phi \quad (2.14)$$

The first term here is achromatic absorption due to an opaque disk of radius R_p (cf. Equation 1.2), whilst the second and third terms account for the influence of the atmosphere. Their role becomes clearer if the polar integral is factored out as

$$\Delta_\lambda = \frac{1}{2\pi}\int_0^{2\pi} \delta_\lambda\,(\phi)\; d\phi \quad (2.15)$$

$$\delta_\lambda(\phi) = \frac{R_p^2 + 2\displaystyle\int_{R_p}^{R_p+H_A} b\left(1 - e^{-\tau_\lambda(b,\,\phi)}\right) db - 2\int_0^{R_p} b\, e^{-\tau_\lambda(b,\,\phi)} db}{R_*^2} \quad (2.16)$$

The latter expression is simply the (1D) transmission spectrum that results from assuming an axially symmetric atmosphere with the same properties as the 2D atmosphere at polar angle ϕ. We therefore see from Equation 2.15 that 2D transmission spectra can be thought of as an azimuthal average of 1D transmission spectra. Expressed in this form, the second term represents an integral of the absorptivity $(1 - e^{-\tau_\lambda})$ of successive planetary annuli (of area $2\pi b\, db$) over the atmosphere above R_p. The third term is a correction, accounting for rays that have sufficiently small τ_λ to transmit thorough the atmosphere *below* R_p via a similar integral over the *transmissivity* $(e^{-\tau_\lambda})$.

[6]Note that $I_{\lambda,\,*}$ cancels, and thus no stellar spectrum is required for this calculation.

2.2.2 Model atmosphere

To compute the integrals over b in Equation 2.16, we require an atmospheric model. This model encodes how the local gas conditions (e.g. pressure, temperature, and composition) vary with altitude. POSEIDON models the terminator region of an exoplanet atmosphere by a discretised grid of layers spaced uniformly in log-pressure. The temperature in each layer is related to pressure by a flexible parametric function. Under the assumptions of hydrostatic equilibrium and the ideal gas law, $r\,(P)$ can then be computed for each layer. The ideal gas law also specifies the number density in each layer, which is required to evaluate the extinction coefficient (see section 2.2.3). Here, I describe each aspect of the calculation to initialise a model atmosphere.

Hydrostatic radial profile

Model atmospheres are divided into 100 layers spaced uniformly in $\log(P)$ from $10^{-6} - 10^{2}$ bar (where 1 bar $= 10^{5}\,\mathrm{Pa} \approx 1\,\mathrm{atm}$). This pressure grid can be related to a radial (altitude) grid by the equation of hydrostatic equilibrium

$$\frac{dP}{dr} = -\frac{G\,M(r)\,\rho(r)}{r^2} = -\rho(r)\,g(r) \tag{2.17}$$

where $M(r)$ is the planet mass interior to r ($\approx M_p$ in the atmosphere), ρ is the local gas mass density, and $g(r)$ is the gravitational field strength. In the second equality, Newton's law of gravitation has been used to define $g(r)$. The mass density is given by

$$\rho(r) = \sum_i n_i(r)\,m_i \equiv n_{\mathrm{tot}}(r)\,\mu(r) \tag{2.18}$$

where n_i and m_i are the *number density* (in m^{-3}) and atomic / molecular mass (in kg) of chemical species i, n_{tot} is the total number density, and μ is the *mean molecular mass*. The ideal gas law relates n_{tot} to the pressure and temperature in a layer via

$$n_{\mathrm{tot}}(r) = \frac{P(r)}{k_B T(r)} \tag{2.19}$$

Substitution of Equations 2.19 and 2.18 into Equation 2.17 yields

$$\frac{dP}{dr} = -\frac{\mu(r)\,g(r)}{k_B T(r)} P(r) = \frac{g_p\,\mu(r)}{k_B T(r)}\,\frac{R_p^2}{r^2} P(r) \tag{2.20}$$

where in the last equality I defined the *surface gravity*[7], $g_p \equiv \frac{GM_p}{R_p^2} = g(r)\frac{r^2}{R_p^2}$. Integration of this equation, using the boundary condition $r(P_{\rm ref}) = R_p$, yields the formal solution

$$r(P) = \left(\frac{1}{R_p} + \frac{1}{R_p^2}\int_{P_{\rm ref}}^{P} \frac{k_B\, T(P)}{g_p\, \mu(P)}\, \frac{1}{P}\, dP\right)^{-1} \tag{2.21}$$

where $P_{\rm ref}$ is the *a priori* unknown pressure at the observed planet radius, R_p, and I have re-expressed T and μ as functions of pressure. The quantity $H \equiv \frac{k_B T}{\mu g_p}$ defines the atmospheric scale height (Equation 1.6). For illustrative purposes, this equation can be solved analytically when T and μ are constant with altitude, yielding

$$r(P) = R_p\left(1 + \frac{H}{R_p}\ln\left(\frac{P}{P_{\rm ref}}\right)\right)^{-1} \approx R_p - H\ln\left(\frac{P}{P_{\rm ref}}\right) \tag{2.22}$$

where in the last equality I expanded for $\frac{H}{R_p} \ll 1$. This shows that pressure (and all quantities proportional to it) roughly fall exponentially with altitude as

$$P(r) \propto n_{\rm tot}(r) \propto \rho(r) \propto e^{-(r-R_p)/H} \tag{2.23}$$

However, real exoplanet atmospheres are not expected to be isothermal. I therefore solve Equation 2.21 numerically for a given pressure-temperature profile, $T(P)$.

Pressure-temperature profile

The temperature in each layer, $T(P)$, is computed via the parametric pressure-temperature (P-T) profile equations given in Madhusudhan & Seager (2009)

$$\begin{aligned}
P &= P_0\, e^{\alpha_1\sqrt{T-T_0}} && (P_0 < P < P_1), && \text{(Region 1)}\\
P &= P_2\, e^{\alpha_2\sqrt{T-T_2}} && (P_1 < P < P_3), && \text{(Region 2)}\\
T &= T_3 && (P > P_3), && \text{(Region 3)}
\end{aligned} \tag{2.24}$$

where P_0 and T_0 are the pressure and temperature at the top of the atmosphere, $P_{1,3}$ and $T_{1,3}$ are specified at the region boundaries, and P_2 and T_2 encode a (potential) temperature inversion point. Temperature gradients are controlled by α_1 and α_2. Taking into account continuity at the region boundaries, the parametric P-T profile employed by POSEIDON is specified by 6 parameters: α_1, α_2, T_0, P_1, P_2, and P_3.

[7] g_p measurements are more precise than M_p for transiting exoplanets (Southworth et al., 2007), so surface gravity is preferred over planet mass for the hydrostatic calculations.

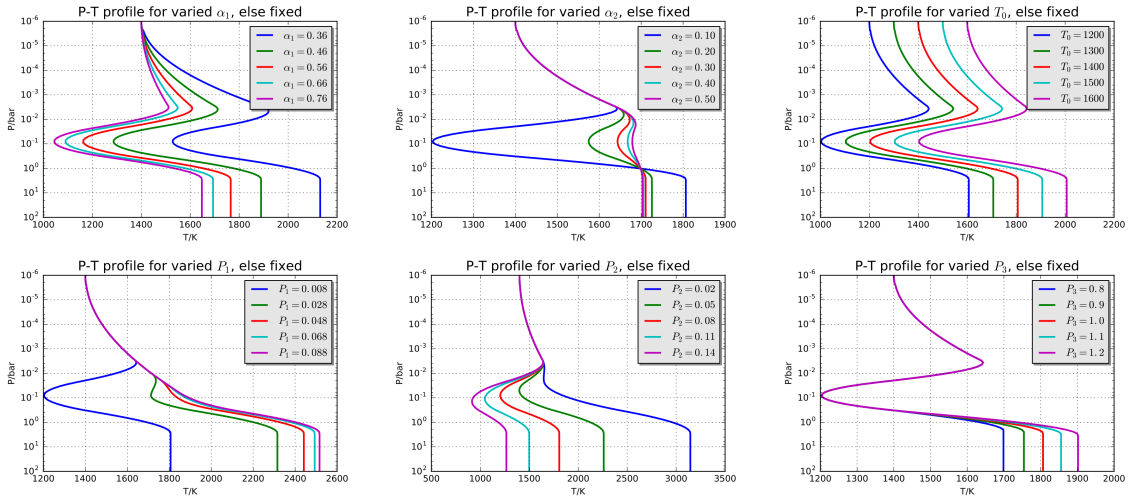


Fig. 2.2 The parametric pressure-temperature (P-T) profile in POSEIDON. From the top-left to bottom-right, $\alpha_1, \alpha_2, T_0, P_1, P_2, P_3$ are varied (via Equation 2.24).

The P-T profile functional form as each parameter is varied is shown in Figure 2.2. We see that this profile is highly flexible and generic. For example, the lower-left panel shows a profile with a thermal inversion (blue) smoothly transitioning to profiles where the temperature monotonically decreases with altitude (once $P_1 > P_2$). The first two panels also show profiles become increasingly isothermal as α_1 and $\alpha_2 \to 1$.

Atmospheric composition

Exoplanet atmospheres can contain a wide range of chemical species. Hot Jupiters are expected to be dominated by H_2 and He (see section 1.3.2), with other gases existing as trace species. Exo-Neptunes and super Earths can span a wider range of compositions, including H_2O- and CO_2-rich atmospheres (Moses et al., 2013b). Common species expected to exist in exoplanet atmospheres include: H_2O, CH_4, NH_3, HCN, CO, CO_2, C_2H_2, H_2S, N_2, Na, and K (Madhusudhan et al., 2016b). Heavy metal oxides (e.g. TiO and VO), are anticipated to be important at higher temperatures (Sharp & Burrows, 2007; Woitke et al., 2018). Cooler, terrestrial, atmospheres, can also contain molecules such as O_2, O_3, N_2O, and NO_2. As we shall see in Chapter 5, all these chemical species, including many others, are included in POSEIDON.

The abundance of a species is quantified by its *mixing ratio*, $X_i \equiv n_i/n_{\rm tot}$, which I assume to be uniform both in altitude and across the terminator (i.e. representing an average limb abundance). This also implies that the mean molecular mass, $\mu = \sum_i X_i\, m_i$, is uniform. With $T(P)$ and μ determined, Equation 2.21 can be solved.

2.2.3 The extinction coefficient

Having specified the local conditions in each layer, and the abundance of each chemical species, the extinction coefficient, κ_λ, can be determined. The wavelength dependence of optical depths, and hence transmission spectra, arises from κ_λ via Equation 2.10. An exoplanet atmosphere presents multiple sources of extinction to stellar rays, with the combined influence of the atmosphere summarised by

$$\kappa_\lambda(r, \phi) = \kappa_{\text{chem},\lambda}(r, \phi) + \kappa_{\text{Rayleigh},\lambda}(r, \phi) + \kappa_{\text{CIA},\lambda}(r, \phi) + \kappa_{\text{cloud},\lambda}(r, \phi) \tag{2.25}$$

where $\kappa_{\text{chem},\lambda}$ is extinction due to photons absorbed by molecules and atoms, $\kappa_{\text{Rayleigh},\lambda}$ is extinction from photons Rayleigh scattered out of the beam, $\kappa_{\text{CIA},\lambda}$ is extinction from collision-induced absorption (CIA), and $\kappa_{\text{cloud},\lambda}$ is the combined extinction from absorption and scattering due to clouds or hazes. The dependence of each coefficient on r and ϕ is short-hand for the local atmospheric conditions (P and T) where the extinction is evaluated. The first three terms are, in principle, known *a priori* for each chemical species from quantum mechanics or laboratory measurements. The final term, due to clouds or hazes, is handled differently due to the unknown nature of exoplanet clouds (as discussed in the next section). I will now briefly describe the opacity sources used in the initial version of POSEIDON (as employed here and in Chapters 3 and 4). In Chapter 5, I will present an extensive update to these opacities which will be used in subsequent chapters.

Molecular and atomic opacities

The extinction coefficient due to molecular or atomic absorption can be written as

$$\kappa_{\text{chem},\lambda}(P, T) = \sum_{i}^{N_{\text{spec}}} n_i(P, T)\, \sigma_{\text{abs,i}}(\lambda, P, T) \tag{2.26}$$

where n_i and $\sigma_{\text{abs,i}}$ are, respectively, the number density and *absorption cross section* of the i_{th} chemical species, evaluated at a given P and T. $\sigma_{\text{abs,i}}$ encodes the probability of one molecule or atom absorbing a photon at wavelength λ (in units of area per species). Multiplication by n_m, the number per unit volume of the same species (determined in section 2.2.2), then yields the extinction due to species i. The total extinction is then the sum over contributions of all chemical species considered in a given model.

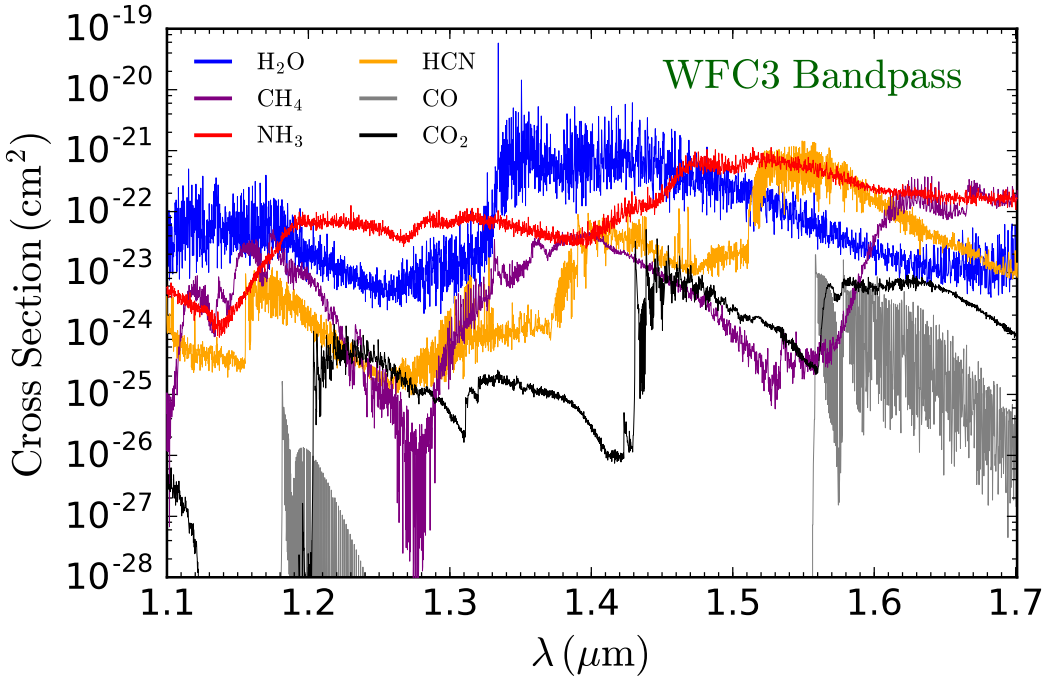


Fig. 2.3 Near-infrared molecular absorption cross sections.
The main spectrally active molecules in the HST WFC3 G141 band pass (1.1–1.7 µm) are shown. Each cross section is depicted at $T = 1400\,\mathrm{K}$ and $P = 10^{-3}\,\mathrm{bar}$ - roughly representative of the upper atmosphere of a hot Jupiter probed in transmission geometry.

The calculation of absorption cross sections from quantum transitions will be explained in detail in Chapter 5. For our present purposes, it suffices to note that $\sigma_{\mathrm{abs,\,i,}}(\lambda,\, P,\, T)$ are pre-calculated inputs, and hence do not need to be derived during a transmission spectrum calculation. For the initial version of POSEIDON, molecular cross sections for H_2O, CH_4, CO, CO_2, NH_3, and HCN were pre-computed following the methodology of Hedges & Madhusudhan (2016) and Gandhi & Madhusudhan (2017) - provided by S. Gandhi. These species have prominent near-infrared absorption features, as shown in Figure 2.3. The line lists (see Chapter 5) used to compute these cross sections come from HITEMP (Rothman et al., 2010) for H_2O, CO, and CO_2, and ExoMol (Tennyson et al., 2016) for CH_4, HCN, and NH_3. The cross sections were produced at a $1\,\mathrm{cm^{-1}}$ spectral resolution, on a pressure and temperature grid ranging from 10^{-4} to 10^{2} bar and 300 to 3500 K. Atomic opacities due to Na and K are also included, following the semi-analytic Lorentzian profiles used in Christiansen et al. (2010) (a more accurate treatment of alkali opacities will be described in Chapter 5). Given these tabulated cross sections, $\sigma_{\mathrm{abs,\,i,}}(\lambda,\, P,\, T)$ is found for each model atmosphere layer by linearly interpolating each cross section in log-P and T to the local conditions.

Rayleigh scattering

Alongside absorption, molecules and atoms can also scatter photons out of the line-of-sight. The extinction coefficient due to Rayleigh scattering can be treated as a pure loss term (as few photons are forward scattered for a Rayleigh phase function), taking the form

$$\kappa_{\text{Rayleigh},\,\lambda}\,(P,\,T) = \sum_{i}^{N_{\text{spec}}} n_i\,(P,\,T)\,\sigma_{\text{Ray,\,i}}(\lambda) \tag{2.27}$$

where $\sigma_{\text{Ray,\,i}}$ is the Rayleigh scattering cross section of the i_{th} chemical species at wavelength λ. This functional form ($\kappa = n\sigma$) is identical to that for molecular and atomic absorption (Equation 2.26), but with the Rayleigh scattering cross section independent of pressure and temperature. For the purposes here (and in Chapters 3 and 4), I consider only Rayleigh scattering from H_2 - the dominant molecule for gaseous exoplanets - with a cross section given by (Dalgarno & Williams, 1962)

$$\sigma_{\text{Ray},\,\text{H}_2}(\lambda) = \frac{8.14 \times 10^{-53}}{\lambda^4} + \frac{1.28 \times 10^{-66}}{\lambda^6} + \frac{1.61 \times 10^{-80}}{\lambda^8} \tag{2.28}$$

where all quantities are in SI units (i.e. $\sigma_{\text{Ray},\,\text{H}_2}$ in units of m^2 / molecule and λ in m). Note that the wavelength dependence in Equation 2.28 is close to λ^{-4} for near-infrared and optical wavelengths, but deviates to become steeper for shorter wavelengths approaching the UV. As we shall see in Chapter 5, where I will generalise this treatment and include Rayleigh scattering from other molecules, the deviation from $\sigma_{\text{Ray}} \propto \lambda^{-4}$ arises from the wavelength dependence of the H_2 refractive index and the non-spherical nature of the molecule.

Collision-induced absorption

Collision-induced absorption is a continuum opacity source arising from temporary dipoles induced by molecular collisions. Notably, even molecules that are ordinarily spectrally inactive (i.e. $\sigma_{\text{abs}} \approx 0$), such as H_2, can contribute to the total extinction via CIA. The relevant extinction coefficient takes the form

$$\kappa_{\text{CIA}\,\lambda}\,(P,\,T) = \sum_{i}^{N_{\text{spec}}} \sum_{j \geqslant i}^{N_{\text{spec}}} n_i\,(P,\,T)\,n_j\,(P,\,T)\,\sigma_{\text{CIA,\,i-j}}(T,\,\lambda) \tag{2.29}$$

where $\sigma_{\text{CIA,\,i-j}}$ is the binary absorption cross section between species i and j. CIA is particularly important at high pressures, as the pair nature of the process results

in $\kappa \propto n_{\rm tot}^2$ (de Wit & Seager, 2013). At wavelengths where other opacity sources are negligible ('windows'), CIA effectively provides a 'floor' to the spectrum limiting the deepest altitude which can be probed. I note that binary absorption cross sections are independent of pressure, and can be found tabulated in the literature (e.g. Richard et al., 2012) rather than calculated.

2.2.4 A generalised 2D cloud & haze prescription

Correctly modelling exoplanet transmission spectra requires the inclusion of clouds and hazes. POSEIDON implements a 2D cloud and haze prescription, considering the possibility of terminators containing both cloudy and clear regions. In cloudy regions, I consider an opaque cloud deck located at $P = P_{\rm cloud}$, below which no electromagnetic radiation may pass - effectively corresponding to the large particle size limit of Mie scattering (e.g Kitzmann & Heng, 2018). This region also contains a haze, distributed uniformly throughout the atmosphere, with an extinction coefficient given by a two-parameter power law (Lecavelier Des Etangs et al., 2008) to account for small scattering particles. The extinction coefficient due to clouds and hazes can then be written as

$$\kappa_{\rm cloud,\,\lambda}\,(r, \phi) = \begin{cases} \tilde{\kappa}_{\rm cloud}\,(\lambda,\, P), & \phi \in \text{cloudy} \\ 0, & \phi \in \text{clear} \end{cases} \tag{2.30}$$

$$\tilde{\kappa}_{\rm cloud}\,(\lambda,\, P) = \begin{cases} n_{\rm H_2}\, a\, \sigma_0 (\lambda/\lambda_0)^{\gamma}, & P < P_{\rm cloud} \\ \infty, & P \geqslant P_{\rm cloud} \end{cases} \tag{2.31}$$

where λ_0 is a reference wavelength (350 nm), σ_0 is the H_2-Rayleigh scattering cross section at the reference wavelength ($5.31 \times 10^{-31}\,\mathrm{m}^2$), a is the 'Rayleigh enhancement factor', and γ is the 'scattering slope'. The first term accounts for the effect of scattering due to small particles of an *a priori* unknown nature, with the parameter γ potentially indicative of the aerosol causing the slope (Pinhas & Madhusudhan, 2017). The second term models a sharp cut-off in transmission, simulating the high optical depths encountered inside clouds in slant-geometry (Fortney, 2005). This prescription is of sufficient generality to reproduce the additional extinction seen in current exoplanet spectra attributable to clouds, including the prospect of inhomogeneous terminator cloud and haze distributions (Line & Parmentier, 2016).

With the extinction coefficient in this form, an important simplification can be made to 2D transmission spectra. Recalling the general expression from Equation 2.15, I first write

$$\Delta_\lambda = \frac{1}{2\pi} \int_0^{2\pi} \delta_\lambda \left(\phi\right) d\phi \approx \sum_{n=1}^{N} \bar{\phi}_n \, \delta_{\lambda,\, n} \tag{2.32}$$

where $\delta_\lambda \left(\phi\right)$ is 1D transmission spectrum of an atmosphere with the same composition, temperature structure, and cloud properties as found at polar angle ϕ (given by Equation 2.16). In the approximation, I group the atmosphere into N regions with similar properties, specified by the reduced polar angular extent of the n_{th} region, $\bar{\phi}_n \equiv \Delta\phi_n / 2\pi$ ($\sum_{n=1}^{N} \bar{\phi}_n \equiv 1$), as visualised in Figure 2.1 (right). A 2D transmission spectrum is therefore effectively a weighted sum of the 1D transmission spectra corresponding to each atmospheric region, $\delta_{\lambda,\, n}$, with the weights given by $\bar{\phi}_n$. Note that two parts of the terminator occurring at different ϕ, but with similar atmospheric properties, can be grouped into the same $\bar{\phi}_n$ bin (due to $\delta_{\lambda,\, n}$ being similar). This symmetry implies, for example, that all clear regions of the atmosphere can be grouped together into a single 'effectively clear' region, even if all clear regions are not spatially continuous.

Given the aforementioned symmetry, and the 2D cloud and haze prescription, I take $N = 2$ and consider one region to be cloudy and the other clear. This yields the simple equation

$$\Delta_\lambda = \bar{\phi}\, \delta_{\lambda,\, \mathrm{cloudy}} + (1 - \bar{\phi})\, \delta_{\lambda,\, \mathrm{clear}} \tag{2.33}$$

where I have defined $\bar{\phi}_1 \equiv \bar{\phi}$, $\delta_{\lambda,\,1} \equiv \delta_{\lambda,\,\mathrm{cloudy}}$, $\delta_{\lambda,\,2} \equiv \delta_{\lambda,\,\mathrm{clear}}$, and used $\bar{\phi}_2 = 1 - \bar{\phi}_1$. This limit is similar to the 'patchy cloud' model considered by Line & Parmentier (2016), but with the inclusion of an inhomogeous haze. The reduced polar angle $\bar{\phi}$ is then equivalent to the total terminator cloud coverage. This cloud prescription is thus encoded by 4 parameters: a, γ, P_{cloud}, and $\bar{\phi}$. Importantly, evaluation of Equation 2.33 only requires one to compute two 1D transmission spectra. I assume in this formulation that the composition and temperature structure is the same in both clear and cloudy regions. Inhomogeneities in T and X_i around the terminator (e.g. Kataria et al., 2016) are neglected, with their retrieved values representing terminator averages. I reserve these inhomogeneities for future study, focusing here on inhomogenous clouds.

Figure 2.4 shows the effect of varying the four cloud parameters on a transmission spectrum generated by POSEIDON. The top panels demonstrate that a and γ encode the strength and gradient of a slope manifesting particularly at visible wavelengths $<$ 0.7µm. In contrast, the cloud deck pressure and terminator cloud coverage strongly influence spectra across the optical and infrared. Raising the cloud deck (lowering

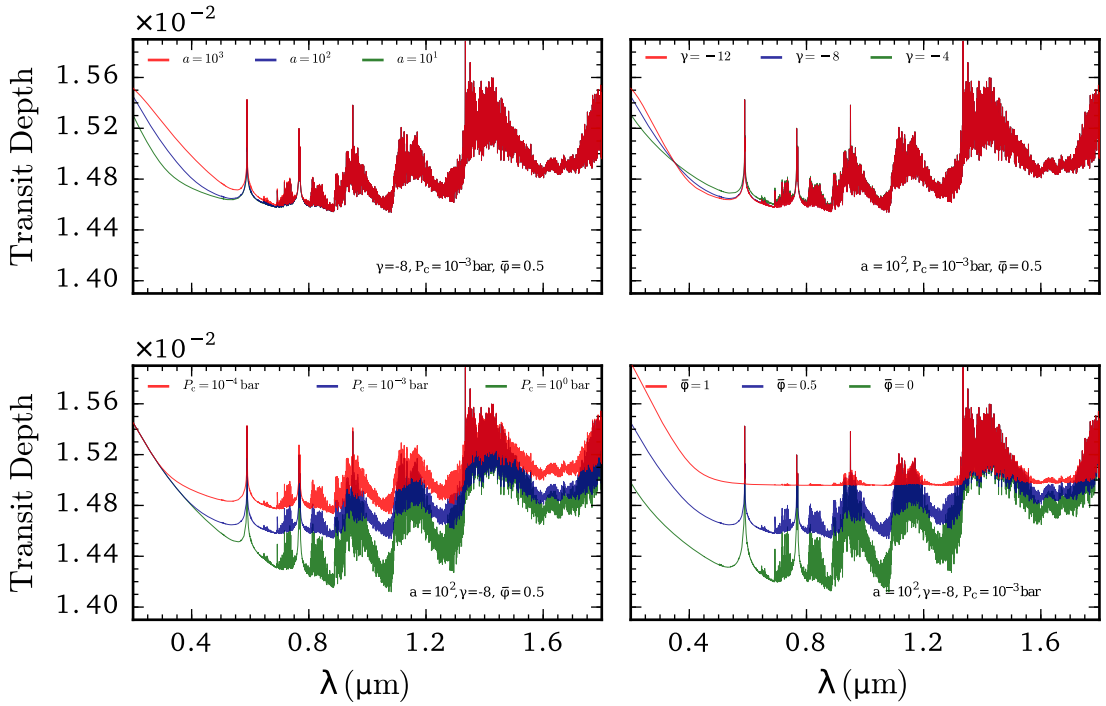


Fig. 2.4 The effect of 2D clouds and hazes on transmission spectra. The impact on a fiducial model transmission spectrum (blue) of changing the cloud and haze parameters in Equations 2.31 and 2.33 is demonstrated. Top: variation caused by the haze parameters, a and γ. Bottom: variation caused by the cloud deck pressure, $P_{\rm cloud}$ (left), and the terminator cloud coverage, $\bar{\phi}$ (right). The fiducial model assumes a roughly solar-composition atmosphere at an isothermal temperature of 1400 K.

$P_{\rm cloud}$) increases the transit depth at all wavelengths (as the $\tau = 1$ surface is pushed to progressively higher altitudes). As $\bar{\phi} \rightarrow 1$, the base-level of the spectrum becomes increasingly flat, until the 'hard surface' limit of a uniform cloud is recovered.

By combining this 2D cloud model with the sources of molecular and atomic extinction, the total extinction coefficient (Equation 2.25) can be evaluated in each layer of the atmosphere. Integration of the extinction coefficient along a slant path produces the optical depth encountered by rays at a given impact parameter (Equation 2.10). Once optical depths have been computed for a grid of impact parameters, a 2D transmission spectrum can be produced via Equations 2.16 and 2.33. This is the essence of the forward model in POSEIDON. As this model is specified in terms of parameters (e.g. the P-T profile parameters and mixing ratios), spectra can be flexibly evaluated for a plethora of different combinations of atmospheric properties. Armed with this model, I now turn to the essence of retrieval: the usage of a statistical algorithm to extract atmospheric properties from an observed spectrum.

2.3 Bayesian atmospheric retrieval

Ultimately, we are interested in revealing the underlying nature of an exoplanet atmosphere (i.e. parameters encoding the P-T profile, chemistry, clouds etc.) from an observed transmission spectrum. An additional question one must assess is the suitability of the forward model itself in light of the data. These two tasks, parameter estimation and model comparison, can be accomplished within a Bayesian framework.

The architecture of the POSEIDON retrieval algorithm is depicted in Figure 2.5. The atmosphere is encoded by a vector of parameters. For a given parameter combination, the forward model outputs a spectrum, convolves it with relevant instrument point

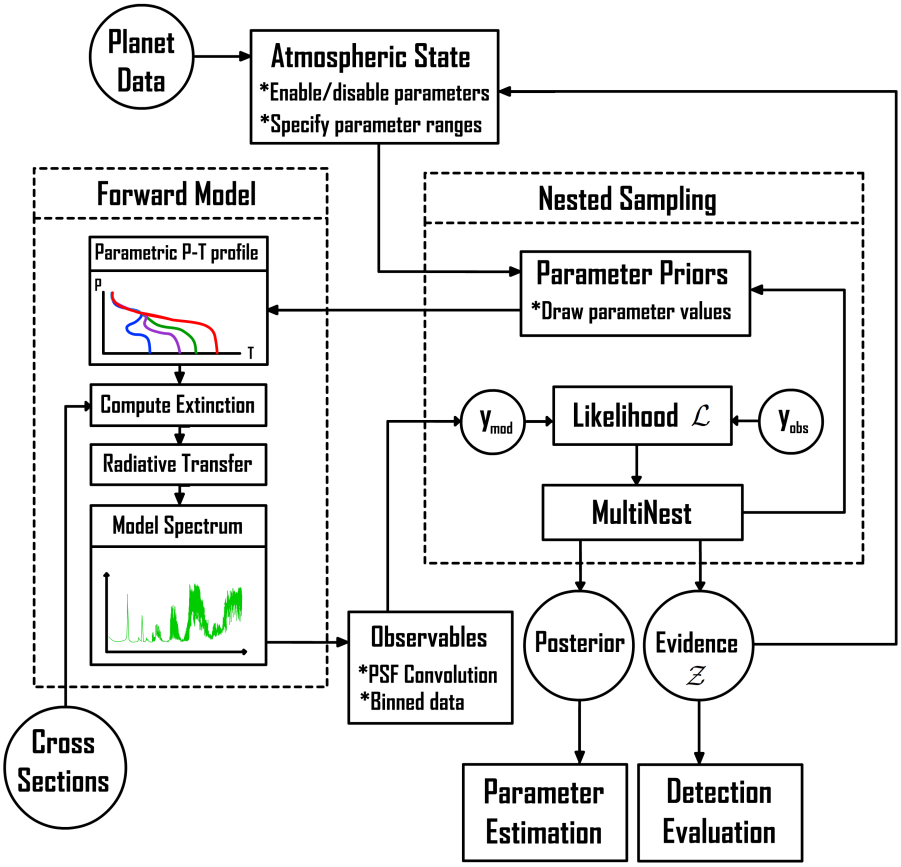


Fig. 2.5 Architecture of the POSEIDON atmospheric retrieval code. A forward model, described in section 2.2, is repeatedly called to generate transmission spectra for different parameter inputs – each selected by a sampling algorithm. The algorithm compares each simulated spectrum to a corresponding observed spectrum, outputting posterior samples and the Bayesian evidence of the model. The posterior allows parameter estimation, whilst the evidence allows Bayesian model comparison.

spread functions (PSF), and integrates over the respective instrument functions to produce simulated data points. At each point in parameter space, these predicted data points, $y_{\rm mod}$, are compared with observed data points, $y_{\rm obs}$, to compute the likelihood of each set of parameters. The likelihood, in turn, informs the choice of the next parameter set by the sampling algorithm. The statistical algorithm allows thorough exploration of the entire multidimensional parameter space, yielding parameter constraints, along with the Bayesian evidence used to quantify the suitability of the model itself. POSEIDON employs a nested sampling algorithm to accomplish this purpose.

I proceed to outline the conversion of a model spectrum into simulated observations in section 2.3.1, explain the forward model parametrisation in section 2.3.2, define the statistical aspects and terminology of atmospheric retrieval in section 2.3.3 and, finally, describe the nested sampling algorithm employed by POSEIDON in section 2.3.4.

2.3.1 Simulating observations

Once a model transmission spectrum, Δ_λ, has been computed, it needs to be rendered compatible with observations by actual instruments. This process has two steps:

1. Convolve Δ_λ with the point spread function (PSF) of each instrument[8] used to observe the target exoplanet. This degrades the high spectral resolution model to the native resolution of the instrument. The relevant convolution operation is expressed as[9]

$$\tilde{\Delta}_{\lambda,\,i} = \int_{\lambda_{l,i}-\epsilon}^{\lambda_{r,i}+\epsilon} \Delta_{\lambda'}\,\mathrm{PSF}\,(\lambda-\lambda')\,\mathrm{d}\lambda' \tag{2.34}$$

   where $\lambda_{l,i}$ and $\lambda_{r,i}$ are, respectively, the left and right bin edges for the i_{th} data point, ϵ is a small interval to account for flux just outside the bin contributing inside after convolution, and PSF is the point spread function of a particular instrument. I take $\epsilon = 2\sigma_{\rm PSF}$ (i.e. two PSF standard deviations).

2. Integrate the convolved spectrum over the instrument transmission function, binning to the same spectral resolution and wavelengths as the observations. This step accounts for variable instrument sensitivity to different λ during the binning process. The i_{th} simulated data point can hence be produced by

[8] This step is not necessary for photometric observations, which proceed to step 2 directly.

[9] Writing the convolution in this form implicitly assumes that the stellar intensity, $I_{\lambda,\,*}$, is sufficiently flat over the wavelength interval of a given data bin (see, e.g., Deming & Sheppard, 2017).

$$y_{\mathrm{mod,\,i}} = \frac{\int_{\lambda_{l,i}}^{\lambda_{r,i}} T(\lambda')\,\tilde{\Delta}_{\lambda',\,i}\,d\lambda'}{\int_{\lambda_{l,i}}^{\lambda_{r,i}} T(\lambda')\,d\lambda'} \tag{2.35}$$

where $T(\lambda)$ is the transmission function of a particular instrument[10].

Simulated observations must be rapidly produced for the purposes of atmospheric retrieval, as a large parameter space of potential atmospheres needs to be explored. A key consideration here is the appropriate wavelength grid on which to solve the radiative transfer equation. In Chapter 5, I will outline the creation of high spectral resolution ($R \sim 10^6$) cross sections. If the resulting opacity database is used at its native resolution, POSEIDON can compute line-by-line[11] transmission spectra. However, line-by-line computations are infeasible for atmospheric retrieval, wherein many millions of spectra are required. Instead, I employ the technique of *opacity sampling*.

Opacity sampling rests on the resolution of observations being many orders of magnitude less than line-by-line. If a model wavelength grid is initialised at an intermediate resolution – higher than the observations, but lower than the cross sections – then the opacity at each model wavelength can be approximated (or sampled) by the nearest wavelength in the cross section array. The radiative transfer calculation is then evaluated on the intermediate resolution wavelength grid, accelerating the computation of a single model. So long as the model grid contains sufficient sampled points within each data point bin, the error vs. the line-by-line case is small.

The opacity sampling technique is illustrated in Figure 2.6. A line-by-line transmission spectrum is fed through Equations 2.34 and 2.35 to produce simulated HST WFC3 observations (at $R \approx 60$). These binned points are considered 'ground-truth'. The left panel shows that cross sections sampled at 50× coarser resolution (R =20,000) produce very similar simulated data points. As the sampling resolution further decreases, differences between the simulated observations grow. However, even at a sampling resolution of R = 1,000 (right panel), differences are comparable to present observational uncertainties. The retrieval studies in this work usually adopt R = 2,000 as an optimal compromise between speed and accuracy (with a 'sanity-check' at R = 10,000 used to verify that reported constraints are consistent with a higher resolution).

[10] A repository of instrument functions can be found at the SVO Filter Profile Service.

[11] A 'line-by-line' model is high enough resolution to resolve the widths of individual spectral lines.

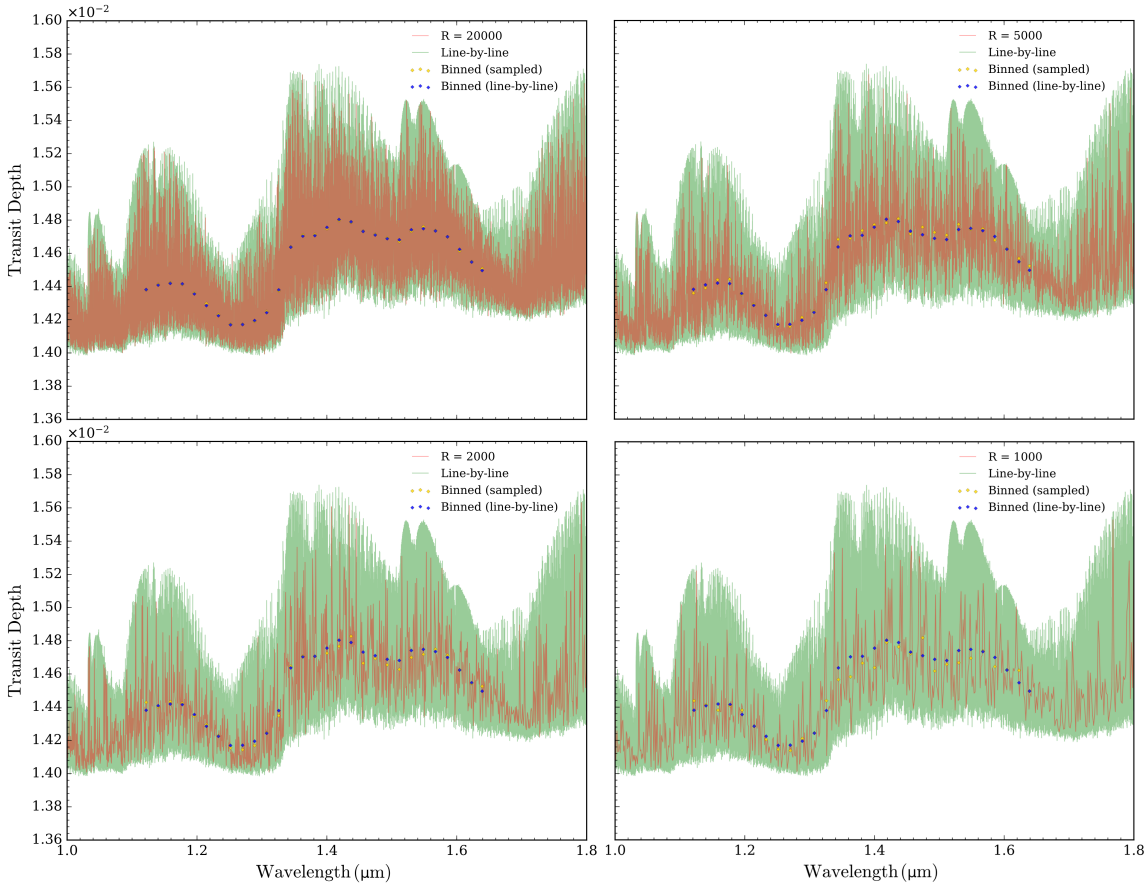


Fig. 2.6 Simulating observations at different spectral resolutions.
The four panels each show a high-resolution ($R \sim 10^6$) transmission spectrum (green) and an identical model with cross sections sampled at a lower resolution (red). From top left to bottom right, the cross sections are sampled at $R = 20000$, 5000, 2000, and 1000. Simulated HST WFC3 data points produced from the line-by-line model (blue diamonds) are compared against those from the sampled model (gold diamonds).

One may wonder why the cross sections are sampled to a lower resolution instead of being averaged over each desired bin. As shown in Figure 2.3 (see also Chapter 5), high-resolution cross sections can vary by orders of magnitude within each bin. A bin-averaged cross section thus results in overestimated optical depths, $\tau_{\text{binned}} \sim n\sigma_{\text{binned}}$, and hence the transmission, $e^{-\tau_{\text{binned}}}$, will be underestimated. From Equation 2.16, we see that using such band-averaged cross sections result in overestimated transit depths. A better approximation can be achieved from using log-averaged cross sections (accounting for the order-of-magnitude variability of the function), but a similar line of reasoning leads to log-averaged cross sections under-estimating the opacity in each bin, and hence systematically under-estimating the transit depth.

An alternative approach, commonly used in solar system atmosphere modelling, is the k-distribution method (e.g. Goody et al., 1989; Lacis & Oinas, 1991). This approach takes advantage of the fact that the radiative transfer equation depends on the magnitude of the atmospheric opacity at a given wavelength, but not the wavelength itself. Different wavelengths with similar opacities in a given bin can therefore be grouped together to form an opacity distribution function. Unlike cross sections, which vary by orders-of-magnitude in a non-trivial manner, the opacity distribution function, and hence its cumulative distribution function, are both relatively smooth functions. Mean spectral quantities in low-resolution bins, such as the transmissivity, can then be readily computed by integrating over the cumulative distribution function of the opacity within each bin. The key advantage here is a dramatic decrease in the number of bins in which the radiative transfer equation needs to be solved.

However, practical applications of the k-distribution method encounter some drawbacks. Strictly speaking, the k-distribution method only holds for a homogeneous atmosphere (i.e. constant P, T, and composition). For realistic atmospheres, one has to assume that a cross section evaluated at two wavelengths which fall in the same cumulative distribution function bin in one layer, $\sigma_{\lambda_{1,2}}$, remain in the same bin (i.e. are 'correlated') throughout the atmosphere. This is the 'correlated-k approximation' (Goody et al., 1989). The correlated assumption breaks down if, for example, σ_{λ_1} increases in strength for higher T more rapidly than σ_{λ_2}, such that when going between different atmospheric layers they no longer fall in the same cumulative opacity bin. Furthermore, combining k-distributions for different gases is not as trivial as adding extinction coefficients (e.g. see Mollière et al., 2015, Appendix B), and relies on assuming the distributions for different gases are spectrally uncorrelated in the output low-resolution wavelength bins. Despite these assumptions, the correlated-k approximation typically achieves binned accuracies within $\sim 5-10\%$ of a line-by-line solution (Lacis & Oinas, 1991; Amundsen et al., 2014; Mollière et al., 2015). A detailed comparison between opacity sampling and correlated-k approaches is beyond the scope of the present work, with opacity sampling used from here on.

2.3.2 Atmospheric parametrisation

The POSEIDON forward model is expressed in terms of a parametric state vector, which encodes the pressure-temperature profile, composition, and cloud / haze properties of a model atmosphere. Where appropriate, additional 'nuisance' parameters, such as relative offsets between datasets, can be added. The combined parameter vector usually contains $\approx 15 - 30$ free parameters, on which I will now elaborate.

The P-T profile can be described either by a single-parameter isotherm or the 6 parameter profile (α_1, α_2, T_0, P_1, P_2, P_3) discussed in section 2.2.2. Solving the hydrostatic equation also requires a reference pressure parameter, $P_{\rm ref}$ (section 2.2.2). I consider $P_{\rm ref}$ to also be a P-T profile parameter (as it is required to compute the radial profile, r). Each P-T profile is therefore encoded by either 2 or 7 parameters.

The atmospheric composition (section 2.2.2) is parametrised by the terminator mixing ratios, X_i, of a range of chemical species. Since the following studies focus on gaseous exoplanets, it can be safely assumed that a large fraction of the atmosphere is composed of H_2 and He (from their known bulk density). I therefore treat H_2 + He as a single gas with a fixed solar ratio ($X_{\rm He}/X_{\rm H_2} = 0.17$) and compute its contribution via[12] $X_{\rm H_2+He} = 1 - \sum_i X_i$ (as the mixing ratios sum to 1 by definition).

The 2D cloud and haze model employed by POSEIDON (section 2.2.4) is described by 4 parameters: a, γ, $P_{\rm cloud}$, and $\bar{\phi}$. Subsets of this general prescription can be used for cloud model comparisons. For example, a 'uniform cloud deck' model has $\bar{\phi}$ fixed to 1 (and may or may not include the two haze parameters).

With a parametric transmission spectrum model, and a prescription to convert the model output into simulated observations, the POSEIDON forward model can now be placed within a Bayesian framework to enable atmospheric retrieval.

2.3.3 Bayesian framework

Consider a set of forward models, M_i, described by a set of physical parameters $\boldsymbol{\theta}$. Our *a priori* expectations on the values of the parameters are encoded in the *prior probability density function*: $p\left(\boldsymbol{\theta}|M_i\right)$. By obtaining a set of observations $\boldsymbol{y}_{\rm obs}$, one can formally update our knowledge on the values of these parameters via *Bayes' theorem*

$$p\left(\boldsymbol{\theta}|\boldsymbol{y}_{\rm obs}\,, M_i\right) = \frac{p\left(\boldsymbol{y}_{\rm obs}|\boldsymbol{\theta}\,, M_i\right)\, p\left(\boldsymbol{\theta}|M_i\right)}{p\left(\boldsymbol{y}_{\rm obs}|M_i\right)} \equiv \frac{\mathcal{L}\left(\boldsymbol{y}_{\rm obs}|\boldsymbol{\theta}\,, M_i\right)\, \pi\left(\boldsymbol{\theta}|M_i\right)}{\mathcal{Z}\left(\boldsymbol{y}_{\rm obs}|M_i\right)} \tag{2.36}$$

[12]This assumes H_2 and He are the only non-parametrised abundant gases in the atmosphere.

where in the equivalency I have defined the conventional notation for the *likelihood function*, $\mathcal{L}$, prior, π, and *Bayesian evidence*, $\mathcal{Z}$. The quantity on the left-hand side is the *posterior probability density function*, which quantifies our knowledge on the values of each parameter in model M_i following an observation.

The priors for each parameter underlying the forward model (section 2.3.2) are given in Table 2.1. I elect for generous 'uninformative' priors. A uniform prior probability is ascribed to parameters expected to vary by less than two orders of magnitude (e.g. T_0), whilst a uniform-in-the-logarithm prior is used for parameters expected to vary over many orders of magnitude (e.g. $P_{\rm ref}$). An additional subtlety in the choice of these priors must be made explicit: since both the mixing ratios, X_i, and the reduced polar extent, $\bar{\phi}_i$, must sum to unity (by definition), in the most general case a flat Dirichlet prior (uniform over a simplex subspace) is most appropriate. Here, I use the fact that we know *a priori* from the bulk density of gaseous exoplanets that the dominant atmospheric component is H_2 + He to treat the remaining gases as trace species with log-uniform priors – this assumption must be relaxed for high mean molecular weight atmospheres, such as those of super-Earths (Benneke & Seager, 2012). Similarly, a terminator divided into two regions needs only a single uniform parameter $\bar{\phi}_1$, with the extent of the second region automatically specified by $\sum_i \bar{\phi}_i = 1$.

The likelihood, $\mathcal{L}\left(\boldsymbol{y}_{\rm obs}|\boldsymbol{\theta}\,, M_i\right)$, is a measure of the plausibility of the forward model producing the observed data for a choice of model parameters. For observations with independently distributed Gaussian errors, the likelihood is given by

$$\mathcal{L}\left(\boldsymbol{y}_{\rm obs}|\boldsymbol{\theta},\, M_i\right) = \prod_{k=1}^{N_{\rm obs}} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{\left[y_{{\rm obs},\,k} - y_{{\rm mod},\,k}\left(\boldsymbol{\theta}\right)\right]^2}{2\sigma_k^2}\right) \tag{2.37}$$

where $N_{\rm obs}$ is the number of observed transit depths, σ_k is the standard deviation (i.e. the *precision* or observational uncertainty) on the k_{th} observed transit depth data point, and $y_{{\rm mod},\,k}\left(\boldsymbol{\theta}\right)$ is the k_{th} simulated model data point (produced by the procedure explained in section 2.3.1).

The Bayesian evidence is the key quantity employed in Bayesian model comparison. From Equation 2.36, we see that it is simply the normalising factor that ensures the integral of the posterior probability density over the entire parameter space evaluates to unity. The evidence can therefore be written as

$$\mathcal{Z}\left(\boldsymbol{y}_{\rm obs}|M_i\right) = \int_{{\rm all}\,\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{y}_{\rm obs}|\boldsymbol{\theta},\, M_i\right) \pi\left(\boldsymbol{\theta}|M_i\right)\, {\rm d}\boldsymbol{\theta} \tag{2.38}$$

Table 2.1 Typical priors for atmospheric retrievals with POSEIDON

Parameter	Prior	Typical range
P-T profile		
$\alpha_{1,2}$	Uniform	0.02 – 2.0 $\mathrm{K}^{-1/2}$
$P_{1,2}$	Log-uniform	Full atmosphere[†]
P_3	Log-uniform	10^{-2} – 10^{2} bar
P_{ref}	Log-uniform	Full atmosphere
T_0	Uniform	T_{eq} dependent
Composition		
X_i	Log-uniform	10^{-12} – 10^{-2}
Clouds		
a	Log-uniform	10^{-4} – 10^{8}
γ	Uniform	-20 – 2
P_{cloud}	Log-uniform	Full atmosphere
$\bar{\phi}$	Uniform[‡]	0 – 1

[†] The P-T profile parametrisation of Madhusudhan & Seager (2009) typically also imposes $P_2 < P_1 < P_3$ for transmission spectra retrievals, so samples violating these additional constraints are rejected in all retrievals in this thesis.

[‡] For more than $N = 2$ terminator regions, a Dirichlet prior is more appropriate.

Notes: 'Full atmosphere' is the modelled pressure range, usually 10^{-6} – 10^{2} bar. The T_0 prior spans a generous range around the estimated planet equilibrium temperature, T_{eq}. X_i represents the set of mixing ratios, i denotes any chemical species with a parametrised abundance. The upper bound for X_i is less consequential for trace species in H_2+He dominated atmospheres, but samples with $\sum_i X_i > 1$ must be rejected.

The utility of the evidence can be qualitatively understood as a 'figure of merit' that is maximised by models with a high likelihood in a compact parameter space (Trotta, 2008). To see this, consider the addition of a new parameter to a model. By extending the dimensionality of the parameter space, the value of the prior probability density, π, will be diluted across this additional volume. The evidence for this more complex model will then only increase if the new volume contains previously unsampled regions of high likelihood. If one wishes to go beyond simply fitting a spectrum to compare different potential models, the Bayesian evidence therefore provides an automatic implementation of Occam's Razor by penalising those models with unjustified complexity.

The quantitative use of the evidence becomes clear when using Bayes' theorem to consider the relative probability of two competing models in light of the data

$$\frac{p\left(M_i|\boldsymbol{y}_{\rm obs}\right)}{p\left(M_j|\boldsymbol{y}_{\rm obs}\right)} = \frac{\mathcal{Z}\left(\boldsymbol{y}_{\rm obs}|M_i\right)}{\mathcal{Z}\left(\boldsymbol{y}_{\rm obs}|M_j\right)} \frac{p\left(M_i\right)}{p\left(M_j\right)} \equiv \mathcal{B}_{ij} \frac{p\left(M_i\right)}{p\left(M_j\right)} = \mathcal{B}_{ij} \tag{2.39}$$

where in the equivalency I defined the *Bayes factor* for model i vs. model j as the ratio of the evidences of the two models. In the last equality, I assume the *a priori* odds ratio of the two models to be unity (given no reason to favour either model initially). We see that $\mathcal{B}_{ij} > 1$ favours model i, whilst $\mathcal{B}_{ij} < 1$ favours model j. Specifically, values of at least $\mathcal{B}_{ij} = 3, 12, 150$ are often interpreted as 'weak', 'moderate', and 'strong' evidence, respectively, in favour of M_i over M_j (Trotta, 2008; Benneke & Seager, 2013). These empirically calibrated evidence thresholds are known as Jeffreys' scale (Jeffreys, 1939). The Bayes factor can be related to the commonly used frequentist measure of sigma-significance by inverting the relations (Sellke et al., 2001)

$$\mathcal{B}_{ij} \leqslant -\frac{1}{e\, p \ln p} \tag{2.40}$$

$$p = 1 - \text{erf}\left(\frac{N_\sigma}{\sqrt{2}}\right) \tag{2.41}$$

where p is the 'p-value', erf is the error function, and N_σ is the 'detection significance'. This mapping converts a Bayes factor into an upper bound on the p-value (valid for $p \leqslant e^{-1}$), and hence a lower bound on the detection significance. From these equations, Bayes factors of 3, 12, and 150 are equivalent to 2.1σ, 2.7σ, and 3.6σ, respectively. These Bayes factors are taken as thresholds, with any $\mathcal{B}_{ij}$ exceeding them referred to as 'weak', 'moderate', or 'strong' evidence, respectively. The phrase 'detection' is usually reserved for detection significances $> 5\sigma$.

Once an adequate model is identified via Bayesian model comparison, one can constrain the parameters of the chosen model. This is accomplished by drawing samples from the posterior (Equation 2.36) and marginalising (integrating) over the full range of the other parameters (as illustrated in Figure 1.19). The resulting probability density histograms of each parameter encode our knowledge of the atmospheric state. Strictly speaking, the Bayesian evidence is not required if parameter estimation is the only goal ($\mathcal{Z}$ only normalises the posterior). However, if the evidence is not computed then one is implicitly assuming that the model chosen is itself 'correct'. I therefore conduct *both* Bayesian model comparisons and parameter estimation in this work.

2.3.4 Nested sampling

Traditional sampling techniques do not permit the computationally efficient evaluation of multidimensional evidence integrals. The technique of *nested sampling* (Skilling, 2004), however, was designed to compute $\mathcal{Z}$ with posterior samples produced as a byproduct. The ability of nested sampling to simultaneously pursue parameter estimation and model comparison has resulted in its adoption as the technique of choice in many recent atmospheric retrieval studies (e.g. Benneke & Seager, 2013; Waldmann et al., 2015a; Line & Parmentier, 2016).

The central concept behind nested sampling is the conversion of Equation 2.38 into a one-dimensional integral

$$\mathcal{Z} = \int_{\text{all}\,\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{y}_{\text{obs}}|\boldsymbol{\theta}\right) \pi\left(\boldsymbol{\theta}\right)\, d\boldsymbol{\theta} \rightarrow \int_0^1 \mathcal{L}(X)\, dX \tag{2.42}$$

where X is the 'prior volume'[13], defined by

$$X(\mathcal{L}^*) = \int_{\mathcal{L}(\boldsymbol{\theta})>\mathcal{L}^*} \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} \tag{2.43}$$

and $\mathcal{L}^*$ is a likelihood threshold. We see that $X(\mathcal{L}^*)$ is the integral of the prior probability density over the region of parameter space satisfying $\mathcal{L} > \mathcal{L}^*$. When $\mathcal{L}^* = 0$, normalisation of the prior implies that $X_{\text{max}} = 1$. Similarly, when $\mathcal{L}^* > \mathcal{L}_{\text{max}}$, the maximum likelihood, we have that $X_{\text{min}} = 0$ by definition. As $X(\mathcal{L}^*)$ is a monotonically decreasing function, the inverse function, $\mathcal{L}(X)$, will also possess this property. This permits the transformation in Equation 2.42.

The evidence can be approximately computed by numerical evaluation of the transformed one-dimensional integral in Equation 2.42 via

$$\mathcal{Z} \approx \sum_{i=1}^{N_{\text{samp}}} \mathcal{L}_i\, w_i \tag{2.44}$$

where N_{samp} is the number of sampled likelihood values, and $w_i = \frac{1}{2}(X_{i-1} - X_{i+1})$ are trapezoidal weights. The set of sampled likelihood values, $\mathcal{L}_i$, and associated prior volumes, X_i, are produced by a nested sampling algorithm. I will now briefly outline the steps involved in nested sampling. For more detailed discussions, I refer the interested reader to Skilling (2004), Feroz & Hobson (2008), or Speagle (2019).

[13]Not to be confused with mixing ratios.

Nested sampling algorithms essentially follow steps similar to the following:

1. Randomly draw a user-specified number of 'live points', $N_{\rm live}$, from the prior.
2. Sort the samples by likelihood, identifying the point of minimum likelihood, $\mathcal{L}_{\rm min}$.
3. Remove the sample with $\mathcal{L}_{\rm min}$ (i.e. make it a 'dead point'). At iteration i, call this point $\mathcal{L}_i$.
4. Draw new samples from the prior until one is found satisfying $\mathcal{L} > \mathcal{L}_{\rm min}$. The new sample becomes 'live', replacing the dead point.
5. Compute X_i for the shrunk volume within the iso-likelihood contour $\mathcal{L} > \mathcal{L}_i$. As the prior volume within each contour is related to the previous volume by a random multiplicative factor less than 1, the probability distribution of X_i yields $\log(X_i) \approx -(i + \sqrt{i})/N_{\rm live}$, and hence $X_i \approx e^{-i/N_{\rm live}}$ is usually assumed.
6. Check whether the remaining contribution to the evidence (estimated by $\Delta\mathcal{Z}_i \approx \mathcal{L}_{\rm max} X_i$) is less than a specified tolerance. If not, return to step 2. If true, terminate iterations.
7. Remove the final set of live points one-by-one according to the current value of $\mathcal{L}_{\rm min}$, setting $N_{\rm live} = N_{\rm live} - 1$ in each step to produce a final additional sequence of $\mathcal{L}_i$ and X_i (for increased accuracy).
8. Evaluate $\mathcal{Z}$ via Equation 2.44 for the full set of recorded $\mathcal{L}_i$ and X_i.
9. Finally, the posterior is constructed from the probabilities of all live and dead points according to

$$p_i = \frac{\mathcal{L}_i w_i}{\mathcal{Z}} \tag{2.45}$$

The progression of a nested sampling algorithm is graphically depicted in Figure 2.7. The left panel shows that iso-likelihood contours, containing nested shells of prior volume, are progressively swept through towards regions in which the highest likelihoods are localised. The right panel visually represents the integration of the likelihood over prior volume, with the evidence given by the area under the curve.

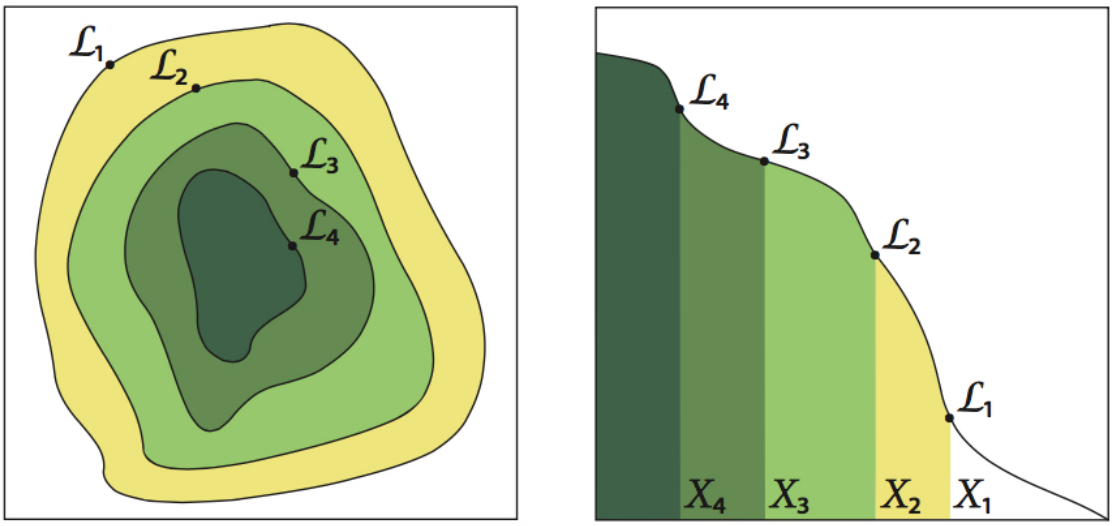


Fig. 2.7 Schematic illustration of the nested sampling algorithm.
Left: iso-likelihood contours sweep through the posterior of a 2D parameter space towards higher likelihoods. Right: integration of the likelihood over prior volume, with the evidence given by the area under the curve. Credit: Feroz et al. (2013).

MultiNest

POSEIDON utilises the multimodal nested sampling algorithm MultiNest (Feroz & Hobson, 2008; Feroz et al., 2009, 2013), implemented via the Python wrapper PyMultiNest (Buchner et al., 2014). MultiNest follows an algorithm similar to that outlined above, but with the live points at each iteration drawn from progressively shrinking ellipsoids. Ellipsoidal sampling increases the acceptance rate (vs. random draws from the full prior) once live points begin to cluster in regions of high likelihood (see Feroz et al., 2009, for a detailed description of this procedure). A notable strength of MultiNest in the context of atmospheric retrieval is its ability to navigate significantly non-Gaussian, degenerate, and non-trivially curved posteriors. MultiNest is also fully parallelised, enabling large-scale cluster computing.

By coupling the POSEIDON forward model to MultiNest, the plausibility of a wide variety of model atmospheres can be assessed (e.g. detections of chemical species and / or clouds). During a typical retrieval analysis, I run POSEIDON for > 10 nested models to evaluate detection significances. For each model, I usually take $N_{\rm live}$ =4,000, typically requiring $\sim 5 \times 10^6$ transmission spectra computations to obtain accurate evidences and robust parameter estimates. With the forward model and retrieval framework established, I proceed to demonstrate their combined effectiveness.

2.4 Validating a new retrieval code

There are two steps to validate a retrieval code: (i) verify the forward model outputs correct spectra; and (ii) successfully retrieve the atmospheric state of a simulated dataset produced by the forward model. Tests of this manner enable the accuracy and reliability of a retrieval code to be established, as well as bringing to light any biases or degeneracies in the results which could impact the interpretation of real spectra.

2.4.1 Forward model validation

Prior to the generation of a simulated dataset, I undertook an internal comparison between the POSEIDON forward model and that utilised in Madhusudhan et al. (2014b). The radiative transfer schemes and parametric P-T profiles in both models are identical, but the latter only produces cloud-free spectra. To permit a direct comparison, POSEIDON was set to simulate a cloud-free atmosphere and I temporarily replaced POSEIDON's cross sections with the same cross sections employed by Madhusudhan et al. (2014b). By ensuring both codes have the same inputs and simulate transmission spectra for an identical atmosphere, any differences must arise from the codes themselves. The comparison yielded agreement over a range of simulated atmospheres, with the only difference being a minor offset ($\sim 10\,\mathrm{ppm}$) due to the higher-order numerical methods employed by POSEIDON.

I then undertook a comparison with an external model from the published literature. Specifically, Robinson (2017) presented (in his Fig. 7) a model transmission spectrum for a simple hot Jupiter with all parameters specified. Namely, a notional planet with $R_p = 1.16\,R_J$, $R_* = 0.78\,R_\odot$, and $M_p = 1.14\,M_J$ (defining the surface gravity, g_p). The atmosphere is isothermal at $T = 1500\,\mathrm{K}$, divided into 126 layers spaced uniformly in $\log(P)$ from $10^{-9} - 10\,\mathrm{bar}$, with $P_{\mathrm{ref}} = 10\,\mathrm{bar}$. The composition is $X_{\mathrm{H_2}} = 0.85$, $X_{\mathrm{He}} = 0.15$, and $X_{\mathrm{H_2O}} = 4 \times 10^{-4}$. This model is a standard benchmark in the literature (e.g. Line et al., 2013; Heng & Kitzmann, 2017).

Figure 2.8 presents the comparison between a transmission spectrum generated by POSEIDON and that shown in Robinson (2017)'s Fig. 7. As the line-by-line spectrum produced by POSEIDON is at a much higher spectral resolution ($R \sim 10^6$ vs. $R \sim 10^2$), I bin the native resolution spectrum down to a comparable resolution ($R = 80$). The agreement between the two models is remarkable, especially considering the opacities are computed using different resolutions, techniques, and line lists (T. Robinson, personal communication). Indeed, the only difference which emerged from

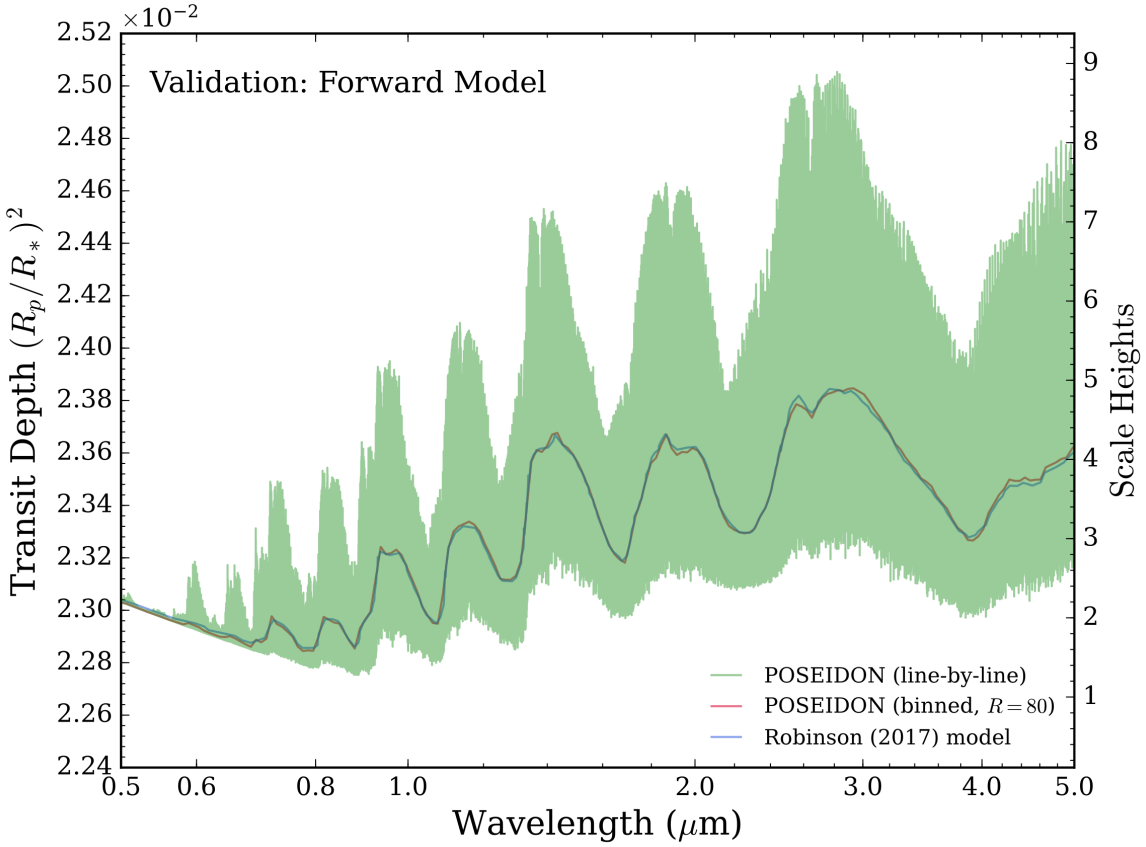


Fig. 2.8 Validating POSEIDON's forward model via an independent model. A line-by-line transmission spectrum computed by POSEIDON (green) is generated using identical parameters to a model spectrum in Robinson (2017) (blue). To render these models comparable, the spectrum from POSEIDON is binned down to a similar spectral resolution of $R = 80$ (red). The two model spectra are in excellent agreement.

this comparison was a differing definition in the value of R_J (I use the equatorial Jovian radius, whilst Robinson (2017) used the volumetric mean radius). This excellent agreement not only validates POSEIDON against the model in Robinson (2017), but also indirectly validates it against a number of codes that have undergone the same comparison (e.g. Line et al., 2013; Irwin et al., 2008; Waldmann et al., 2015a).

2.4.2 Retrieval validation

I now proceed to demonstrate the typical results from an application of POSEIDON to retrieve a transmission spectrum dataset. The goal here is to start with a synthetic dataset, based on a known model spectrum, and use POSEIDON to retrieve the underlying model parameters. This will reveal how well the parameters can be recovered.

I first generated a moderate resolution ($R \approx 10,000$) transmission spectrum. The parameters are chosen to resemble the properties of HD 209458b. The P-T profile was chosen to possess a monotonically decreasing temperature with altitude, such that the temperature in the photosphere ($P \sim 10^{-2}$ bar) is around the planetary equilibrium temperature ($T_{\rm phot} \approx 1400$ K). The molecular composition is chosen to resemble solar abundances in chemical equilibrium. A terminator cloud coverage of 40% was ascribed, with cloudy regions consisting of a high-altitude cloud deck at $P_{\rm cloud} = 0.1$ mbar subsumed in a uniform-with-altitude haze. The values of the parameters used to produce the spectrum are given in the embedded table in Figure 2.10.

With the spectrum generated, I produced a synthetic dataset at a precision commensurate with currently available observations. Following the procedure in section 2.3.1, I convolved the model spectrum with the PSFs for the HST STIS (G430/G750) and WFC3 (G141) instruments and integrated over the corresponding instrument functions to produce a set of low-resolution binned simulated data points. The wavelength locations of the simulated data and binning resolution was chosen to match those of the real transmission spectrum observations of HD 209458b presented in Sing et al. (2016). Gaussian scatter at the levels of 120 ppm and 40 ppm are added to the visible and near-infrared simulated observations, respectively - commensurate with the real STIS and WFC3 data precisions for this planet. This combined visible + near-infrared dataset served as the input to POSEIDON.

For the purposes of this validation test, retrievals were run with 16 parameters. That is, 7 for the P-T profile, 4 for clouds / hazes, and 5 compositional parameters (H_2O, CH_4, NH_3, HCN, and Na+K[14]). During these retrievals, transmission spectra were evaluated at 2,000 wavelengths sampled uniformly from 0.2–2.0 µm. I initialised multiple nested sampling runs with $N_{\rm live}$ =1,000 - 8,000 to confirm consistency. I find that 4,000 live points offers an optimum trade-off between minimising Bayesian evidence computation error ($\Delta \ln \mathcal{Z} \approx 0.05$) and reducing the time necessary to reach the termination criteria. A typical retrieval with 4,000 live points generates $\sim 5 \times 10^6$ spectra, taking $\lesssim 10$ hr on a standard 4-core 3 GHz desktop computer. The following results are from the 4,000 live point retrieval.

Figure 2.9 compares the true spectrum and P-T profile underlying the synthetic data to those retrieved by POSEIDON. The median retrieved spectrum is in excellent agreement with the true spectrum, remaining coincident from 0.3–2.0 µm to < 40 ppm precision. The WFC3 band pass is constrained even more tightly, typically to < 20 ppm

[14]Na+K is here treated as a single gas in a solar ratio. This assumption is relaxed in later chapters.

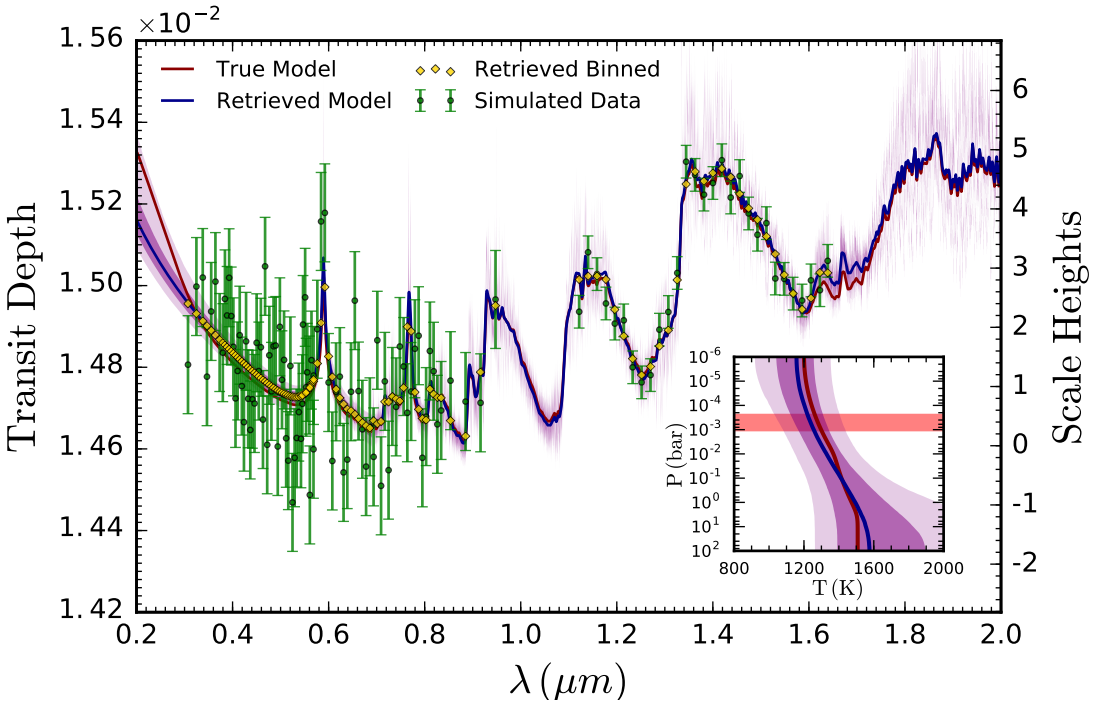
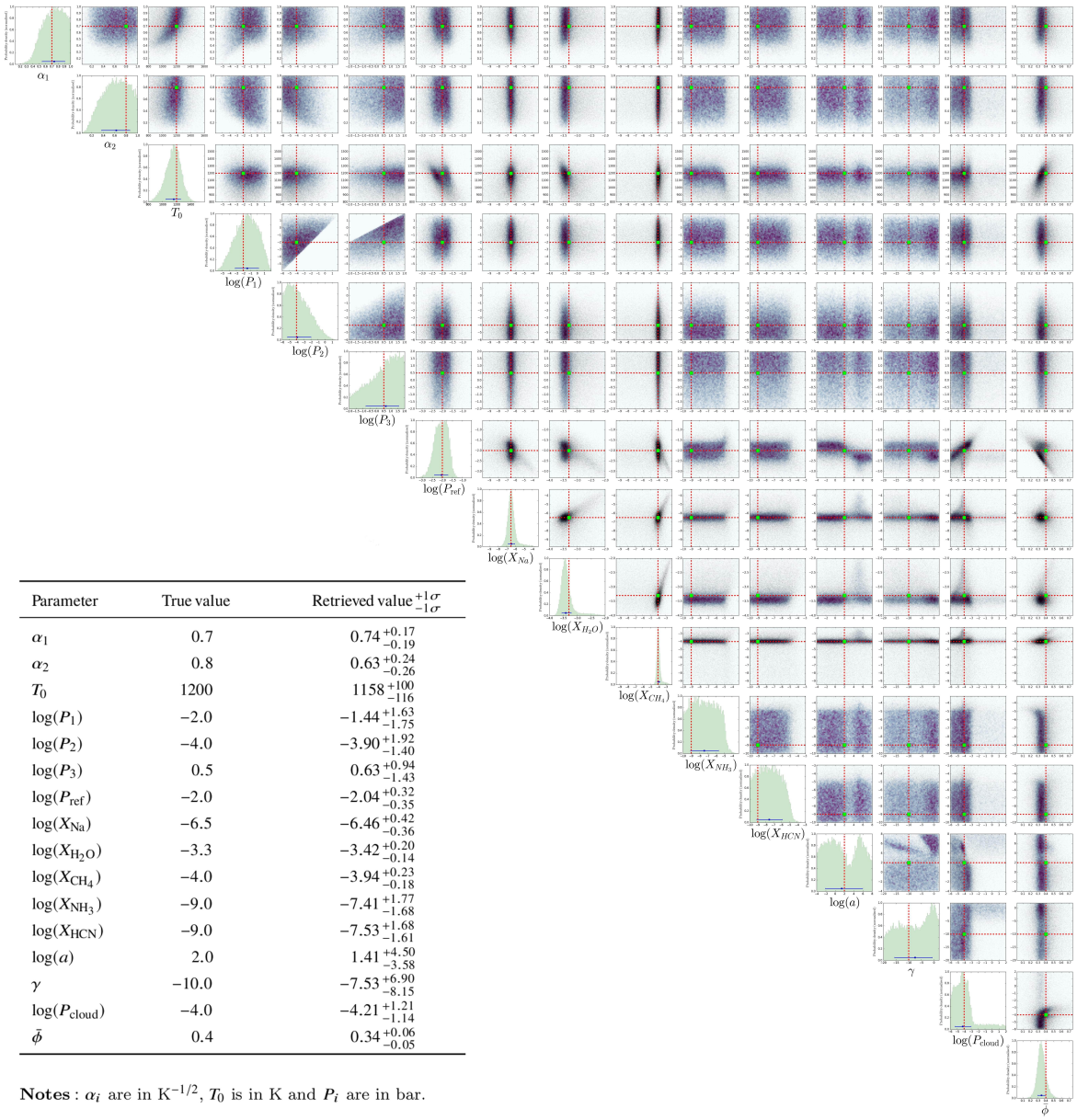


Fig. 2.9 Retrieving a synthetic spectrum and P-T profile with POSEIDON. A moderate resolution ($R \approx$10,000) model transmission spectrum (red) is converted into synthetic WFC3 and STIS data points (green). A retrieval of this data produces posterior samples, from which a median retrieved spectrum (blue), 1σ confidence region (dark purple shading), and 2σ confidence region (light purple shading) are constructed. The true and retrieved spectra have been Gaussian smoothed for clarity. The median retrieved model, binned to the data resolution, is shown by gold diamonds. The right y-axis shows equivalent scale heights above $(R_p/R_*)^2$. Inset: the retrieved terminator P-T profile. The true profile (red) and median retrieved profile (blue) lie within the 1σ confidence region. The horizontal red band roughly corresponds to the $\tau \approx 1$ pressure range at 1.5 µm(a near-infrared photosphere proxy).

precision. The only region of significant deviation is the UV below 0.3 µm, where no data points inform the retrieval. The true P-T profile agrees with median retrieved profile to < 50 K. The 1σ profile extent is tightly constrained to ~ 100 K, with the contours matching the shape of the profile with altitude. As expected, the constraints become tighter around the infrared photosphere ($P \sim 10^{-2}$ bar), and expand for pressures away from those probed in transmission geometry. This demonstrates that, given sufficiently high-precision transmission spectra, the shape of terminator P-T profiles can be correctly inferred. I will further explore differences between retrievals with a flexible P-T profile vs. an assumed isotherm in Chapters 6 and 7.



Parameter	True value	Retrieved value $^{+1\sigma}_{-1\sigma}$
α_1	0.7	$0.74^{+0.17}_{-0.19}$
α_2	0.8	$0.63^{+0.24}_{-0.26}$
T_0	1200	1158^{+100}_{-116}
$\log(P_1)$	−2.0	$-1.44^{+1.63}_{-1.75}$
$\log(P_2)$	−4.0	$-3.90^{+1.92}_{-1.40}$
$\log(P_3)$	0.5	$0.63^{+0.94}_{-1.43}$
$\log(P_{\rm ref})$	−2.0	$-2.04^{+0.32}_{-0.35}$
$\log(X_{\rm Na})$	−6.5	$-6.46^{+0.42}_{-0.36}$
$\log(X_{\rm H_2O})$	−3.3	$-3.42^{+0.20}_{-0.14}$
$\log(X_{\rm CH_4})$	−4.0	$-3.94^{+0.23}_{-0.18}$
$\log(X_{\rm NH_3})$	−9.0	$-7.41^{+1.77}_{-1.68}$
$\log(X_{\rm HCN})$	−9.0	$-7.53^{+1.68}_{-1.61}$
$\log(a)$	2.0	$1.41^{+4.50}_{-3.58}$
γ	−10.0	$-7.53^{+6.90}_{-8.15}$
$\log(P_{\rm cloud})$	−4.0	$-4.21^{+1.21}_{-1.14}$
$\bar{\phi}$	0.4	$0.34^{+0.06}_{-0.05}$

Notes : α_i are in $\mathrm{K}^{-1/2}$, T_0 is in K and P_i are in bar.

Fig. 2.10 Posterior from a synthetic data retrieval with POSEIDON. This posterior distribution results from an atmospheric retrieval of the synthetic data in Figure 2.9. The corner plot depicts correlations between pairs of retrieved parameters and marginalised histograms for each parameter. The true parameter values are indicated by red dashed lines and green squares, with the median retrieved values and $\pm 1\sigma$ constraints shown by blue error bars on each histogram. Table inset: summary of the true parameter values underlying the synthetic data, along with the median retrieved values and 1σ errors. All parameters are correctly retrieved to within 1σ.

The full posterior distribution is shown in Figure 2.10. It is seen that POSEIDON correctly retrieves all the parameters used to generate the synthetic dataset to within

1σ. The most tightly constrained parameters are the abundances of H_2O and CH_4, at ≈ 0.2 dex precision. Where absorption signatures of a molecular species are not deemed necessary to explain the data, such as with the low NH_3 and HCN abundances, the posterior retains the flat shape of the prior below an established upper bound. The mode of the cloud deck pressure is coincident with the true value, with the terminator cloud fraction sharply localised. The haze properties, however, remain relatively unconstrained. This is unsurprising, as light transmitting through the relatively small fraction of the model atmosphere above the 0.1 mbar cloud deck will be relatively insensitive to the haze (at the noise level of the simulated optical data). Indeed, the small tails in the chemical abundance posteriors towards higher values can be seen as arising from a weak correlation with the possibility of strong ($a \approx 10^6$) hazes. For data where the scattering slope can be readily resolved, this tail is not present.

Breaking the cloud-composition degeneracy

An important conclusion from the posterior distribution in Figure 2.10 is that the chemical inferences are relatively independent of the values of the cloud parameters. This can be seen from the roughly horizontal correlations between the mixing ratios of the constrained chemical species and the cloud parameters. This lack of a strong cloud-composition degeneracy is in striking contrast to the classical picture of the influence of a uniform cloud deck on transmission spectra (e.g. Benneke, 2015). However, the cloud model used by POSEIDON does not necessarily require *uniform* cloud decks. By accounting for the possibility of partially cloudy (and therefore partially clear) terminators, POSEIDON has the ability to 'peer beneath the clouds' and obtain *simultaneous* precise constraints on compositions and cloud properties from transmission spectra. Qualitatively, this ability is afforded by the clear region of the terminator, which enables retrievals to disentangle the reference pressure from the cloud deck pressure; hence breaking the degeneracy between clouds and chemistry.

To reiterate: ***non-uniform terminator cloud coverage can enable precise determination of chemical abundances from transmission spectra***. This finding, first published in MacDonald & Madhusudhan (2017a), has been reaffirmed by subsequent retrieval studies using the 2D cloud and haze prescription introduced in this chapter (Pinhas et al., 2019; Welbanks & Madhusudhan, 2019). As most gaseous exoplanets are expected to have partially cloudy, rather than fully cloudy, terminators (e.g. Parmentier et al., 2013), I propose that inhomogenous clouds are a natural solution to the problem of inferring the composition of cloudy exoplanet atmospheres.

2.5 Summary

In this chapter, I have presented a new approach to the atmospheric retrieval of exoplanets. The algorithm herein developed provides a general framework to extend traditional 1D retrieval techniques to permit the extraction of 2D terminator properties from exoplanet transmission spectra. The key results from this chapter are as follows:

1. I have created a new atmospheric retrieval code, *POSEIDON*, capable of extracting precise atmospheric properties from transmission spectra of cloudy exoplanets.

2. POSEIDON exploits the insight that exoplanet clouds are likely to be inhomogeous to break cloud-composition degeneracies.

3. POSEIDON's radiative transfer have been validated against independent codes, demonstrating correct computation of transmission spectra.

4. An initial test retrieval on synthetic Hubble observations showed that precise constraints on exoplanet atmosphere compositions, temperature structures, and cloud properties can be obtained with current observations of cloudy exoplanets.

The following chapters will now turn to apply POSEIDON to interpret real observations of exoplanet transmission spectra.

Chapter 3

HD 209458b in New Light

3.1 The archetypal hot Jupiter

HD 209458b was the first exoplanet discovered by the transit method (Charbonneau et al., 2000). As a prototypical hot Jupiter, with $T_{\rm eq} \approx 1400\,{\rm K}$, it was quickly recognised as an ideal target for transmission spectroscopy (Seager & Sasselov, 2000). Within two years of this prediction, the atmosphere of HD 209458b was detected by a Hubble Space Telescope (HST) transmission spectrum, revealing Na absorption (Charbonneau et al., 2002). The years since have seen HD 209458b serve as a proving ground for atmospheric characterisation, igniting the fledgling field of exoplanetary atmospheres. Various molecular species containing carbon and oxygen have been claimed in its atmosphere, including H_2O, CH_4, CO, and CO_2 (Barman, 2007; Swain et al., 2009; Madhusudhan & Seager, 2009; Snellen et al., 2010; Deming et al., 2013). Of these claimed detections, H_2O has been robustly verified by WFC3 spectroscopy (Deming et al., 2013) and CO by high-resolution Doppler spectroscopy (Snellen et al., 2010).

High-precision spectral observations of HD 209458b renders this planet an important case study for detailed atmospheric reconnaissance by atmospheric retrieval techniques. The WFC3 transmission spectrum of HD 209458b (Deming et al., 2013) was first retrieved by Madhusudhan et al. (2014a). They reported a H_2O abundance of $\sim$0.01-0.05$\times$ the prediction for a solar abundance atmosphere in chemical equilibrium ($\sim 5\times10^{-3}$), despite the host star containing a super-solar metallicity ($[O/H]_* = 0.092 \pm 0.036$, Brewer et al. (2016)). The low inferred H_2O abundance suggests interesting implications for the formation and migration of this planet (Madhusudhan et al., 2014b). However, this retrieval assumed a cloud-free atmosphere. Benneke (2015) revisited the same spectrum with a retrieval including clouds - albeit with the assumption of

radiative-convective equilibrium and an imposed C-N-O chemical network - and inferred a composition instead consistent with solar abundances. This latter view was further reinforced by Sing et al. (2016), who used transmission spectra of ten hot Jupiters to claim that clouds and hazes, not sub-solar H_2O, sufficiently explain the spectra. However, the work of Sing et al. (2016) only considered a small grid of forward models assuming chemical equilibrium, rather than conducting a retrieval analysis. In another effort, Tsiaras et al. (2016b) detected H_2O but was unable to robustly constrain its abundance. Suggestions of solar H_2O abundances have since been called into question. Barstow et al. (2017) performed retrievals of the Sing et al. (2016) datasets and found that nine of their ten planets possess sub-solar H_2O abundances (for retrievals not assuming chemical equilibrium) – with HD 209458b the driest at $\sim 0.01 - 0.02\times$ solar. These competing lines of evidence leaves the question of whether or not the atmosphere of HD 209458b has a water-deficit unsettled.

Given that HD 209458b has one of the most precise transmission spectra presently available, it is a natural first port of call for the new retrieval paradigm introduced in Chapter 2. In this chapter, I seek to resolve the differing conclusions on the nature of HD 209458b's atmosphere through the first application of POSEIDON. Using a state-of-the-art nested sampling algorithm, I extensively explore the model parameter space in a fully Bayesian retrieval framework - with $> 10^8$ spectra generated in the course of this investigation. Furthermore, unlike the retrievals of Madhusudhan et al. (2014a) and Barstow et al. (2017), I additionally include the nitrogen-bearing molecules NH_3 and HCN in these retrievals. This is due to their strong absorption features in the 1.1–1.7 µm spectral range (see Figure 2.3) covered by the HST WFC3 primary grism.

In what follows, I describe the retrieval configuration in section 3.2. I retrieve the atmospheric properties of HD 209458b in section 3.3. Finally, in section 3.4 I discuss the implications of these results for the nature of the archetypal hot Jupiter.

3.2 Retrieval configuration: HD 209458b

The modelling and retrieval methodology applied to HD 209458b's transmission spectrum is broadly as outlined in Chapter 2. Prior to presenting the results, I first outline specific aspects of the retrieval analysis as they apply to this planet.

3.2.1 Observations

I consider the visible & near-infrared transmission spectrum of HD 209458b presented in Sing et al. (2016). This spectrum consists of three sets of Hubble grism observations from the STIS and WFC3 instruments: (i) HST STIS G430 ($N_{\rm data} = 51$, $\lambda = 0.3072$–$0.5685\,\mu$m, $R \approx 100$, $\sigma \approx 95\,$ppm), HST STIS G750 ($N_{\rm data} = 42$, $\lambda = 0.5453$–$0.9475\,\mu$m, $R \approx 80$, $\sigma \approx 100\,$ppm), and WFC3 G141 ($N_{\rm data} = 29$, $\lambda = 1.1217$–$1.6391\,\mu$m, $R \approx 75$, $\sigma \approx 35\,$ppm). The combined spectrum offers 122 data points with nearly continuous spectral coverage from 0.31–1.64 µm, making this one of the highest quality transmission spectra presently available. The observations are shown in Figure 3.1.

3.2.2 Model parametrisation

POSEIDON was configured to use up to 16 parameters: 7 for the terminator P-T profile, 5 for the terminator chemistry ($X_{\rm H_2O}$, $X_{\rm CH_4}$, $X_{\rm NH_3}$, $X_{\rm HCN}$, and $X_{\rm Na}$), and a 4-parameter 2D cloud and haze prescription (see section 2.3.2). Note that I have restricted the choice of molecules to those expected to dominate absorption in the WFC3 band pass, thus molecules such as CO and CO_2 are not included in this initial application (retrievals with additional molecules will be presented in subsequent chapters). For simplicity, I also elected not to describe $X_{\rm K}$ by an additional parameter, instead parametrising $X_{\rm Na}$ and fixing $X_{\rm K}$ via an assumed solar ratio $\frac{X_{\rm Na}}{X_{\rm K}} \approx 6.4 \times 10^{-2}$ (Asplund et al., 2009). I note that some authors refer to the Na parameter used here by $X_{\rm Na+K}$. The priors applied for each parameter are given in Table 3.1. Finally, the physical properties of HD 209458b and its parent star (serving as fixed inputs) are given by: $R_p = 1.359\,R_J$, $g_p = 9.192\,{\rm ms}^{-2}$, and $R_* = 1.155\,R_\odot$ (Deming et al., 2013; Sing et al., 2016).

3.2.3 Retrieval strategy

I conducted a suite of retrievals spanning multiple cloud models. Four different cloud/haze prescriptions were considered: (i) patchy clouds and haze, (ii) patchy clouds without hazes, (iii) uniform clouds and hazes, and (iv) a clear atmosphere. A Bayesian model comparison allowed the identification of which cloud model is favoured by the observations. Within each cloud prescription, I then performed multiple retrievals with one or more chemical species selectively removed to evaluate their detection significances. In total, this amounted to 8 independent retrievals for each cloud model, each requiring $\sim 5 \times 10^6$ spectra to be computed, such that an unprecedented $> 10^8$ spectra were explored. The total time required amounted to ~ 800 CPU hours.

Table 3.1 Atmospheric retrieval priors: HD 209458b

Parameter	Prior	Range
P-T profile		
$\alpha_{1,2}$	Uniform	0.02 – 1 $K^{-1/2}$
$P_{1,2}$	Log-uniform	10^{-6} – 10^{2} bar
P_3	Log-uniform	10^{-2} – 10^{2} bar
$P_{\rm ref}$	Log-uniform	10^{-4} – 10^{2} bar
T_0	Uniform	800 – 1600 K
Composition		
X_i	Log-uniform	10^{-10} - 10^{-2}
Clouds		
a	Log-uniform	10^{-4} – 10^{8}
γ	Uniform	-20 – 2
$P_{\rm cloud}$	Log-uniform	10^{-6} – 10^{2} bar
$\bar{\phi}$	Uniform	0 – 1

3.3 The atmosphere of HD 209458b

I now report the results from the first application of POSEIDON to an observed exoplanet transmission spectrum. First, in section 3.3.1, a Bayesian model comparison allows the properties of clouds/hazes present at HD 209458b's terminator to be revealed. Secondly, detections of chemical species, their associated abundances, and the sensitivity of said abundances to various cloud models, are presented in section 3.3.2. Finally, in section 3.3.3, I offer constraints on HD 209458b's terminator temperature structure.

3.3.1 Cloud properties

I find strong evidence for the presence of inhomogenous clouds at the terminator of HD 209458b. Table 3.2 summarises the results of a Bayesian model comparison, which indicates a 4.5σ preference for the patchy cloud model over the uniform cloud model. The evidence for patchy clouds emerges from the broadened wings of the H_2O features around 1.2 µm and 1.4 µm, shown in Figure 3.1 (see also Figure 2.4), resulting in a better fit to the infrared data than models without patchy clouds can achieve. The

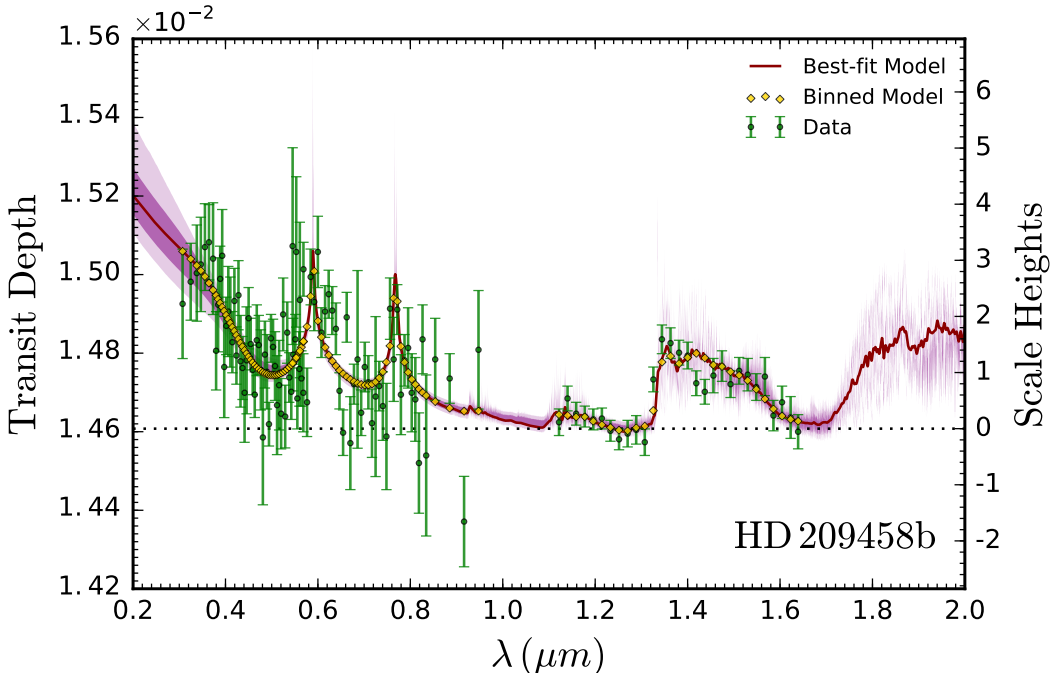


Fig. 3.1 Retrieved transmission spectrum of the hot Jupiter HD 209458b. The observed transit depths (Sing et al., 2016) are plotted as green error bars. The best-fitting model spectrum retrieved by POSEIDON (for the favoured patchy cloud model) is depicted by the dark red curve (Gaussian smoothed for clarity). The median retrieved spectrum, binned to the resolution of the data, is shown as yellow diamonds. 1σ (dark purple) and 2σ (light purple) confidence regions (at $R \approx 10000$) are derived from 10,000 random posterior samples. The black dotted line indicates $(R_p/R_*)^2$. The number of equivalent scale heights above this reference baseline is computed with respect to the median retrieved photosphere temperature (see section 3.3.3).

cloud-free model is ruled out to $> 5\sigma$ when compared to patchy clouds and to 3.4σ when compared to uniform clouds. These detection significances can be understood by examining the posterior of the patchy cloud retrieval (Figure 3.2), which reveals a cloud fraction of $\bar{\phi} = 0.57^{+0.07}_{-0.12}$. Given that the cloud fraction is closer to 1 than 0, it is unsurprising that the Bayesian evidence of the uniform cloud model exceeds that of the cloud-free model. Even after marginalising over all other parameters, the posterior probability distribution of $\bar{\phi}$ is inconsistent with values of 0 or 1 (5σ range: $0.14 - 0.77$); reinforcing the large penalty the Bayesian evidence suffers when forced to consider models fixed at these values. As one final assessment of the presence of patchy clouds, I considered an additional model where $\bar{\phi}$ was fixed to the value from the parameter draw with the maximum likelihood / minimum χ^2 from the patchy cloud retrieval ($\bar{\phi} \approx 0.47$). This 'fixed fraction' model possesses the largest Bayesian evidence amongst the considered cloud models, reinforcing that patchy clouds are favoured by the data.

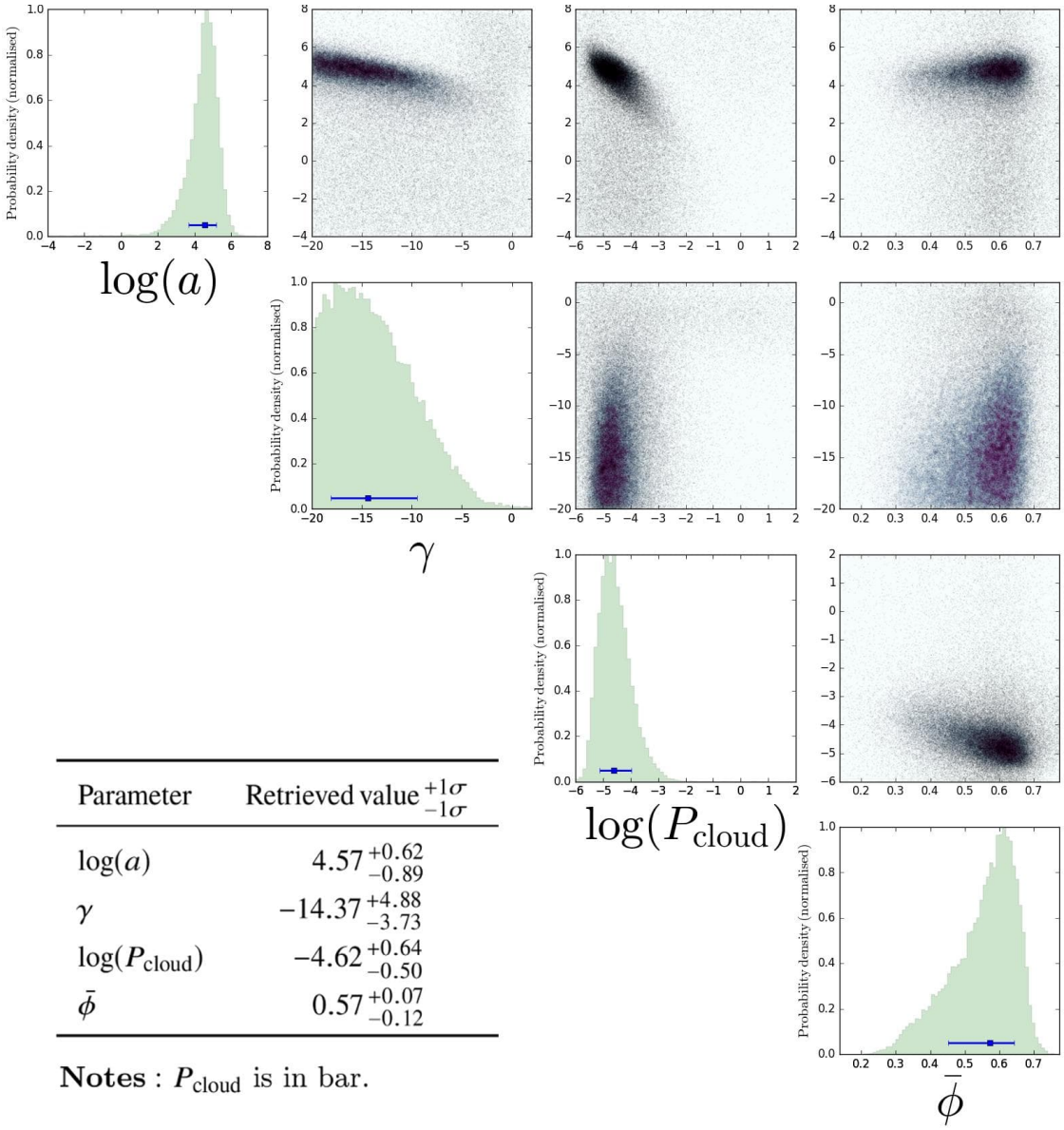
Table 3.2 Model comparison: terminator cloud distribution on HD 209458b

Model	Evidence $\ln(\mathcal{Z}_i)$	Best-fit $\chi^2_{r,\,\mathrm{min}}$	Bayes Factor $\mathcal{B}_{0i}$	Significance of Ref.
Patchy clouds	953.16	1.45	Ref.	Ref.
Uniform clouds	944.91	1.52	3.8×10^3	4.5σ
Cloud-free	940.47	1.62	3.3×10^5	5.4σ
No haze	949.57	1.57	36	3.2σ
Fixed fraction	953.60	1.44	0.65	N/A

Notes : The 'fixed fraction' model has $\bar{\phi}$ held at the best fit value from the 'patchy clouds' model (0.47). The 'no haze' model is a patchy cloud model that considers only H_2-Rayleigh scattering in cloudy regions. $\chi^2_{r,\mathrm{min}}$ is the minimum reduced chi-square ($\chi^2/[N_{\mathrm{data}} - N_{\mathrm{params}}]$). The significance indicates the degree of preference for the reference model, highlighted in bold, over each alternative model. N/A indicates no (or negative) evidence ($\mathcal{B}_{\mathrm{ij}} \lesssim 1$) for the reference model.

A high-altitude ($\sim 0.01 - 0.1$ mbar) cloud deck is inferred in the cloudy region of the terminator. Above this cloud deck, I report moderate evidence (3.2σ) of high-altitude hazes. The necessity of strongly enhanced Rayleigh scattering ($\sim$ 5,000-100,000 $\times$ H_2-Rayleigh) is visibly apparent in Figure 3.1 from the steep increase in the transit depth towards shorter wavelengths. In addition to the high strength of the Rayleigh enhancement factor, a, the scattering slope, γ, is remarkably negative, tending to prefer values towards the lower edge of the prior (see Figure 3.2). The presence of such a strong scattering signature at these altitudes suggests the presence of small-grained haze particles with large scattering cross sections. Two possibilities for the origin of this upper-atmosphere signature come to mind: (i) incredibly light particles, capable of being lofted to these altitudes by vertical mixing; and / or (ii) continuous replenishment of the haze at high-altitude (e.g. by photochemical reactions).

From these results, it is clear that patchy cloud models are preferred by the observations. In what follows, I therefore select the 'fixed fraction' cloud model, unless otherwise stated. This model maximises the Bayesian evidence, and thus holds the greatest sway in light of the data. The full posterior for this model is displayed in Figure 3.3. The first point to note is that the values of the remaining cloud parameters (a, γ, and P_{cloud}), along with their associated errors, remain consistent to within 1σ of the values from the 'patchy clouds' model (shown in Figure 3.2). This indicates that, for this spectrum, allowing the cloud fraction to vary as a free parameter does

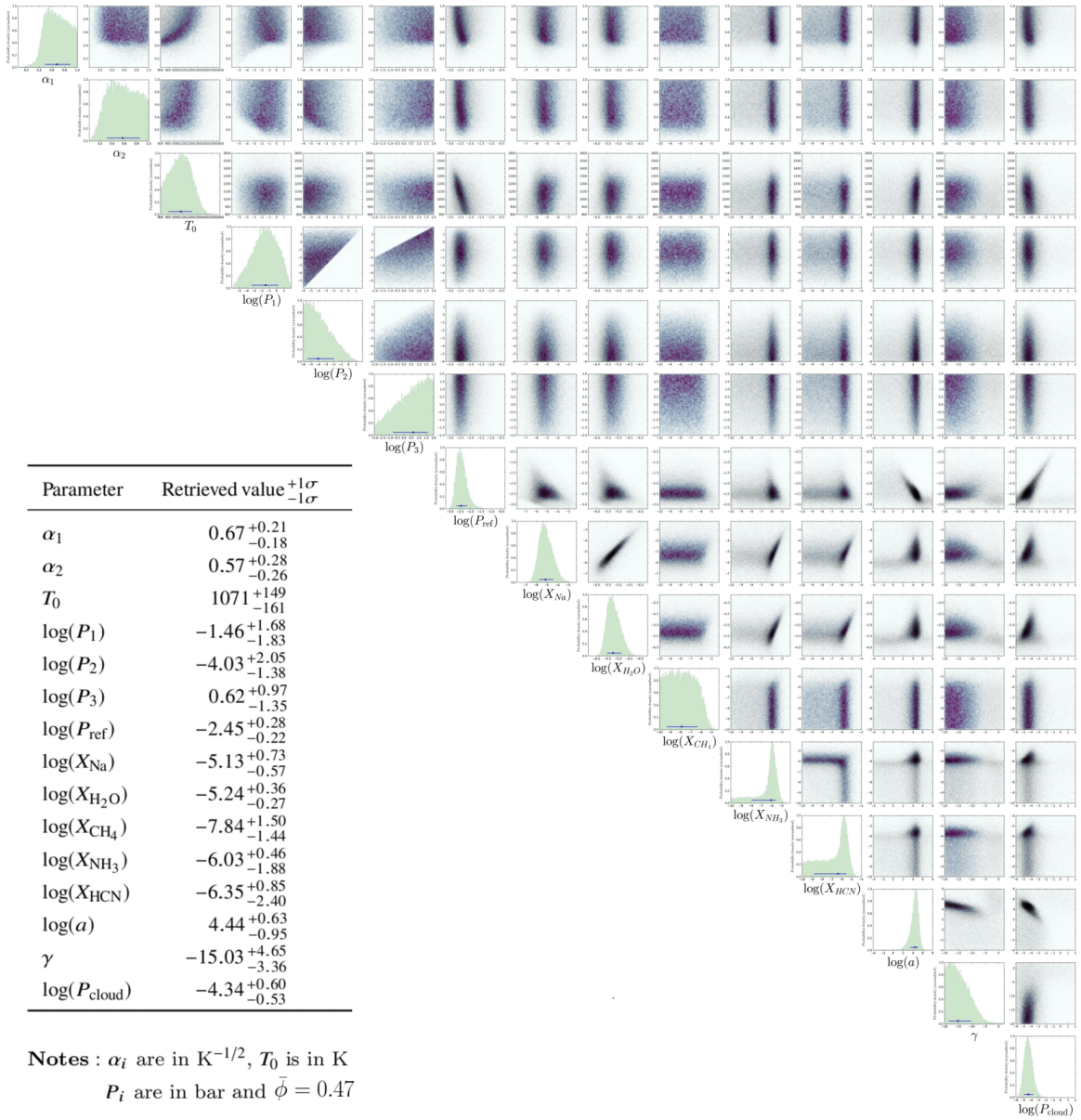


Parameter	Retrieved value $^{+1\sigma}_{-1\sigma}$
$\log(a)$	$4.57^{+0.62}_{-0.89}$
γ	$-14.37^{+4.88}_{-3.73}$
$\log(P_{\rm cloud})$	$-4.62^{+0.64}_{-0.50}$
$\bar{\phi}$	$0.57^{+0.07}_{-0.12}$

Notes : $P_{\rm cloud}$ is in bar.

Fig. 3.2 Retrieved cloud properties at HD 209458b's terminator.
The posterior subset shown here corresponds to the 'patchy clouds' model from Table 3.2. The corner plot depicts correlations between pairs of retrieved parameters and marginalised histograms for the value of each parameter. The median retrieved values and $\pm 1\sigma$ constraints are shown on each histogram by blue error bars. Table inset: summary of the median retrieved cloud parameters and 1σ confidence levels.

not overly affect inferences for the remaining cloud parameters (compared to when $\bar{\phi}$ is fixed at the best fit value). In order to illustrate the importance of selecting the cloud model that is supported by the data, in the next section I will demonstrate how inferred chemical abundances crucially depend upon the assumed cloud model.



Parameter	Retrieved value $^{+1\sigma}_{-1\sigma}$
α_1	$0.67^{+0.21}_{-0.18}$
α_2	$0.57^{+0.28}_{-0.26}$
T_0	1071^{+149}_{-161}
$\log(P_1)$	$-1.46^{+1.68}_{-1.83}$
$\log(P_2)$	$-4.03^{+2.05}_{-1.38}$
$\log(P_3)$	$0.62^{+0.97}_{-1.35}$
$\log(P_{\rm ref})$	$-2.45^{+0.28}_{-0.22}$
$\log(X_{\rm Na})$	$-5.13^{+0.73}_{-0.57}$
$\log(X_{\rm H_2O})$	$-5.24^{+0.36}_{-0.27}$
$\log(X_{\rm CH_4})$	$-7.84^{+1.50}_{-1.44}$
$\log(X_{\rm NH_3})$	$-6.03^{+0.46}_{-1.88}$
$\log(X_{\rm HCN})$	$-6.35^{+0.85}_{-2.40}$
$\log(a)$	$4.44^{+0.63}_{-0.95}$
γ	$-15.03^{+4.65}_{-3.36}$
$\log(P_{\rm cloud})$	$-4.34^{+0.60}_{-0.53}$

Notes : α_i are in $\rm K^{-1/2}$, T_0 is in K
P_i are in bar and $\bar{\phi} = 0.47$

Fig. 3.3 Full posterior from POSEIDON's retrieval of HD 209458b.
The corner plot depicts correlations between pairs of retrieved parameters and marginalised histograms for the values of each parameter extracted by the retrieval. Note that the abundances of H_2O and Na (hence also K, via their assumed abundance ratio) are tightly constrained, relatively independent of the cloud parameters. This verifies POSEIDON's ability to break cloud-composition degeneracies (see section 2.4.2). The median retrieved values and $\pm 1\sigma$ constraints are shown by blue error bars. Table inset: median retrieved values and 1σ confidence levels for each parameter, following marginalisation over the other 14 dimensions of the parameter space.

3.3.2 Chemical composition

Atomic and molecular detections

I confirm previous detections of Na and H_2O (e.g. Charbonneau et al., 2002; Deming et al., 2013) in the transmission spectrum of HD 209458b. Table 3.3 displays the statistical significances of each considered chemical species. This model comparison establishes detections of H_2O and Na+K at 9.1σ and 7.3σ confidence, respectively. CH_4 is not detected, though I establish at $> 10\sigma$ confidence that the *combination* of CH_4 or H_2O is required. The higher significance of this combination is due to the overlapping absorption features of these two molecules at 1.15 µm and 1.40 µm – removing just H_2O then allows CH_4 to partially compensate, such that the individual H_2O significance is slightly lower (see Benneke & Seager, 2013, for a discussion).

I additionally report strong evidence of nitrogen chemistry (NH_3+HCN) at 3.7σ confidence. The Bayes factor between the models with and without nitrogen chemistry is 186 - corresponding to 'strong evidence' on Jeffreys' scale (see section 2.3.3). The evidence of NH_3+HCN is robust to the consideration of models without patchy clouds, rising to 4.9σ and 7.7σ for uniformly cloudy and cloud-free models, respectively.

Table 3.3 Model comparison: atmospheric composition of HD 209458b

Model	Evidence $\ln(\mathcal{Z}_i)$	Best-fit $\chi^2_{r,\,\mathrm{min}}$	Bayes Factor $\mathcal{B}_{0i}$	Significance of Ref.
Full Chemistry	953.60	1.44	Ref.	Ref.
No H_2O+CH_4	904.62	2.35	1.9×10^{21}	10.1σ
No H_2O	914.62	2.14	8.4×10^{16}	9.1σ
No Na+K	928.92	1.93	5.2×10^{10}	7.3σ
No NH_3+HCN	948.37	1.53	186	3.7σ
No NH_3	952.80	1.44	2.2	1.9σ
No HCN	953.35	1.42	1.3	1.4σ
No CH_4	954.01	1.42	0.7	N/A

Notes : The 'full chemistry' model includes opacity due to H_2, He, Na, K, H_2O, CH_4, NH_3, and HCN. It corresponds to the 'fixed fraction' cloud model in Table 3.2. $\chi^2_{r,\,\mathrm{min}}$ is the minimum reduced chi-square ($\chi^2/(N_{\mathrm{data}} - N_{\mathrm{params}})$). The significance indicates the degree of preference for the reference model, highlighted in bold, over each alternative model. N/A indicates no (or negative) evidence ($\mathcal{B}_{\mathrm{ij}} \lesssim 1$) for the reference model.

When considering the patchy cloud model preferred by the observations (see Table 3.2), it is not possible to differentiate between the presence of NH_3 or HCN. This is due to patchy clouds manifesting in the transmission spectrum via a broadening of the wings of spectral bands, as mentioned previously. The broadened H_2O features then better match the infrared data, such that either NH_3 or HCN can provide the additional opacity required to optimally match the data. The evidence of nitrogen chemistry in the transmission spectrum of HD 209458b will be further elucidated after the presentation of abundance constraints.

If one considers alternative cloud models (with lower Bayesian evidences), the conclusions drawn about nitrogen chemistry are altered. In particular, in both the cloud-free and uniformly cloudy models the NH_3-HCN degeneracy does not manifest, with only NH_3 being indicated by the retrieval analysis. If one chooses these models instead, the significance for NH_3 alone rises from 1.9σ (for patchy clouds) to 4.9σ / $\mathcal{B}_{0i} = 22{,}000$ (for cloud-free atmospheres) and 3.3σ / $\mathcal{B}_{0i} = 58$ (for uniform clouds), respectively. On the other hand, HCN is not inferred for these models. These points illustrate that the choice of cloud model can have a major influence on chemical detection significances. However, as patchy cloud models can encompass both cloud-free and uniform cloud atmospheres, the significances from the preferred patchy cloud model are the most stringent and reliable. I therefore quote these significances as the main results in this chapter. Without considering patchy clouds, the more nuanced picture of an NH_3-HCN degeneracy would not have manifested, and one could be led to (erroneously) believe that NH_3 had been detected. I therefore stress that care must be taken to identify the most appropriate cloud model for any given dataset (e.g. by a Bayesian model comparison), *before* reporting chemical detections. With this caveat in mind, I now proceed to discuss abundance constraints for the inferred chemical species.

Abundance constraints

I present constraints on the chemical abundances at HD 209458b's terminator in Figure 3.4. I focus first on the abundances corresponding to the preferred patchy cloud model (green histograms). The abundances reported are amongst the most precise ever obtained from an exoplanet transmission spectrum. The H_2O abundance, in particular, is constrained to ~ 0.3 dex. This precise abundance demonstrates POSEIDON's ability to break cloud-composition degeneracies (see section 2.4.2) on a real transmission spectrum of a cloudy exoplanet atmosphere. The manner in which assumed the cloud distribution influences abundance inferences will be discussed at the end of this section.

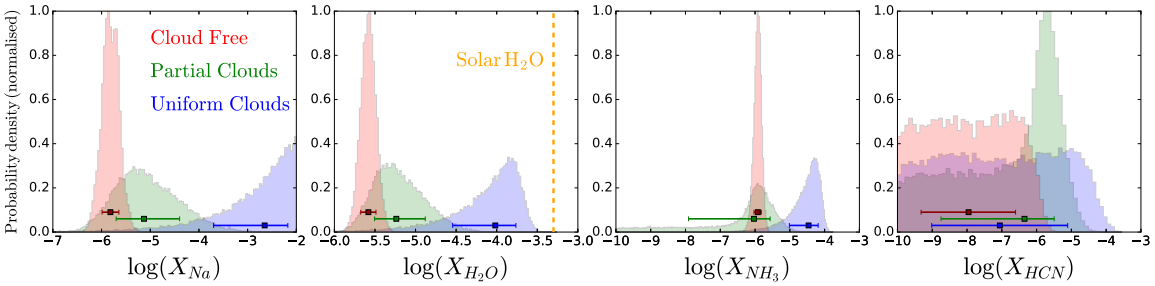


Fig. 3.4 Retrieved chemical abundances at HD 209458b's terminator. The red, green and blue histograms are inferences from the 'cloud-free', 'fixed fraction' and 'uniform clouds' models of Table 3.2. H_2O, Na+K, and NH_3 / HCN are inferred in all three models. All three models are inconsistent with a solar water abundance (dashed orange line) at $> 3\sigma$ confidence. The cloud-free and patchy cloud models are consistent with solar Na ($\log X_{\rm Na} \approx -5.76$, Asplund et al. (2009)). The retrieved NH_3 and HCN abundances are enhanced by $\gtrsim 100\times$ above chemical equilibrium expectations for a solar-N abundance (see section 3.3.2). The uniformly cloudy model is biased to higher abundances (as discussed in section 3.4.1), whilst the cloud-free model is consistent with the abundances from the preferred partial cloud model. However, the cloud-free model underestimates the uncertainty in the derived abundances. The posteriors from the 'patchy clouds' model in Table 3.2 is similar to that of the 'fixed fraction' model, so it is omitted for clarity. CH_4 is only constrained by an upper limit for all the conducted retrievals (e.g. Figure 3.3), so is not shown here.

The terminator of HD 209458b is inconsistent with a solar H_2O abundance. This is established at $> 5\sigma$ confidence for both patchy cloud and cloud-free models and at $> 3\sigma$ confidence for uniform clouds. The retrieved H_2O abundance, $\log(X_{\rm H_2O}) = -5.24^{+0.36}_{-0.27}$, is remarkably consistent with values reported by Madhusudhan et al. (2014b) $\left(-5.27^{+0.65}_{-0.16}\right)$ and Barstow et al. (2017) (-5.3 to -5.0). This is unsurprising, as the observed spectrum (Figure 3.1) shows the amplitude of the H_2O absorption feature at 1.4 µm is only 2 scale-heights – whereas a solar composition atmosphere at a similar temperature and cloud fraction would give $\sim$ 5 scale heights (e.g. Figure 2.9).

The Na abundance can also be reasonably well-constrained ($\sim$ 0.6 dex), despite the $\sim$ 120 ppm errors in the optical STIS data. While this serves as an important demonstration of principle, I caution against reading too much into the retrieved values. This is due to the simplicity of the present treatment (especially given Na and K were here combined into a single parameter). Precise alkali abundance constraints, and their implications for atmospheric chemistry, is an interesting avenue for future studies.

The abundances of both NH_3 and HCN show a sharp peak at mixing ratios of $\sim 10^{-6}$, with a tail towards lower abundances. The NH_3 abundance is the best constrained:

$\log(X_{\mathrm{NH_3}}) = -6.03^{+0.46}_{-1.88}$ $(0.01 - 2.7$ ppm). The tails stem from the fact that either of these nitrogen-bearing species can explain the observed absorption features - if one has a high abundance, the other will have a low abundance and *vice versa*. Ultimately, it is this long tail that prevents a unique determination of which of these species best explains the additional absorption attributed to nitrogen chemistry. It can be seen from the lower probability density in the tail of the NH_3 posterior in Figure 3.4 that the presence of NH_3 is marginally preferred over HCN when considering partial cloud coverage (this can also be seen from the higher Bayes factor for NH_3 in Table 3.3).

The NH_3-HCN abundance degeneracy is not present when considering cloud-free or uniformly cloudy models. Both of these models feature posteriors with well-constrained NH_3 abundances but only an upper limit HCN - explaining the NH_3 evidence observed in these models. Given the combination of these high NH_3 detection significances, and the coincident peak of its abundance distribution between the cloud-free and patchy cloud models, I suggest that NH_3 is the source of the inferred nitrogen chemistry. However, it will ultimately require future observations to resolve this ambiguity and definitively detect a specific nitrogen-bearing molecule.

Nitrogen chemistry

Given the consistent indications of nitrogen chemistry across all considered cloud models, I now proceed to identify the absorption features giving rise to these signatures. In Figure 3.5, I show the effect on the best-fitting transmission spectrum $\left(\log X_{[\mathrm{H_2O,\,CH_4,\,NH_3,\,HCN}]} = [-5.21,\, -8.63,\, -5.72,\, -8.39]\right)$ of removing opacity due to NH_3 and HCN. Given that NH_3 is the dominant nitrogen-bearing molecule for the best-fitting model spectrum, the difference between the spectra with and without the influence of nitrogen chemistry amounts to an assessment of the impact of NH_3 on the near-infrared transmission spectrum.

The primary impact of NH_3 absorption in the WFC3 band pass is to raise the transit depth of HD 209458b by $\sim 5 \times 10^{-5}$ compared to what would be expected from pure H_2O absorption over the spectral range ~ 1.45–$1.7\,\mu$m. A secondary feature of magnitude $\sim 1 \times 10^{-5}$ is seen between ~ 1.2–$1.3\,\mu$m. These absorption features are readily identified by an examination of the NH_3 cross section (Figure 2.3), which is seen to dominate that of H_2O over these two regions. The necessity of additional absorption is evident from the data itself, as there are 4 data points elevated by 2σ and one point elevated by 3σ above the model without nitrogen chemistry.

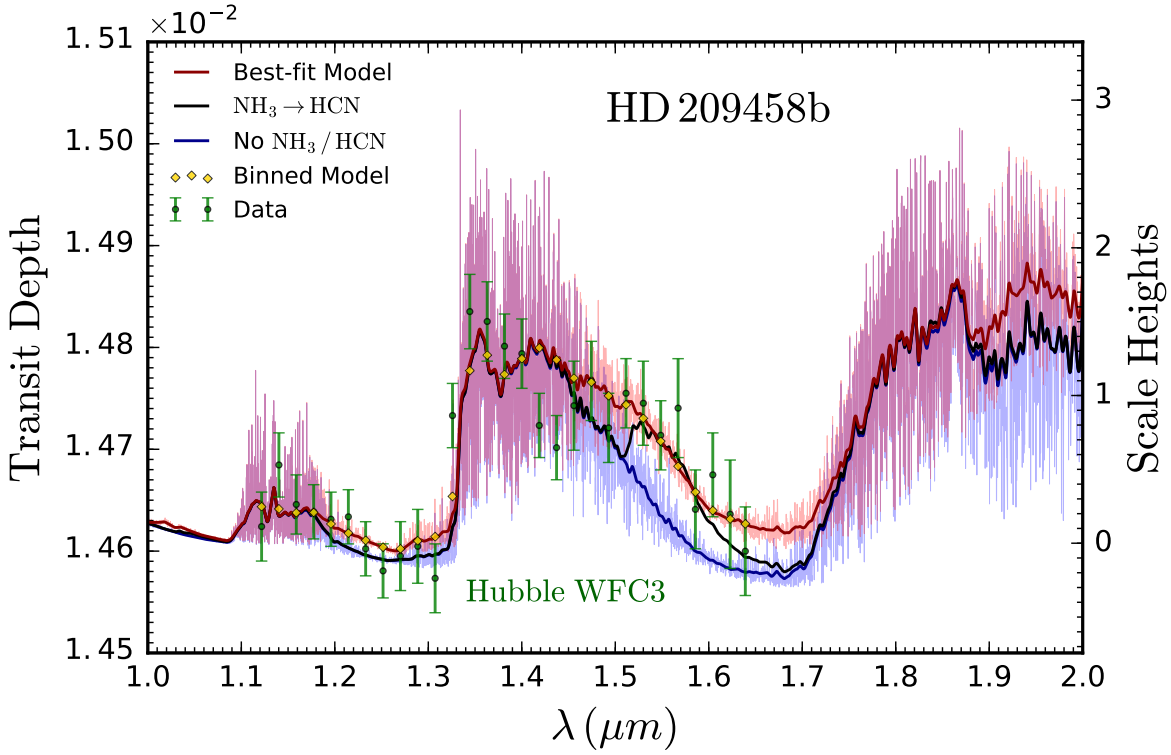


Fig. 3.5 Evidence of nitrogen chemistry at the terminator of HD 209458b. The observed WFC3 transit depths (green error bars) are compared against the best-fitting retrieved transmission spectrum (red), which contains 2 ppm NH_3. The yellow diamonds are the best-fit spectrum binned to the resolution of the data. The blue spectrum shows how the best-fit spectrum changes when NH_3 opacity is disabled. The black spectrum is similar to the best-fit model, but with the abundances of NH_3 and HCN interchanged (such that HCN becomes the dominant nitrogen-bearing molecule). Each spectrum is plotted both at $R \approx 10000$ and as Gaussian smoothed curves (to aid comparison with the lower resolution data). Exploration of an extensive parameter space of differing compositions, temperature structures, and clouds/hazes, establishes that the model including the combination of NH_3 and HCN (corresponding to the red spectrum) is preferred over models with no nitrogen chemistry at 3.7σ confidence. The primary evidence for nitrogen-bearing species comes from the additional absorption over the 1.45–1.70 µm spectral range.

I now offer suggestions for how to distinguish between NH_3 and HCN in transmission spectra where nitrogen chemistry (i.e. their combination) is detected. The difficulty inherent in this task is demonstrated in Figure 3.5, where the black curve shows the effect of interchanging the abundances of NH_3 and HCN, such that HCN becomes the dominant nitrogen-bearing molecule. HCN causes an increase in the transit depth that

is similar to that of NH_3 from $\sim$ 1.53–1.6 µm, though it deviates from NH_3 outside this spectral range. This suggests that the NH_3-HCN degeneracy may be lifted by higher resolution (and precision) observations in three regions of the WFC3 band pass: (i) 1.2–1.32 µm, (ii) 1.46–1.52 µm, and (iii) 1.6–1.7 µm. Given the difference between spectra featuring NH_3 vs. those featuring HCN are on the order of the error bars (35 ppm), this is clearly pushing the frontier of current observational capabilities. These ideas will be further developed in Chapter 4, in which I develop a strategy to uniquely detect NH_3 and HCN in cloudy exoplanet atmospheres.

Another potential avenue to improve detection significances of NH_3 and / or HCN with WFC3 observations may be rooted in the prevalence of clouds at an exoplanet's terminator. Namely, the sharp NH_3 posterior for the cloud-free model shown in Figure 3.4 suggests that genuinely cloud-free atmospheres may allow highly robust detections of NH_3 (and possibly also HCN). This possibility, and the influence of clouds more broadly, will be considered further in what follows.

The influence of clouds on retrieved chemical abundances

I shall now explore the extent to which the cloud model assumed by a retrieval can influence inferred chemical abundances. An innate challenge in distinguishing between NH_3 and HCN in partially cloudy transmission spectra has already been identified. We have further seen that their detection significances and abundance constraints depend crucially on the cloud model employed by the retrievals.

In general, cloud-free abundances are artificially well constrained at lower values than those inferred by the preferred patchy cloud model (though they remain consistent within 1σ). However, this suggests that exoplanets with genuinely low cloud coverage may permit strong detections with precise abundance constraints. This is especially evident in the case of nitrogen chemistry, where the cloud-free model clearly identifies the presence of NH_3 at $\approx 5\sigma$ and constrains its abundance to $\log(X_{\mathrm{NH_3}}) = -5.92^{+0.10}_{-0.11}$ (i.e. 0.1 dex precision). I hence suggest that, even given the precision of current observations, precise abundance determinations of nitrogen-bearing molecules can be obtained for exoplanets with low overall terminator cloud coverage.

In contrast, the uniform cloud model tends to overestimate chemical abundances. This is strikingly apparent in the case of Na, which favours unphysical values tending towards the upper limit of the prior (10^{-2}). The biasing of abundances to erroneously high values under the assumption of uniform clouds is a consequence of a degeneracy between clouds, hazes, the reference pressure, and chemical abundances. For uniform

terminator cloud coverage, the cloud deck pressure, $P_{\rm cloud}$, and the pressure at the radius of the planet, $P_{\rm ref}$, exist on a line in parameter space without a unique solution. Defining $P_{\rm ref}$ at radii different to R_p merely offsets the line of degeneracy. This family of solutions determines the baseline of the spectrum (Figure 3.1, black dotted line). When $P_{\rm ref}$ and $P_{\rm cloud}$ are lowered along this line of solutions, both the amplitude of spectral features and the Rayleigh enhancement factor can rise to produce an identical spectrum. This biasing to higher abundances is an artefact of the *assumption* of uniform clouds. I verified this by running a retrieval on simulated data with an input Na abundance of 10^{-6}, assuming uniform clouds, and found similar behaviour to that seen in Figure 3.4. I therefore caution against the blind application of uniform cloud models. Indeed, the results in Figure 3.4 suggest that cloud-free models are a better option if one solely wishes to estimate the chemical abundances of a transiting exoplanet.

I have demonstrated here that partial cloud coverage breaks the degeneracy between clouds and chemistry (see section 2.4.2) imposed artificially by the assumption of uniform clouds. Alternatively, this degeneracy may be broken by assuming *a priori* knowledge of the scattering slope in the optical (e.g Benneke & Seager, 2012) or of the molecular/condensate chemistry (e.g. Kreidberg et al., 2015). The attraction of partial clouds is that we do not have to make such assumptions, allowing the direct retrieval of the cloud, haze, composition, and temperature properties of an exoplanet atmosphere *simultaneously*. Though patchy clouds can enable precise determination of molecular abundances, and are preferred by the Bayesian evidence at high significance, they complicate the interpretation of nitrogen chemistry. Specifically, patchy clouds tend to broaden the slopes of molecular features (Figure 2.4, bottom right), such that the transit depth 'gap' filled by nitrogen chemistry over $\sim$ 1.45–1.70 µm (Figure 3.5) is shrunk by the broader 1.4 µm H_2O absorption feature.

Given these compositional and cloud inferences, I now proceed to present the retrieved P-T profile from the terminator of HD 209458b. The combination of all three of these properties is required to build a coherent picture of the conditions at the terminator of this exoplanet atmosphere.

3.3.3 Temperature structure

The retrieved temperature structure for the terminator of HD 209458b favours a temperature gradient. If the data supported an isothermal profile we would expect to see $\alpha_{1,2}$ tending towards larger values, whereas Figure 3.3 shows they have maximum *a posteriori* values lower than the upper bound (where atmospheres are essentially isothermal). Whilst unsurprising from physical arguments and GCM simulations, this point merits emphasis. It is often assumed in retrievals of transmission spectra that with currently available data: (i) it is not possible to retrieve the shape of the terminator P-T profile; and (ii) an isothermal profile does not overly affect the inferred abundances. Here, I demonstrate the invalidity of the first assumption for high-precision HST data, which was also examined by Barstow et al. (2013). For a critical examination of the second assumption, I refer the reader to Rocchetto et al. (2016).

Figure 3.6 shows the retrieved P-T profile of HD 209458b's terminator. A proxy for the near-infrared photosphere (the 1σ spread of the $\tau = 1$ surface at 1.5 µm) is highlighted, as this indicates the pressure range probed at a wavelength with significant molecular absorption[1] (cf. Figure 3.5). The temperature is not constant across the photosphere, with a gradient of ~ 50 K. By assuming an isothermal profile, this behaviour, and its effect on molecular cross sections, will not be captured. As expected intuitively, a tight temperature constraint is obtained around photospheric pressures: $T_{\rm phot} = 1221^{+131}_{-138}$ K. The confidence regions naturally expand away from pressures directly probed by the observations – especially for pressures $\gtrsim 100$ mbar, which we would usually expect to start becoming opaque due to collision-induced absorption. The retrieved $T_{\rm phot}$ is some 200 K cooler than the planetary equilibrium temperature ($T_{\rm eq} \approx 1450$ K). This is unsurprising, given that transmission spectra probe high-altitudes in the cooler terminator region. Again, I emphasise that such inferences are made possible by the high-precision transmission spectrum.

This profile represents the average terminator P-T profile. Strictly speaking, we expect this to be composed of two underlying profiles: a cooler profile in the cloudy terminator region and a warmer profile in the clear region. This correspondence arises from condensates tending to form in cooler regions, where the P-T profile may intersect additional condensation curves than in the warmer region (Marley et al., 2013). This P-T profile averaging also explains the relatively high-altitude photosphere, as the

[1] In Chapter 6, I will use a more general photosphere definition to account for a range of wavelengths.

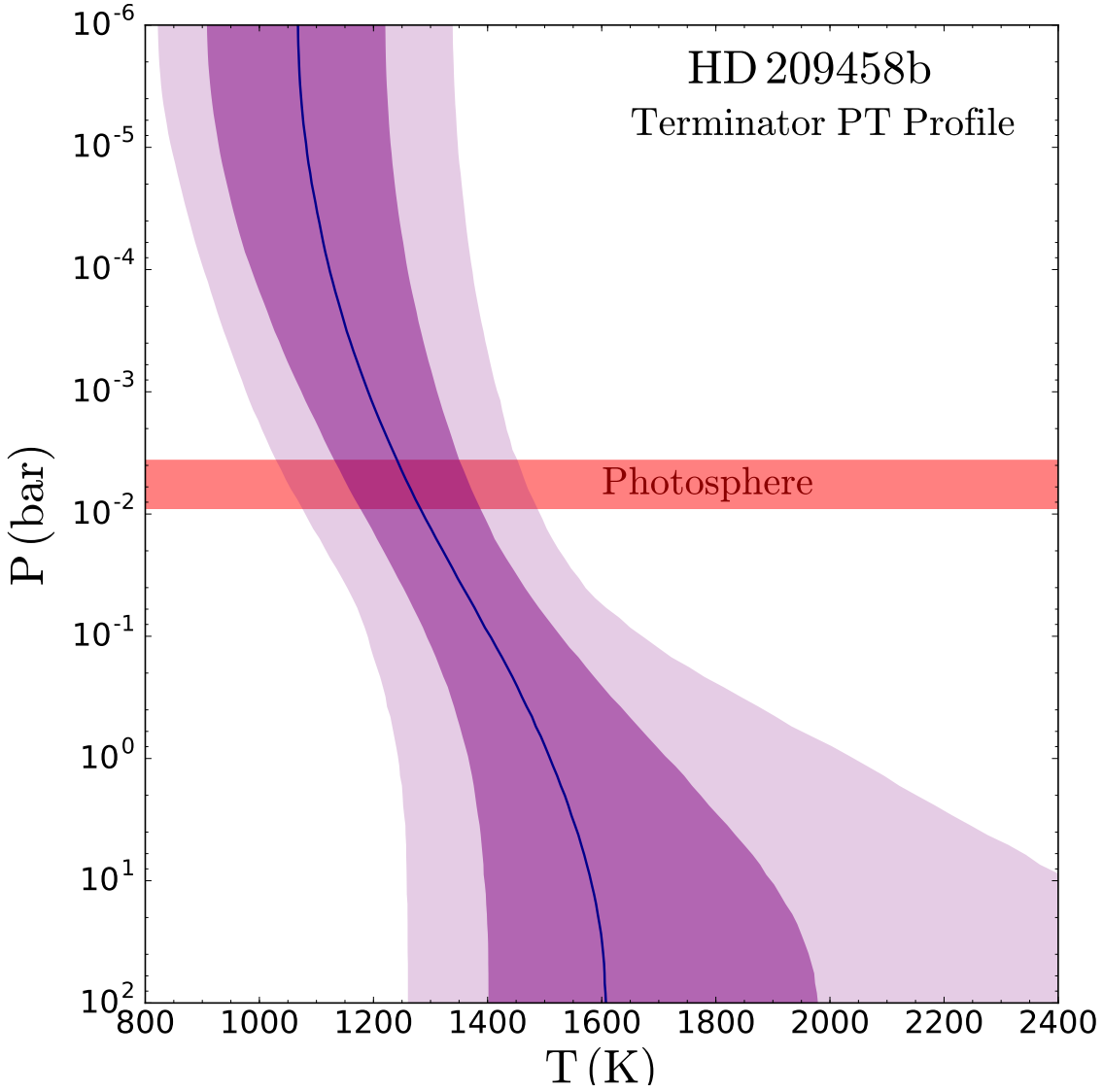


Fig. 3.6 Retrieved temperature structure at HD 209458b's terminator. The median retrieved terminator P-T profile (dark blue curve) is shown, along with the corresponding 1σ (dark purple shading) and 2σ (light purple shading) confidence regions for the temperature in each layer. The confidence regions and median profile are derived from 10,000 random posterior samples. The approximate 1σ extent of the near-infrared photosphere (here proxied by $\tau = 1$ at 1.5 µm) is shown by the red horizontal band. The median temperature corresponding to this photosphere region, 1221 K, is approximately 200 K below the planetary equilibrium temperature, with a temperature gradient of ~ 50 K across the photosphere.

opaque cloud deck inferred at $P_{\rm cloud} \approx 0.01 - 0.1$ mbar in the cloudy region of the terminator combines with the cloud-free region to determine the $\tau = 1$ surface.

3.4 Implications

The results from this retrieval analysis shed new light into the enigmatic atmosphere of the canonical hot Jupiter HD 209458b. I will now discuss the implications of these findings in the wider context of exoplanet studies. In turn, I examine the connection between nitrogen chemistry and chemical disequilibrium in section 3.4.1, implications for hot Jupiter formation in section 3.4.2, the debate over HD 209458b's H_2O abundance in section 3.4.3, and finally, in section 3.4.4, the nature of the clouds present in this hot Jupiter atmosphere.

3.4.1 Disequilibrium nitrogen chemistry

Given what is known from solar system and brown dwarf studies, the evidence presented here for nitrogen chemistry in an exoplanet atmosphere should come as no surprise. In particular, NH_3 is present both on Jupiter – in the form of high altitude clouds (Sato & Hansen, 1979) – and in brown dwarf atmospheres (e.g. Roellig et al., 2004; Saumon et al., 2006; Cushing et al., 2008). Recently, Line et al. (2015) demonstrated that NH_3 can be detected on brown dwarfs using low-resolution near-infrared data, but noted a lack of obvious spectral features leading to their detection. I demonstrated in Figure 3.5 that the absorption features contributing to the NH_3 inference in the near-infrared are located over the ranges $\sim$ 1.45–1.7 µm and $\sim$ 1.2–1.3 µm. Additionally, the fact that Madhusudhan et al. (2016a) reported sharp NH_3 abundance constraints in a population of three T-type brown dwarfs using WFC3 data alone further supports that it is possible to detect NH_3 using WFC3 spectra. I also note that the nitrogen-bearing molecule HCN has been suggested, at low statistical significance, as a potential component in the atmosphere of the super-Earth 55 Cancri e (Tsiaras et al., 2016a). However, as noted in section 3.3.2, the detection significance of nitrogen chemistry is sensitive to the chosen cloud model – with the lowest confidence arising for models with patchy clouds. This raises the possibility that additional physical mechanisms, not considered in the models used here, could produce a similar effect to that which I attribute to nitrogen chemistry. To obtain unambiguous detections of nitrogen chemistry will require the concerted efforts of increasingly sophisticated retrieval forward models and additional high-precision observations. This combination will enable the definitive resolution of differences between models with and without nitrogen chemistry.

Abundance constraints for nitrogen-bearing molecules opens a new window into exoplanetary composition and atmospheric dynamics. The NH_3 abundance I infer ($\approx$ 1 ppm) represents a $\gtrsim 100\times$ enhancement over the expectation for an atmosphere with a solar nitrogen abundance at ≈ 1200 K in thermochemical equilibrium (Moses et al., 2011). This suggests non-equilibrium processes may prove necessary, transporting NH_3 from regions where higher abundances naturally exist. One such avenue is transport-induced quenching, whereby regions where the characteristic dynamical timescale ($\tau_{\rm dyn}$) is less than the chemical reaction timescale ($\tau_{\rm chem}$) reflect the abundance from the 'quench level' where $\tau_{\rm dyn} = \tau_{\rm chem}$. For HD 209458b, $\tau_{\rm dyn} \sim 10^5$ s for both horizontal and vertical advection (Cooper & Showman, 2006), whereas in the terminator photosphere ($P \sim 10$ mbar, $T \sim 1200$ K) $\tau_{\rm chem} \sim 10^{13}$ s for $NH_3 \rightarrow N_2$ conversion (Zahnle & Marley, 2014). If horizontal quenching dominates, the NH_3 abundance would be expected to follow the chemical equilibrium value characteristic of the dayside, where $\tau_{\rm chem}$ is shorter (Agúndez et al., 2012). If, however, vertical quenching dominates, the NH_3 abundance in the observable photosphere will reflect that of the chemical equilibrium abundance at the altitude where $\tau_{\rm chem} \sim 10^5$ s, which occurs around pressures ≈ 1 bar (Moses et al., 2011). The abundance estimate of $\approx$1 ppm obtained here is remarkably consistent with that predicted at the terminator of HD 209458b by vertical quenching models (Moses et al., 2011), using nominal temperature and atmospheric mixing profiles from GCMs (Showman et al., 2009), assuming a solar nitrogen abundance. A wider range of parameters (e.g. mixing strengths, nitrogen abundance, etc.) beyond those currently assumed in forward models (e.g. Moses et al., 2011; Venot et al., 2013) could also potentially explain the same. This suggests that atmospheric NH_3 abundance constraints across a wide variety of exoplanets could provide a powerful diagnostic of the frequency and strength of non-equilibrium transport processes.

3.4.2 Formation conditions

Recent years have seen increased interest in utilising elemental ratios as formation diagnostics (see section 1.3.3). In particular, the C/O ratio is often invoked in attempts to constrain planetary formation and migration pathways relative to the snow lines of major condensates (e.g. Öberg et al., 2011; Madhusudhan et al., 2014c). Nitrogen chemistry offers complementary diagnostics, as enhanced N/H ratios are anticipated for planets forming further out in protoplanetary disks (Piso et al., 2016). Indeed, Piso et al. (2016) suggested that the N/O ratio for planets forming in outer disks could be significantly enhanced relative to the stellar value and to the C/O ratio.

The abundance constraints on H_2O and NH_3 presented here suggest a scenario where the planet formed far out in the disk, before migrating inwards to its present location by dynamical scattering. The sub-solar H_2O mixing ratio, concluded despite the consideration of clouds, is inconsistent with an atmospheric composition of solar elemental abundances. Explaining this low H_2O abundance requires either a significantly sub-solar overall metallicity, or a super-solar C/O ratio (Madhusudhan et al., 2014a). On the other hand, as discussed above, the observed NH_3 abundance is consistent with disequilibrium chemistry and a nearly solar nitrogen abundance (Moses et al., 2011). Therefore, a consistent possible explanation for both the low H_2O and high NH_3 abundances is a super-solar C/O *and* a super-solar N/O ratio; the metallicity can be solar for all elements except oxygen. This composition can be achieved if the planet formed beyond the CO_2 or CO snow lines, accreted mostly gas (Öberg et al., 2011; Madhusudhan et al., 2014b; Piso et al., 2016), and migrated to its current orbit by disk-free mechanisms (Madhusudhan et al., 2014b) or formed via pebble accretion (Madhusudhan et al., 2017). This is further supported by the fact that the host star HD 209458 is super-solar in metallicity, including O, which means that it would be infeasible to obtain such a low planetary O abundance if it migrated through the disk and accreted planetesimals (e.g. Brewer et al., 2016; Mordasini et al., 2016).

3.4.3 Solar vs. sub-solar H_2O

The robust result presented here that the terminator of HD 209458b is depleted in H_2O relative to solar values runs contrary to the claim asserted by Sing et al. (2016). By not performing a retrieval, explicitly imposing thermochemical equilibrium, assuming isothermal P-T profiles, and considering only global clouds/hazes, their models induce sufficient *a priori* biases as to render their conclusions unreliable. Indeed, the inadequacy of this forward model approach has already been shown by Barstow et al. (2017), who performed retrievals on the same dataset and obtained a sub-solar H_2O abundance ($0.01 - 0.02\times$ solar) in excellent agreement with that found here and by Madhusudhan et al. (2014a). Though our H_2O abundances agree, the retrievals of Barstow et al. (2017) are somewhat limited by the usage of an optimal estimation algorithm, which explored only a limited volume of parameter space (3,600 spectra vs. our 10^8) on a pre-defined grid of temperature profiles and cloud properties. Furthermore, the lack of marginalisation over parameters or Bayesian evidence computation afforded by such an algorithm renders it impractical for parameter estimation or Bayesian model comparison, such as I have conducted here.

Another line of evidence into HD 209458b's H_2O abundance is dayside emission spectroscopy, as recently pursued by Line et al. (2016). They found a broad range of possible H_2O abundances: 0.06-10 × solar (i.e. including substantially sub-solar, as well as super-solar, H_2O abundances to 1σ). However, as they point out, the inferences are hampered by an anomalously high CO_2 abundance which is strongly correlated with the H_2O abundance. This is a well recognised problem in emission spectra retrievals (Heng & Lyons, 2016) which future work needs to investigate. This degeneracy does not, however, influence transmission spectra retrievals as considered in this thesis.

Other sophisticated retrieval analyses have relied on making *a priori* model assumptions to break cloud-composition degeneracies when inverting transmission spectra. Benneke (2015), who also used a nested sampling algorithm as in the present study, claimed a solar H_2O abundance for the terminator of HD 209458b. However, their approach explicitly imposed a vast array of *a priori* physics and chemistry: (i) P-T profiles are not retrieved, with radiative-convective equilibrium assumed; (ii) C-N-O chemical pathways are enforced; (iii) clouds, assumed to be composed of $MgSiO_3$, $MgFeSiO_4$ and SiC, are constructed using a model inspired by that of Ackerman & Marley (2001), and (iv) clouds are assumed to be uniform across the terminator.

More generally, the imposition of *a priori* assumptions has been used in retrievals of transmission spectra for several exoplanets. In such cases, chemical and/or radiative equilibrium is enforced to derive elemental O and C abundances (e.g., Benneke, 2015; Kreidberg et al., 2015). These approaches, more akin to forward models, undermine the ability of a retrieval to accomplish its fundamental goal: to infer the properties of an atmosphere with an absolute *minimal* set of assumptions. Succinctly put: we have shown that, even considering clouds and hazes, a sub-solar H_2O abundance at the terminator is essential to explain the low-amplitude spectral features of HD 209458b.

3.4.4 Cloud properties

The inference of a partially cloudy atmosphere along HD 209458b's terminator complements observations of inhomogeneous clouds in both the solar system and brown dwarfs. On Earth and Jupiter, a banded cloud structure arises from atmospheric convection cells transporting gas parcels vertically, where clouds form upon crossing the relevant condensation curve, with the dry air carried to a different latitude where the formation of clouds is suppressed (de Pater & Lissauer, 2001). A similar mechanism has been postulated to induce latitudinally inhomogeneous clouds in brown dwarfs (e.g., Marley et al., 2010), with observational evidence recently emerging (e.g., Buenzli et al., 2012).

Inhomogeneous cloud distributions have similarly been predicted to be common across the terminator region of hot Jupiters (Parmentier et al., 2016). The physical mechanism here is the day-night temperature contrast on tidally locked planets. This drives a super-rotating equatorial jet, which in turn raises the temperature of the eastern terminator by hundreds of K above that of the cooler western terminator (Showman & Guillot, 2002), where clouds are then more likely to form. Interestingly, I infer the properties of the cloudy region to consist of extremely high-altitude ($\approx 0.01 - 0.1$ mbar) clouds with a strongly scattering haze above the deck. The temperatures at these altitudes (see Figure 3.6) are less than those at which photochemical hazes are expected to form ($\sim 1000 - 1100$ K, see Zahnle et al., 2009; Moses, 2014), so it is possible that the cloud deck or hazes I infer may be photochemical in origin. This possibility will require detailed photochemical models to explore its plausibility.

The asymmetric cloud properties at HD 209458b's terminator naturally raises the possibility of asymmetry in other properties. For example, Kataria et al. (2016) predicted the eastern limb of HD 209458b's terminator should be warmer than the western limb by around 200 K, with NH_3 enhanced by an order of magnitude on the western limb. The retrievals presented here do not consider these effects, with the ability to disentangle the influence of clouds and chemistry contingent on both a sufficiently long spectral baseline (i.e. optical + near-infrared data) and the planet itself not possessing completely uniform clouds ($\bar{\phi} \approx 1$). Ultimately, the constraints derived by a retrieval algorithm are specific to the framework of the assumed models, and it remains to be investigated if inter-terminator differences between P-T profiles or abundances can be extracted using current or near-future transmission spectra.

Though future facilities, such as the James Webb Space Telescope (JWST), will revolutionise our understanding of exoplanetary atmospheres, much can still be accomplished with currently available high-precision HST spectra. The results presented here have re-affirmed that the key to obtaining precise chemical abundances from transmission spectra of cloudy exoplanets is rooted in the partially cloudy nature of their terminator. In other words, the light transmitting through cloud-free regions facilitates breaking many apparent degeneracies between clouds and chemistry. As discussed in Chapter 2, and explicitly demonstrated here for real exoplanet observations, clouds need not be an insurmountable issue for sufficiently high-precision HST transmission spectra. As we approach a decade since the advent of exoplanet atmospheric retrieval, the time has come to move beyond one-dimensional models. As we enter the golden age of retrieval, the future is inherently multi-dimensional.

3.5 Summary

In this chapter, I have demonstrated an application of the new atmospheric retrieval code POSEIDON to the visible and near-infrared transmission spectrum of the hot Jupiter HD 209458b. This retrieval analysis, published in MacDonald & Madhusudhan (2017a), is the most extensive of any conducted beforehand (involving the exploration of an unprecedented 10^8 spectra). The major results from this chapter are as follows:

1. The H_2O abundance on the terminator of HD 209458b is $30 - 100\times$ sub-solar (5-15 ppm). The sub-solar conclusion is robust to the assumed cloud model, even considering two-dimensional cloud/haze distributions.

2. Partially cloudy models are favoured by a Bayesian model comparison over both uniformly cloudy (4.5σ) and cloud-free (5.4σ) models. The terminator cloud fraction is constrained to $57^{+7}_{-12}\%$.

3. I report the first evidence of nitrogen chemistry in an exoplanet atmosphere – established at $> 3.7\sigma$ confidence. Both cloud-free and uniformly cloudy models identify NH_3 as the probable cause of nitrogen-induced absorption observed from 1.45–1.7 µm. The NH_3 abundance is constrained to 0.01-2.7 ppm.

4. Scattering due to high-altitude ($P < 0.05$ mbar) hazes is inferred at 3.2σ.

5. The terminator temperature structure of a transiting exoplanet can be constrained using high-precision HST transmission spectra. The temperature in the line-of-sight near-infrared photosphere is constrained to 1221^{+131}_{-138} K.

This chapter has demonstrated that a wealth of information may be extracted from currently available HST transmission spectra. In particular, current observations are already capable of inferring, at high-significance, the presence of 2D cloud distributions. As exoplanet atmospheres are inherently multidimensional, treating them as one-dimensional will at best miss key insights, and at worse incur unnecessary degeneracies. Precision measurements of the nature of exoplanet atmospheres are therefore already possible when high-precision observations meet sophisticated retrieval algorithms.

Chapter 4

Signatures of New Molecules in Hot Jupiter Atmospheres

4.1 The new chemical frontier

The chemical characterisation of exoplanetary atmospheres is rapidly entering a golden era. Robust detections of C, H, and O-bearing molecules from infrared spectroscopy are now commonplace (e.g. Snellen et al., 2010; Deming et al., 2013; Sing et al., 2016) – see also the summaries in section 1.4. As we saw in Chapter 3, precise HST WFC3 spectra can provide high-significance detections of H_2O in hot Jupiter atmospheres. However, H_2O is not the only molecule with strong features in the $\sim$ 1.1–1.7 µm WFC3 range. Strong features due to other molecules, such as CH_4, NH_3, and HCN, can also occur at these wavelengths (e.g. Figure 2.3). Similarly, visible transmission spectra have offered detections of Na and K (e.g. Snellen et al., 2008; Wilson et al., 2015; Sing et al., 2015), but these are not the only absorbing chemical species at visible wavelengths. In particular, metal oxides, such as TiO and VO, have been predicted to be especially prominent visible absorbers in highly-irradiated exoplanet atmospheres (Hubeny et al., 2003; Fortney et al., 2008). There is clearly a need to consider a wide variety of potential chemical species when retrieving exoplanet atmospheres.

In Chapter 3, I identified the first evidence of nitrogen chemistry (NH_3 / HCN) in a hot Jupiter atmosphere. This finding opens a variety of new avenues for transmission spectroscopy of exoplanets. In particular, it offers the intriguing possibility that signatures of new molecular species, previously unseen in exoplanet atmospheres, are accessible to already-existing observational facilities. In further exploring this, and associated ideas, some natural questions emerge:

- Is evidence of nitrogen chemistry also present in other hot Jupiters?
- What are the key factors governing the detectability of nitrogen chemistry?
- Can signatures of NH_3 and HCN be disentangled to produce unique detections?
- Are other molecules detectable from low-resolution transmission spectra with current space or ground-based instruments?
- What do the retrieved abundances of these new molecules tell us about the nature of exoplanet atmospheres?

In this chapter, I tackle these questions by pursuing two lines of inquiry. First, in section 4.2, I undertake a search for infrared spectral signatures of nitrogen chemistry in a population of hot Jupiters. Secondly, in section 4.3, I seek to identify the first signatures of heavy-element molecules in an exoplanet atmosphere. Finally, the implications of these two research directions are examined in section 4.4.

4.2 Nitrogen chemistry in hot Jupiters

I begin by generalising the preliminary study into signatures of nitrogen chemistry presented in Chapter 3 (which focused on HD 209458b's atmosphere). In what follows, the presence of NH_3 and HCN in hot Jupiter atmospheres is motivated in section 4.2.1. The detectability of nitrogen chemistry in hot Jupiter transmission spectra is explored in section 4.2.2. Finally, in section 4.2.3, I present the results of a search for signatures of nitrogen chemistry in 9 hot Jupiter atmospheres.

4.2.1 Theoretical motivation

Nitrogen chemistry is expected to exist in exoplanetary atmospheres (Burrows & Sharp, 1999; Lodders & Fegley, 2002). However, the anticipated equilibrium abundances of such species in the upper atmospheres of hot Jupiters are small: $\sim 10^{-7}$ and $\sim 10^{-8}$ for NH_3 and HCN respectively – assuming solar composition, C/O = 0.5, and N/O = 0.2 at $\sim 1500\,\mathrm{K}$ (Madhusudhan, 2012; Heng & Tsai, 2016). Detecting such trace abundances is impractical with current observations, often leading to the exclusion of such molecules from exoplanetary spectral analyses. However, observable nitrogen chemistry may occur under some circumstances. One avenue is enhanced elemental ratios: HCN abundances increase by $\sim 10^4$ for C/O $\gtrsim 1$ (Madhusudhan, 2012); both

NH_3 and HCN weakly increase with N/O (Heng & Tsai, 2016). Such enhanced ratios could be remnants of planetary formation (Öberg et al., 2011; Madhusudhan et al., 2014c; Mordasini et al., 2016; Piso et al., 2016).

Alternatively, disequilibrium chemistry can enhance NH_3 and HCN abundances by $\gtrsim 2$ orders of magnitude over equilibrium expectations at altitudes probed by transmission spectra (Zahnle et al., 2009; Moses et al., 2011, 2013a; Venot et al., 2012). There are two principle disequilibrium avenues: (i) transport-induced quenching (e.g., via vertical mixing), and (ii) photochemistry (e.g., by UV irradiation). Quenching occurs in atmospheric regions where a dynamical transport process is characteristically faster than a certain chemical reaction (e.g. $N_2 + 3\,H_2 \rightleftharpoons 2\,NH_3$). The transport process then fixes the chemical abundances to equilibrium values from atmospheric regions where local conditions result in a commensurate chemical reaction timescale. For NH_3 and HCN, this occurs in the deep atmosphere ($P \sim 1$ bar, Moses et al., 2011), where equilibrium abundances are considerably higher. Vertical mixing then dredges-up these molecules to the upper atmosphere.

Photochemistry can enhance HCN abundances (at the expense of NH_3, CH_4, and N_2) for pressures $\lesssim 10^{-3}$ bar (Zahnle et al., 2009; Moses et al., 2011). Photochemical deviations should become more pronounced for molecules produced by photochemistry, and those sensitive to it, on lower temperature planets, due to the deeper quench points and slower reaction rates impeding attempts to drive products back towards equilibrium (Moses et al., 2011). These conclusions are relatively insensitive to the C/O ratio (Moses et al., 2013a). An atmosphere subjected to extreme photochemistry may then display abundant HCN and depleted NH_3 in the photosphere, whilst one with strong vertical mixing and minimal photochemistry could display abundant NH_3 and / or HCN.

The impact of disequilibrium nitrogen chemistry on transmission spectra has been considered in forward modelling studies. Shabram et al. (2011) identified HCN absorption features at ~ 1.5, 3.3 and 7 µm, suggesting that the *James Webb Space Telescope* (JWST) NIRSpec prism will be able to observe the former two. Moses et al. (2011) commented that disequilibrium nitrogen chemistry could potentially reach detectable levels, and therefore strongly recommended including HCN and NH_3 within spectral analyses. Without including these disequilibrium products, as is somewhat common in atmospheric retrievals, the prospect of detecting nitrogen chemistry has been artificially quenched.

Towards observing nitrogen chemistry in hot Jupiters

In Chapter 3, I identified a slope from ∼ 1.5–1.7 µm in the WFC3 transmission spectrum of HD 209458b, suggesting NH_3 or HCN as possible contributors. At the precision of the data, either molecule provided reasonable fits. However, as discussed in section 3.3.2, qualitatively different WFC3 features become apparent at higher spectral resolution: an 'NH_3 shoulder' redwards of the 1.4 µm H_2O feature, vs. a 'HCN peak' around 1.55 µm. The NH_3 feature appears to have been missed in previous spectral studies. There are two primary reasons for this: (i) NH_3 is often not included as an opacity source when modelling hot Jupiter spectra (e.g. Deming et al., 2013; Madhusudhan et al., 2014b; Barstow et al., 2017); (ii) models assuming chemical equilibrium, or an assumed chemical network, as part of the retrieval (e.g Benneke, 2015; Sing et al., 2016) tend to have low NH_3 abundances. Incomplete opacity data below ∼ 3 µm (e.g., see Shabram et al., 2011, their Fig. 5) could also contribute, as many studies have not used the latest high-temperature NH_3 and HCN line lists (Yurchenko et al., 2011; Barber et al., 2014; Tennyson et al., 2016). The evidence I presented for nitrogen chemistry in HD 209458b has since motivated retrievals to include nitrogen chemistry for other planets. For example, Kilpatrick et al. (2018) observed an apparent absorption feature at 1.55 µm in WASP-63b's transmission spectrum. Atmospheric retrievals by four different groups identified this as consistent with super-solar HCN.

4.2.2 Detectability of nitrogen chemistry

Here, I examine the optimum near-infrared regions to search for nitrogen chemistry. A logical place to begin is a comparison of the cross sections of NH_3 and HCN to that of the dominant near-infrared absorber, H_2O. I then explore how various atmospheric properties influence the prevalence of NH_3 and HCN absorption signatures. Finally, I consider how these findings can be employed by ground and space based facilities to uniquely detect NH_3 and HCN in exoplanet atmospheres.

NH_3 & HCN absorption features

Figure 4.1 contrasts the cross sections of NH_3 and HCN to that of H_2O from 1–5 µm at 1000 K, 1500 K, and 2000 K. We see that it is possible for the cross sections of nitrogen-bearing molecules to exceed that of H_2O by ∼ 2 orders of magnitude at wavelengths where H_2O possesses local absorption minima. The WFC3 G141 band pass contains NH_3 and HCN features around ∼ 1.5–1.6 µm, along with a weaker unique

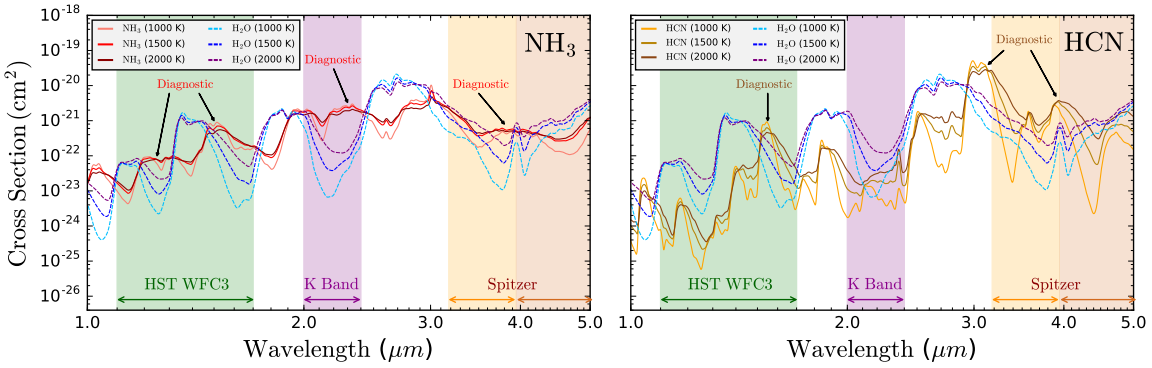


Fig. 4.1 Near-infrared NH_3 and HCN absorption cross sections. Left: NH_3 cross section (red, solid) compared to H_2O (blue, dashed), at $T = 1000$ K, 1500 K, and 2000 K (darker colours imply higher temperatures) and $P = 1$ mbar. Right: HCN cross section (orange, solid) compared to H_2O under the same conditions. All cross sections have been smoothed for clarity. The shaded regions indicate the spectral ranges / band passes for WFC3 G141, K-band, and Spitzer IRAC. The wavelength range shown is observable by JWST NIRSpec (or NIRISS+NIRCam) – see Beichman et al. (2014). Diagnostic strong NH_3 and HCN absorption features are annotated.

NH_3 feature at $\sim 1.2\,\mu$m. NH_3 possesses a prominent feature at $\sim 2.2\,\mu$m (K-band), whilst HCN has an especially strong feature at $\sim 3.1\,\mu$m. Both molecules absorb at $4\,\mu$m, between the two Spitzer IRAC photometric bands. The K-band NH_3 feature is a powerful diagnostic, coinciding with minima for both H_2O and HCN. The cross section contrast between NH_3 / HCN and H_2O tends to increase at lower temperatures, suggesting that lower temperature planets may possess amplified nitrogen chemistry features (as further examined later in this section). HCN features peak more sharply than NH_3 features at lower temperatures, which can enable unique identification in regions of overlapping absorption (e.g. in the WFC3 G141 band pass).

Factors influencing detectability

The relative strengths of absorption cross sections are not the only consideration governing the imprint of nitrogen chemistry into transmission spectra. I now consider differences between model transmission spectra with and without opacity due to NH_3 and HCN. By defining the difference between two such model spectra as the *transit depth excess*, I explore how these differences vary across a grid of atmospheric models computed by POSEIDON. Specifically, I consider how signatures of nitrogen chemistry are influenced by the *relative* abundances NH_3/H_2O and HCN/H_2O, the atmospheric temperature, and across the transition from clear to cloudy atmospheres.

Consider a fiducial hot Jupiter system with the following properties: $R_p = 1.2\,R_J$, $R_* = R_\odot$, and $g_p = 10\,\mathrm{ms}^{-2}$. The model planet has an atmosphere at $T = 1500\,\mathrm{K}$ (isothermal with height), with volume mixing ratios (besides NH_3 and HCN) representative of chemical equilibrium: $\log(X_{H_2O})$ = -3.3, $\log(X_{CH_4})$ = -6.0, $\log(X_{CO})$ = -3.3, $\log(X_{CO_2})$ = -7.0 (Madhusudhan, 2012). These 'background' abundances are held constant throughout this section. The planet is taken to have a high-altitude cloud deck at $P_{\mathrm{cloud}} = 1\,\mathrm{mbar}$, with a terminator cloud fraction of 50%. The reference pressure, P_{ref}, is set at 10 mbar. This model atmosphere is then 'injected' with an amount of nitrogen chemistry given by either X_{NH_3}/X_{H_2O} or $X_{HCN}/X_{H_2O} = 0.1$, producing two reference models with either enhanced NH_3 or HCN.

I compute transmission spectra for both the 'enhanced NH_3' and 'enhanced HCN' reference models, with and without opacity due to nitrogen chemistry enabled. The inset of Figure 4.2 demonstrates (for the NH_3 reference model) how the spectrum is altered when nitrogen chemistry opacity is disabled – with the shaded difference region defining the transit depth excess. This calculation is then repeated for perturbed versions of each reference model to explore the effects of NH_3 and HCN abundances, atmospheric temperature, and cloud fraction, on the prominence of nitrogen chemistry signatures. The influence of each of these three factors is considered in turn.

Figure 4.2 (top) demonstrates the transit depth excess to be strongly correlated with the relative mixing ratios of each nitrogen-bearing species to water. This trend can be understood by considering relative extinction coefficients. For example, $\kappa_{HCN}/\kappa_{H_2O} = (X_{HCN}/X_{H_2O})\ \sigma_{HCN}/\sigma_{H_2O}$, and so the relative importance of HCN opacity vs. H_2O increases in proportion with the relative mixing ratio. This expression also highlights the importance of their relative cross sections. Since the cross sections of NH_3 and HCN are rarely enhanced by more than 100× over H_2O from 1–5 µm (Figure 4.1), absorption signatures start to become negligible for X_{NH_3}/X_{H_2O} or $X_{HCN}/X_{H_2O} \lesssim 10^{-2}$ (corresponding to $\kappa_{NH_3}/\kappa_{H_2O}$ or $\kappa_{HCN}/\kappa_{H_2O} \lesssim 1$). However, when the abundances of NH_3 or HCN become comparable to H_2O, a plethora of absorption features with amplitudes $\gtrsim 300\,\mathrm{ppm}$ emerge throughout the near-infrared.

Figure 4.2 (middle), illustrates the influence of atmospheric temperature on the transit depth excess. Two effects compete as temperatures decrease: (i) the H_2O cross section exhibits deeper minima (Figure 4.1); and (ii) the atmospheric scale height decreases. The combined effect is for many NH_3 / HCN features to initially intensify from 2000 K → 1500 K. The strongest features (which saturate the atmosphere) then dampen from 1500 K → 1000 K due to the overall height of the atmosphere decreasing.

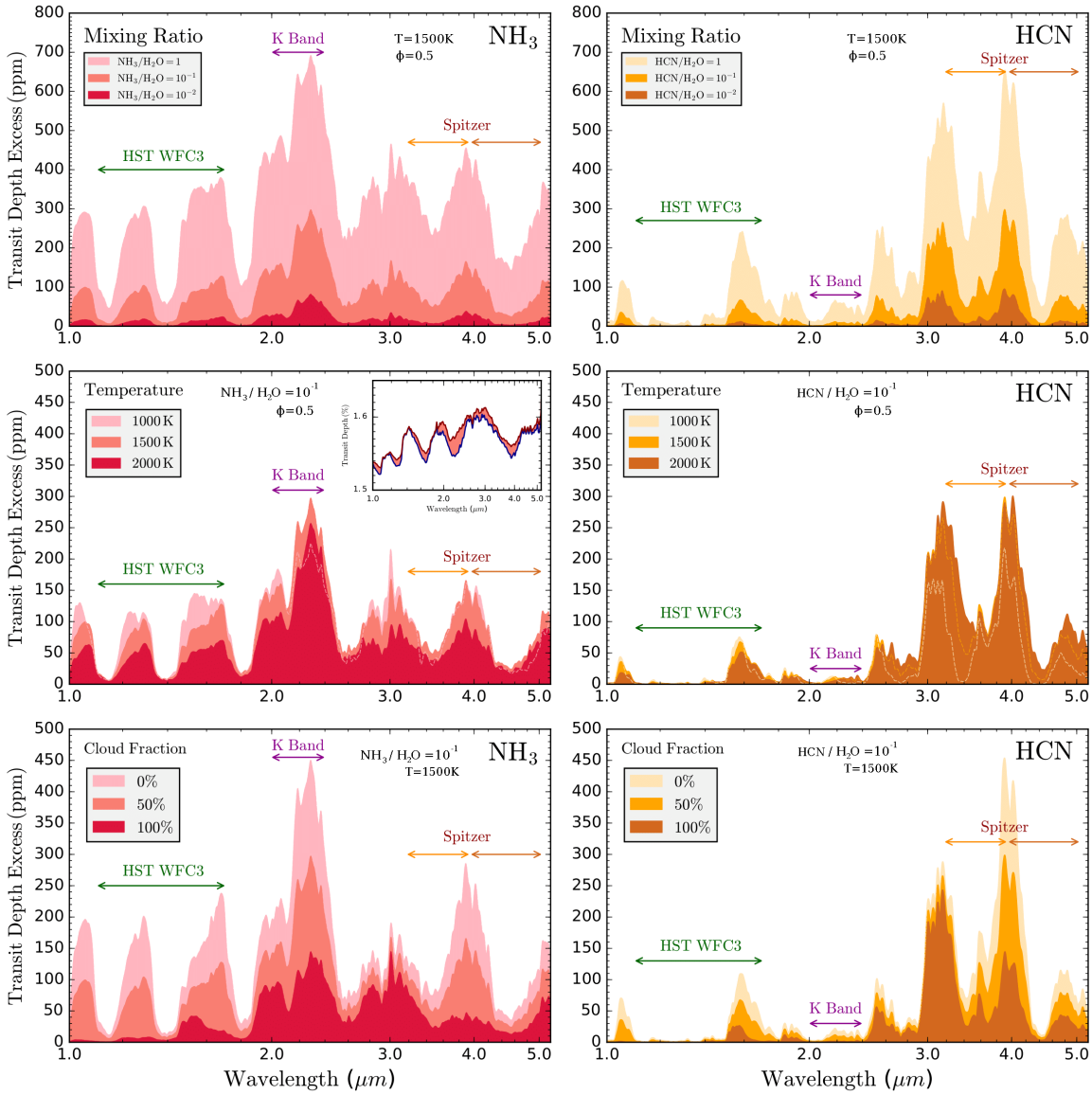


Fig. 4.2 Factors influencing NH_3 & HCN signatures in transmission spectra. A reference model with enhanced nitrogen chemistry (see text for details) is perturbed in mixing ratio, temperature, and cloud fraction. The 'transit depth excess' results from subtracting a transmission spectrum including nitrogen chemistry from an identical model with nitrogen chemistry opacity disabled (as demonstrated in the inset). All transmission spectra are computed at $R = 10^4$, before being smoothed for clarity. Left: transit depth excess variations for a reference model with NH_3. Right: transit depth excess variations for a reference model with HCN. Dashed lines indicate covered regions. The WFC3, K-band, and Spitzer IRAC band passes are indicated.

Weaker features, such as NH_3 at 1.4 µm, can continue to strengthen with decreased temperature in line with the changing relative cross sections. Generally, HCN features become sharper for cooler temperatures, as expected from the cross sections (Figure 4.1), whilst NH_3 features retain similar widths. Overall, nitrogen chemistry absorption features remain potent over a wide range of temperatures expected in hot Jupiter atmospheres ($\sim$ 1000-2000 K) – especially in the WFC3 G141 band pass for cooler temperatures. Note in particular that strong absorption in the K-band ($\sim$ 2.2 µm) is a robust NH_3 diagnostic for $T \gtrsim 1000$ K, whilst the $\sim$ 3.1 µm and 4.0 µm HCN features are prominent for $T \gtrsim 1500$ K. Signatures of nitrogen chemistry, if present, may therefore be detectable even for higher temperature hot Jupiters.

Figure 4.2 (bottom), demonstrates that increasing terminator cloud coverage (i.e. $\bar{\phi} \rightarrow 1$) generally dampens absorption contrasts. This is unsurprising, as a high-altitude cloud deck with near-complete terminator coverage indiscriminately blocks electromagnetic radiation. Despite this, the strongest absorption features (K-band for NH_3, and $\sim$ 3.1 µm and 4.0 µm for HCN) can remain prominent even for uniform terminator clouds (darkest shading in Figure 4.2). Increased dampening can result from clouds at higher altitudes (lower pressures) than that considered here, though it is unclear theoretically if grey cloud decks can exist at $P_{\rm cloud} < 1$ mbar (Fortney et al., 2010; Parmentier et al., 2013). Absorption features located near H_2O cross section minima strengthen considerably as the terminator becomes cloud-free; NH_3 or HCN, rather than clouds, then become the dominant opacity source at these wavelengths. Where H_2O absorption is also prominent, (e.g. 3.1 µm), signatures of nitrogen chemistry are less sensitive to the cloud fraction. This change in the relative amplitudes of certain NH_3 or HCN features (e.g. signatures at 3.1 µm vs. 4.0 µm) offers an avenue to constrain terminator cloud fractions using near-infrared transmission spectra.

A Strategy to Uniquely Detect NH_3 and HCN

Figure 4.2 demonstrates that signatures of nitrogen chemistry, often $\gtrsim$ 100 ppm, are detectable with HST WFC3 spectra. In particular, absorption at $\sim$ 1.2 µm and in the K-band uniquely indicates NH_3. HCN absorbs strongly at $\sim$ 3.1 µm and 4.0 µm. I suggest ground-based K-band photometry or spectroscopy as a promising avenue to assess the presence of NH_3. Null detections in K-band could rule out NH_3, whilst suggesting HCN as the cause of excess 1.55 µm WFC3 absorption. Furthermore, robust detections of HCN via the $\sim$ 3.1 µm and 4.0 µm features will be feasible with JWST.

4.2.3 Evidence of nitrogen chemistry in hot Jupiter spectra

Having identified nitrogen chemistry absorption features, I proceed to search for these signatures in observed transmission spectra across a population of hot Jupiters. I apply a uniform atmospheric retrieval analysis to 9 hot Jupiter transmission spectra, spanning visible to near-infrared wavelengths ($\sim$ 0.3–5 µm). After briefly describing the transmission spectra observations and retrieval approach, I present constraints on nitrogen chemistry for the most promising candidate exoplanets.

Observations

Hot Jupiter transmission spectra were obtained from the compilation presented in Sing et al. (2016). They provide uniformly reduced observations for 10 hot Jupiters, including HST STIS and WFC3 spectra along with Spitzer IRAC photometry. I focus on the 8 planets for which data from all three instruments are available: WASP-12b, WASP-17b, WASP-19b, WASP-31b, HAT-P-1b, HAT-P-12b, HD 189733b, and HD 209458b. I also include the WFC3 spectrum of WASP-63b (Kilpatrick et al., 2018), due to suggestive indications of HCN absorption around 1.55 µm.

Retrieval approach

The configuration of POSEIDON used for these retrievals is broadly the same as those for HD 209458b in Chapter 3 (described in section 3.2). However, one notable difference is that the spectral range of the forward models considered here must extend to cover the Spitzer IRAC bands at $\sim$ 3.6 µm and $\sim$ 4.5 µm. Transmission spectra are therefore generated from 0.2–5.2 µm at $R = 1000$. As CO and CO_2 absorb strongly in these bands, they must also be included as parameters in the retrievals. I also describe Na and K by separate parameters (instead of a single parameter as in chapter 3). To ensure a uniform analysis, all priors are as given in Table 3.1 (with the exception of T_0, for which the upper limit is set at $T_{\rm eq} + 200$ K).

The models considered here have a maximum of 19 parameters: 6 for the P-T profile, 8 for the mixing ratios (Na, K, H_2O, CH_4, NH_3, HCN, CO, and CO_2), 4 for clouds / hazes, and $P_{\rm ref}$. An initial retrieval, using all 19 parameters, was performed on the transmission spectrum of each planet. Wherever the best-fitting retrieved spectrum featured a WFC3 transit depth excess (due to NH_3 or HCN) > 30 ppm, the planet was designated a nitrogen chemistry candidate. This resulted in three candidates with plausible suggestions of nitrogen chemistry: WASP-31b, WASP-63b, and HD 209458b.

For the three candidate planets, a further set of three retrievals were conducted: (i) including NH_3, but no HCN; (ii) including HCN, but with no NH_3; and (iii) without nitrogen chemistry. I also provide the posterior distributions from the full retrieval for each of these planets online[1]. The Bayesian evidences from these retrievals enable nested model comparisons, yielding detection significances of NH_3 or HCN. The best-fitting spectra from retrievals with and without nitrogen chemistry additionally provide visual evidence of which data (if any) is the source of inferred nitrogen chemistry.

Results: inferences of nitrogen chemistry

Figure 4.3 displays best-fitting spectra from the retrievals with and without nitrogen chemistry. WASP-31b and HD 209458b favour models including NH_3, whilst WASP-63b favours models including HCN. WASP-31b features a strong NH_3 transit depth excess of ~ 400 ppm at ~ 1.55 µm. WASP-63b exhibits a HCN transit depth excess of ~ 200 ppm at similar wavelengths. HD 209458b has the weakest signature of nitrogen chemistry, with a transit depth excess of ~ 50 ppm attributed to NH_3. The blue curves in Figure 4.3 demonstrate the ability of a retrieval without nitrogen chemistry to compensate by adjusting other atmospheric parameters. This somewhat obscures the NH_3 signal for HD 209458b, but demonstrates for WASP-31b and WASP-63b that retrievals without nitrogen chemistry struggle to match the observations.

Uniquely identifying nitrogen-bearing species is challenging at the resolution and precision of present WFC3 observations – largely due to overlapping NH_3 and HCN absorption features. This difficulty is particularly apparent for HD 209458b, as we saw in Chapter 3. In this chapter, I have been more conservative in evaluating evidence for nitrogen chemistry (by marginalising over cloud fractions, instead of fixing $\bar{\phi}$). As of such, signatures of nitrogen chemistry in HD 209458b are less significant against this more stringent criteria. However, the inclusion of Spitzer data provides additional information from the different shapes of NH_3 and HCN absorption at these wavelengths (Figure 4.1). As a result, the NH_3-HCN degeneracy seen for HD 209458b in Chapter 3 is lifted, with NH_3 identified as the favoured molecule. On the other hand, unique identification of nitrogen-bearing molecules is far easier for planets with higher NH_3 or HCN abundances relative to H_2O (Figure 4.2). This is the case for both WASP-31b and WASP-63b, where strong WFC3 features permit unique identification of signatures attributable to specific nitrogen-bearing molecules (as shown in Figure 4.4).

[1] http://doi.org/10.5281/zenodo.1014847

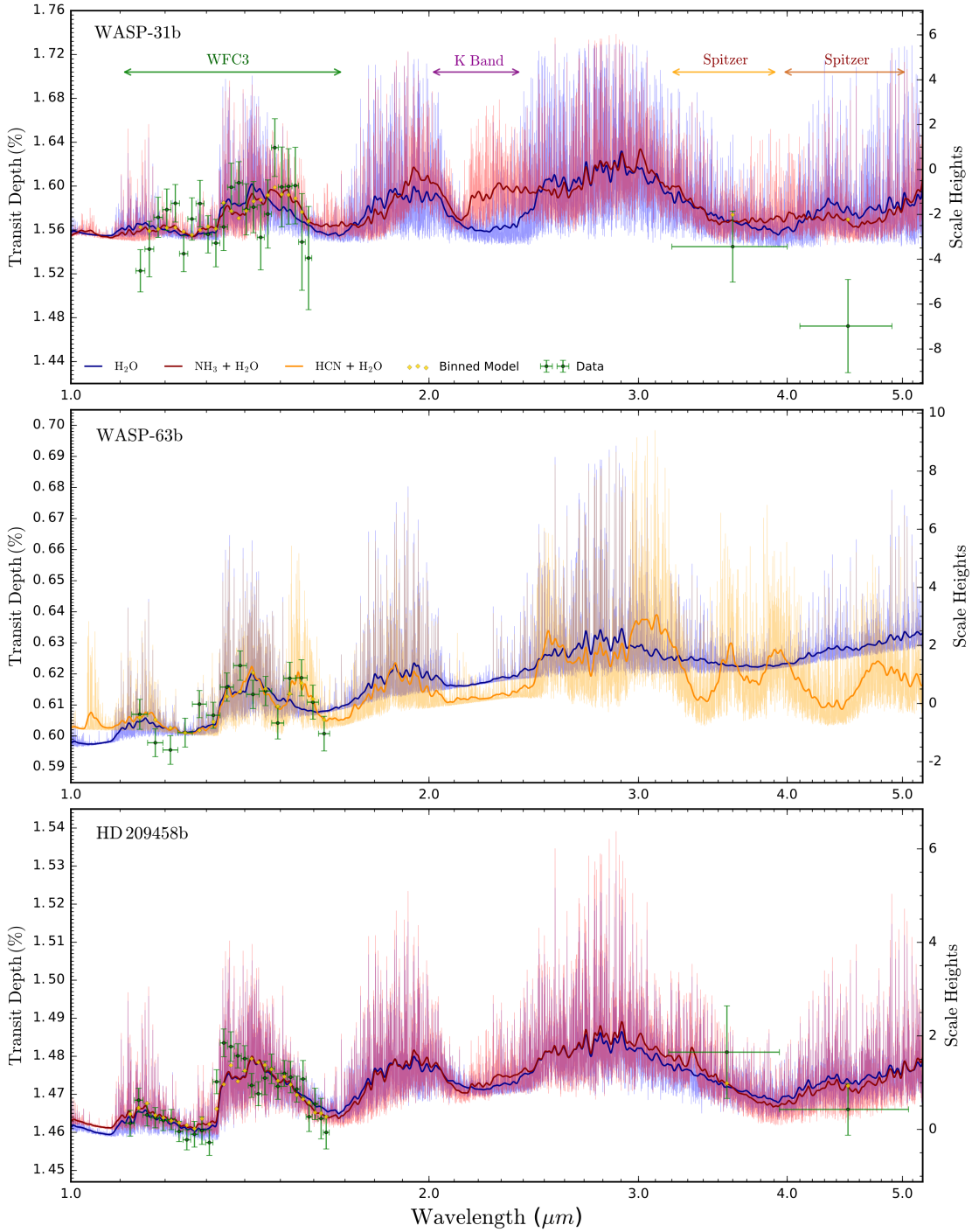


Fig. 4.3 Transmission spectra of nitrogen chemistry candidate planets. Best-fitting transmission spectra (at $R = 5000$) from three retrievals are shown: no nitrogen chemistry (blue), including NH_3 (red), and including HCN (orange). The dark curves are smoothed representations of each spectrum. The observations (green error bars) and binned points for the best-fitting spectrum (gold diamonds) are over-plotted.

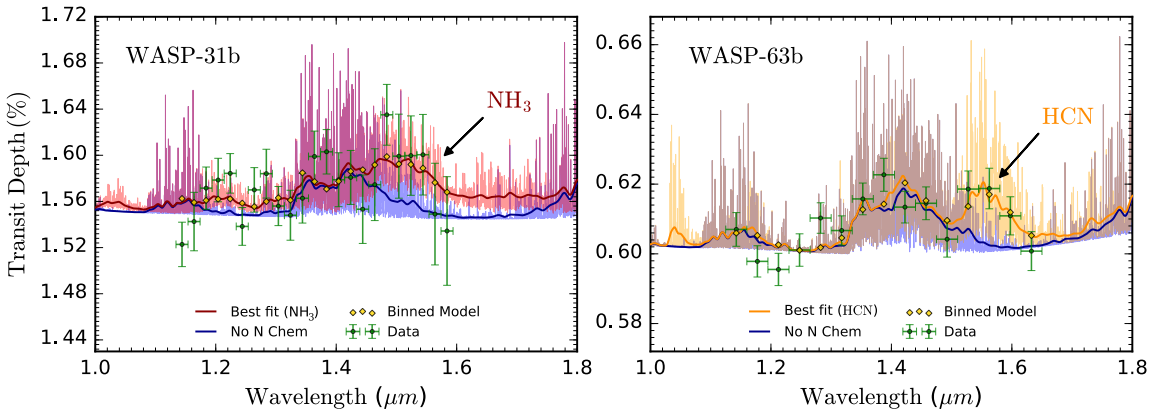


Fig. 4.4 Evidence of nitrogen chemistry in hot Jupiter transmission spectra. Left: evidence of NH_3 in WASP-31b (2.2σ). Right: evidence of HCN in WASP-63b (2.3σ). The blue spectra are otherwise identical to the best-fitting spectra, but with opacity due to nitrogen chemistry disabled. The dark curves are smoothed representations of each spectrum. The observations (green error bars) and binned points for the best-fitting spectrum (gold diamonds) are over-plotted.

I report weak evidence of NH_3 in WASP-31b (2.2σ). A model comparison between retrievals with NH_3 + HCN (i.e. the full retrieval) and without NH_3, establishes a Bayes factor[2] of 3.8, identifying NH_3, within the framework of the present model, as the only scenario explaining the $\sim$ 400 ppm WFC3 feature around 1.5 µm. Previous studies of WASP-31b were unable to fit this feature, either due to the non-inclusion of NH_3 in their models (Sing et al., 2015; Barstow et al., 2017) or by assuming chemical equilibrium (Sing et al., 2016). The retrieval conducted here without nitrogen chemistry (Figure 4.3, blue) similarly struggles to fit this feature. If these data points are indeed explained by NH_3, I predict a $\sim$ 500 ppm K-band NH_3 feature (see Figure 4.3). If confirmed, this represents the first inference of NH_3 in a transiting exoplanet atmosphere.

I further report weak evidence of HCN in WASP-63b (2.3σ, Bayes factor = 4.7), due to a $\sim$ 200 ppm peak around 1.55 µm. This result has been confirmed by independent retrievals (at varying significance levels) using other retrieval codes (Kilpatrick et al., 2018). If HCN is responsible for the observed WFC3 excess absorption, I predict a $\sim$ 400 ppm feature will manifest near 3.1 µm, along with a $\sim$ 200 ppm feature near 4.0 µm (see Figure 4.3). Note that these detection significances include marginalisation over the entire parameter space, including NH_3-HCN and cloud-composition degeneracies; therefore the transmission spectra of WASP-31b and WASP-63b *cannot* be adequately fit without disequilibrium nitrogen chemistry.

[2]All Bayes factors given here are for *unique* species, not the combination of NH_3 + HCN.

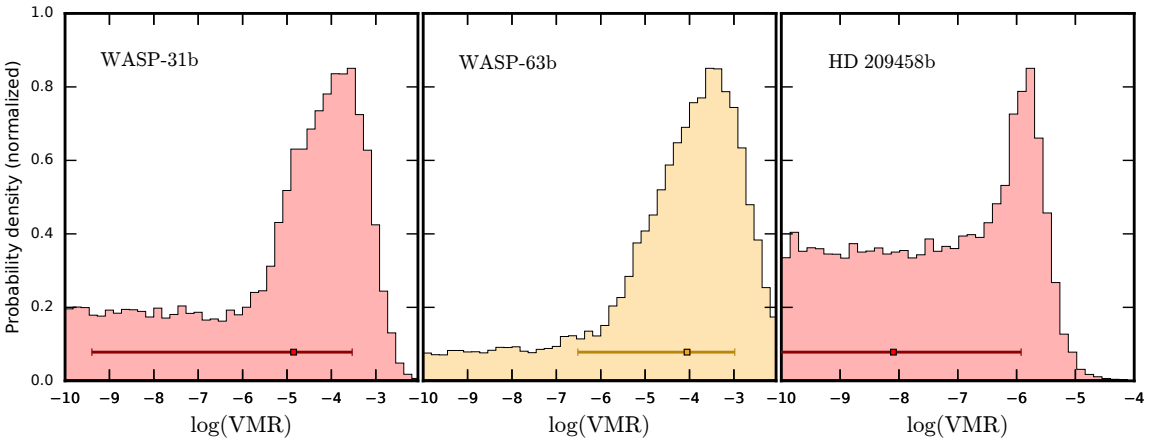


Fig. 4.5 Retrieved NH_3 and HCN abundances from hot Jupiter spectra. The histograms give the marginalised posterior distributions for the NH_3 and HCN volume mixing ratios ('VMR', i.e. X_{NH_3} and X_{HCN}). The nitrogen chemistry abundances retrieved by POSEIDON, from left to right, are derived from the transmission spectra of WASP-31b, WASP-63b, and HD 209458b. The posteriors are coloured according to the molecular signature favoured by their spectra: NH_3 (red) or HCN (orange).

Constraints on the mixing ratios of NH_3 and HCN for each planet are shown in Figure 4.5. The maximum *a posteriori* solutions show abundances enhanced by $\sim$ 3-4 orders of magnitude over equilibrium expectations for WASP-31b and WASP-63b, and $\sim$ 1 order of magnitude for HD 209458b. The larger abundances for these two planets result in more prominent signatures of nitrogen chemistry (see Figure 4.3). The derived abundance constraints are relatively weak due to the lack of lower bounds. These abundance 'tails' arise from degeneracies, as I will now discuss.

Resolving degenerate solutions

The limited wavelength range of current observations permits a range of possible interpretations. This is particularly true for WASP-63b, where the lack of Spitzer or visible data precludes determining the spectral continuum. With low resolution or limited precision data, retrievals have flexibility in adjusting other parameters to partially compensate for removing NH_3 or HCN. For example, molecular abundances can be degenerate with terminator cloud coverage or with each other. However, present observations are of sufficient quality to distinguish NH_3 and HCN features from CH_4 (which absorbs at $\sim 1.7\,\mu m$, rather than $\sim 1.55\,\mu m$ for NH_3 or HCN).

Despite degenerate solutions arising from WFC3 observations, model differences arising at longer wavelengths can resolve these degeneracies. Considering Figure 4.3, one can see that K-band spectra of WASP-31b or Spitzer observations of WASP-63b can further differentiate between solutions with and without nitrogen chemistry. Such observations would offer tighter constraints on their NH_3 and HCN abundances, respectively. HD 209458b is more challenging, as the low inferred NH_3 abundance only predicts K-band absorption of ~ 25 ppm due to NH_3. However, the demonstration in this chapter that the addition of the two (relatively imprecise) Spitzer points was sufficient to resolve the NH_3-HCN degeneracy from Chapter 3 is an important demonstration of the constraining power of observations covering multiple absorption features over a wide spectral baseline. Ultimately, observations in K-band and at 3.1 µm or 4.0 µm are critical to resolving model degeneracies. Looking to the future, the entire 1–5 µm spectral range will be covered by JWST, at an enhanced spectral resolution and precision, holding great promise for robust detections of disequilibrium nitrogen chemistry, and precise abundance constraints, in the years to come.

4.3 Detection of a heavy molecule in a hot Jupiter

Atmospheric retrieval of space-based infrared spectra provides a powerful means to both detect and constrain the abundances of light molecular species. I showed in Chapter 3 that HST WFC3 spectra can detect H_2O, whilst in section 4.2 I further illustrated that near-infrared spectra can provide evidence of NH_3 and HCN. Here, I will demonstrate the promise of *ground-based* optical transmission spectroscopy, which offers a window into heavy-element molecular chemistry in exoplanet atmospheres.

In what follows, I first summarise the theoretical importance of metal oxides in hot Jupiter atmospheres, along with searches for them, in section 4.3.1. I propose a strategy to detect metal oxides using ground-based observations of hot Jupiters in section 4.3.2. Following this strategy, a Very Large Telescope (VLT) FORS2 hot Jupiter transmission spectrum of unprecedented quality is introduced in section 4.3.3. Finally, in section 4.3.4, I apply POSEIDON to this spectrum, and in so doing present the first robust detection of a heavy molecule in an exoplanet atmosphere.

4.3.1 The long search for metal oxides

Heavy metal oxides, in particular TiO, have been known to exist in the atmospheres of cool stars ($T_{\rm eff} < 4000\,{\rm K}$) for over a century (Fowler, 1904). Indeed, the stellar spectral class M is defined by the onset of prominent TiO features (Morgan et al., 1943). Cooler M dwarfs, and L-type brown dwarfs, additionally show signatures of VO and metal hydrides, such as CaH and FeH (Bessell, 1991; Kirkpatrick, 2005). The observational ease in detecting these heavy-element molecules arises from their strong opacity at visible wavelengths (Sharp & Burrows, 2007).

Early theoretical studies of hot Jupiter atmospheres highlighted the potential importance of metal oxides. Hubeny et al. (2003) showed that TiO and VO should exist for the conditions present in hot Jupiters (assuming chemical equilibrium), potentially causing temperature inversions in their dayside atmospheres. Fortney et al. (2008) used the presence (or absence) of TiO and VO opacity to define two distinct classes of exoplanet atmospheres. Those with TiO/VO ($T_{\rm eff} \gtrsim 1500\,{\rm K}$) were predicted to possess emission features in eclipse spectra (due to dayside temperature inversions, see section 1.4.2), whilst transmission spectra with TiO/VO would show strong absorption features at visible wavelengths (from ~ 0.4–$0.9\,\mu{\rm m}$). This link between metal oxides and inversions has come to be known as the *TiO/VO inversion hypothesis*.

Early searches for metal oxides in hot Jupiter atmospheres were motivated by tentative evidence for thermal inversions. Excess thermal emission in Spitzer observations of HD 209458b was interpreted as evidence for a thermal inversion (Burrows et al., 2007; Knutson et al., 2008). Many of the early attempts to verify the TiO/VO inversion hypothesis therefore focused on HD 209458b. Désert et al. (2008) searched for TiO and VO in the HST STIS transmission spectrum of HD 209458b. They inferred broadband absorption between the Na and K doublets attributable to an unknown absorber, possibly TiO or VO, but no specific spectral signatures were identified (largely due to the low-spectral resolution data). Subsequent atmospheric studies of HD 209458b cast doubt on both the presence of TiO and a dayside temperature inversion. Hoeijmakers et al. (2015) used high-resolution Doppler spectroscopy to report a non-detection of TiO in the transmission spectrum of HD 209458b (though they noted inaccurate TiO line positions in theoretical template spectra hinders detecting this molecule by cross correlation). Furthermore, recent re-analyses of Spitzer observations no longer favour an inversion on the dayside of HD 209458b (Diamond-Lowe et al., 2014; Line et al., 2016). Many searches for TiO in transmission spectra of other exoplanets have similarly returned null results (Sing et al., 2013; Gibson et al., 2013; Bento et al., 2014).

Non-detections of TiO and VO across multiple hot Jupiters require the abundances of these species to be lower than initially predicted. One explanation for the lack of prominent metal oxide features is condensation of these species from the upper atmosphere. Spiegel et al. (2009) showed that strong vertical mixing is required to maintain heavy species like TiO in the observable upper atmosphere. Parmentier et al. (2013) used a 3D model of HD 209458b to demonstrate that condensation in atmospheric cold traps can deplete the atmosphere of TiO, such that the molecule would no longer be observable. An alternative explanation is that atmospheres with high C/O ratios have less O available to form metal oxides (Madhusudhan et al., 2011).

Recently, suggestive evidence for metal oxides has emerged for two hot Jupiters. Haynes et al. (2015) observed the dayside eclipse spectrum of WASP-33b, finding 3 data points consistent with TiO emission features. Evans et al. (2016) observed the transmission spectrum of WASP-121b, noting that 3 data points in the visible are consistent with TiO or VO absorption (though a haze could also produce similar absorption). Despite these tantalising signs, and over a decade of searches, a definitive detection of metal oxides in exoplanet atmospheres has remained elusive.

4.3.2 A strategy to detect metal oxides in hot Jupiters

Previous attempts to detect metal oxides in hot Jupiter atmospheres have faced two key challenges:

1. Low-resolution spaced-based spectroscopy cannot resolve absorption bands of metal oxides or hydrides. Even if broadband absorption / emission features are inferred, they cannot be attributed to a specific species (e.g. Désert et al., 2008; Haynes et al., 2015; Evans et al., 2016).

2. High-resolution ground-based spectroscopy relies on accurate spectral line positions (at $R \sim 10^5 - 10^6$) to detect molecules. Most TiO line lists (e.g. Schwenke, 1998) fall short of this accuracy, resulting in unreliable template spectra for cross correlation analyses (Hoeijmakers et al., 2015).

In this section, I will demonstrate that an alternative approach, intermediate resolution ($R \sim 100 - 200$) ground-based transmission spectroscopy, offers a solution to both of these issues. To illustrate this point, I show the cross sections of TiO and VO (at $T = 2000\,\mathrm{K}$ and $P = 1\,\mathrm{mbar}$) in Figure 4.6. It is seen that HST STIS (and WFC3 G102) covers the spectral range with the strongest TiO and VO absorption features.

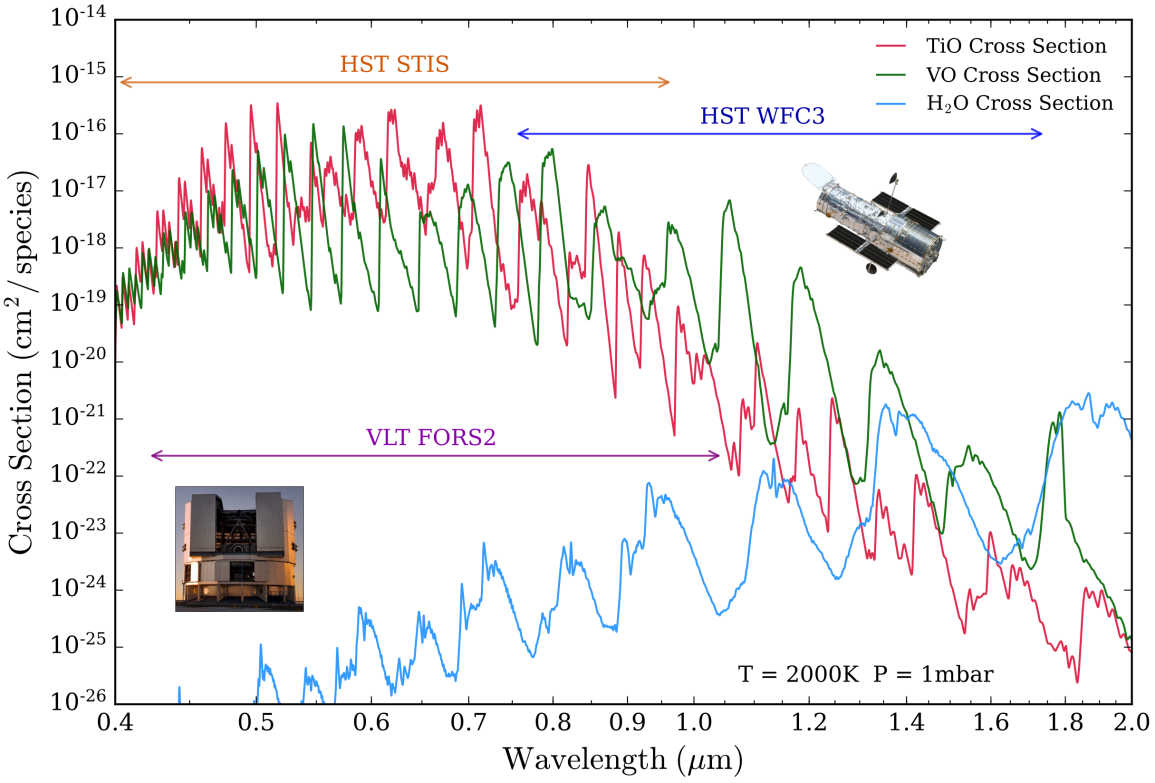


Fig. 4.6 Cross sections of the metal oxides TiO and VO.
The cross sections of TiO (red) and VO (green) are compared against H_2O (light blue). All cross sections are shown at $T = 2000\,\mathrm{K}$, $P = 1\,\mathrm{mbar}$, and have been smoothed from line-by-line resolution for clarity. The spectral ranges of three instruments are shown: HST STIS (orange arrows), HST WFC3 (dark blue arrows), and VLT FORS2 (purple arrows). Note that each of the plotted ranges comprise multiple grisms (e.g. WFC3 G102 and G141), and the lower range of STIS G430 can extend to $\sim 0.3\,\mu\mathrm{m}$.

However, whilst this excess absorption is clearly distinct from lighter molecules such as H_2O, the TiO and VO cross sections are indistinguishable at low spectral resolutions (e.g. the STIS spectrum of WASP-19b has 6 data points over this entire range, Sing et al., 2016). At somewhat higher resolutions ($R \gtrsim 100$), the band structure[3] of TiO and VO seen in Figure 4.6 can be resolved, allowing unique detections of metal oxides.

The spectral resolution required to detect metal oxides (and possibly metal hydrides) can be reached with the VLT FORS2 instrument. I demonstrate this in the next two sections. I first present the FORS2 optical transmission spectrum of the hot Jupiter WASP-19b, before retrieving its atmospheric composition with POSEIDON.

[3]Inaccurate line positions are less problematic at lower-resolutions, as line intensities are more important for determining band structure.

4.3.3 Transmission spectroscopy of a hot Jupiter with FORS2

WASP-19b is a highly-irradiated inflated hot Jupiter. Its physical properties are given by: $R_p = 1.31\,R_J$, $M_p = 1.15\,M_J$, and $g_p = 14.3\,\mathrm{ms}^{-2}$ (Hebb et al., 2010). WASP-19b orbits a late G-type star ($R_* = 0.993\,R_\odot$) with one of the shortest orbital periods amongst known hot Jupiters (19 hr / $a = 0.016$ au). The close proximity to its parent star results in an effective temperature $\gtrsim 2000$ K (Wong et al., 2016), rendering it a promising target for transmission spectroscopy.

From Nov. 2014 – Feb. 2016, three transits of WASP-19b were observed with the VLT FORS2 instrument. Each transit was observed with a grism covering a different wavelength range: 600B (0.44–0.61 µm), 600RI (0.541–0.846 µm), and 600z (0.745–1.03 µm). The reduction of these three sets of observations to produce a complete optical transmission spectrum is described in detail in Sedaghati et al. (2017). Figure 4.7 shows the VLT FORS2 optical transmission spectrum of WASP-19b. The

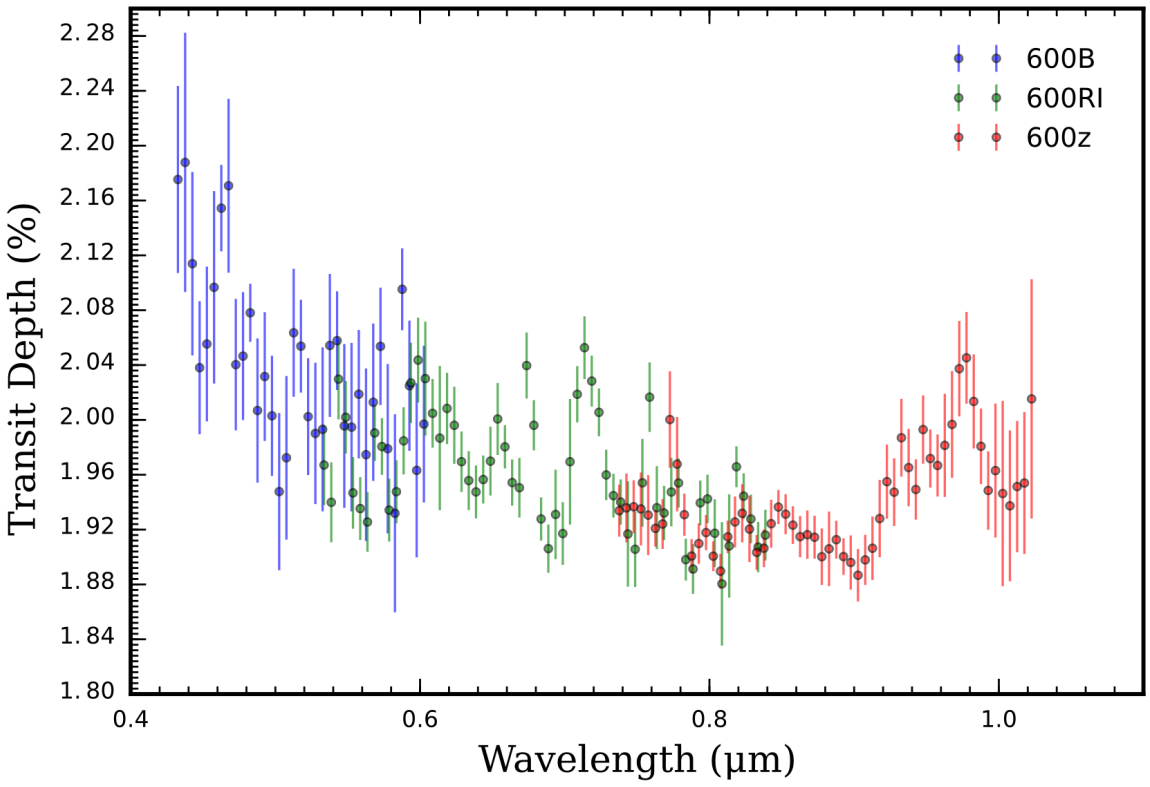


Fig. 4.7 VLT FORS2 optical transmission spectrum of WASP-19b. The blue, green, and red data points correspond to VLT FORS2 observations using the 600B, 600RI, and 600z grisms, respectively. The combined transmission spectrum consists of 155 data points with continuous coverage from 0.44–1.03 µm.

data from each grism has the following properties: 600B ($N_{\rm data}$ = 35, $R \approx 105$, $\sigma \approx 530$ ppm), 600RI ($N_{\rm data}$ = 62, $R \approx 140$, $\sigma \approx 240$ ppm), and 600z ($N_{\rm data}$ = 58, $R \approx 180$, $\sigma \approx 260$ ppm). The combined transmission spectrum offers 155 data points with continuous coverage from 0.44–1.03 µm. Even by eye, suggestions of absorption features are apparent. This spectrum represents an ideal test case to demonstrate atmospheric retrieval using a high-quality ground-based optical transmission spectrum.

4.3.4 The inferno world with titanium skies

Here, I present the results of an atmospheric retrieval analysis on the VLT FORS2 transmission spectrum of WASP-19b with POSEIDON. The specific retrieval configuration applied to WASP-19b is first outlined, in particular the inclusion of new opacity sources, before the resulting inferred atmospheric composition is revealed.

Retrieval configuration

The substructure apparent in WASP-19b's transmission spectrum (see Figure 4.7) motivates the inclusion of a variety of potential visible absorbers. In the retrievals presented so far (e.g. section 4.2.3), the only sources of visible opacity considered were Na, K, and clouds/hazes. These opacity sources cannot match the feature at ~ 0.72 µm, nor the substructure from 0.6–0.9 µm. I therefore extended POSEIDON to include metal oxides / hydrides with prominent visible absorption features, namely: TiO, VO, AlO, TiH, NaH, MgH, CrH, CaH, ScH and FeH (Sharp & Burrows, 2007) - the cross sections were provided by S. Gandhi. These 10 additional molecules offer additional confidence that any detections are due to a specific metal oxide / hydride.

With these molecules included, the retrievals applied to WASP-19b have a maximum of 25 parameters. The first 24 parameters are standard: 6 for the P-T profile, 13 for the mixing ratios (Na, K, H_2O, TiO, VO, AlO, TiH, NaH, MgH, CrH, CaH, ScH and FeH), 4 for clouds/hazes, and $P_{\rm ref}$. Given the high resolution of the spectrum, I additionally allow for the possibility of a systematic relative wavelength shift between the model and data - as often employed in similar resolution brown dwarf retrievals (Line et al., 2015; Burningham et al., 2017) - providing an additional 'nuisance' parameter, $\delta\lambda$. I note that there is no need to separately parametrise CH_4, NH_3, HCN, CO, or CO_2, as these molecules possess negligible absorption features for $\lambda < 1$ µm. Model transmission spectra are computed at 2000 wavelengths, uniformly spaced between 0.4–1.1 µm, for each sampled point in the 25-dimensional parameter space.

Table 4.1 Atmospheric retrieval priors: WASP-19b

Parameter	Prior	Range
P-T profile		
$\alpha_{1,2}$	Uniform	0.02 – 1 $\mathrm{K}^{-1/2}$
$P_{1,2}$	Log-uniform	10^{-6} – 10^{2} bar
P_3	Log-uniform	10^{-2} – 10^{2} bar
P_{ref}	Log-uniform	10^{-6} – 10^{2} bar
T_0	Uniform	1200 – 2600 K
Composition		
X_i	Log-uniform	10^{-12} - 10^{-1}
Clouds		
a	Log-uniform	10^{-4} – 10^{8}
γ	Uniform	-50 – 2
P_{cloud}	Log-uniform	10^{-6} – 10^{2} bar
$\bar{\phi}$	Uniform	0 – 1
Other		
$\delta\lambda$	Uniform	-100 – 100 Å

The priors used for each parameter are summarised in Table 4.1. Note in particular two key changes vs. the retrievals presented in earlier chapters: (i) the lower bound on X_i has been decreased to 10^{-12}, and (ii) the lower bound for γ has been decreased to -50. The former reflects that metal oxides and metal hydrides can still contribute significant opacity for mixing ratios $\sim 10^{-10}$ (the previous lower bound). The latter is due to initial test retrievals for this planet converging to values of γ at the lower edge of the prior (when set to -20). Additionally, wavelength shifts of up to ± 100 Å are considered. All other priors are justified as before (see section 2.3.3).

Results: the atmospheric composition of WASP-19b

I initially conducted a round of retrievals to assess which model components are necessary to explain WASP-19b's transmission spectrum. A 25-dimensional 'reference' retrieval, covering the entire parameter space previously described, was first conducted. The resulting posterior displayed only upper bounds for all chemical species besides Na, K, H_2O, TiO, and MgH, with the cloud fraction tending towards 1 (i.e. 100%

cloud/haze coverage). With respect to this reference model, I then ran two nested retrievals to assess the preferred cloud model: (i) an identical retrieval without a cloud deck, and (ii) a retrieval without haze. Removing the haze incurred a substantial penalty to the Bayesian evidence ($\ln \mathcal{Z} = 987.2 \rightarrow 970.2$, or $\mathcal{B}_{0i} = 2.0 \times 10^7$), establishing a 6.2σ detection of haze. Removing clouds resulted in a largely unchanged evidence, indicating the data does not support the presence, or absence, of a cloud deck. Finally, I conducted a retrieval to assess whether the wavelength shift parameter, $\delta\lambda$, is necessary. Without the wavelength shift, the Bayesian evidence suffers a large penalty ($\ln \mathcal{Z} = 987.2 \rightarrow 970.6$) equivalent to a 6.1σ detection. Indeed, the retrieved wavelength shift is tightly constrained at a non-zero value: $\delta\lambda = -72.8^{+10.0}_{-11.3}$ Å (for comparison, the data bin width is 100 Å). The necessity of this parameter suggests a systematic wavelength shift, unaccounted for during data reduction, remains present in the data.

A second round of retrievals was conducted to assess the importance of the remaining atmospheric components. Given the conclusions from the first round of retrievals, the new reference model contained Na, K, H_2O, TiO, MgH, and uniform clouds/hazes (i.e. $\bar{\phi}$ fixed to 1). This reference model contained 16 parameters. A nested model comparison was then conducted by removing each chemical species from the reference model and computing the resulting Bayesian evidence change. When K and MgH were removed, little change, or a very slight increase, in $\ln \mathcal{Z}$ was observed – indicating that, at best, their presence is only marginally suggested by the data. However, changes in $\ln \mathcal{Z} > 4$ ($\mathcal{B}_{0i} > 55$) were observed when Na, H_2O, or TiO were removed, indicating support for their presence at significances $> 3.3\sigma$. The lack of support for MgH (or any other metal hydride in the first round of retrievals) demonstrates that the low-wavelength ($\lesssim 0.5\,\mu$m) observations, despite suggestions of substructure, are optimally explained by a haze. Therefore, the only chemical species required to explain the FORS2 transmission spectrum of WASP-19b are Na, H_2O, and TiO.

Finally, a minimal reference model containing just Na, H_2O, TiO, haze, and a cloud deck was considered. The cloud deck parameter ensures that derived compositional constraints account for cloud-chemistry degeneracies, which can manifest for uniform cloud models. This final reference model, specified by 14 parameters, results in a Bayesian evidence of $\ln \mathcal{Z} = 990.71$. This notable increase over the initial reference model (cf. 987.2) reaffirms that the additional complexity incurred by other metal oxides/hydrides is not justified in light of the present observations. I then conducted a final series of four nested retrievals to compute detection significances for Na, H_2O, TiO, and the haze. The resulting Bayesian model comparison is shown in Table 4.2.

Table 4.2 Model comparison: atmospheric components of WASP-19b

Model	Evidence $\ln(\mathcal{Z}_i)$	Best-fit $\chi^2_{r,\,\min}$	Bayes Factor $\mathcal{B}_{0i}$	Significance of Ref.
Reference	990.71	1.63	Ref.	Ref.
No H_2O	961.83	2.05	3.5×10^{12}	7.9σ
No TiO	963.17	2.06	9.1×10^{11}	7.7σ
No Haze	965.33	1.95	1.0×10^{11}	7.4σ
No Na	986.36	1.65	77	3.4σ

Notes: The 'reference' model includes opacity due to H_2, He, Na, H_2O, and TiO, along with a parametrised cloud/haze prescription. $\chi^2_{r,\,\min}$ is the minimum reduced chi-square ($\chi^2/(N_{\rm data} - N_{\rm params})$). The significance indicates the degree of preference for the reference model, highlighted in bold, over each alternative model.

I report strong detections of H_2O (7.9σ), TiO (7.7σ), and a uniform haze (7.4σ) enveloping the terminator. The H_2O detection confirms a previous HST WFC3 detection (Huitson et al., 2013). I also find moderate evidence of Na (3.4σ). To demonstrate the contribution of each of these atmospheric components to WASP-19b's transmission spectrum, the overall best-fitting spectrum[4] from the final reference model is shown in Figure 4.8. Considering each spectral region in turn, we see that the 600B (blue) data are consistent with a strong scattering slope and the Na resonance doublet. The oscillatory features in the 600RI (green) data are due to TiO absorption features. Finally, the 600z data are explained by a combination of TiO and H_2O absorption for $\lambda \lesssim 0.92\,\mu$m, while the large feature around 1 µm is due to H_2O.

Constraints on the volume mixing ratios and haze properties at WASP-19b's terminator from the final reference retrieval are presented in Figure 4.9. The derived chemical abundances (to 1σ) are as follows: H_2O (18-1,300 ppm), TiO (0.015-1.1 ppb), and Na (0.028-140 ppm). The H_2O and Na abundances span a wide range, broadly consistent with expectations from a solar composition atmosphere in chemical equilibrium (Madhusudhan, 2012). The TiO abundance, however, would be expected to be ~ 100 ppb ($\log X_{\rm TiO} \sim -7$) at the derived 1 mbar temperature ($2,350^{+168}_{-314}$ K) under chemical equilibrium (Woitke et al., 2018). The derived TiO abundance is instead 3 orders of magnitude lower, with a sufficiently precise constraint to establish it as

[4]For reference, the best-fitting spectrum has H_2O, TiO, and Na abundances of 180 ppm, 0.12 ppb, and 6.4 ppm, respectively; a haze with a = 470,000 and γ = -30; a wavelength shift of $\delta\lambda = -73.6$ Å; and $P_{\rm ref} = 0.93$ bar. Confidence ranges for each parameter are given in the text.

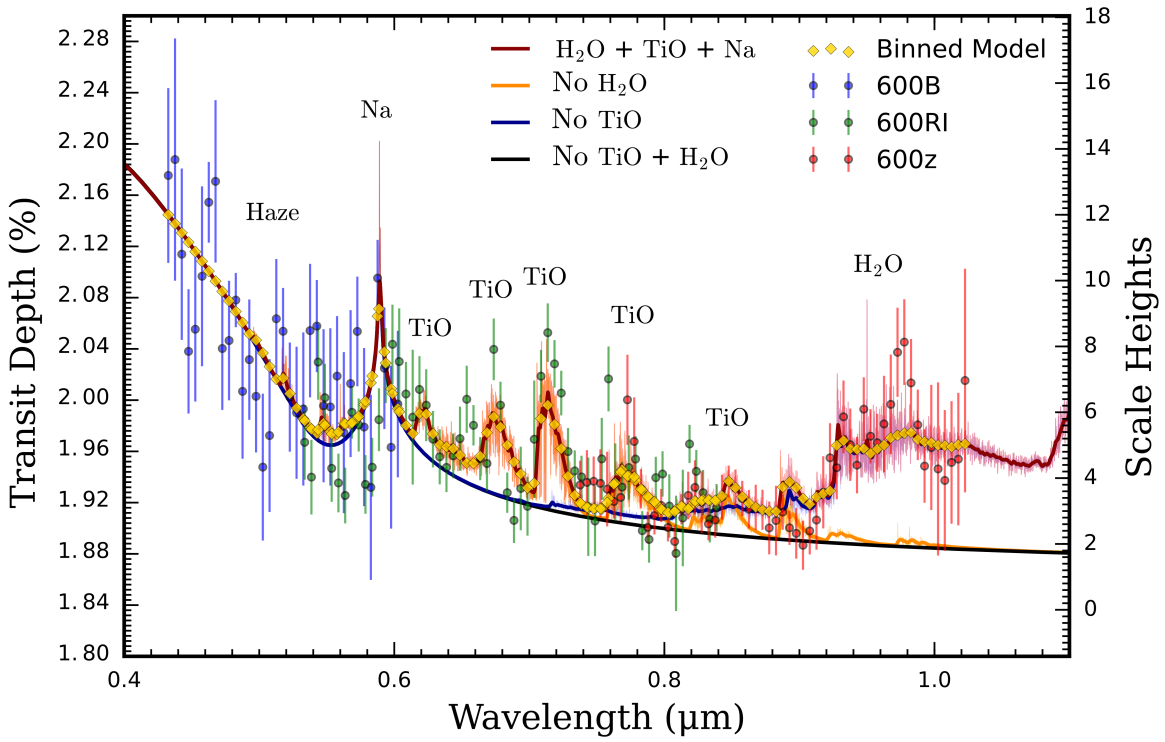


Fig. 4.8 Retrieved transmission spectrum of WASP-19b.
The blue, green, and red data points correspond to VLT FORS2 observations using the 600B, 600RI and 600z grisms. A corrective wavelength shift of 73.6 Å between the model and the data has been applied (see the text). The best-fitting spectrum derived by POSEIDON (red) includes opacity from H_2O, TiO, Na, and a global haze. Three additional spectra demonstrate disabling various opacity sources from the best-fitting model. In turn, these spectra show the effect of removing H_2O (orange), TiO (blue), and H_2O + TiO (black). Each spectrum is plotted both at $R \approx 3000$ and as a smoothed representation (dark curves with corresponding colours). Binned model points for the best-fitting spectrum (gold diamonds) are over-plotted.

sub-solar at $> 5\sigma$ confidence. Finally, the detected haze is $\sim$100,000$\times$ stronger than Rayleigh scattering from H_2, follows a power-law with an exponent of $\gamma = -26^{+4}_{-5}$, and is consistent with 100% coverage across the terminator. This slope, in particular, far exceeds what would be expected from Mie scattering of many known condensates (e.g. Pinhas & Madhusudhan, 2017). A potential alternative explanation for such steep slopes could be uncorrected stellar activity – as suggested recently by Espinoza et al. (2019). However, as described in Sedaghati et al. (2017), corrections for both occulted and unocculted star spots were applied in deriving the WASP-19b spectrum analysed here. Nevertheless, as the scattering slope here is uncorrelated with the retrieved abundances, its physical origin does not alter compositional inferences.

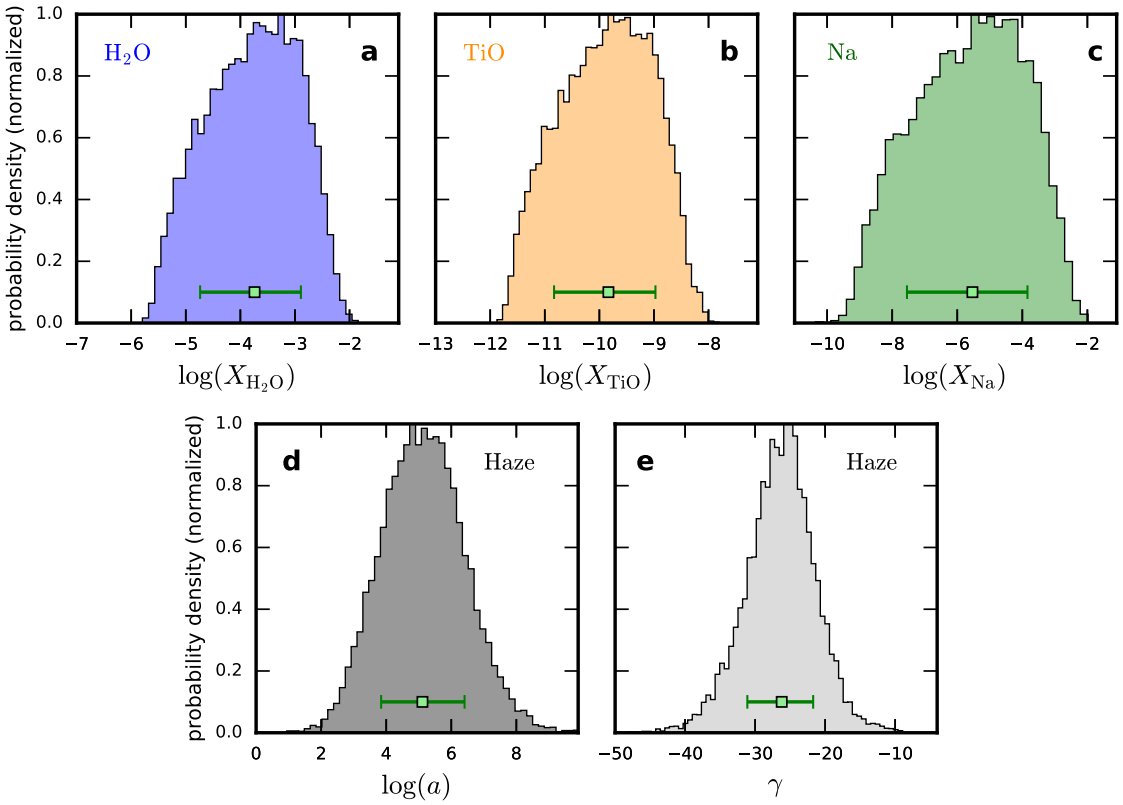


Fig. 4.9 Retrieved abundances and haze properties of WASP-19b.
The histograms give marginalised posterior distributions from the third stage reference retrieval (see the text). Top row: constraints on the volume mixing ratios of H_2O, TiO, and Na at WASP-19b's terminator. Bottom row: posterior distributions for the haze Rayleigh enhancement factor, a, and scattering slope, γ (see section 2.2.4).

I briefly note an important point about the precision of the obtained chemical abundances. It has recently been claimed that accurate H_2O abundances cannot be extracted from HST WFC3 transmission spectra, unless R_p and $P_{\rm ref}$ are assumed (Heng & Kitzmann, 2017). This degeneracy may be lifted by observations of cloud-free atmospheres at visible wavelengths, as an observed H_2-Rayleigh slope constrains $P_{\rm ref}$ (Heng & Kitzmann, 2017). Here, I have demonstrated that constraints of < 1 dex can be placed on individual molecular abundances when retrieving visible wavelength transmission spectra of hazy atmospheres, despite leaving $P_{\rm ref}$ as a free parameter. The ability to accomplish this stems from the data at $\lambda \lesssim 0.5\,\mu$m, which sample the continuum slope due to the haze. This continuum normalises the spectrum, constraining $P_{\rm ref}$, and thereby allows the degeneracy with molecular abundances to be partially collapsed. I stress that observations at short wavelengths, away from spectral features, are essential to constraining abundances of individual species.

4.4 Implications

I will now briefly discuss the implications of the new molecular signatures in hot Jupiter atmospheres presented in this chapter. I first consider hot Jupiter nitrogen chemistry in section 4.4.1, before turning to the detection of titanium oxide in section 4.4.2.

4.4.1 Detecting nitrogen chemistry in hot Jupiters

Detections of nitrogen chemistry in hot Jupiter atmospheres will open a new window into disequilibrium chemistry and planetary formation mechanisms. In particular, high NH_3 abundances are indicative of vertical mixing, with abundance measurements providing constraints on the eddy diffusion coefficient (Moses et al., 2011). High HCN abundances, on the other hand, can indicate vertical mixing, an enhanced C/O ratio, photochemistry, or some combination of these factors. A potential discriminator to identify photochemistry would be a high HCN abundance coupled with an absence of CH_4 and NH_3 (Zahnle et al., 2009; Moses et al., 2011; Venot et al., 2012).

In section 4.2, I demonstrated that nitrogen-bearing molecules can be observed with HST WFC3 spectra. Signatures of NH_3 or HCN can exceed ~ 100 ppm in the WFC3 G141 band pass, especially when NH_3 or HCN abundances are enhanced by disequilibrium mechanisms to exceed $\sim 10^{-1} \times H_2O$. Clouds can dampen these features somewhat, but they remain prominent even for high-altitude uniform cloud decks at 1 mbar. It is therefore important to include NH_3 and HCN in atmospheric retrieval analyses of hot Jupiter spectra – as recently demonstrated by Kilpatrick et al. (2018). Furthermore, ground-based photometry/spectroscopy in the K-band at ~ 2.2 µm can contribute enormously to resolving NH_3-HCN degeneracies arising from WFC3 spectra alone. In short, current observations, whether from space or ground, are already capable of probing nitrogen chemistry in hot Jupiter atmospheres.

I have shown that evidence of disequilibrium nitrogen chemistry is already present in archival transmission spectra of hot Jupiters. The retrieved NH_3 and HCN abundances for WASP-31b and WASP-63b, respectively, are enhanced over equilibrium values by $\sim$ 3-4 orders of magnitude. Such high values suggest that chemical equilibrium is violated in hot Jupiter atmospheres, and should not be imposed *a priori* in atmospheric retrievals. As many existing disequilibrium chemistry models predict NH_3 / HCN abundances enhanced by $\sim 1-2$ orders of magnitude above equilibrium expectations, future theoretical studies will be required to explore the feasibility of scenarios capable of producing the NH_3 and HCN abundances inferred here.

In the future, transmission spectra with JWST will be ideally suited to observing a plethora of NH_3 and HCN features. Noise floors of ~ 10 ppm precision, potentially achievable with NIRISS / NIRSpec (Beichman et al., 2014), should allow robust detections and abundance constraints for nitrogen-bearing molecules. In particular, ambiguities between NH_3 and HCN may be resolved by observing strong NH_3 absorption at ~ 1.6 µm and ~ 2.2 µm (K-band) or strong HCN absorption at ~ 3.1 µm and 4.0 µm. Such observations will enable unique detections of NH_3 and HCN in many exoplanetary atmospheres. Indeed, I will directly demonstrate the ability to detect molecules, and constrain their abundances, with JWST observations in Chapter 7. Ultimately, the results presented here from current transmission spectra can be considered an early preview into the world of nitrogen chemistry.

4.4.2 The first detection of TiO in an exoplanet atmosphere

The detection of TiO in WASP-19b's atmosphere – the first detection of a metal oxide in an exoplanet atmosphere – brings a decade-long search to a close (see section 4.3.1). The presence of gas phase TiO is consistent with theoretical expectations for high-temperature ($\gtrsim 2000$ K) heavy-element chemistry in hot Jupiter atmospheres (Fortney et al., 2008). This detection demonstrates the importance of visible-wavelength molecular opacity in retrievals of exoplanet transmission spectra. In particular, this result demonstrates the utility of ground-based transmission spectroscopy for pursuing detailed molecular surveys of exoplanetary atmospheres at visible wavelengths.

The strong visible opacity contributed by TiO could have substantial effects on the temperature structure and circulation of WASP-19b's atmosphere. TiO, if present in large enough quantities, may cause a dayside thermal inversion (Hubeny et al., 2003; Fortney et al., 2008) which could be observable in high-precision thermal emission spectra obtainable with HST and JWST in the future. The increased opacity resulting from the presence of TiO could also lead to strong day-night temperature gradients in the atmosphere (Showman et al., 2009). Future observations of spectral features in emission (rather than absorption) in dayside spectra of WASP-19b, complementing the present detection of TiO at the terminator, would confirm a critical link between dayside inversions and the presence of TiO.

This work also represents the first abundance constraint for a heavy-element molecule in an exoplanet atmosphere. The robust finding of a sub-solar TiO abundance (to $> 5\sigma$ confidence) could explain why a previous analysis of a HST WFC3 spectrum of WASP-19b was unable to establish the existence of TiO (Huitson et al., 2013). Indeed,

the sub-solar TiO abundance found here could represent the first observational evidence for theories suggesting TiO depletion via condensation in cold traps (Spiegel et al., 2009; Parmentier et al., 2013). Further elucidating this possibility will require detections and abundance constraints for TiO across a greater population of hot Jupiters. With a larger sample size, future studies may then be able to conduct comparative planetology of heavy-element chemistry, perhaps searching for correlations between properties such as $T_{\rm eq}$ and $X_{\rm TiO}$ to provide an important observational test for TiO condensation theories and 3D GCM models of hot Jupiter atmospheres.

These results have paved the way towards detailed characterisation of high temperature, heavy-element, chemistry in exoplanetary atmospheres. Since this work was published in Sedaghati et al. (2017), a detection of TiO has been reported in a second exoplanet, WASP-33b, from high-resolution dayside Doppler spectroscopy (Nugroho et al., 2017). The metal oxide VO has also been detected in WASP-121b's transmission and emission spectrum (Evans et al., 2017, 2018). A collection of heavy atoms and ions, including Ti, Ti^+, and Fe, have also recently been detected in the transmission spectrum of the ultra-hot Jupiter KELT-9b from high-resolution terminator Doppler spectroscopy (Hoeijmakers et al., 2018a, 2019). With this flood of recent detections, we have entered the era of heavy-element chemistry in exoplanet atmospheres.

4.5 Summary

In this chapter, I have extended the atmospheric retrieval approach from Chapter 3 to consider the detectability of new molecules in hot Jupiter atmospheres. Through a series of atmospheric retrievals on transmission spectra of 9 hot Jupiters, I have identified the first evidence of nitrogen chemistry in hot Jupiter atmospheres. I have also established the first detection of a metal oxide in an exoplanetary atmosphere. In what follows, points 1-3 summarise the major results from section 4.2, whilst points 4-6 summarise the principal findings from section 4.3.

1. NH_3 and HCN can be detected in exoplanet atmospheres using current HST WFC3 or ground-based K-band transmission spectra. Absorption features due to these species can often exceed 100 ppm.

2. I report weak evidence of NH_3 in WASP-31b's atmosphere (2.2σ) and HCN in WASP-63b's atmosphere (2.3σ).

3. The retrieved NH_3 and HCN abundances are enhanced by $\sim 3-4$ orders of magnitude over equilibrium expectations, providing observational evidence for disequilibrium nitrogen chemistry in hot Jupiter atmospheres.

4. I report a conclusive detection of titanium oxide (TiO) in the atmosphere of the hot Jupiter WASP-19b (at 7.7σ confidence). The TiO abundance is sub-solar to $> 5\sigma$ confidence.

5. Additional components of WASP-19b's atmosphere include: H_2O (7.9σ), a strongly scattering haze (7.4σ), and Na (3.4σ).

6. I have demonstrated that ground-based visible transmission spectra of hot Jupiter atmospheres with VLT FORS2 can detect heavy-element molecules.

The two stories in this chapter demonstrate complementary paths in the unfolding narrative of exoplanetary science. On the one hand, I have verified a decade-old theoretical prediction by detecting TiO in a hot Jupiter atmosphere. On the other, I have shown that hot Jupiter spectra can show signatures of nitrogen chemistry, despite the common assumption of chemical equilibrium largely precluding the observation of these molecules. Taken together, these findings illustrate that though some theoretical predictions for exoplanet atmospheres will surely be vindicated, the unexpected should also be embraced as we seek to elucidate the nature of these distant worlds.

Chapter 5

An Opacity Database for Sub-Stellar Atmospheres

5.1 Chemical fingerprints in alien skies

One of the most important inputs to exoplanet atmosphere models is wavelength-dependent opacity. The opacity of an atmosphere, arising from quantum interactions between photons and matter, determines how radiation is absorbed or scattered. Crucially, the opacity contributed by each molecule or atom has a unique spectral fingerprint, enabling the identification of chemical species with known opacities from a spectrum (given sufficiently high-quality observations). As atmospheric retrievals take opacities as fixed inputs, assumed *a priori* known, the accuracy and completeness of opacity databases are key to the successful characterisation of exoplanetary atmospheres.

In previous chapters, we have already seen examples where the range of opacities considered during retrievals can change the interpretation of exoplanet spectra. In Chapters 3 and 4, I found that the inclusion of NH_3 and HCN opacity in retrieval analyses provides (previously unidentified) evidence for these species in three hot Jupiter spectra. The degeneracy between NH_3 and HCN absorption features (at low spectral resolution) also demonstrates the importance of considering a wide range of molecules in retrievals. For if only one species in a degenerate pair is considered, retrievals may overestimate detection significances (e.g. if only HCN opacity were considered, a retrieval would be unable to 'fold in' the possibility of NH_3 as an alternative explanation). Similarly, the detection of TiO in WASP-19b (section 4.3) required opacity due to a range of alternative metal oxides and hydrides to be considered, ensuring that the detected spectral features were attributable to TiO at high-significance.

Recently, the range of exoplanet atmospheres for which spectra can be observed has grown beyond hot Jupiters. Consequently, the opacities required to model and retrieve the full diversity of observable exoplanet atmospheres is poised to dramatically expand. Gaseous planets with dayside temperatures approaching those of cool stars, so-called 'ultra-hot Jupiters', are expected to possess many atomic and ionic species in their atmospheres (e.g. Kitzmann et al., 2018; Arcangeli et al., 2018) – as recently demonstrated by a detection of Fe, Fe^+, and Ti^+ in the atmosphere of KELT-9b (Hoeijmakers et al., 2018a). High-temperature ($T_{eq} \gtrsim 2000\,K$) rocky planets, or 'lava worlds', are thought to possess atmospheres with quite different compositions to hot Jupiters, featuring molecules such as SO_2, HF, HCl, and OH (Tennyson & Yurchenko, 2017). Besides these new classes of hot worlds, the push to characterise cooler, lower-mass, exoplanets closer to the terrestrial regime necessitates the consideration of additional opacity sources. The characterisation of exoplanets in and around the habitable regime requires opacity due to absorption and Rayleigh scattering from potential bulk atmosphere gases (e.g. N_2, O_2, CO_2), and even prospective biosignatures (e.g. O_3, CH_4, N_2O) (Kaltenegger, 2017).

Given these demands for new molecular, atomic, and ionic opacities, in this chapter I will introduce an extensive update to the opacity database included in POSEIDON. The main goals behind this update are:

1. The inclusion of opacities important for modelling terrestrial planets, super-Earths, warm Neptunes, and ultra-hot Jupiters.

2. Raising the spectral resolution of molecular cross sections to $R \sim 10^6$ (i.e. $10 - 100\times$ higher than those used in Chapters 3 and 4), such that the POSEIDON forward model can generate template spectra for high-resolution Doppler spectroscopy (e.g. with ESPRESSO).

3. Employ the latest theoretical and experimental line lists - ensuring the maximal completeness, accuracy, and temperature coverage currently possible.

4. Update the spectral line broadening of molecules and atoms to account for a H_2 and He background gas, where data is available, when modelling gaseous exoplanet spectra (the opacities in previous chapters assumed air broadening).

As we have yet to see how the cross sections underlying extinction coefficients (see section 2.2.3) are calculated, I will first detail the procedure used to create different types of absorption and scattering cross sections.

In what follows, I outline how one goes from a list of quantum transitions to a molecular cross section in section 5.2. A new technique to rapidly calculate molecular cross sections is introduced in section 5.3. The calculation of atomic cross sections and continuum opacities is described in sections 5.4 and 5.5, respectively. Finally, in section 5.6, the newly expanded POSEIDON opacity database is presented.

5.2 Molecular opacities

Absorption cross sections arise from the combination of a large number of quantum mechanical transitions. Electronic, rotational, and vibrational transitions all contribute to molecular cross sections. At the elevated temperatures of many exoplanet atmospheres ($\gtrsim 1000\,\mathrm{K}$), a variety of excited states can be populated. Accurate computation of cross sections therefore requires the inclusion of typically $\sim 10^7 - 10^{10}$ transitions for many molecules of interest (Tennyson & Yurchenko, 2018).

Consider a single transition between two molecular energy levels. The contribution of this transition to the total cross section at wavenumber $\tilde{\nu}$ ($\equiv 1/\lambda$) is given by[1]

$$\delta\sigma_j(\tilde{\nu},\, P,\, T) = S_j(T)\, f(\tilde{\nu},\, \tilde{\nu}_{0,\,j},\, P,\, T) \tag{5.1}$$

where j denotes the transition, $S_j(T)$ is the 'line intensity' (encoding the strength of the transition, in units of cm species^{-1}), f is the 'line profile' (describing the shape of the spectral line, in units of cm), and $\tilde{\nu}_{0,\,j}$ is the central wavenumber of the transition (in cm^{-1}). I will now describe how these two terms are related to properties of a given transition, before summarising the databases where one can find lists of these properties for spectral lines of many molecules.

5.2.1 Line intensities

Assuming local thermodynamic equilibrium (LTE), the line intensity can be written as (Rothman et al., 1998; Yurchenko et al., 2018d)

$$S_j(T) = \frac{A_j\, g_j^{\mathrm{up}}}{8\pi c\, \tilde{\nu}_{0,\,j}^2}\, \frac{\exp\left(-c_2\, \tilde{E}_j^{\mathrm{low}}/T\right)}{Q(T)}\, \left(1 - \exp\left(-\frac{c_2\, \tilde{\nu}_{0,\,j}}{T}\right)\right) \tag{5.2}$$

[1]By convention, the spectroscopic inputs in all line databases are given in cgs units. For consistency with the literature, the cross section formulae presented here are also written in cgs units.

where A_j (in s^{-1} species^{-1}) is the Einstein A coefficient of the transition, g_j^{up} is the upper level degeneracy, c is the speed of light (in $\mathrm{cm\,s}^{-1}$), $c_2 = hc/k_B$ is the second radiation constant (in cm K), $\tilde{E}_j^{\mathrm{low}} = E_j^{\mathrm{low}}/hc$ is the energy term value of the lower level (in cm^{-1}), and $Q(T)$ is the partition function. The three terms in this expression can be understood as arising from: (1) the intrinsic strength of the transition per species; (2) the (Boltzmann) fraction of the species occupying the lower energy level; and (3) a correction for stimulated emission. The partition function is defined as

$$Q(T) = \sum_i g_i \exp\left(-c_2\,\tilde{E}_i/T\right) \tag{5.3}$$

where the sum is carried out over all significantly occupied energy levels i. An alternative method to compute $S_j(T)$, should the line intensity and partition function be known at a reference temperature, T_{ref}, is obtained via the scaling

$$S_j(T) = S_j(T_{\mathrm{ref}})\,\frac{Q(T_{\mathrm{ref}})}{Q(T)}\,\frac{\exp\left(-c_2\tilde{E}_j^{\mathrm{low}}/T\right)}{\exp\left(-c_2\tilde{E}_j^{\mathrm{low}}/T_{\mathrm{ref}}\right)}\,\frac{1-\exp\left(-c_2\,\tilde{\nu}_{0,\,j}/T\right)}{1-\exp\left(-c_2\,\tilde{\nu}_{0,\,j}/T_{\mathrm{ref}}\right)} \tag{5.4}$$

This latter formalism is typically used with $T_{\mathrm{ref}} = 296\,\mathrm{K}$ (i.e. scaling pre-tabulated line intensities and partition functions from standard Earth temperature).

5.2.2 Broadening of spectral lines

The shape of a molecular transition is given by the line profile, $f(\tilde{\nu},\,\tilde{\nu}_{0,\,j},\,P,\,T)$. Spectral lines are not monochromatic (i.e. delta functions) and always possess a width in the frequencies (and hence wavelengths) where photons can excite the transition. A variety of physical effects influence the width of spectral lines. On a fundamental level, the minimum width a spectral line can possess comes from the Heisenberg uncertainty principle. The finite lifetime of any excited state leads to a finite energy uncertainty, such that a line must have a (full) width $\Delta\tilde{\nu}_j > A_j/2\pi c$. However, as other (external) processes can induce much greater broadening, this natural line width is often neglected for molecular transitions. The two sources of line broadening relevant to exoplanet atmospheres are thermal (Doppler) broadening and pressure (collisional) broadening.

Thermal broadening

Thermal broadening results from the spread of velocities in a gas. Incoming photons are Doppler shifted in the reference frame of a molecule, such that different frequency photons can be absorbed in a transition when thermal motion is present. For a gas following a Maxwell-Boltzmann velocity distribution, the line profile is a Gaussian

$$f_{\mathrm{D}}(\tilde{\nu} - \tilde{\nu}_{0,j}) = \sqrt{\frac{\ln 2}{\pi}} \frac{1}{\alpha_{\mathrm{D}}} \exp\left(-\frac{\ln 2\,(\tilde{\nu} - \tilde{\nu}_{0,j})^2}{\alpha_{\mathrm{D}}^2}\right) \tag{5.5}$$

where

$$\alpha_{\mathrm{D}} = \sqrt{\frac{2 \ln 2\, k_{\mathrm{B}} T}{m}}\, \frac{\tilde{\nu}_{0,j}}{c} \tag{5.6}$$

is the half-width at half-maximum (HWHM) of the Doppler profile (in cm^{-1}) and m is the mass of the molecule. We see that thermal broadening increases for higher temperatures and for transitions at larger wavenumbers (lower wavelengths).

Pressure broadening

Pressure broadening results from collisions between a molecule and a background gas. Conceptually, it can be understood as arising from random collisions shortening the average lifetime of states. Via the Heisenberg uncertainty principle, the uncertainty in the energy (hence wavenumber) of all transitions thereby increases in the presence of collisions. Pressure broadening results in a Lorentzian line profile

$$f_{\mathrm{L}}(\tilde{\nu} - \tilde{\nu}_{0,j}) = \frac{1}{\pi} \frac{\gamma_{\mathrm{L}}}{(\tilde{\nu} - \tilde{\nu}_{0,j})^2 + \gamma_{\mathrm{L}}^2} \tag{5.7}$$

where

$$\gamma_{\mathrm{L}} = \sum_p \gamma_{\mathrm{L,p}}^0\, X_p \left(\frac{T_0}{T}\right)^{n_{\mathrm{L,p}}} \frac{P}{P_0}. \tag{5.8}$$

is the HWHM of the Lorentzian profile (in cm^{-1}). This general expression accounts for multiple perturbing (collisional) species p, with $\gamma_{\mathrm{L,p}}^0$ the HWHM at a reference pressure P_0 (typically 1 bar) and temperature T_0 (typically 296 K) due to p, X_p the perturber volume mixing ratio, and $n_{\mathrm{L,p}}$ the temperature exponent for broadener p. The two pressure broadening parameters, γ_{L}^0 and n_{L}, are functions of the transition quantum numbers (especially the lower level total angular momentum, J). For the present work, I consider broadening due to either H_2 and He (in solar proportions, X_{H_2} = 85%, X_{He} = 15%) or air - depending on the available broadening data.

Combined thermal and pressure broadening

In general, each line will exhibit some degree of thermal and pressure broadening. The combined influence of these two broadening mechanisms can be expressed by the convolution $f_{\rm D} * f_{\rm L}$, which defines the Voigt line profile

$$f_{\rm V}(\tilde{\nu} - \tilde{\nu}_{0,\,j}) \equiv \int_{-\infty}^{+\infty} f_{\rm D}(\tilde{\nu}' - \tilde{\nu}_{0,\,j})\, f_{\rm L}(\tilde{\nu} - \tilde{\nu}')\, d\tilde{\nu}' \tag{5.9}$$

Substituting Equations 5.5 and 5.7 gives

$$f_{\rm V}(\tilde{\nu} - \tilde{\nu}_{0,\,j}) = \sqrt{\frac{\ln 2}{\pi^3}}\, \frac{\gamma_{\rm L}}{\alpha_{\rm D}} \int_{-\infty}^{+\infty} \frac{\exp\left(-\ln 2\,(\tilde{\nu}' - \tilde{\nu}_{0,\,j})^2/\alpha_{\rm D}^2\right)}{(\tilde{\nu} - \tilde{\nu}')^2 + \gamma_{\rm L}^2}\, d\tilde{\nu}' \tag{5.10}$$

A change of variables according to

$$x = \sqrt{\ln 2}\, \left(\frac{\tilde{\nu} - \tilde{\nu}_{0,\,j}}{\alpha_{\rm D}}\right) \tag{5.11}$$

$$y = \sqrt{\ln 2}\, \left(\frac{\gamma_{\rm L}}{\alpha_{\rm D}}\right) \tag{5.12}$$

$$t = \sqrt{\ln 2}\, \left(\frac{\tilde{\nu}' - \tilde{\nu}_{0,\,j}}{\alpha_{\rm D}}\right) \tag{5.13}$$

allows Equation 5.10 to be written in the simplified form

$$f_{\rm V}(\tilde{\nu} - \tilde{\nu}_{0,\,j}) = \frac{1}{\alpha_{\rm D}} \sqrt{\frac{\ln 2}{\pi}}\, \left(\frac{y}{\pi} \int_{-\infty}^{+\infty} \frac{e^{-t^2}}{(x - t)^2 + y^2}\, dt\right) \tag{5.14}$$

where

$$K(x, y) \equiv \frac{y}{\pi} \int_{-\infty}^{+\infty} \frac{e^{-t^2}}{(x - t)^2 + y^2}\, dt \tag{5.15}$$

defines a dimensionless function. Though this integral cannot be evaluated analytically, its numerical evaluation is aided by recognising that the complex Faddeeva function

$$W(z) \equiv \frac{i}{\pi} \int_{-\infty}^{+\infty} \frac{e^{-t^2}}{z - t}\, dt \tag{5.16}$$

has real and imaginary components given by

$$W(z) = K(x, y) + i\, L(x, y) \tag{5.17}$$

where

$$L(x, y) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{(x - t)\, e^{-t^2}}{(x - t)^2 + y^2}\, dt \tag{5.18}$$

The utility here is that efficient numerical packages are available to compute the Faddeeva function. The Voigt profile for a given transition is then

$$f_{\mathrm{V}}(\tilde{\nu} - \tilde{\nu}_{0,\,j}) = \frac{1}{\alpha_{\mathrm{D}}} \sqrt{\frac{\ln 2}{\pi}}\, Re[W(x + i\, y)] \tag{5.19}$$

with x and y specified in terms of $\tilde{\nu}$, $\tilde{\nu}_{0,\,j}$, α_{D}, and γ_{L} via Equations 5.11 and 5.12. The HWHM, in turn, are specified as functions of pressure and temperature via Equations 5.6 and 5.8.

5.2.3 Molecular line list databases

The parameters needed to calculate $S_j(T)$ and $f_{\mathrm{V}}(\tilde{\nu},\, \tilde{\nu}_{0,\,j},\, P,\, T)$ for all transitions j comprise a *line list*. Line list databases commonly used for exoplanet atmosphere models include HITRAN (Gordon et al., 2017), HITEMP (Rothman et al., 2010), and ExoMol (Tennyson et al., 2016). HITRAN specialises in low-temperature ($\sim 300\,\mathrm{K}$) line lists valid for Earth atmosphere and solar system applications. HITEMP extends the temperature coverage of HITRAN (for a limited number of molecules) by including weaker transitions which become more important for $T > 1000\,\mathrm{K}$. ExoMol provide an ever-expanding collection of molecular line lists designed for high-temperature ($\lesssim 4000\,\mathrm{K}$) applications especially well-suited for exoplanet atmospheres. HITRAN and HITEMP provide the parameters to evaluate $S_j(T)$ via Equation 5.4, whilst ExoMol provide the parameters for Equation 5.2. All three databases provide a collection of pressure broadening parameters (mainly for air, H_2, and He).

Using a line list, the profile and intensity for each transition can be determined. The total cross section at wavenumber $\tilde{\nu}$ can then be computed by summing over all transitions

$$\sigma(\tilde{\nu},\, P,\, T) = \sum_j S_j(T)\, f_{\mathrm{V}}(\tilde{\nu},\, \tilde{\nu}_{0,\,j},\, P,\, T) \tag{5.20}$$

satisfying $|\tilde{\nu} - \tilde{\nu}_{0,\,j}| \leqslant \Delta\tilde{\nu}_{\mathrm{cut}}$, where $\Delta\tilde{\nu}_{\mathrm{cut}}$ is a line wing cutoff accounting for sub-Voigt far line wings (see e.g. Grimm & Heng, 2015, for a discussion). The f_{V} here is hence understood to be re-normalised to 1 due to the truncation of the line wings. I will now turn to the practical matter of efficiently computing molecular cross sections.

5.3 Rapid computation of molecular cross sections

The calculation of molecular cross sections (or opacities) represents a substantial computational challenge. Hundreds of millions of Voigt profiles must be evaluated, with each profile computed at thousands of points, across a grid of pressures and temperatures. For example, computing a cross section from the BT2 H_2O line list (Barber et al., 2006), containing 500 million lines and occupying $\approx$ 10 GB of file space, is a moderate challenge, whilst the latest ExoMol CH_4 line list (Yurchenko et al., 2017), containing 34 billion lines and occupying nearly 1 TB, is a colossal undertaking.

Previous works have employed many strategies to render cross section calculations feasible in a reasonable time frame. Hedges & Madhusudhan (2016) apply an intensity cutoff, such that all lines with $S_j(T) < 10^{-30}\,\mathrm{cm\ species^{-1}}$ are neglected (with an even higher cutoff applied for large line lists, e.g. CH_4). Grimm & Heng (2015) follow a brute force technique using graphics processing units (GPUs), enabling computation of 10^5 Voigt profiles per second. Min (2017) developed a random sampling technique to approximate Voigt profiles, reaching speeds of 3.5×10^5 profiles per second. Most recently, Yurchenko et al. (2018d), recognising evaluating Voigt profiles is the computation bottleneck, introduced the 'vectorised Voigt' method to approximate Voigt profiles. The vectorised Voigt method uses the observation that the line wings of Voigt profiles are relatively insensitive to the value of α_D. Given this, a set of template Voigt profiles, $f_\mathrm{V}^\mathrm{ref}(\Delta\tilde{\nu}, \alpha_\mathrm{D} = 1\,\mathrm{cm^{-1}}, \gamma_\mathrm{L})$, where $\Delta\tilde{\nu}$ is the wavenumber distance from the line centre and a range of γ_L is spanned, can be pre-computed. The wings of all line profiles beyond $4\,\mathrm{cm^{-1}}$ are then set equal to the template $f_\mathrm{V}^\mathrm{ref}$ for the relevant γ_L. Computationally expensive Voigt profile evaluations (e.g. via Equation 5.19) then need only be evaluated in the line cores ($\Delta\tilde{\nu} < 4\,\mathrm{cm^{-1}}$) for each specific transition.

Here, I present a generalisation of the vectorised Voigt method. The method I propose removes the need to compute Voigt profiles for individual transitions in the line cores, substantially decreasing run times for opacity calculations. The *Generalised Vectorised Voigt* (GVV) method utilises pre-computed Voigt profiles and their derivatives to rapidly calculate molecular cross sections from billions of lines at high-resolution. A typical application reaches speeds of 2×10^5 Voigt profiles per second on a single CPU core, with much greater speeds when parallelised. The code implementing the GVV method is written in Python, whilst other opacity codes are generally written in C++ or FORTRAN. This new method has been used to create an extensive database of molecular opacities using state-of-the-art line lists and pressure broadening data.

5.3.1 The Generalised Vectorised Voigt (GVV) algorithm

The essential idea behind the GVV method is that, at fixed P and T, $\alpha_\mathrm{D} = \alpha_\mathrm{D}(\tilde{\nu}_{0,j})$ and $\gamma_\mathrm{L} \approx \gamma_\mathrm{L}(J^\mathrm{low})$, where J^low is the total rotational quantum number of the lower energy level. The dependence on J^low arises from the dependence of the pressure broadening parameters, $\gamma^0_{\mathrm{L,p}}$ and $n_{\mathrm{L,p}}$ (see Equation 5.8), on the quantum numbers of the states involved in a transition[2]. As J^low can only take discrete values, γ_L is known *a priori* for all transitions. Recognising that, for typical application to exoplanet atmospheres, $\tilde{\nu}_{0,j}$ ranges from $200 - 25000\,\mathrm{cm}^{-1}$ (0.4–50 µm), or $\approx$ two orders of magnitude, Voigt profiles can be pre-computed for a grid of $\alpha^\mathrm{ref}_\mathrm{D}$ spaced logarithmically over this wavenumber range. A set of template Voigt profiles, $f^\mathrm{ref}_\mathrm{V} = f_\mathrm{V}(\Delta\tilde{\nu}^\mathrm{ref}, \alpha^\mathrm{ref}_\mathrm{D}, \gamma_\mathrm{L})$, are thereby constructed. For a given transition, to zeroth-order, the line profile can be approximated by the template with γ_L corresponding to the J^low of the transition and $\alpha^\mathrm{ref}_{\mathrm{D,i}} \approx \alpha_\mathrm{D}(\tilde{\nu}_{0,j})$ (i.e. the closest pre-computed value, i, to the actual α_D at the transition wavenumber $\tilde{\nu}_{0,j}$). However, here I use an improved approximation to accurately capture the sensitivity of the line core shape to the precise value of α_D.

To first order, an arbitrary Voigt profile can be approximated by a series expansion about a nearby pre-computed template profile by

$$f_\mathrm{V}(\Delta\tilde{\nu},\, \alpha_\mathrm{D},\, \gamma_\mathrm{L}) \approx f^\mathrm{ref}_\mathrm{V} + \left(\frac{\partial f_\mathrm{V}}{\partial \Delta\tilde{\nu}}\right)^\mathrm{ref}_{\alpha_\mathrm{D},\,\gamma_\mathrm{L}} \delta\Delta\tilde{\nu} + \left(\frac{\partial f_\mathrm{V}}{\partial \alpha_\mathrm{D}}\right)^\mathrm{ref}_{\Delta\tilde{\nu},\,\gamma_\mathrm{L}} \delta\alpha_\mathrm{D} + \left(\frac{\partial f_\mathrm{V}}{\partial \gamma_\mathrm{L}}\right)^\mathrm{ref}_{\Delta\tilde{\nu},\,\alpha_\mathrm{D}} \delta\gamma_\mathrm{L} \quad (5.21)$$

where $\Delta\tilde{\nu} = \tilde{\nu} - \tilde{\nu}_{0,j}$ is the distance from the line centre, $f^\mathrm{ref}_\mathrm{V} = f_\mathrm{V}(\Delta\tilde{\nu}^\mathrm{ref},\, \alpha^\mathrm{ref}_\mathrm{D},\, \gamma^\mathrm{ref}_\mathrm{L})$ is a template Voigt profile, $\delta\Delta\tilde{\nu} = \Delta\tilde{\nu} - \Delta\tilde{\nu}^\mathrm{ref}$, $\delta\alpha_\mathrm{D} = \alpha_\mathrm{D} - \alpha^\mathrm{ref}_\mathrm{D}$, and $\delta\gamma_\mathrm{L} = \gamma_\mathrm{L} - \gamma^\mathrm{ref}_\mathrm{L}$. All partial derivatives are evaluated at the reference wavenumber separation, Doppler HWHM, and Lorentzian HWHM. In the GVV method proposed here, this expression can be further simplified. First, as the actual Lorentzian widths correspond precisely to the reference widths, we have $\gamma_\mathrm{L} = \gamma^\mathrm{ref}_\mathrm{L}$ and hence the final term may be dropped. Secondly, in the case where the computational wavenumber grid (on which the cross sections are computed) matches the reference wavenumber grid of the template profiles, $\Delta\tilde{\nu} = \Delta\tilde{\nu}^\mathrm{ref}$ and the first partial derivative may be dropped. However, this term must be kept if the computational grid spacing changes with wavenumber (as is the case in the present algorithm, for reasons which will shortly be elucidated). The most important term to determine is thus the derivative with respect to α_D.

[2]Pressure broadening can also depend on other quantum numbers, such as J^up and vibrational quantum numbers. However, currently available pressure broadening data often only preserves the strong J^low dependence (Tennyson et al., 2016; Barton et al., 2017).

Differentiation of Equation 5.14 w.r.t. α_D gives

$$\left(\frac{\partial f_\mathrm{V}}{\partial \alpha_\mathrm{D}}\right)_{\Delta\tilde{\nu},\,\gamma_\mathrm{L}} = -\frac{1}{\alpha_\mathrm{D}^2}\sqrt{\frac{\ln 2}{\pi}}\,K + \frac{1}{\alpha_\mathrm{D}}\sqrt{\frac{\ln 2}{\pi}}\left(\frac{\partial K}{\partial \alpha_\mathrm{D}}\right)_{\Delta\tilde{\nu},\,\gamma_\mathrm{L}} \tag{5.22}$$

Application of the chain rule allows the derivative in the second term to be written as

$$\left(\frac{\partial K}{\partial \alpha_\mathrm{D}}\right)_{\Delta\tilde{\nu},\,\gamma_\mathrm{L}} = \left(\frac{\partial K}{\partial x}\right)_y \left(\frac{\partial x}{\partial \alpha_\mathrm{D}}\right)_{\Delta\tilde{\nu},\,\gamma_\mathrm{L}} + \left(\frac{\partial K}{\partial y}\right)_x \left(\frac{\partial y}{\partial \alpha_\mathrm{D}}\right)_{\Delta\tilde{\nu},\,\gamma_\mathrm{L}} \tag{5.23}$$

Recalling the definitions of x and y (Equations 5.11 and 5.12), this expression can be written solely in terms of dimensionless derivatives of K via

$$\left(\frac{\partial K}{\partial \alpha_\mathrm{D}}\right)_{\Delta\tilde{\nu},\,\gamma_\mathrm{L}} = -\frac{x}{\alpha_\mathrm{D}}\left(\frac{\partial K}{\partial x}\right)_y - \frac{y}{\alpha_\mathrm{D}}\left(\frac{\partial K}{\partial y}\right)_x \tag{5.24}$$

These derivatives of K can be evaluated by recalling the complex Faddeeva function, $W(z) = K(x,y) + i\,L(x,y)$. From the definitions of K (Equation 5.15) and L (Equation 5.18), it can be seen that the Cauchy–Riemann equations

$$\begin{aligned}\left(\frac{\partial K}{\partial x}\right)_y &= \left(\frac{\partial L}{\partial y}\right)_x \\ \left(\frac{\partial K}{\partial y}\right)_x &= -\left(\frac{\partial L}{\partial x}\right)_y\end{aligned} \tag{5.25}$$

are satisfied, ensuring that $W(z)$ is complex differentiable. From the definition of $W(z)$ (Equation 5.16), it follows that

$$\frac{dW}{dz} = -\frac{i}{\pi}\int_{-\infty}^{+\infty}\frac{e^{-t^2}}{(z-t)^2}\,dt \tag{5.26}$$

Integrating this equation by parts then yields the identity

$$\frac{dW}{dz} = \frac{2\,i}{\sqrt{\pi}} - 2\,z\,W(z) \tag{5.27}$$

Moreover, differentiating Equation 5.16 w.r.t. x (at constant y) results in the same integral as Equation 5.26, such that we can also write

$$\frac{dW}{dz} = \left(\frac{\partial W}{\partial x}\right)_y \tag{5.28}$$

where

$$\left(\frac{\partial W}{\partial x}\right)_y = \left(\frac{\partial K}{\partial x}\right)_y + i\left(\frac{\partial L}{\partial x}\right)_y \tag{5.29}$$

Equation 5.28 allows Equation 5.29 to be equated to 5.27, yielding

$$\left(\frac{\partial K}{\partial x}\right)_y + i\left(\frac{\partial L}{\partial x}\right)_y = \frac{2\,i}{\sqrt{\pi}} - 2\,z\,W(z) \tag{5.30}$$

Substitution of $z = x + i\,y$ and $W(z) = K(x,y) + i\,L(x,y)$ and rearranging gives

$$\left(\frac{\partial K}{\partial x}\right)_y + i\left(\frac{\partial L}{\partial x}\right)_y = 2\,(yL - xK) + 2\,i\left(\frac{1}{\sqrt{\pi}} - (yK + xL)\right) \tag{5.31}$$

Using the relation $\partial L/\partial x = -\partial K/\partial y$ (Equation 5.25), this expression can be written entirely in terms of derivatives of K. From the real part of Equation 5.31, we have

$$\left(\frac{\partial K}{\partial x}\right)_y = 2\,(yL - xK) \tag{5.32}$$

while the imaginary part yields

$$\left(\frac{\partial K}{\partial y}\right)_x = 2\left(yK + xL - \frac{1}{\sqrt{\pi}}\right) \tag{5.33}$$

These two expressions can now be substituted into Equation 5.24 to obtain

$$\left(\frac{\partial K}{\partial \alpha_\mathrm{D}}\right)_{\Delta\tilde{\nu},\,\gamma_\mathrm{L}} = \frac{2}{\alpha_\mathrm{D}}\left(x^2K - 2xyL - y^2K + \frac{y}{\sqrt{\pi}}\right) \tag{5.34}$$

Finally, substituting this equation into Equation 5.22 produces

$$\left(\frac{\partial f_\mathrm{V}}{\partial \alpha_\mathrm{D}}\right)_{\Delta\tilde{\nu},\,\gamma_\mathrm{L}} = \frac{2}{\alpha_\mathrm{D}^2}\sqrt{\frac{\ln 2}{\pi}}\left(\frac{y}{\sqrt{\pi}} + \left[x^2 - y^2 - \frac{1}{2}\right]K - 2xyL\right) \tag{5.35}$$

At first, it appears this expression requires two non-analytic integrals to be evaluated (Equations 5.15 and 5.18). However, as computing the template Voigt profiles requires $W(z)$ to be evaluated, the real and imaginary parts of $W(z)$ (i.e. K and L) are already known. Equation 5.35 can therefore be trivially evaluated and stored alongside the template profiles $f_\mathrm{V}^\mathrm{ref}$. This specific example illustrates the more general principle that derivatives of the Voigt function can be written recursively in terms of the real and imaginary components of a complex function (Heinzel, 1978; Schreier, 1992).

A similar derivation (using Equations 5.32 and 5.33) yields the derivative of the Voigt function w.r.t. $\Delta\tilde{\nu}$

$$\left(\frac{\partial f_{\rm V}}{\partial \Delta\tilde{\nu}}\right)_{\alpha_{\rm D},\,\gamma_{\rm L}} = \frac{2}{\alpha_{\rm D}^2}\frac{\ln 2}{\sqrt{\pi}}\left(yL - xK\right) \tag{5.36}$$

and the derivative w.r.t. $\gamma_{\rm L}$

$$\left(\frac{\partial f_{\rm V}}{\partial \gamma_{\rm L}}\right)_{\alpha_{\rm D},\,\gamma_{\rm L}} = \frac{2}{\alpha_{\rm D}^2}\frac{\ln 2}{\sqrt{\pi}}\left(yK + xL - \frac{1}{\sqrt{\pi}}\right) \tag{5.37}$$

While the final derivative is not used here, it is included for completeness in anticipation of future applications (e.g. to account for $\gamma_{\rm L}$ perturbations due to other quantum numbers). Should second, or higher, order derivatives be desired, they can similarly be derived and expressed as functions of K and L (though the equations become more cumbersome). For the present applications, I have found a first order approximation achieves a sufficient balance between accuracy and computational time.

To compute the template Voigt profiles, a computational grid, $\Delta\tilde{\nu}^{\rm ref}$, must be introduced. This grid must be sufficiently fine to resolve the shape of each transition to ensure their opacity is well-sampled. Consider the HWHM of a Voigt profile, which can be approximated, to 0.02% accuracy, by (Olivero, 1977)

$$\gamma_{\rm V} \approx 0.5346\,\gamma_{\rm L} + \sqrt{0.2166\,\gamma_{\rm L}^2 + \alpha_{\rm D}^2} \tag{5.38}$$

Following Hedges & Madhusudhan (2016), I consider a grid spacing of $\gamma_{\rm V}/6$ to be sufficient to resolve a line profile. However, even at a fixed P and T, $\gamma_{\rm V}$ will not be the same for each spectral line. Indeed, $\gamma_{\rm V} = \gamma_{\rm V}(J^{\rm low}, \tilde{\nu}_{0,\,j})$ due to the dependence of $\gamma_{\rm L}$ on the lower level rotational quantum number and $\alpha_{\rm D}$ on the transition central wavenumber. $\gamma_{\rm V}$ tends to decrease for higher $J^{\rm low}$, whilst it always increases with $\tilde{\nu}_{0,\,j}$ (as $\alpha_{\rm D}$ is directly proportional to $\tilde{\nu}_{0,\,j}$). To ensure all spectral lines are well-resolved, I divide the computational domain into three regions, with a grid spacing of

$$d\tilde{\nu}^{\rm ref}_{[1,2,3]} = \min\left(\frac{\gamma_{\rm V}\left(\gamma_{\rm L}^{\rm min},\,\tilde{\nu}_0^{\rm ref} = [10^2, 10^3, 10^4]\,{\rm cm}^{-1}\right)}{6},\, d\tilde{\nu}_{\rm out}\right) \tag{5.39}$$

where $[1, 2, 3]$ denotes the first, second, and third grids, $\gamma_{\rm L}^{\rm min}$ is the minimum Lorentzian HWHM (usually at the maximum $J^{\rm low}$), $\tilde{\nu}_0^{\rm ref}$ is a set of three reference wavenumbers corresponding to each grid, and $d\tilde{\nu}_{\rm out}$ is the output wavenumber resolution for molecular

cross sections (here, $0.01\,\mathrm{cm}^{-1}$). Note that I use $d\tilde{\nu}$ to denote grid spacing, to avoid confusion with $\Delta\tilde{\nu}$ (the displacement of $\tilde{\nu}$ from the line centre). This equation, evaluated at a specific P and T, sets the grid spacing for the line profile and cross section calculations. The first grid covers $\tilde{\nu} < 10^3\,\mathrm{cm}^{-1}$, the second covers $10^3 \leqslant \tilde{\nu} < 10^4\,\mathrm{cm}^{-1}$, and the third covers $\tilde{\nu} \geqslant 10^4\,\mathrm{cm}^{-1}$. Equation 5.39 also defines three reference Voigt widths, $\gamma^{\mathrm{ref}}_{\mathrm{V},[1,2,3]}$, which are used to set the line wing cutoffs for the template Voigt profiles on each grid via

$$\Delta\tilde{\nu}^{\mathrm{cut}}_{[1,2,3]} = \min\left(500\,\gamma^{\mathrm{ref}}_{\mathrm{V},[1,2,3]},\, 30\,\mathrm{cm}^{-1}\right) \tag{5.40}$$

where the maximum cutoff value ($30\,\mathrm{cm}^{-1}$) corresponds to $500\,\gamma_{\mathrm{V}}$ when $\gamma_{\mathrm{V}} = 6\,d\tilde{\nu}_{\mathrm{out}} = 0.06\,\mathrm{cm}^{-1}$. As discussed in Hedges & Madhusudhan (2016), this cutoff ensures the bulk of the line opacity is captured, while terminating prior to the regime of sub-Lorentzian behaviour. Combining Equations 5.39 and 5.40, we see that each template Voigt profile needs to be computed at 3001 wavenumber points (line centre inclusive) up to the wing cutoff. Due to the symmetry of the Voigt profile, only one wing need be computed.

The reasoning behind using three grids, as opposed to a single grid for one value of $\tilde{\nu}^{\mathrm{ref}}_0$, is twofold. First, for low pressures where thermal broadening dominates, profiles with $\tilde{\nu}_{0,\,j} \ll \tilde{\nu}^{\mathrm{ref}}_0$ will be under-sampled whilst those with $\tilde{\nu}_{0,\,j} \gg \tilde{\nu}^{\mathrm{ref}}_0$ will be over-sampled. In the Doppler broadening limit, narrow Gaussian profiles initialised on a single computational grid at, say, $\tilde{\nu}^{\mathrm{ref}}_0 = 500\,\mathrm{cm}^{-1}$ (e.g. Hedges & Madhusudhan, 2016; Gandhi & Madhusudhan, 2017) have a spacing at $10{,}000\,\mathrm{cm}^{-1}$ of $d\tilde{\nu} = \gamma^{\mathrm{ref}}_{\mathrm{V}}/6 \approx \gamma_{\mathrm{V}}/120$ (i.e. $20\times$ finer than needed for the increased thermal widths at higher wavenumbers). A single grid is therefore computationally inefficient for low pressures. Secondly, transitions with $\tilde{\nu}_{0,\,j} \gg \tilde{\nu}^{\mathrm{ref}}_0$ can have a Voigt width $\gamma_{\mathrm{V}} \gg \gamma^{\mathrm{ref}}_{\mathrm{V}}$ (i.e. greater than the reference Voigt width corresponding to $\tilde{\nu}^{\mathrm{ref}}_0$). As the line wing cutoff is determined by $500\,\gamma^{\mathrm{ref}}_{\mathrm{V}}$, transitions at large wavenumbers can experience opacity loss due to the applied wing cutoff being much smaller than $500\,\gamma_{\mathrm{V}}$ for a specific transition. Defining three different $\gamma^{\mathrm{ref}}_{\mathrm{V}}$ values for progressively larger $\tilde{\nu}^{\mathrm{ref}}_0$ compensates for this issue by extending the line wing cutoff. Finally, I note that differences between γ_{V} and $\gamma^{\mathrm{ref}}_{\mathrm{V}}$, due to taking only $\gamma^{\mathrm{min}}_{\mathrm{L}}$ in initialising the grid spacing and wing cutoffs, are a smaller effect. For example, for H_2O at 1400 K the difference between $\gamma^{\mathrm{min}}_{\mathrm{L}}$ ($J^{\mathrm{low}} = 50$) and $\gamma^{\mathrm{max}}_{\mathrm{L}}$ ($J^{\mathrm{low}} = 0$) is $< 40\%$ (Tennyson et al., 2016) - with differences between intermediate values of J^{low} smaller. With $d\tilde{\nu}^{\mathrm{ref}}_{[1,2,3]}$ specified, template Voigt profiles can be constructed.

The template Voigt profiles, $f_{\rm V}^{\rm ref} = f_{\rm V}(\Delta\tilde{\nu}^{\rm ref}, \alpha_{\rm D}^{\rm ref}, \gamma_{\rm L})$, are pre-computed at the beginning of a cross section calculation. The template Doppler widths, $\alpha_{\rm D}^{\rm ref}$, are log-uniformly spaced from $\alpha_{\rm D}(\tilde{\nu} = 1\,{\rm cm}^{-1})$ to $\alpha_{\rm D}(\tilde{\nu} = 30000\,{\rm cm}^{-1})$ at 500 points - such that neighbouring values differ by $\approx 1\%$. The wavenumber corresponding to each $\alpha_{\rm D}^{\rm ref}$ (Equation 5.6) determines which grid spacing $d\tilde{\nu}_{[1,2,3]}^{\rm ref}$ (Equation 5.39) and line wing cutoff $\Delta\tilde{\nu}_{[1,2,3]}^{\rm cut}$ (Equation 5.40) is appropriate (according to which grid $\tilde{\nu}_{0,\,j}$ lies on). For $\gamma_{\rm L}$, the range of $J^{\rm low}$ with known broadening parameters is typically in the range 10-50. The total number of template profiles is thus $N_{\alpha_{\rm D}} \times N_{\gamma_{\rm L}}$ (e.g. $500 \times 50 = 25,000$ for H_2O), with each evaluated at 3001 wavenumbers. The initialisation stage requires the Voigt function (Equation 5.19) and its derivatives (Equations 5.35 and 5.36) to be evaluated and stored at $\approx 10^7 - 10^8$ points (typically taking 30-300 s).

With the template Voigt line profiles in hand, the GVV algorithm to compute a molecular cross section proceeds as follows. For each transition, j, in a line list:

1. Compute $S_j(T)$ using Equation 5.2 (for ExoMol) or Equation 5.4 (for HITRAN). If necessary, extrapolate $Q(T)$ using a 5th order spline.

2. Identify which of the three computational grids $\tilde{\nu}_{0,\,j}$ lies on. Grid 1 covers $\tilde{\nu} < 10^3\,{\rm cm}^{-1}$, grid 2 covers $10^3 \leqslant \tilde{\nu} < 10^4\,{\rm cm}^{-1}$, and grid 3 covers $\tilde{\nu} \geqslant 10^4\,{\rm cm}^{-1}$.

3. Place $\tilde{\nu}_{0,\,j}$ at the closest wavenumber on the relevant grid.

4. Identify the template Voigt profile, $f_{{\rm V},({\rm i},{\rm k})}^{\rm ref}$, with $\alpha_{{\rm D},\,{\rm i}}^{\rm ref} \approx \alpha_{{\rm D},\,{\rm j}}$ and $\gamma_{{\rm L},\,{\rm k}} = \gamma_{\rm L}(J_j^{\rm low})$. For transitions with $J_j^{\rm low} > J_{\rm max}^{\rm low}$ (i.e. the maximum value for which broadening parameters are available), the template with $\gamma_{\rm L}(J_{\rm max}^{\rm low})$ is taken.

5. If the profile range (i.e. $\tilde{\nu}_{0,\,j} - \Delta\tilde{\nu}_{[1,2,3]}^{\rm cut} \rightarrow \tilde{\nu}_{0,\,j} + \Delta\tilde{\nu}_{[1,2,3]}^{\rm cut}$) falls entirely on one computational grid: working outwards from the line centre, use the 1st and 3rd terms of 5.21 to adjust the template to the desired $\alpha_{{\rm D},\,{\rm j}}$. This utilises the pre-computed $\partial f_{\rm V}/\partial\alpha_{\rm D}$ at the expansion point.

6. Should the profile wings cross grid boundaries, also use the 2nd term of Equation 5.21 to correct for the new grid $\tilde{\nu}$ points being misaligned compared to the template - essentially using $\partial f_{\rm V}/\partial\Delta\tilde{\nu}$ to interpolate to the new grid spacing.

7. At each point along the adjusted profile, increment $\sigma(\tilde{\nu})$ by $\delta\sigma_j(\tilde{\nu}) = S_j(T)\, f(\Delta\tilde{\nu})$.

Finally, the cross section array $\sigma(\tilde{\nu})$ is binned to an output resolution of $0.01\,{\rm cm}^{-1}$, ranging from 200 to 25,000 $\,{\rm cm}^{-1}$ (i.e. 0.4–50 µm) - corresponding to $R = 10^6$ at 1 µm. The calculation can then be repeated for each (P, T) where a cross section is required.

5.3.2 Validating the GVV method

Figure 5.1 compares a H_2O cross section computed using the GVV method with a traditional calculation. Specifically, I compare the outputs from the GVV method with that from the ExoCross code (Yurchenko et al., 2018d). Each calculation uses the BT2 H_2O line list (Barber et al., 2006), evaluates the cross section at $T = 1000\,\mathrm{K}$ and $P = 1\,\mathrm{bar}$, uses the same pressure broadening parameters (S. Yurchenko, personal communication, 2018), and outputs to identical wavenumber grids. Keeping these aspects of the calculation identical serves to validate both the GVV method and the code I created to implement it. The lower panel in Figure 5.1 shows the fractional difference between the two cross sections over the 0.4–50.0 um spectral range. The median fractional difference is $< 10^{-2}$, which I consider excellent agreement.

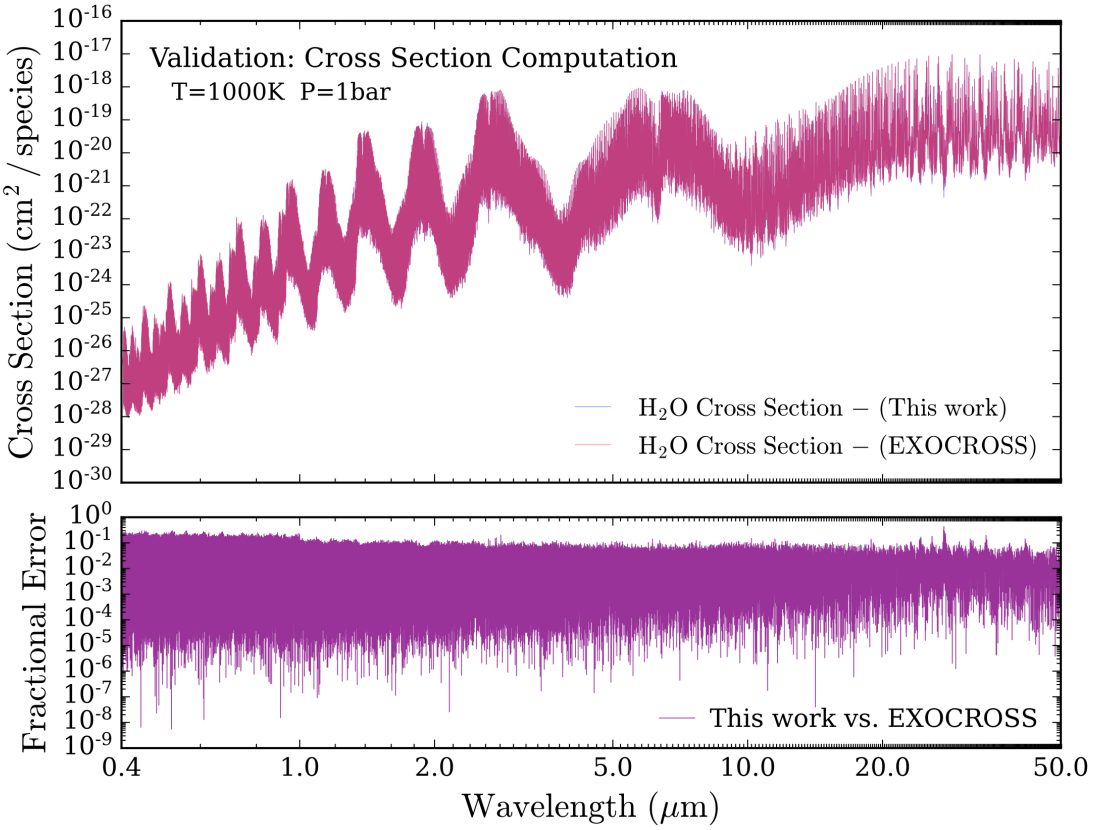


Fig. 5.1 Validation of the Generalised Vectorised Voigt (GVV) method. Top: a H_2O cross section computed by the GVV method (blue) is compared against an independent calculation using the ExoCross code (Yurchenko et al., 2018d) (red). Bottom: fractional difference between the H_2O cross sections calculated by each method. The median fractional difference is $\lesssim 10^{-2}$, indicating excellent agreement.

The influence of different calculation choices is illustrated in Figure 5.2. Here, I additionally compare the aforementioned cross sections with the output of a third independent code (M. Marley, personal communication, 2018). This third cross section uses weaker pressure broadening parameters, resulting in sharper transitions with a greater core strength but weaker continuum (due to profile area conservation). The lower panel demonstrates that the choice of different broadening parameters results in a fractional difference $\sim 10\times$ greater than that between the GVV method and ExoCross. At present, there is often little reason to prefer one set of broadening parameters over another, demonstrating that factors like the chosen broadening parameters are a greater source of error than approximations built into the GVV method.

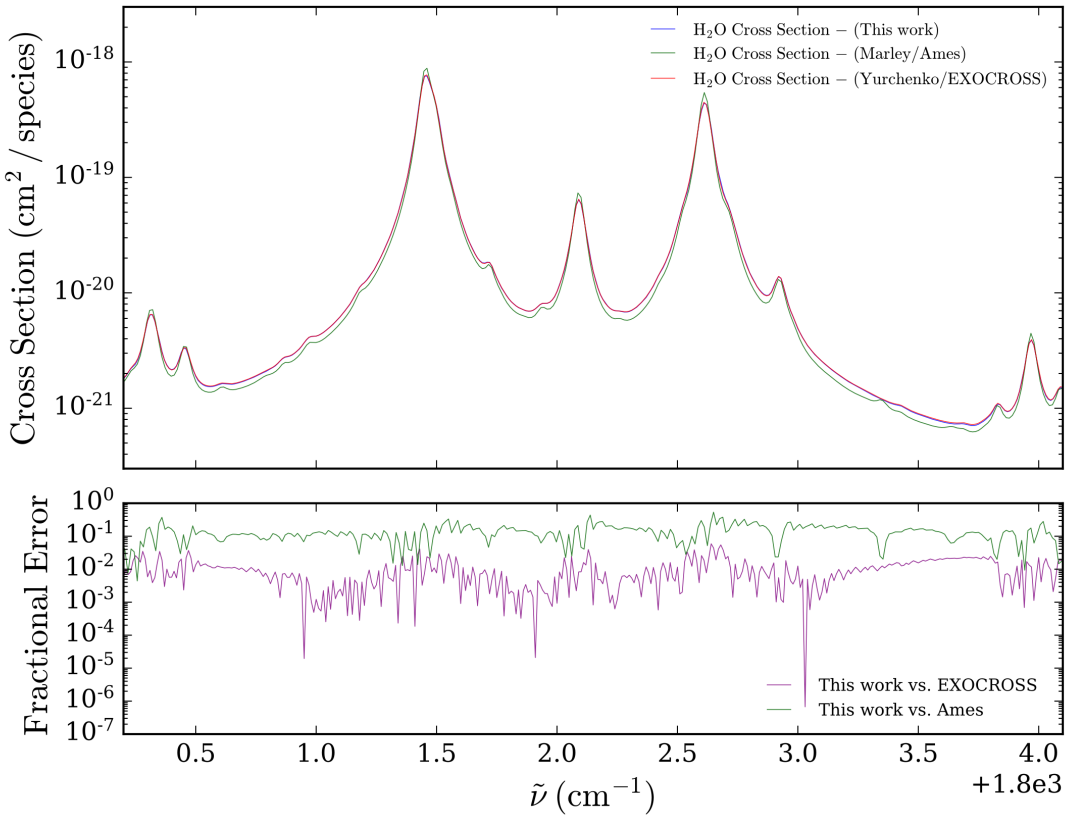


Fig. 5.2 Cross section comparison for different pressure broadening.
Top: zoomed-in view of the H_2O cross section around 5.56 µm, with individual line profiles resolved. Three calculations are compared. The GVV method (blue) and ExoCross (red) use the same pressure broadening parameters, whilst the third (green) uses different broadening parameters (M. Marley, personal communication, 2018). Bottom: fractional differences between the cross sections calculated by each method. The difference rises by a factor of ~ 10 when different broadening parameters are used.

5.4 Atomic opacities

The computation of cross sections for atoms, both neutral and ionised, follows similar principles as outlined above for molecules. However, the absence of rotational or vibrational transitions results in vastly smaller line lists ($\lesssim 10^3 - 10^5$). The GVV method is therefore not necessary for atomic opacity calculations, where Voigt profiles can be separately computed for each transition. However, there are some unique considerations for atomic opacities: (i) atomic transitions are not included in molecular line list databases (e.g. ExoMol or HITEMP), so one must use databases such as NIST (Kramida et al., 2018) or VALD (Ryabchikova et al., 2015); (ii) atomic line databases store different transition properties to molecular databases; and (iii) strong atomic resonance lines have large opacities in the far line wings where Voigt profiles are no longer an accurate description. I will address each of these aspects in turn.

5.4.1 Atomic line intensities

Atomic line databases do not usually store Einstein A coefficients, A_j. Instead, atomic databases provide $\log(g_j^{\rm low} f_{\rm osc,\,j})$, where $g_j^{\rm low}$ is the lower level degeneracy and $f_{\rm osc,\,j}$ is the oscillator strength of transition j (not to be confused with line profiles). Oscillator strengths and Einstein A coefficients are related by

$$g_j^{\rm low} f_j = \left(\frac{m_e\, c}{8\pi^2\, e^2\, \tilde{\nu}^2}\right) g_j^{\rm up} A_j \tag{5.41}$$

where m_e and e are the mass and charge of the electron (cgs units). Substitution of Equation 5.41 into Equation 5.2 gives the line intensity of an atomic transition

$$S_j(T) = \frac{\pi e^2\, g_j^{\rm low} f_{\rm osc,\,j}}{m_e\, c^2} \frac{\exp\left(-c_2\, \tilde{E}_j^{\rm low}/T\right)}{Q(T)} \left(1 - \exp\left(-\frac{c_2\, \tilde{\nu}_{0,\,j}}{T}\right)\right) \tag{5.42}$$

where the three terms have the same physical interpretation as in Equation 5.2.

5.4.2 Pressure broadening for atoms

Atomic databases also do not provide pressure broadening parameters in the form of $\gamma^0_{\rm L,\,p}$ and $n_{\rm L,\,p}$, as required to compute $\gamma_{\rm L}$ (via Equation 5.8). Instead, they tabulate $\Gamma_{\rm vdW,\,j}$, the van der Waals broadening coefficient for transition j, which can be related

to $\gamma_{\rm L}$ by (Sharp & Burrows, 2007)

$$\gamma_{\rm L} = \frac{\Gamma_{\rm vdW,\,j}}{4\pi c} \sum_p \left(\frac{m_H(m+m_p)}{m_p(m+m_H)}\right)^{\frac{3}{10}} \left(\frac{\alpha_p}{\alpha_H}\right)^{\frac{2}{5}} \left(\frac{T}{1000\,{\rm K}}\right)^{\frac{3}{10}} X_p\, n_{\rm tot} \tag{5.43}$$

where m_H, m, and m_p are, respectively, the masses of the hydrogen atom, the atomic species experiencing line broadening, and the perturbing species causing the broadening; α_p and α_H are the polarisabilities of the perturber and hydrogen, respectively, and $n_{\rm tot}$ is the total local number density of atoms and molecules. To make the comparison between this expression and Equation 5.8, I substitute the ideal gas law ($n_{\rm tot} = P/k_BT$) into Equation 5.43 to make the temperature and pressure dependence explicit. After various rearrangements and unit conversions (e.g. dyne/cm^2 into bar), I obtain

$$\gamma_{\rm L} = \sum_p \gamma^0_{\rm L,\,p}\, X_p \left(\frac{296\,{\rm K}}{T}\right)^{\frac{7}{10}} \left(\frac{P}{1\,{\rm bar}}\right). \tag{5.44}$$

where

$$\gamma^0_{\rm L,\,p} = 2.2593427 \times 10^7\, \Gamma_{\rm vdW,\,j} \left(\frac{\tilde{m}_H(\tilde{m}+\tilde{m}_p)}{\tilde{m}_p(\tilde{m}+\tilde{m}_H)}\right)^{\frac{3}{10}} \left(\frac{\alpha_p}{\alpha_H}\right)^{\frac{2}{5}} \tag{5.45}$$

and $\tilde{m}_i \equiv m_i/u$ are masses expressed in atomic mass units, u. With Equation 5.44 in precisely the same form as Equation 5.8, we see that pressure broadening of atomic lines has $n_{\rm L,\,p} = 0.7$ for all perturbing gases. Equations 5.44 and 5.45 allow the Lorentzian HWHM, $\gamma_{\rm L}$ (in cm^{-1}), to be computed for broadening due to arbitrary background gases. As before, in this work I assume the broadening is due to H_2 and He in solar proportions ($X_{\rm H_2} = 85\%$, $X_{\rm He} = 15\%$).

However, for some transitions in atomic line databases a value of $\Gamma_{\rm vdW,\,j}$ is not available. This is particularly true for most non-resonant alkali metal lines in VALD. In this case, I use an analytic approach to derive approximate pressure broadening coefficients. In the impact theory of van der Waals interactions, the HWHM of a spectral line is given approximately by (Peach, 1981; Dimitrijević & Peach, 1990)

$$\gamma_{\rm L} = \frac{7.9008}{2\pi c} \sum_p \bar{v}_p^{\frac{3}{5}}\, C_{6,\,p}^{\frac{2}{5}}\, X_p\, n_{\rm tot} \tag{5.46}$$

where

$$\bar{v}_p = \left[\frac{8k_BT}{\pi}\left(\frac{m+m_p}{m\,m_p}\right)\right]^{\frac{1}{2}} \tag{5.47}$$

is the mean relative perturber velocity and $C_{6,p}$ is the van der Waals interaction coefficient. Under the assumption of a hydrogen-like (Coulombic) potential, as is the case for alkali metals, $C_{6,p}$ can be expressed as (Unsöld, 1955; Hindmarsh et al., 1967)

$$C_{6,p} = \frac{e^2}{\hbar}\,\alpha_p\,\left(\langle r^2_{\rm up}\rangle - \langle r^2_{\rm low}\rangle\right) \tag{5.48}$$

where $\hbar \equiv h/2\pi$ is the reduced Planck's constant and $\langle r^2_{\rm up}\rangle$ and $\langle r^2_{\rm low}\rangle$ are the mean-squared electron-nucleus separation for the upper and lower states of the transition. For hydrogenic wavefunctions, the squared separation for state k is (Bates & Damgaard, 1949)

$$\langle r^2_k\rangle = \frac{n^{*^2}_k}{2(Z+1)^2}\left(5n^{*^2}_k + 1 - 3l_k(l_k+1)\right)a_0^2 \tag{5.49}$$

where Z is the total charge of the atom, a_0 is the Bohr radius, and l_k is the orbital angular momentum of state k. n^*_k is the effective principal quantum number, given by

$$n^{*^2}_k = 13.6\,{\rm eV}\,\frac{({\rm Z}+1)^2}{{\rm E}_\infty - {\rm E_k}} \tag{5.50}$$

where E_∞ is the ionisation energy of the atom and E_k is the energy of the state k (both in units of eV). Substituting Equations 5.48, 5.47, and the ideal gas law into Equation 5.46, I finally obtain (after some algebra and unit conversions)

$$\gamma_{\rm L} = \sum_p \gamma^0_{\rm L,\,p}\,X_p\left(\frac{296\,{\rm K}}{T}\right)^{\frac{7}{10}}\left(\frac{P}{1\,{\rm bar}}\right). \tag{5.51}$$

where

$$\gamma^0_{\rm L,\,p} = 0.1972\left(\frac{\tilde{m}+\tilde{m}_p}{\tilde{m}\,\tilde{m}_p}\right)^{\frac{3}{10}}\left(\frac{\alpha_p}{\alpha_H}\right)^{\frac{2}{5}}\left(\frac{\langle r^2_{\rm up}\rangle_j}{a_0^2} - \frac{\langle r^2_{\rm low}\rangle_j}{a_0^2}\right)^{\frac{2}{5}} \tag{5.52}$$

The overall procedure for deriving $\gamma_{\rm L}$ for atomic transitions is therefore as follows:

1. Check if a van der Waals coefficient, $\Gamma_{\rm vdW}$, is available for transition j.
2. If $\Gamma_{\rm vdW,\,j}$ is known: use Equations 5.45 and 5.44 to determine $\gamma_{\rm L}$.
3. If $\Gamma_{\rm vdW,\,j}$ is unknown and the atom has a single valence electron: use Equations 5.50 and 5.49 to check if $\langle r^2_{\rm up}\rangle_j - \langle r^2_{\rm low}\rangle_j > 0$ (verifies validity of hydrogenic approx.).
4. If hydrogenic approximation holds: use Equations 5.52 and 5.51 to determine $\gamma_{\rm L}$.
5. If none of the above holds (usually rare): set $\gamma_{\rm L} = 0$.

VALD also provides natural broadening coefficients, $\Gamma_{\mathrm{nat,j}}$, which can be combined with pressure broadening to obtain a total Lorentzian HWHM, γ_{tot}, via

$$\gamma_{\mathrm{tot}} = \gamma_{\mathrm{L}} + \gamma_{\mathrm{nat}} = \gamma_{\mathrm{L}} + \frac{\Gamma_{\mathrm{nat,j}}}{4\pi c} \tag{5.53}$$

In Voigt profile calculations, γ_{tot} can be used in place of γ_{L}. Although $\gamma_{\mathrm{nat}} \ll \gamma_{\mathrm{L}}$ (except for the lowest pressures), it is included for completeness in atomic opacity calculations.

5.4.3 Sub-Voigt line profiles

Finally, for the far line wings of the Na and K resonance lines I adopt a sub-Lorentzian (and hence sub-Voigt) profile prescription. Following Baudino et al. (2015), the line profile for resonance transitions is taken as

$$f(\Delta\tilde{\nu},\, \alpha_{\mathrm{D}},\, \gamma_{\mathrm{L}}) = \begin{cases} f_{\mathrm{V}}(\Delta\tilde{\nu},\, \alpha_{\mathrm{D}},\, \gamma_{\mathrm{L}}), & \Delta\tilde{\nu} < \Delta\tilde{\nu}_{\mathrm{d}} \\ f_{\mathrm{V}}(\Delta\tilde{\nu}_{\mathrm{d}},\, \alpha_{\mathrm{D}},\, \gamma_{\mathrm{L}}) \left(\frac{\Delta\tilde{\nu}_{\mathrm{d}}}{\Delta\tilde{\nu}}\right)^{\frac{3}{2}} e^{\left(-\frac{\Delta\tilde{\nu}}{\tilde{\nu}_F}\left[\frac{hc\,\Delta\tilde{\nu}}{k_B T}\right]\right)}, & \Delta\tilde{\nu} \geqslant \Delta\tilde{\nu}_{\mathrm{d}} \end{cases} \tag{5.54}$$

where $\Delta\tilde{\nu}_{\mathrm{d}}$ is a detuning frequency, which sets the wavenumber beyond which sub-Voigt behaviour begins, and $\tilde{\nu}_F$ is a fitting parameter used to match this profile with the results of Burrows & Volobuyev (2003). The resonance line detuning frequencies are taken as $\Delta\tilde{\nu}_{\mathrm{d}} = 30\,(T/500\,\mathrm{K})\,\mathrm{cm}^{-1}$ for Na and $20\,(T/500\,\mathrm{K})\,\mathrm{cm}^{-1}$ for K (Burrows et al., 2000). $\tilde{\nu}_F$ is taken as $5000\,\mathrm{cm}^{-1}$ for Na and $1600\,\mathrm{cm}^{-1}$ for K (Baudino et al., 2015). The resonance line profiles are calculated up to a cutoff of $\Delta\tilde{\nu}^{\mathrm{cut}} = 9000\,\mathrm{cm}^{-1}$, with all other lines treated as Voigt profiles with the cutoff specified by Equation 5.40.

5.5 Continuum opacity sources

Continuum opacities, such as Rayleigh scattering and collision-induced absorption (CIA), are especially important to the interpretation of transmission spectra. By providing a spectral continuum independent of the mixing ratios of trace species (de Wit & Seager, 2013), their inclusion is vital to break degeneracies which can otherwise manifest between mixing ratios and the reference pressure parameter (Benneke & Seager, 2012; Griffith, 2014; Heng & Kitzmann, 2017; Welbanks & Madhusudhan, 2019). I will now describe the calculation of Rayleigh scattering cross sections and note the source of CIA used herein.

5.5.1 Rayleigh scattering

The calculation of Rayleigh scattering cross sections is relatively straightforward (compared to molecular and atomic absorption cross sections). In general, they can be expressed, for a chemical species i, as (Sneep & Ubachs, 2005)

$$\sigma_{\mathrm{Ray,\,i}}\left(\tilde{\nu}\right) = \frac{24\pi^3\,\tilde{\nu}^4}{n_{\mathrm{ref}}^2}\left(\frac{\eta_i^2\left(\tilde{\nu}\right)-1}{\eta_i^2\left(\tilde{\nu}\right)+2}\right)^2 F_{\mathrm{k,\,i}}\left(\tilde{\nu}\right) \tag{5.55}$$

where $\eta_i(\tilde{\nu})$ is the refractive index of species i, n_{ref} is a reference number density corresponding to the same conditions (P and T) where $\eta_i(\tilde{\nu})$ is measured, and $F_{\mathrm{k,\,i}}\left(\tilde{\nu}\right)$ is the 'King correction factor'. Here, all refractive indices are either measured or scaled to standard conditions ($T = 273.15\,\mathrm{K}$, $P = 1.01325\,\mathrm{bar}$), such that $n_{\mathrm{ref}} = 2.6867811 \times 10^{19}\,\mathrm{cm}^{-3}$ (known as the Loschmidt constant $\equiv$ 1 amagat).

To make clear that Equation 5.55 is independent of number density, one can use the Lorentz-Lorenz relation (Jackson & Fox, 1999)

$$\bar{\alpha}_i\left(\tilde{\nu}\right) = \frac{3}{4\pi n_{\mathrm{ref}}}\left(\frac{\eta_i^2\left(\tilde{\nu}\right)-1}{\eta_i^2\left(\tilde{\nu}\right)+2}\right) \tag{5.56}$$

which relates the mean molecular / atomic polarisability of the scattering species, $\bar{\alpha}_i\left(\tilde{\nu}\right)$ (in cm^3), to the refractive index. This equation allows the refractive index dependence in Equation 5.55 to be re-expressed in terms of polarisability via

$$\sigma_{\mathrm{Ray,\,i}}\left(\tilde{\nu}\right) = \frac{128}{3}\pi^5\,\bar{\alpha}_i^2\left(\tilde{\nu}\right)\tilde{\nu}^4\,F_{\mathrm{k,\,i}}\left(\tilde{\nu}\right) \tag{5.57}$$

which makes clear that the Rayleigh scattering cross section depends only on the properties of the scattering species, and not on the local gas properties[3].

The King correction factor (King, 1923) encodes the increase in Rayleigh scattering for non-spherical charge distributions. It can be expressed as (Sneep & Ubachs, 2005)

$$F_{\mathrm{k,\,i}}\left(\tilde{\nu}\right) = 1 + \frac{2}{9}\left(\frac{\alpha_i^{\parallel}\left(\tilde{\nu}\right)-\alpha_i^{\perp}\left(\tilde{\nu}\right)}{\bar{\alpha}_i\left(\tilde{\nu}\right)}\right)^2 \tag{5.58}$$

where $\alpha_i^{\parallel}$ and $\alpha_i^{\perp}$ are, respectively, the components of the polarisability tensor parallel and perpendicular to the axis of symmetry of the scattering species. The difference

[3] $\bar{\alpha}_i$ can depend slightly on temperature, but there is often little data available to quantify this.

between these two quantities is also called the polarisability anisotropy. Note that $F_{\mathrm{k,i}}(\tilde{\nu}) = 1$ for spherical atoms (e.g. He), but is otherwise always greater than 1 (though only at the $\sim$ 1-10% level). In practice, $F_{\mathrm{k,i}}(\tilde{\nu})$ is often given in terms of fitting functions (w.r.t. $\tilde{\nu}$) for the components of α_i (e.g. Hohm, 1994) or $F_{\mathrm{k,i}}(\tilde{\nu})$ itself, as derived from experiments (e.g. Sneep & Ubachs, 2005).

In calculating Rayleigh cross sections, I use wavelength-dependent refractive indices and King correction factors wherever available (via Equation 5.55). Should a refractive index not be known, static polarisabilities are obtained from the CRC handbook (Rumble, 2018) and substituted into Equation 5.57. When no King correction factor is available, I take $F_{\mathrm{k,i}}(\tilde{\nu}) = 1$. Finally, I note that the wavelength dependence in Equations 5.55 and 5.57 is close to λ^{-4}, but deviates to become steeper for shorter wavelengths due to the λ dependence of $\eta(\lambda)$ and $F_k(\lambda)$ (explaining the functional form for the H_2 Rayleigh scattering cross section previously seen in Equation 2.28).

5.5.2 Collision-induced absorption

The pair-process of collision-induced absorption has been previously discussed in section 2.2.3. As mentioned there, CIA does not require separate opacity calculations, as binary absorption cross sections (in units of $\mathrm{cm}^5\,\mathrm{molecule}^{-2}$) can be found tabulated in HITRAN (Richard et al., 2012; Karman et al., 2019). For this opacity update, I have included H_2-N_2 and H_2-CH_4 CIA (alongside H_2-H_2 and H_2-He used in previous chapters). Future updates will include other sources of CIA, especially those important for terrestrial atmospheres (e.g. O_2-O_2, Yan et al., 2015).

5.6 The POSEIDON opacity database

Following the methods outlined above, I have created an extensive database of opacity sources relevant to exoplanetary atmospheres. This database is visually represented in Figure 5.3, wherein many prominent sources of visible and near-infrared opacity are compared. The five main takeaways are: (i) atomic resonance transitions have the strongest visible cross sections; (ii) heavy metal oxides and hydrides have large visible opacities; (iii) light molecules dominate the infrared opacity; (iv) Rayleigh scattering cross sections becomes increasingly important at short visible / near-UV wavelengths; and (v) collision-induced absorption is an important infrared opacity source, (especially around 2.5 µm). Recall that the actual influence of each molecule or atom is governed

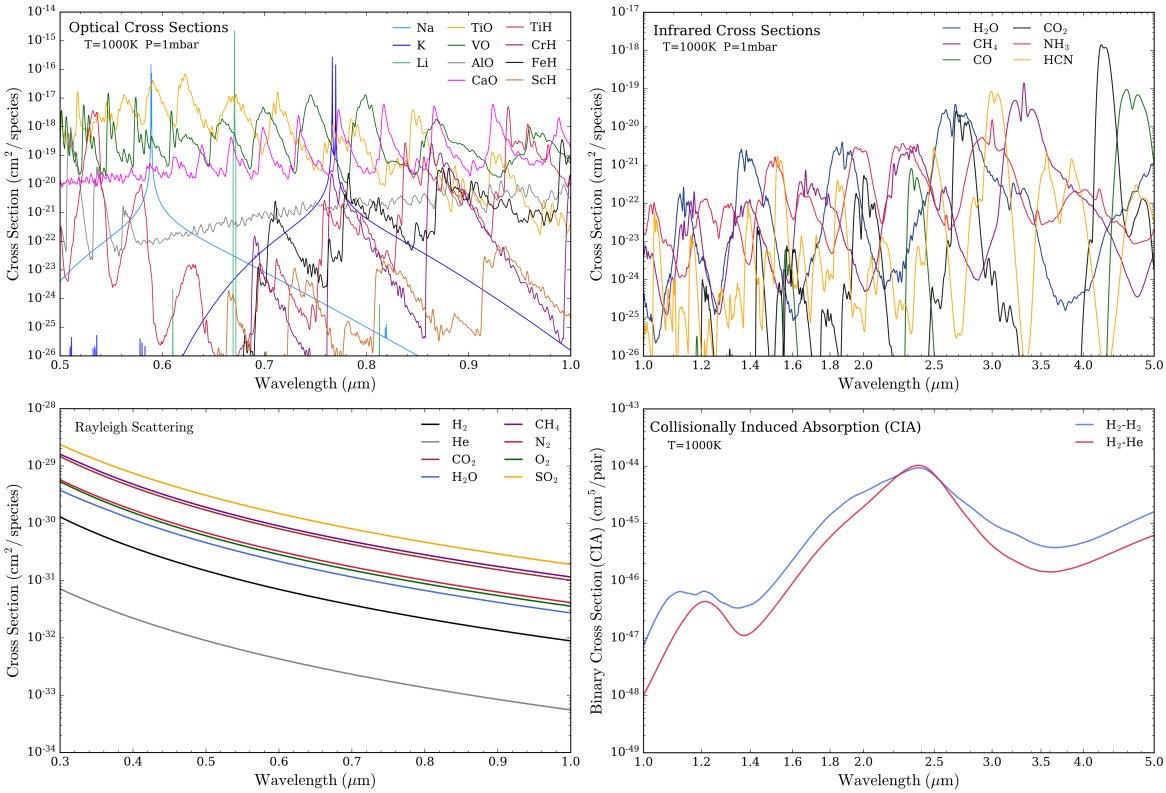


Fig. 5.3 A catalogue of prominent opacity sources included in POSEIDON. Top: selected visible (left) and near-infrared (right) cross sections included in the POSEIDON opacity database. All cross sections shown are calculated at $T = 1000\,\mathrm{K}$ and $P = 1\,\mathrm{mbar}$ - roughly representative of the upper atmosphere of a cooler hot Jupiter / warm exo-Neptune probed by transmission spectra. All cross sections have been smoothed for clarity, with molecular cross section shown at $R \sim 1000$ and atomic cross sections shown at $R \sim 5000$ (to resolve resonance doublets). Bottom: selected continuum opacity sources, including Rayleigh scattering (left) at visible wavelengths and collision-induced absorption (right) in the near-infrared.

by its extinction coefficient, either $\propto n\sigma$ (for absorption / scattering cross sections) or $\propto n^2\sigma$ (for CIA), and therefore a low-abundance species with a strong cross section can rival the extinction of a high-abundance species with a weak cross section. Nevertheless, this plot of the intrinsic absorption / scattering strength of different molecules and atoms provides an intuitive picture into prominent opacity sources.

The molecular and atomic cross sections comprising the POSEIDON opacity database are pre-computed separately to transmission spectra calculations. Each cross section is calculated at 2,480,001 wavenumbers for 9 pressures and 18 temperatures. Table 5.1 summarises all molecules and atoms for which cross sections have currently been calculated, along with the P, T, and $\tilde{\nu}$ grids they span. The temperature and

Table 5.1 Molecular and atomic cross sections supported by POSEIDON

Molecules	H_2O, CH_4, NH_3, HCN, CO, CO_2, C_2H_2, PH_3, SO_2, H_2S, HS, N_2, O_2, O_3, OH, NO, N_2O, NO_2, TiO, VO, AlO, SiO, CaO, TiH, CrH, FeH, ScH, AlH, SiH, BeH, CaH, MgH, LiH, NaH, CH, NH, H_3^+
Atoms†	Na, K, Li, Rb, Cs, Fe, Fe^+, Ti, Ti^+
Grid	
T (K)	100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000, 3500
P (bar)	10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}
$\tilde{\nu}$ (cm^{-1})	$200 - 25000$ (@ $0.01\,\mathrm{cm}^{-1}$)
λ (µm)	$0.4 - 50$ ($R = 10^6$ @ 1 µm)

† Monatomic ions are also referred to as atoms.

Notes: All cross sections are computed and stored on the same T, P, and $\tilde{\nu}$ grid. For each species, cross sections are computed at $N_T \times N_P = 18 \times 9 = 162$ points. $\lambda\,(\mu\mathrm{m}) = \frac{10^4}{\tilde{\nu}(\mathrm{cm}^{-1})}$. After mapping $\sigma_{\tilde{\nu}}$ onto the corresponding λ grid, σ_λ has a decreasing spectral resolution with increasing wavelength ($R = 2 \times 10^6$ @ 0.4 µm $\rightarrow 2.5 \times 10^4$ @ 50 µm).

pressure range has been selected to cover the full range of conditions expected to be probed in spectral observations of exoplanet atmospheres, with the wavenumber range chosen to cover all current (and anticipated) wavelengths in which exoplanet atmospheres are observed. The full database is 25 GB in size, storing $\log_{10}(\sigma)$ values as 32 bit floats for efficiency. Prior to a transmission spectrum calculation, the database is read in species-by-species (for a user-specified list of molecules and atoms) and mapped onto a model wavelength grid (see section 2.3.1). Similarly, the continuum opacity sources (Rayleigh & CIA) included in POSEIDON are summarised in Table 5.2.

The molecular line lists and partition functions used to compute the POSEIDON opacity database primarily come from the ExoMol database (Tennyson et al., 2016). Where high-temperature line lists are not currently available, HITRAN-2016 line lists are used (Gordon et al., 2017). Care has been taken to ensure that the latest and most complete line lists are used. However, in some cases line lists are only guaranteed to be complete up to a certain temperature (e.g. 1500 K for NH_3, Yurchenko et al., 2011). In such cases, the computed cross sections may be less reliable for higher temperatures. Pressure broadening parameters for H_2 and He as a function of J^{low} are used wherever available. Where these are not available, air broadening parameters are obtained by grouping HITRAN broadening parameters according to J^{low} and averaging over other

Table 5.2 Continuum opacity sources included in POSEIDON

Rayleigh	$\eta\,(\tilde{\nu})$ or $\bar{\alpha}\,(\tilde{\nu})$	$F_k\,(\tilde{\nu})$
H_2	Hohm (1994)	Hohm (1994)
He	Cuthbertson & Cuthbertson (1932)	1.0
" "	Mansfield & Peck (1969)	1.0
H_2O	Hill & Lawrence (1986)	Simos et al. (2006)
CO_2	Hohm (1994)	Hohm (1994)
CH_4	Sneep & Ubachs (2005)	Sneep & Ubachs (2005)
NH_3	Hohm (1994)	Hohm (1994)
N_2	Sneep & Ubachs (2005)	Sneep & Ubachs (2005)
O_2	Hohm (1994)	Hohm (1994)
N_2O	Hohm (1994)	Hohm (1994)
O_3	Rumble (2018)	Brasseur & De Rudder (1986)
CO	Rumble (2018)	Bogaard et al. (1978)
H_2S	Rumble (2018)	Bogaard et al. (1978)
SO_2	Rumble (2018)	Bogaard et al. (1978)
C_2H_2	Rumble (2018)	Bogaard et al. (1978)
Others	Rumble (2018)	1.0
CIA	$\sigma_{\mathrm{CIA}}\,(\tilde{\nu})$	$\tilde{\nu}$ range (cm^{-1})
H_2-H_2	Richard et al. (2012)	20 - 10,000
H_2-He	Richard et al. (2012)	20 - 20,000
H_2-N_2	Richard et al. (2012)	0.02 - 1886
H_2-CH_4	Richard et al. (2012)	0.02 - 1946

Notes: He has $F_k = 1$ due to spherical symmetry. All species without $F_k\,(\tilde{\nu})$ data are taken to have $F_k = 1.0\ \forall\ \tilde{\nu}$. CIA outside of the tabulated $\tilde{\nu}$ range is set to zero.

quantum numbers. If air broadening is not available, the prescription from Sharp & Burrows (2007) is applied: $\gamma_{\mathrm{L}}^{0} = (0.0493 - 0.000986 \min\left[J^{\mathrm{low}}, 30\right])$, $n_{\mathrm{L}} = 0$. The full list of molecular line lists used is given in Table 5.3.

The atomic line lists used are taken from the collated compilations in VALD3 (Ryabchikova et al., 2015). All atomic partition functions are obtained from Barklem & Collet (2016). Pressure broadening from H_2 and He is applied to all atomic species, as described in the preceding sections. A full summary of the original line list sources (within VALD3) used to generate each atomic cross section is given in Table 5.4.

Table 5.3 Molecular line lists used in the POSEIDON opacity database

Molecule	Line list	References	Broadening
H_2O	POKAZATEL	Polyansky et al. (2018)	H_2 + He
CH_4	34to10	Yurchenko et al. (2017)	H_2 + He
NH_3	BYTe	Yurchenko et al. (2011)	H_2 + He
HCN	Harris	Barber et al. (2014)	H_2 + He
CO	Li15	Li et al. (2015)	H_2 + He
CO_2	CDSD-4000	Tashkun & Perevalov (2011)	H_2 + He
C_2H_2	ASD-1000	Lyulin & Perevalov (2017)	H_2 + He
PH_3	SAlTY	Sousa-Silva et al. (2015)	H_2 + He
SO_2	ExoAmes	Underwood et al. (2016)	H_2 + He
H_2S	AYT2	Azzam et al. (2016)	Air
HS	SNaSH	Yurchenko et al. (2018b)	Air
N_2	HITRAN-N2	Gordon et al. (2017)	Air
O_2	HITRAN-O2	Gordon et al. (2017)	Air
O_3	HITRAN-O3	Gordon et al. (2017)	Air
OH	Brooke16	Brooke et al. (2016)	Air
NO	NOname	Wong et al. (2017)	Air
N_2O	NOSD-1000	Tashkun et al. (2016)	Air
NO_2	NDSD-1000	Lukashevskaya et al. (2016)	Air
TiO	ToTo	McKemmish et al. (2019)	SB07
VO	VOMYT	McKemmish et al. (2016)	SB07
AlO	ATP	Patrascu et al. (2015)	SB07
SiO	EBJT	Barton et al. (2013)	SB07
CaO	VBATHY	Yurchenko et al. (2016)	SB07
TiH	Burrows05	Burrows et al. (2005)	SB07
CrH	Burrows02	Burrows et al. (2002)	SB07
FeH	Wende	Wende et al. (2010)	SB07
ScH	LYT	Lodi et al. (2015)	SB07
AlH	AlHambra	Yurchenko et al. (2018c)	SB07
SiH	SiGHTLY	Yurchenko et al. (2018a)	SB07
BeH	Darby-Lewis	Darby-Lewis et al. (2018)	SB07
CaH	Yadin-CaH	Yadin et al. (2012)	SB07
“ ”	Li12	Li et al. (2012)	SB07
MgH	Yadin-MgH	Yadin et al. (2012)	SB07
“ ”	Gharib-Nezhad	Gharib-Nezhad et al. (2013)	SB07
LiH	CLT	Coppola et al. (2011)	SB07
NaH	Rivlin	Rivlin et al. (2015)	SB07
CH	Masseron	Masseron et al. (2014)	SB07
NH	Brooke14	Brooke et al. (2014)	SB07
H_3^+	MiZATeP	Mizus et al. (2017)	$\gamma_L^0 = 0.07$, $n_L = 0.5$

Notes: partition functions are obtained from the same references (except for TiO, which uses Schwenke, 1998). Air broadening parameters are obtained from averaging HITRAN γ_{air} and n_{air} values over J^{low}. SB07 refers to the Sharp & Burrows (2007) prescription for metal oxides (their Eq.15). For H_3^+, default ExoMol broadening is used.

Table 5.4 Atomic line lists used in the POSEIDON opacity database

Atom	Line list	References	Broadening
Na	VALD3	Wiese et al. (1966); Kurucz & Peytremann (1975)	H_2 + He
" "	VALD3	Lindgård & Nielson (1977); Ralchenko et al. (2010)	H_2 + He
K	VALD3	Wiese et al. (1966); Kurucz & Peytremann (1975)	H_2 + He
" "	VALD3	Kurucz (2014)	H_2 + He
Li	VALD3	Sengupta (1975); Lindgård & Nielson (1977)	H_2 + He
" "	VALD3	Radziemski et al. (1995); Yan & Drake (1995)	H_2 + He
Rb	VALD3	Warner (1968); von der Goltz et al. (1984)	H_2 + He
Cs	VALD3	Warner (1968)	H_2 + He
Fe	VALD3	Barklem et al. (2000); Kurucz (2014)	H_2 + He
Fe^+	VALD3	Raassen & Uylings (1998); Kurucz (2014)	H_2 + He
" "	VALD3	Barklem & Aspelund-Johansson (2005)	H_2 + He
Ti	VALD3	Martin et al. (1988); Nitz et al. (1998)	H_2 + He
" "	VALD3	Barklem et al. (2000); Kurucz (2014)	H_2 + He
Ti^+	VALD3	Pickering et al. (2001); Wood et al. (2013)	H_2 + He
" "	VALD3	Kurucz (2014)	H_2 + He

Notes: VALD3 (Ryabchikova et al., 2015) provides collated compilations of atomic line lists from the original reference sources listed above. All atomic partition functions were obtained from Barklem & Collet (2016). For the Na and K resonance lines, a sub-Voigt prescription is employed (Burrows & Volobuyev, 2003; Baudino et al., 2015).

Finally, I note that these cross sections can be readily interpolated to any P-T pair within the pre-computed range. Upon initialising a model atmosphere, all cross sections are first linearly interpolated to the 100 layer pressure grid (see section 2.2.2). These cross sections are then interpolated onto a 'fine' temperature grid (typically with $\Delta T_{\text{fine}} = 10\,\text{K}$) according to the exponential model from Hill et al. (2013)

$$\sigma_{\tilde{\nu}}(T) = \sigma_{\tilde{\nu}}(T_1)\, e^{b_{\tilde{\nu}}/T_1}\, e^{b_{\tilde{\nu}}/T} \tag{5.59}$$

where

$$b_{\tilde{\nu}} = \left(\frac{1}{T_2} - \frac{1}{T_2}\right)^{-1} \ln\left(\frac{\sigma_{\tilde{\nu}}(T_1)}{\sigma_{\tilde{\nu}}(T_2)}\right) \tag{5.60}$$

Here, T_1 and T_2 are the temperatures bounding T and $\sigma_{\tilde{\nu}}(T_{1,2})$ are the (pressure-interpolated) cross sections corresponding to these temperatures. The error from this interpolation procedure is typically less than line list uncertainties (Barton et al., 2017). Storing these interpolated cross sections in memory removes the need for pressure and temperature interpolation during a retrieval, substantially decreasing total run times.

5.7 Summary

In this chapter, I have extended the range of molecular and atomic species POSEIDON can model. This lays the groundwork for atmospheric retrievals of a greater compositional diversity of potential exoplanet atmospheres. The key results from this chapter are as follows:

1. I have developed a new technique to rapidly compute molecular cross sections: the *Generalised Vectorised Voigt* (GVV) method. This removes the need to calculate Voigt profiles during line-by-line cross section computations, resulting in highly efficient and rapid calculations.

2. Using the GVV method, an extensive database of molecular opacities has been computed for a wide range of conditions relevant to exoplanet atmospheres.

3. POSEIDON now includes atomic opacities beyond Na and K, with the alkali resonance line profiles improved to account for sub-Voigt wings.

4. The resulting high-resolution ($R \sim 10^6$) molecular and atomic opacity database now allows POSEIDON to compute line-by-line radiative transfer models of exoplanet transmission spectra.

5. Rayleigh scattering cross sections have been derived for many species (beyond the previously considered H_2) – a vital input for models of high mean molecular weight atmospheres.

With this opacity database in hand, the next chapter will see POSEIDON move beyond the realm of hot Jupiters to interpret the transmission spectrum of a cooler, Neptune-mass, planet.

Chapter 6

The Metal-Rich Atmosphere of the Exo-Neptune HAT-P-26b

6.1 From hot Jupiters to exo-Neptunes

The most precise chemical abundance constraints from exoplanetary atmospheres have arisen from studies of hot Jupiters. As we have seen in previous chapters, the extended atmospheres of hot Jupiters, largely due to their high temperatures, renders their atmospheres especially viable for transmission spectroscopy. Studies of the atmospheric compositions of hot Jupiters have generally focused on reporting H_2O abundances, largely due to its prominence in many near-infrared spectra. Forward modelling and retrieval studies have derived H_2O abundances implying a range of O/H ratios, from nearly solar ($\sim 0.05\%$) (Kreidberg et al., 2014b; Line et al., 2016; Sing et al., 2016) to significantly sub-solar (Madhusudhan et al., 2014c; Barstow et al., 2017; Pinhas et al., 2019). For example, I demonstrated in Chapter 3 that the hot Jupiter HD 209458b has a sub-solar H_2O abundance (with a precise constraint of ~ 0.3 dex).

By contrast, the atmospheric composition of lower mass exoplanets, including exo-Neptunes and super-Earths, has proven challenging to constrain. In the solar system, lower mass planets possess atmospheres with an increasing fraction of heavy elements – termed the atmospheric *metallicity*. The metallicity of the solar system giants is commonly expressed in terms of the atmospheric C/H ratio, as derived from CH_4 abundances. For Jupiter and Saturn, C/H is $\sim 4\times$ solar and $\sim 10\times$ solar, respectively (Atreya et al., 2018), whilst the ice-giants Uranus and Neptune are $\sim 80\times$ solar (Karkoschka & Tomasko, 2011; Sromovsky et al., 2011). This trend is consistent with the core-accretion theory of planet formation (Pollack et al., 1996). If low mass

exo-Neptunes form in the same manner, they are anticipated to contain substantially H_2O enriched atmospheres due to accretion of water-rich planetesimals (Fortney et al., 2013). Alternatively, *in situ* formation close to the parent star, resulting in minimal contamination by planetesimals, should lead to H_2/He dominated atmospheres similar in composition to the stellar photosphere (Rogers et al., 2011). H_2O abundances thereby offer insights into the accretion history and physical properties of the original planetesimal building blocks. Measuring the composition of exo-Neptunes thus provides a powerful avenue to differentiate between planet formation mechanisms.

Until recently, constraints on exo-Neptune metallicities have proven relatively inconclusive. GJ 436b possesses a flat transmission spectrum, frustrating attempts to measure its composition. One scenario is GJ 436b's atmosphere is dominated by high-altitude clouds (Knutson et al., 2014a), with another possibility being a high metallicity due to accretion of rocky planetesimals - as suggested by its dayside emission spectrum (Madhusudhan & Seager, 2011; Moses et al., 2013b; Morley et al., 2017). The first detection of H_2O in an exo-Neptune, HAT-P-11b, was reported by Fraine et al. (2014), though the derived metallicity (1-700$\times$ solar, to 3σ) is consistent with both a nearly pure H_2/He envelope and a wide range of core accretion scenarios.

Recently, the transmission spectrum of the exo-Neptune HAT-P-26b has provided a high-significance H_2O detection (Wakeford et al., 2017). HAT-P-26b has a 4.23 day orbit, $M_p = 18.6\,M_\oplus$, $R_p = 6.33\,R_\oplus$ (Hartman et al., 2011; Wakeford et al., 2017), and $g_p = 4.47\,\mathrm{ms}^{-2}$. The combination of low gravity and high temperature ($T_{\rm eq} = 990\,\mathrm{K}$) results in an extended atmosphere ideal for transmission spectroscopy. The first observations of HAT-P-26b's transmission spectrum, using *Magellan* and *Spitzer*, were obtained by Stevenson et al. (2016). They reported evidence of H_2O, though their data could not differentiate between a high metallicity ($\sim 100\times$ solar) clear atmosphere and a solar metallicity atmosphere with a 10 mbar cloud deck. Wakeford et al. (2017) obtained additional visible and infrared observations with HST. This spectrum enabled them to report the first well-constrained metallicity for an exo-Neptune: O/H = $4.8^{+21.5}_{-4.0}\times$ solar (to 1σ) – notably smaller than the $\sim 60\times$ solar expected for a planet of this mass from core-accretion scenarios (Fortney et al., 2013).

In this chapter, I present a comprehensive atmospheric retrieval analysis, utilising all available observations, to derive the composition, metallicity, and other properties of HAT-P-26b's atmosphere. As the lowest mass exoplanet with a detected spectral feature in transit, HAT-P-26b is presently our best window into formation mechanisms of Neptune-mass exoplanets.

In what follows, I first describe the retrieval configuration, including the input observations, applied to HAT-P-26b in section 6.2. I present derived atmospheric properties for HAT-P-26b in section 6.3. Finally, in section 6.4, I discuss the implications for both HAT-P-26b itself and the wider population of exoplanetary atmospheres.

6.2 Retrieval configuration: HAT-P-26b

The retrieval configuration applied to the transmission spectrum of HAT-P-26b follows similar principles to the applications in Chapters 3 and 4. However, as this is the first application of POSEIDON to a Neptune-mass exoplanet, I will describe here salient aspects as they apply to the retrieval of exo-Neptune atmospheres.

6.2.1 Observations

The transmission spectrum of HAT-P-26b comprises 50 observations from both ground and space-based facilities. Magellan LDSS-3C observations (Stevenson et al., 2016) are complemented by HST STIS, HST WFC3, and Spitzer IRAC observations (Wakeford et al., 2017). The combined set of observations offer continuous wavelength coverage across the visible and near-infrared, from 0.5–1.6 µm, with the Spitzer IRAC channels additionally covering the regions surrounding 3.6 µm & 4.5 µm, as shown in Figure 6.1. The STIS G750L data cover 0.5–1.0 µm, with mean spectral resolution $R \approx 20$ and mean precision ~ 130 ppm. The LDSS-3C observations cover 0.7–1.0 µm, with $R \approx 70$ and ~ 155 ppm precision. The WFC3 observations come from two grisms, G102 and G141, covering 0.8–1.1 µm and 1.0–1.6 µm, respectively, with corresponding $R \approx 40$ and 60, and precisions of ~ 70 ppm and 50 ppm. The Spitzer IRAC photometric observations have precisions of ~ 130 ppm at 3.6 µm and 170 ppm at 4.5 µm.

This analysis considers these observations as a given input. The data reduction of these transit observations is discussed in detail in Stevenson et al. (2016); Wakeford et al. (2016); Sing et al. (2016); Wakeford et al. (2017). I note that Spitzer observations are available from both Stevenson et al. (2016) and Wakeford et al. (2017), but I employ only those from Wakeford et al. (2017) to ensure consistent data reduction across the space-based observations. To account for the possibility of differing normalisations between the ground-based and space-based observations (e.g. due to stellar variability or differing reduction procedures), which has not required attention in previous chapters, I allow for a relative vertical offset between these two datasets during retrieval.

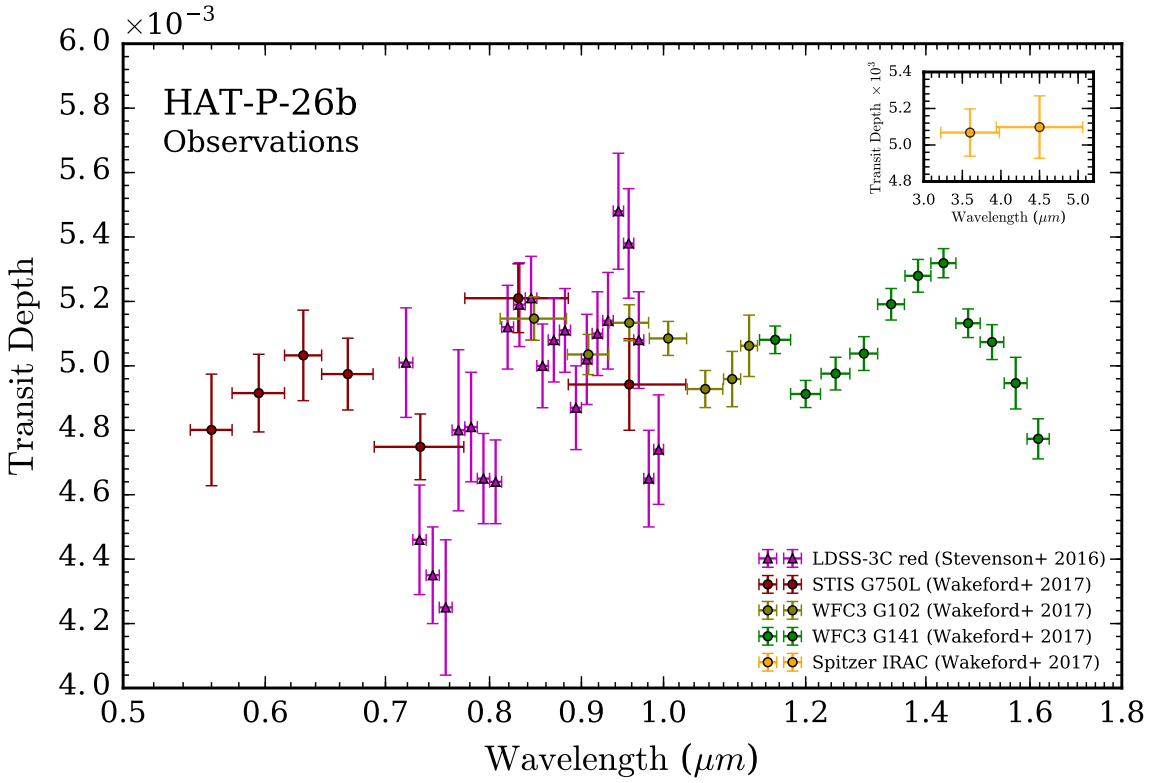


Fig. 6.1 The observed transmission spectrum of HAT-P-26b.
Each transit depth marker is shaped according to literature source – triangles for Stevenson et al. (2016), circles for Wakeford et al. (2017) – and coloured according to instrument mode. Horizontal bars indicate bin widths, vertical error bars correspond to 1σ precisions. Inset: Spitzer observations further into the infrared.

6.2.2 Model parametrisation

A wide range of potential chemical species and compositions need to be considered when modelling the atmospheres of exo-Neptunes (Madhusudhan et al., 2016b). On the one-hand, atmospheres could be H_2-He dominated with other chemical species, such as H_2O and CO, present as trace gases – as for hot Jupiters with near-solar or sub-solar metallicity. Alternatively, a wide range of high mean molecular weight atmospheres, especially H_2O or CO_2 rich compositions, are also a possibility at higher metallicities (Moses et al., 2013b).

I therefore consider many prospective sources of infrared and visible opacity. Standard carbon, oxygen, and nitrogen-bearing species with well-known infrared absorption features are included: H_2O, CH_4, NH_3, HCN, CO, CO_2, and C_2H_2. H_2 and He are treated as a single gaseous species, with a fixed solar H_2/He ratio of 0.17 assumed.

Given the suggestions of substructure in HAT-P-26b's visible-wavelength transmission spectrum (see Figure 6.1), I also consider chemical species with prominent visible cross sections. In particular, in section 4.3 we saw the importance of heavy-element molecules as sources of visible opacity. I therefore include alkali metals, metal oxides, and metal hydrides. Specifically, I consider: Na, K, Li, TiO, VO, AlO, CaO, TiH, CrH, FeH, and ScH. Absorption cross sections for these species are shown in Figure 5.3 (top left). The criterion for selecting these particular visible absorbers, out of the wide range of species included in POSEIDON's opacity database (Table 5.1), was taken to be $\sigma_{\rm vis} > 10^{-23}\,{\rm cm}^2$ at the equilibrium temperature of HAT-P-26b ($\sim 1000\,{\rm K}$) – where 'vis' indicates wavelengths covering the visible portion of the observations (~ 0.5–$1.0\,\mu{\rm m}$). This somewhat arbitrary criterion renders the dimensionality of the parameter space tractable for exploration by removing species with negligible absorption cross sections. I have, however, verified that including additional species with weaker cross sections do not modify the results presented in this chapter.

In considering this wide range of potential opacity sources, the retrievals conducted for HAT-P-26b have a maximum of 30 free parameters. 18 parameters describe the mixing ratios of the chemical species discussed above. 11 parameters are common across retrievals in previous chapters, namely 6 for the P-T profile + $P_{\rm ref}$ (section 2.2.2) and 4 for 2D clouds/hazes (section 2.2.4). A vertical relative offset, $\delta_{\rm rel}$, is also allowed between the Stevenson et al. (2016) and Wakeford et al. (2017) observations (as discussed in section 6.2.1). Model transmission spectra are computed at a constant resolution of $R = 2000$ from 0.4–5.2 µm for each sampled point in the parameter space.

The priors utilised for HAT-P-26b are summarised in Table 6.1. Compared to previous chapters, I have made four changes. First, $\alpha_{1,2}$ have an upper bound of $2\,{\rm K}^{-1/2}$ to better allow extremely isothermal P-T profiles to be recovered. Secondly, the upper atmosphere limit has been moved from 10^{-6} bar to 10^{-7} bar, ensuring that strong lines (e.g. Na) are not saturated at the $R = 2000$ model resolution. All lower limits for P-T profile pressure parameters reflect this change. Thirdly, I have re-parameterised T_0 (i.e. the temperature at the highest altitude) into a new parameter called $T_{1\,{\rm mbar}}$. This serves as a more intuitive proxy for the temperature at a typical pressure level probed by the observations. Finally, mixing ratios upper limits have been set to $10^{-0.3}$ ($\approx 50\%$), with the remainder assumed to comprise H_2 and He in solar proportions (any samples with $\sum_i X_i > 1$ are rejected to avoid negative H_2 and He abundances). Additionally, relative offsets of up to $\pm 1000\,{\rm ppm}$ are considered. All other priors are justified as in previous chapters (e.g. section 2.3.3).

Table 6.1 Atmospheric retrieval priors: HAT-P-26b

Parameter	Prior	Range
P-T profile		
$\alpha_{1,2}$	Uniform	0.02 – 2.0 $\mathrm{K}^{-1/2}$
$P_{1,2}$	Log-uniform	10^{-7} – 10^{2} bar
P_3	Log-uniform	10^{-2} – 10^{2} bar
P_{ref}	Log-uniform	10^{-7} – 10^{2} bar
$T_{1\,\mathrm{mbar}}$	Uniform	100 – 1400 K
Composition		
X_i	Log-uniform	10^{-14} - $10^{-0.3}$
Clouds		
a	Log-uniform	10^{-4} – 10^{8}
γ	Uniform	-20 – 2
P_{cloud}	Log-uniform	10^{-6} – 10^{2} bar
$\bar{\phi}$	Uniform	0 – 1
Other		
δ_{rel}	Uniform	-1000 – 1000 ppm

Notes: All samples with $\sum_i X_i > 1$ are rejected as unphysical.

6.2.3 Retrieval strategy

I conducted over 50 atmospheric retrievals in the course of this investigation. For reporting parameter constraints, I utilise the complete model including all 30 parameters described above. This ensures that any degeneracies, such as between clouds and composition, are captured via posterior marginalisation. As a guiding principle, I also employ nested Bayesian model comparisons to identify the simplest model compatible with the current observations. This enables direct assessment of model complexity by evaluating the statistical significance of each model component (e.g. detection significances of chemical species, clouds, etc). In total, $\gtrsim$ 250 million spectra were generated during this study (amounting to $\sim$ 2000 CPU hours), making this the most extensive atmospheric retrieval analysis for any exoplanet conducted to date.

6.3 The atmosphere of HAT-P-26b

I now present the inferred atmospheric properties of HAT-P-26b. I begin in section 6.3.1 with a brief discussion on the overall quality of fit between the model spectra and the observations. I then describe the retrieved atmospheric composition in section 6.3.2, constraints on the temperature structure in section 6.3.3, and cloud property constraints in section 6.3.4. Finally, in section 6.3.5, I identify the minimal model complexity necessary to explain the present observations.

6.3.1 Retrieved transmission spectrum

Figure 6.2 shows the retrieved transmission spectrum of HAT-P-26b. This corresponds to the 'full' 30-dimensional model described in section 6.2.2, including a wide range of chemical species, non-isothermal P-T profiles, clouds and hazes, and a relative offset between the Stevenson et al. (2016) and Wakeford et al. (2017) datasets. The best-fitting model spectrum, binned to the resolution of the observations, lies within 1σ for 66% of the observations (33 out of 50, yielding $\chi^2_r = 3.55$). Much of the spectral structure in the infrared is explained by H_2O, as concluded by Wakeford et al. (2017). The attribution of spectral features to the substructure at visible wavelengths, well-fit by our models, to specific chemical species is detailed in section 6.3.2. For completeness, the full posterior distribution from this retrieval is also shown in Figure 6.3.

Before deriving atmospheric properties, I first assessed if a relative offset, $\delta_{\rm rel}$, was necessary. An identical retrieval was therefore ran with the raw data alone, with a Bayesian model comparison conducted between this model and the full model. The result was a Bayes factor of 7.86 in favour of the full model (i.e. weak evidence at 2.6σ significance). Hence I include $\delta_{\rm rel}$ as a parameter in all subsequent retrievals. The probable reason for the retrieved offset, $\delta_{\rm rel} = 116 \pm 35$ ppm, is the likelihood penalty induced by the cluster of LDSS-3C observations around 0.72 µm (and to a lesser extent around 0.99 µm), which are much lower than other data points and are otherwise difficult to explain. For example, the data around 0.72 µm are far lower than would be expected of pure H_2 Rayleigh scattering without any additional sources of opacity. Stevenson et al. (2016) has previously identified these points as outliers, which can potentially be attributed to fringing (Wakeford et al., 2017). Nevertheless, the retrieved atmospheric parameters are consistent even without using a relative offset.

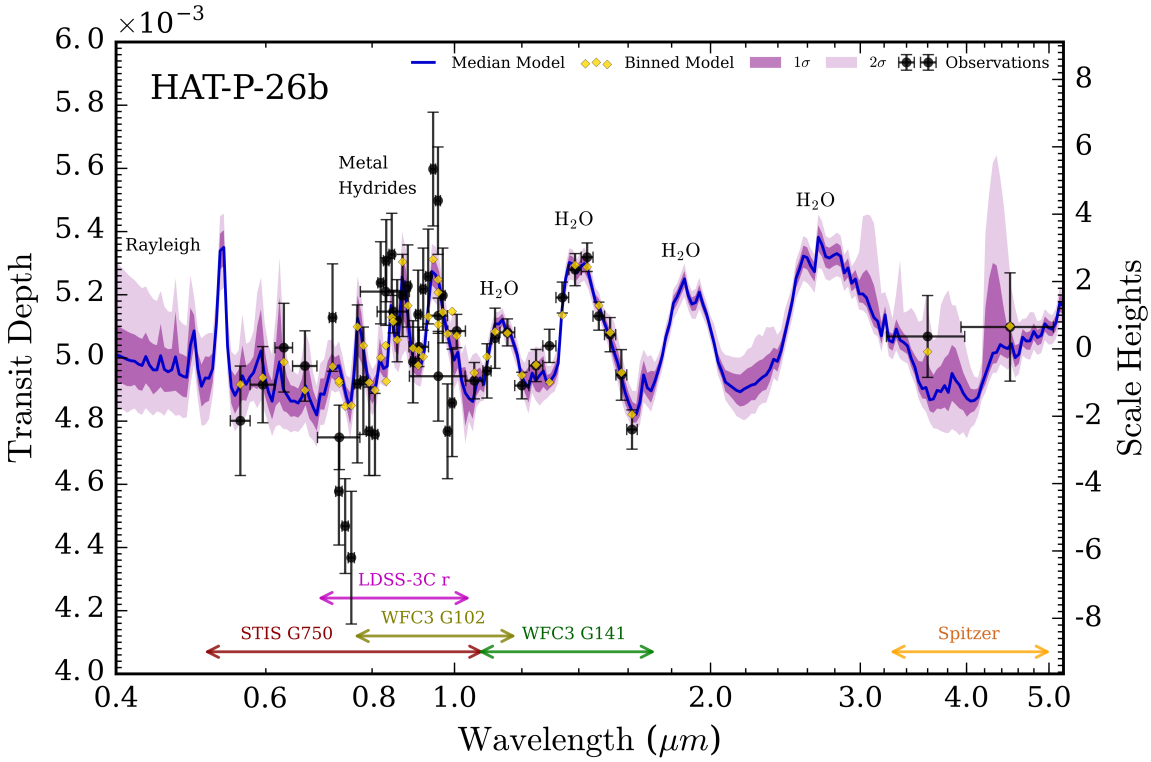
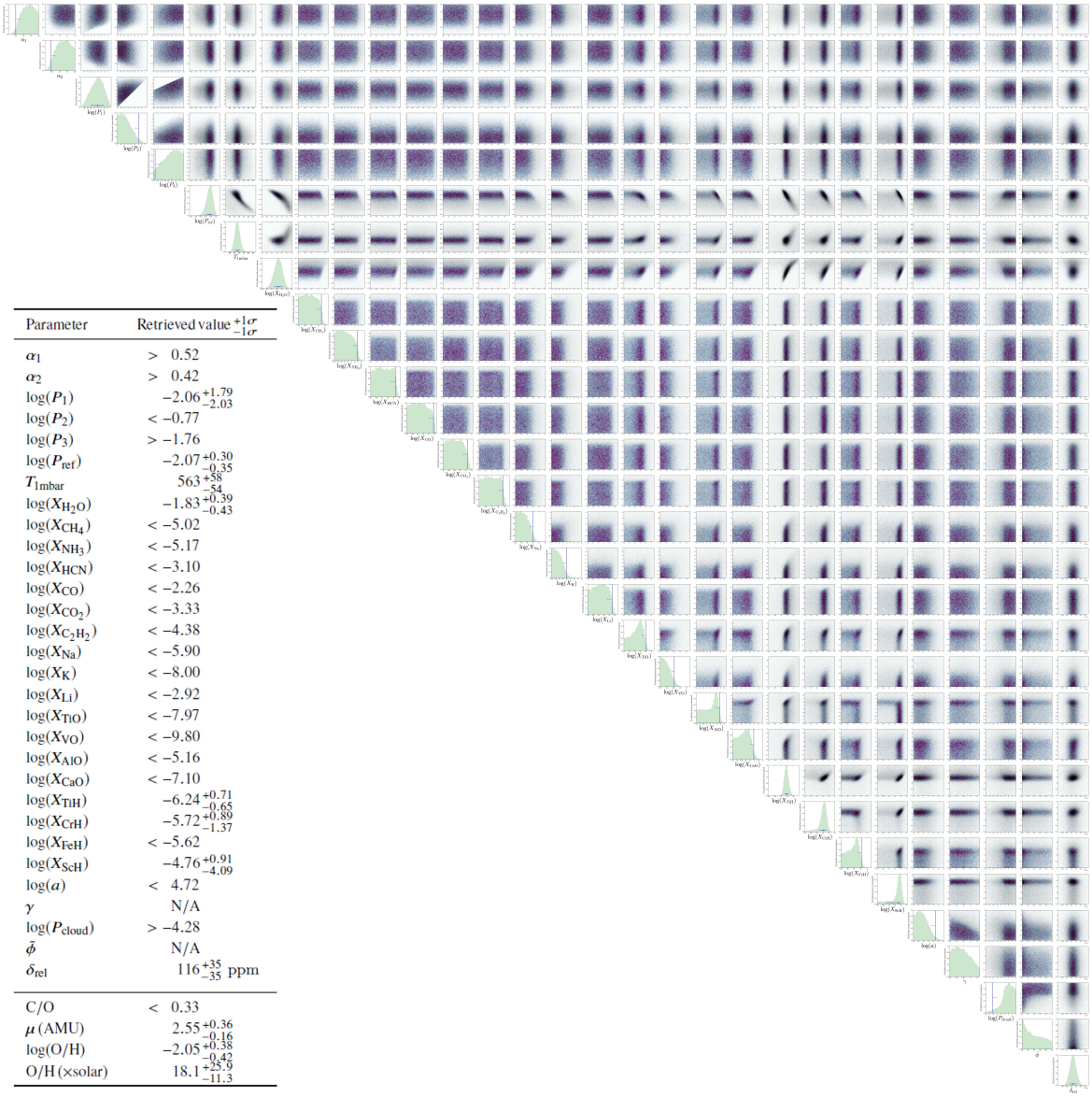


Fig. 6.2 The retrieved transmission spectrum of exo-Neptune HAT-P-26b. The observed transit depths are shown by black circles, with corresponding instrument modes and spectral ranges shown. A relative offset of +116 ppm has been applied to the LDSS-3C red observations (based on the median retrieved $\delta_{\rm rel}$). The median model spectrum obtained by POSEIDON is shown in blue (binned to $R = 100$ for clarity), along with 1σ and 2σ confidence regions in purple, derived from 30,000 random posterior samples. The best-fitting model, binned to the resolution of the data, is shown by gold diamonds. Prominent absorption and scattering features are labelled.

6.3.2 Chemical composition

I shall now describe the atmospheric composition of HAT-P-26b obtained from the retrieval analysis. The compositional results are presented in four stages. First, I establish which chemical species are indicated by the observations, along with their associated statistical significances. Secondly, I examine the specific features in the retrieved spectrum attributed to each indicated species. Thirdly, I report constraints on the abundances of each chemical species. Finally, I present constraints on derived atmospheric properties, such as the metallicity and C/O ratio.



Parameter	Retrieved value $^{+1\sigma}_{-1\sigma}$
α_1	> 0.52
α_2	> 0.42
$\log(P_1)$	$-2.06^{+1.79}_{-2.03}$
$\log(P_2)$	< -0.77
$\log(P_3)$	> -1.76
$\log(P_{\rm ref})$	$-2.07^{+0.30}_{-0.35}$
$T_{\rm 1mbar}$	563^{+58}_{-54}
$\log(X_{\rm H_2O})$	$-1.83^{+0.39}_{-0.43}$
$\log(X_{\rm CH_4})$	< -5.02
$\log(X_{\rm NH_3})$	< -5.17
$\log(X_{\rm HCN})$	< -3.10
$\log(X_{\rm CO})$	< -2.26
$\log(X_{\rm CO_2})$	< -3.33
$\log(X_{\rm C_2H_2})$	< -4.38
$\log(X_{\rm Na})$	< -5.90
$\log(X_{\rm K})$	< -8.00
$\log(X_{\rm Li})$	< -2.92
$\log(X_{\rm TiO})$	< -7.97
$\log(X_{\rm VO})$	< -9.80
$\log(X_{\rm AlO})$	< -5.16
$\log(X_{\rm CaO})$	< -7.10
$\log(X_{\rm TiH})$	$-6.24^{+0.71}_{-0.65}$
$\log(X_{\rm CrH})$	$-5.72^{+0.89}_{-1.37}$
$\log(X_{\rm FeH})$	< -5.62
$\log(X_{\rm ScH})$	$-4.76^{+0.91}_{-4.09}$
$\log(a)$	< 4.72
γ	N/A
$\log(P_{\rm cloud})$	> -4.28
$\bar{\phi}$	N/A
$\delta_{\rm rel}$	116^{+35}_{-35} ppm
C/O	< 0.33
μ (AMU)	$2.55^{+0.36}_{-0.16}$
log(O/H)	$-2.05^{+0.38}_{-0.42}$
O/H (×solar)	$18.1^{+25.9}_{-11.3}$

Fig. 6.3 Posterior distribution from the 'full' retrieval of HAT-P-26b. The corner plot depicts correlations between pairs of retrieved parameters and marginalised histograms for the values of each parameter extracted by POSEIDON. The median retrieved values and $\pm1\sigma$ constraints are shown on each histogram by blue error bars. The table inset summarises the statistical inferences. Where parameters have clear upper and lower bounds, the median retrieved values and $\pm1\sigma$ confidence levels are given. Otherwise, parameter constraints are expressed as 2σ upper or lower bounds. N/A indicates redundant parameters (γ due to the non-detection of haze and $\bar{\phi}$ due to the non-detection of a cloud deck). Derived atmospheric properties are shown at the bottom of the table, where C/O is given as a 2σ upper bound.

Detections

I establish at high significance that the atmosphere of HAT-P-26b is primarily H_2+He dominated, with H_2O as a secondary component. H_2O is detected at 7.2σ confidence via a nested Bayesian model comparison, in agreement with the previous detection of H_2O by Wakeford et al. (2017). Notably, H_2O is the only considered species that explains the broad $\sim 1.2\,\mu$m and $\sim 1.4\,\mu$m absorption features present in the observations (see Figure 6.2). Following the confirmation of H_2O, I directly computed the significance of H_2+He as the background gas by considering a model with no H_2+He, H_2O as the dominant gas, and all other species as trace gases. This model poorly fit the observations – primarily due to the resulting high mean molecular weight of ~ 18 atomic mass units – enabling it to be established, at 7.6σ confidence, that H_2+He must be the dominant background gas.

The statistical significances of other chemical species were similarly established by nested model comparisons, as summarised in Table 6.2. The majority of potential trace gases have little or negative evidence supporting their presence. As a reminder, $\mathcal{B}_{ij} \lesssim 1$ indicate that a given chemical species is not necessary to explain the observations (see section 2.3.3). However, there are three exceptions: TiH ($\mathcal{B}_{ij} = 122$ / 3.6σ), CrH ($\mathcal{B}_{ij} = 3$ / 2.1σ), and ScH ($\mathcal{B}_{ij} = 2$ / 1.8σ). On the Jeffreys' scale, these correspond to moderate evidence for TiH, weak evidence for CrH, and marginal evidence for ScH. Note that these significances automatically account for the possibility of overlapping spectral features, as the nested model comparisons fully map the parameter space without each considered species to explore alternative explanations.

The positive evidence supporting three metal hydrides motivated additional tests. I first sought to establish whether the retrievals indicated metal hydrides uniquely, or a combination of metal hydrides and metal oxides. This was motivated by the presence of 'L-shaped' degeneracies in the posterior between some metal oxides and hydrides (e.g. between ScH and AlO, as seen in the lower right of Figure 6.3). I therefore ran a retrieval with all four metal oxides (TiO, VO, AlO, and CaO) removed, and another with all four metal hydrides (TiH, CrH, FeH, and ScH) removed. By removing metal oxides, the Bayesian evidence improved ($\ln(\mathcal{Z}_i)$: $352.26 \rightarrow 354.08$) and the minimum reduced chi-square decreased ($\chi^2_{r,\mathrm{min}}$: $3.55 \rightarrow 3.04$). However, upon removing metal hydrides the Bayesian evidence suffered a substantial penalty ($\ln(\mathcal{Z}_i)$: $352.26 \rightarrow 345.66$) and the minimum reduced chi-square increased ($\chi^2_{r,\mathrm{min}}$: $3.55 \rightarrow 3.79$), both due to the worse fit at visible wavelengths. Taken together, these indicate no evidence for metal oxides, but strong evidence ($\mathcal{B}_{ij} = 722$ / 4.1σ) in favour of metal hydrides.

Table 6.2 Model comparison: atmospheric composition of HAT-P-26b

Model	Evidence $\ln(\mathcal{Z}_i)$	Best-fit $\chi^2_{r,\,\min}$	Bayes Factor $\mathcal{B}_{0i}$	Significance of Ref.
Full Chemistry	352.26	3.55	Ref.	Ref.
No H_2+He	325.59	6.32	3.82×10^{11}	7.6σ
No H_2O	328.12	6.03	3.03×10^{10}	7.2σ
No CH_4	352.41	3.39	0.86	N/A
No NH_3	352.64	3.37	0.68	N/A
No HCN	352.45	3.44	0.82	N/A
No CO	352.34	3.39	0.92	N/A
No CO_2	352.24	3.40	1.01	N/A
No C_2H_2	352.39	3.41	0.88	N/A
No Na	352.70	3.41	0.64	N/A
No K	352.97	3.37	0.49	N/A
No Li	352.43	3.41	0.84	N/A
No TiO	352.47	3.47	0.81	N/A
No VO	352.29	3.41	0.35	N/A
No AlO	352.25	3.40	1.01	N/A
No CaO	352.59	3.44	0.71	N/A
No TiH	347.45	4.08	122	3.6σ
No CrH	351.16	3.54	3.01	2.1σ
No FeH	352.43	3.46	0.84	N/A
No ScH	351.57	3.48	2.00	1.8σ
No ScH+AlO	350.25	3.72	7.44	2.5σ
No M-Oxides	354.08	3.04	0.16	N/A
No M-Hydrides	345.66	3.79	732	4.1σ

Notes : The 'Full Chemistry' reference model includes opacity due to H_2, He, H_2O, CH_4, NH_3, HCN, CO, CO_2, C_2H_2, Na, K, Li, TiO, VO, AlO, CaO, TiH, CrH, FeH, and ScH. The 'No M-Oxides' model has TiO, VO, AlO, and CaO removed; the 'No M-Hydrides' model has TiH, CrH, FeH, and ScH removed. $\chi^2_{r,\,\min}$ is the minimum reduced chi-square ($\chi^2/(N_{\rm data} - N_{\rm params})$). The significance indicates the degree of preference for the reference model, highlighted in bold, over each alternative model. N/A indicates no (or negative) evidence ($\mathcal{B}_{\rm ij} \lesssim 1$) supporting a given chemical species.

In order to assess the robustness of these chemical inferences to the data sources used, I also computed a series of retrievals using only the consistently-reduced space-based observations (i.e. HST STIS+WFC3 and Spitzer). In this case, the statistical significances for each inferred species become: H_2+He (6.7σ), H_2O (9.2σ), CrH (2.2σ), ScH (1.8σ), and metal hydrides (2.4σ). The presence of TiH is not supported without the Magellan LDSS-3C data, indicating that the evidence for this species arises from the ground-based observations (for reasons we will see in the next section). However, CrH and ScH are still inferred using only the space-based observations. Though the combined statistical significance of metal hydrides is reduced by only using the space-based data, the independent corroboration from the ground-based data supporting the presence of a visible opacity source attributable to metal hydrides strengthens the case for this atmospheric interpretation.

Spectral signatures

Figure 6.4 demonstrates how the observations can be attributed to signatures of specific molecules. To clearly highlight spectral features, the model shown only includes opacity from species with Bayes factors exceeding unity in Table 6.2 – specifically, I show the best-fitting spectrum from the 'minimal complexity' model to be discussed in section 6.3.5. The spectrum is underlain by a continuum due to H_2 Rayleigh scattering in the visible and H_2+He CIA in the infrared. Above this continuum, the contributions of specific molecules are shown, with especially prominent features labelled.

The infrared transit depths are best explained by the combination of H_2O and ScH. Specifically, the broad absorption features around 1.15 µm and 1.4 µm in the Hubble WFC3 G141 band pass are attributed to H_2O, whilst elevated absorption in windows between H_2O features (1.06 µm, 1.25 µm, and 1.6 µm) are attributed to ScH opacity. TiH also contributes around 1.09 µm. Similarly, in the Spitzer band passes H_2O features occur around 3.0 µm and 4.5 µm and ScH contributes around 3.9 µm. The low transit depth of the 4.5 µm Spitzer IRAC point is the main reason for non-detections of CO or CO_2. Despite ScH contributing to the overall spectral fit, its unique statistical significance is low (1.8σ). This is partly due to a degeneracy with AlO, which has many overlapping features in the near-infrared with ScH (not shown), providing an alternative explanation for excess absorption between H_2O features. I verified this by running a retrieval with neither ScH nor AlO present, finding an increased significance of 2.5σ for the combination (Table 6.2). Note that the identification of these signatures is despite fully accounting for the possibility of clouds (as discussed further in section 6.3.4).

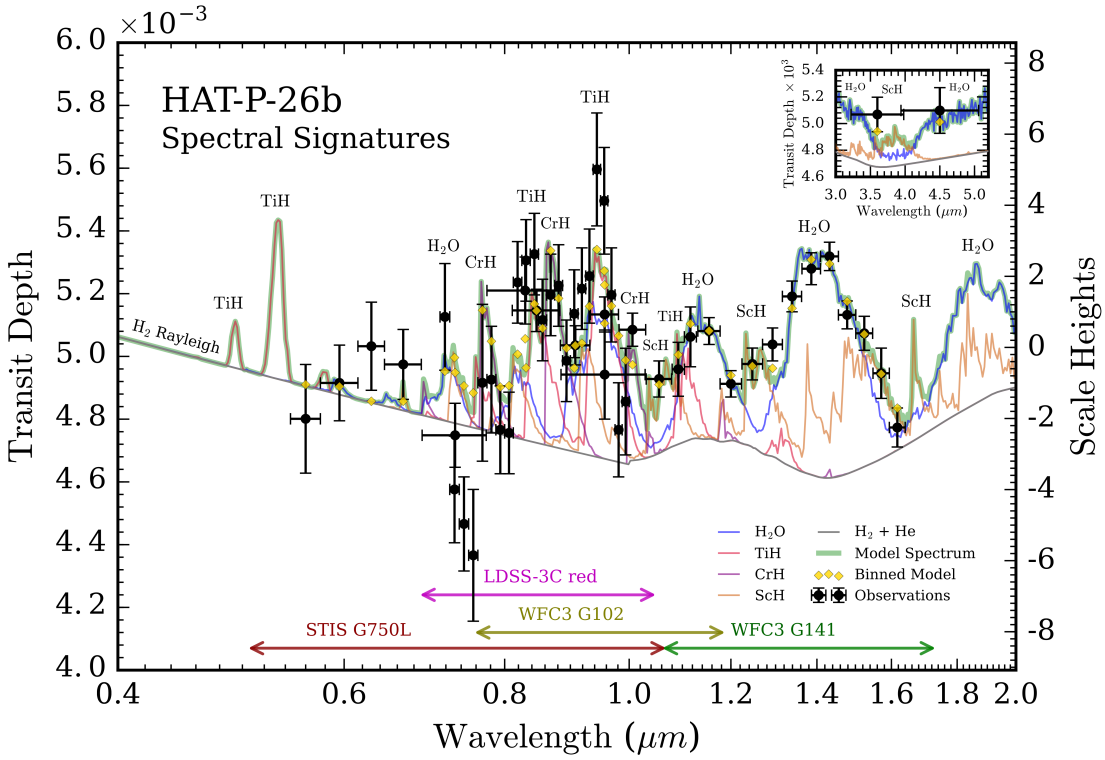


Fig. 6.4 Chemical signatures in the transmission spectrum of HAT-P-26b. The green shading indicates the best-fitting spectrum, binned to $R = 300$, from the 'minimal complexity' model (see section 6.3.5). The grey line shows the spectral continuum from H_2 and He alone. Other lines display contributions of each retrieved chemical species above the H_2+He continuum. Prominent absorption and scattering features are labelled. The observations are shown by black error bars. A relative offset of +125ppm has been applied to the LDSS-3C data (based on the median retrieved $\delta_{\rm rel}$). The best-fitting model, binned to the data resolution, is shown by gold diamonds. Inset: the spectral region surrounding Spitzer band passes.

The optical transit depths yield the primary evidence for metal hydrides. 15 data points from $\sim$ 0.84–1.0 µm are elevated by at least 1σ, and in some cases 3σ, above what can be explained by H_2+He and H_2O alone (Figure 6.4, blue line). I note that the LDSS-3C data (Stevenson et al., 2016) and WFC3 G102/G141 data (Wakeford et al., 2017) find good concordance over this spectral region, contributing to the $> 4\sigma$ combined evidence for metal hydrides. Prominent features due to TiH arise around 0.84 µm and 0.96 µm, while CrH has features around 0.78 µm, 0.88 µm, and 1.02 µm. The inference of TiH is mostly due to elevated LDSS-3C data, though the HST data is also consistent with TiH around 0.84 µm. This best-fitting spectrum also offers a testable prediction for the presence of TiH: a search for absorption from the strong TiH feature at 0.54 µm (as could be seen by future HST STIS G430 observations).

Abundance constraints

Constraints on the abundances of each constituent of HAT-P-26b's atmosphere are presented in Figure 6.5. I show only the posterior distributions for species whose presence was supported by the model comparison in Table 6.2. These histograms are the result of marginalising over a wide range of potential atmospheric scenarios, including: degeneracies due to overlapping spectral features, clouds, reference-pressure normalisation, and non-isothermal temperature structures. This ensures that the quoted constraints are as conservative as the model assumptions allow. For non-detected species, upper bounds may still be placed on their abundances through ruling out values above which spectral features not apparent in the observations would have manifested (see Figure 6.3, table inset, for these 2σ upper bounds).

The derived bounded abundance constraints are as follows. The most well-constrained mixing ratios are those of H_2+He and H_2O, correspondingly: $X_{H_2+He} = 0.985^{+0.010}_{-0.022}$ and $\log(X_{H_2O}) = -1.83^{+0.39}_{-0.43}$. The latter, with a median precision of 0.41 dex, represents the most precise water abundance measurement for an exo-Neptune to date. For the inferred metal hydrides, the mixing ratio constraints are: $\log(X_{TiH}) = -6.24^{+0.71}_{-0.65}$, $\log(X_{CrH}) = -5.72^{+0.89}_{-1.37}$, and $\log(X_{ScH}) = -4.76^{+0.91}_{-4.09}$. I note that the tail to lower abundances for CrH is caused by the data points around 0.8 µm having sufficiently tall 1σ uncertainties so as to encompass both the H_2+He continuum and the possible contribution of CrH (see Figure 6.4). Similarly, the long tail in the ScH abundance posterior manifests due to a degeneracy with AlO absorption, as discussed previously. These tails are reflected in the low marginalised significances for these two species (2.1σ and 1.8σ for CrH and ScH, respectively).

The solar photospheric abundances of O, Ti, Cr, and Sc are log(M/H) = -3.31, -7.05, -6.36, -8.85, respectively (Asplund et al., 2009). We thus see the H_2O abundance is significantly super-solar (3σ lower limit of $\log X_{H_2O} = -3.00$). Similarly, the TiH abundance is $\sim 6\times$ above the solar Ti abundance (consistent to 2σ). CrH is consistent with the solar Cr abundance. The mode of the ScH posterior is enhanced by $\sim 10^4$ over the solar Sc abundance (though the tail remains consistent within 1σ). However, due to the low significance of ScH, I advise caution in interpreting this value. The plausibility of these inferred metal hydrides will be discussed in detail in section 6.4.3.

I briefly note that there has been considerable discussion in the literature regarding the ability to obtain precise molecular abundance constraints from transmission spectra (e.g. Benneke & Seager, 2012; Griffith, 2014; Heng & Kitzmann, 2017). Concerns have been raised that retrievals of low-resolution near-infrared transmission spectra can

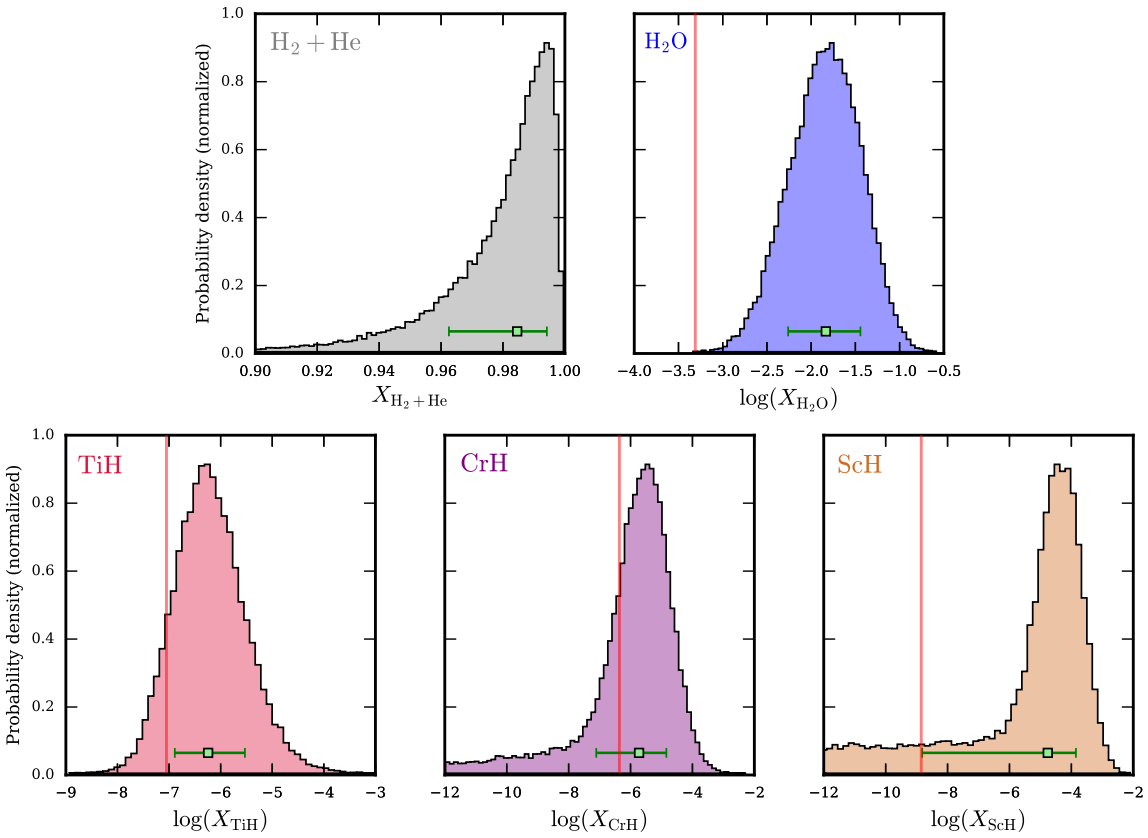


Fig. 6.5 Retrieved chemical abundances at HAT-P-26b's terminator. The histograms show the posteriors for the volume mixing ratios of each chemical species identified by the model comparison in Table 6.2. H_2+He is detected at 7.6σ and H_2O is detected at 7.2σ. The trace gases TiH, CrH, and ScH are suggested at 3.6σ, 2.1σ, and 1.8σ, respectively. The metal hydride combination has a significance of 4.1σ. The abundance of the bulk component, H_2+He, is given in a linear scale, whilst the secondary component (H_2O) and trace gases are given on a log-scale. The green error bars give the median derived abundances and $\pm 1\sigma$ confidence levels. Vertical red lines indicate the solar O, Ti, Cr, and Sc abundances (Asplund et al., 2009).

suffer from a degeneracy between R_p, $P_{\rm ref}$, and X_i. This is not an issue for the present retrievals. The long spectral baseline from the visible to the infrared allows simultaneous constraints on X_i and $P_{\rm ref}$, achieved via measurements of molecular absorption features and an independent continuum. In particular, H_2 Rayleigh scattering in the visible and H_2-H_2 CIA in the near-infrared provide independent continua in their respective spectral ranges (de Wit & Seager, 2013) (see Figure 6.4). The corresponding posteriors (e.g. Figure 6.3) hence show only a weak correlation between each X_i and $P_{\rm ref}$. This degeneracy therefore does not impede the ability to obtain high-precision abundance measurements for HAT-P-26b's atmosphere.

Metallicity, mean molecular weight, and C/O

I now present constraints on derived bulk properties of HAT-P-26b's atmosphere. From the posterior distribution of each chemical species, statistical estimates of the atmospheric metallicity, C/O ratio, and mean molecular weight can be obtained. Metallicity is commonly defined as the number ratio of elements with atomic number > 2 compared to that of hydrogen, relative to the solar value for the same elemental ratio (i.e. $\frac{\mathrm{M/H}}{\mathrm{M/H}_\odot}$, where 'M' is any element heavier than He). The mean molecular weight is the average mass of the chemical species comprising the atmosphere (commonly expressed in atomic mass units). The C/O elemental ratio provides an important diagnostic of the overall atmospheric composition (e.g. Madhusudhan, 2012).

The only elemental ratio that has been robustly determined for all four solar system giant planets is C/H (via the CH_4 abundance, see Atreya et al., 2018). This difficulty is due to other refractory species condensing out of the gas phase at low temperatures. C is thus taken as a proxy for 'M' in deriving solar system giant metallicities. For hot Jupiters, whose higher temperatures yield spectra often dominated by H_2O (Sing et al., 2016), the metallicity has instead been inferred via the atmospheric O/H ratio. In what follows, I similarly take the term 'metallicity' to mean $\frac{\mathrm{O/H}}{\mathrm{O/H}_\odot}$.

Previous studies have mainly taken two approaches to estimate atmospheric metallicities. Some studies treat the metallicity as a free, directly retrieved, parameter and, assuming chemical equilibrium, derive 'chemically-consistent' mixing ratios (e.g. Fraine et al., 2014; Wakeford et al., 2017). Others retrieve the H_2O mixing ratio directly from a spectrum and divide by the H_2O mixing ratio expected of a solar-composition atmosphere at the retrieved temperature under chemical equilibrium (e.g. Kreidberg et al., 2014b). The latter approach, often referred to as 'free retrieval' does not directly impose chemical equilibrium in deriving the atmospheric H_2O abundance itself, but implicitly assumes chemical equilibrium in metallicity estimation by dividing by a 'solar' mixing ratio which does assume chemical equilibrium.

Here I take a different approach, free from the assumption of chemical equilibrium. I derive the O/H posterior probability distribution by drawing samples directly from the posterior of each molecular mixing ratio. For a given sample, we have

$$\mathrm{O/H} = \frac{X_{\mathrm{H_2O}} + X_{\mathrm{CO}} + 2X_{\mathrm{CO_2}} + X_{\mathrm{TiO}} + ...}{2X_{\mathrm{H_2}} + 2X_{\mathrm{H_2O}} + 4X_{\mathrm{CH_4}} + 3X_{\mathrm{NH_3}} + X_{\mathrm{HCN}} + ...} \tag{6.1}$$

where '...' on the numerator indicates all other species containing O atoms, and '...' on the denominator indicates all other species containing H atoms. The metallicity is

then obtained by dividing this elemental ratio by the present-day solar photosphere O/H ratio ($O/H_\odot = 4.90 \times 10^{-4}$, Asplund et al. (2009)). When running a retrieval where one or more of these species is not included, their mixing ratios are set to 0. A posterior distribution for the metallicity is then constructed from the ensemble of O/H ratios derived from the samples. The C/O ratio is similarly constructed by

$$C/O = \frac{X_{CH_4} + X_{HCN} + X_{CO} + X_{CO_2}}{X_{H_2O} + X_{CO} + 2X_{CO_2} + X_{TiO} + X_{VO} + X_{AlO}} \tag{6.2}$$

though only upper limits can be placed with non-detection of carbon-bearing species. Finally, the atmospheric mean molecular weight is

$$\mu = \sum_i \frac{X_i \, m_i}{m_u} \tag{6.3}$$

where m_i is the mass of each chemical species and m_u is the atomic mass unit. For reference, hydrogen-dominated atmospheres with a solar H_2/He ratio have $\mu \approx 2.3$.

The derived posteriors constructed in this manner are shown in Figure 6.6. The median values and $\pm 1\sigma$ constraints for each property are also provided in the table inset of Figure 6.3. An important caveat merits mention: only gas phase chemistry in the observable atmosphere is reflected by these derived properties; the formation of condensates at altitudes below those probed by the transmission spectrum could lead to the inferred values differing from the well-mixed values in the deep atmosphere.

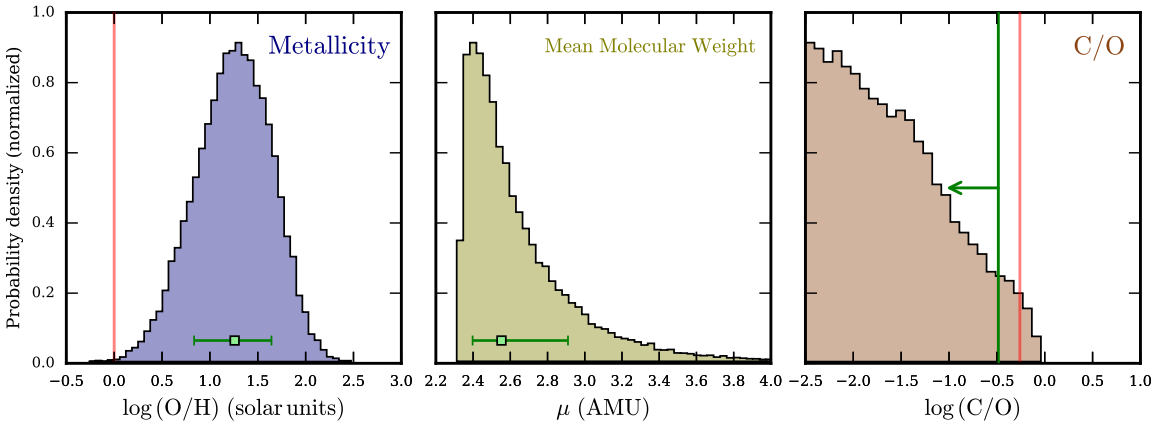


Fig. 6.6 Derived properties of HAT-P-26b's atmosphere.
The histograms show posterior distributions for the metallicity, mean molecular weight, and C/O ratio. Metallicity and C/O are given on a logarithmic scale. 'AMU' denotes atomic mass units. Median derived values and $\pm 1\sigma$ confidence levels are shown by green error bars. The vertical red lines indicate the solar metallicity and C/O.

I find the atmosphere of HAT-P-26b to be metal-enriched. The derived metallicity, $18.1^{+25.9}_{-11.3}\times$ solar, revises the previous inference obtained from space-based spectra alone ($= 4.8^{+21.5}_{-4.0}\times$ solar, Wakeford et al. (2017)). Though these metallicities are consistent to 1σ, the higher median and increased precision of the present value alters the interpretation. Specifically, whilst the previously inferred metallicity is consistent with a solar value, I rule out this possibility to $> 3\sigma$ confidence (lower limit, $1.3\times$ solar). This revised metallicity, the most precise obtained for an exo-Neptune to date, is lower than the solar system mass-metallicity expectation ($\sim 60\times$ solar), but consistent to 2σ. The implications of these findings in light of mass-metallicity trends will be discussed in section 6.4.1, along with the formation of HAT-P-26b in section 6.4.2.

Constraints on the mean molecular weight and C/O are less informative. I find an atmosphere marginally heavier than a solar H_2-He mixture with $\mu = 2.55^{+0.36}_{-0.16}$ (but consistent to 2σ). This enhanced value is due to the molecular weight contribution of the $1.5^{+2.1}_{-0.9}$% H_2O comprising the atmosphere. Non-detections of carbon-bearing species place a 2σ upper bound on the C/O ratio of 0.33, implying a sub-solar C/O (cf. 0.55). However, I caution that this conclusion is drawn primarily from the 4.5 µm Spitzer point, and future infrared observations will be required to meaningfully constrain the C/O ratio (see Chapter 7). Having presented constraints on HAT-P-26b's atmospheric composition, I now turn to the inferred temperature structure and cloud properties.

6.3.3 Temperature structure

The retrieved terminator P-T profile and 1 mbar temperature are shown in Figure 6.7. $T_{1\mathrm{mbar}}$ is constrained to 563^{+59}_{-55} K. This is cooler than the equilibrium temperature of HAT-P-26b ($\approx$ 1000 K), which reflects the transmission spectrum probing the upper altitudes of the terminator. I note that the $T_{1\mathrm{mbar}}$ posterior and the μ posterior (Figure 6.6) are consistent, as the same random sample draws are used to construct each figure. I emphasise this point, as both these variables combine to influence the local scale height ($k_B T/\mu g$) and hence the amplitudes of spectral features.

The retrieved P-T profile exhibits a temperature gradient across the photosphere. The pressure range probed by the infrared observations ($\sim 10^{-4} - 10^{-1}$ bar, see Figure 6.7, left) is approximated from the $\tau \approx 1$ pressures probed at wavelengths inside a prominent H_2O absorption feature and in the neighbouring continuum – specifically, pressures corresponding to the data points with the maximum (1.432 µm) and minimum (1.617 µm) infrared transit depths. Across the photosphere, the median P-T profile exhibits a temperature gradient of $\sim$ 80 K, though gradients as high as $\sim$ 200 K down to

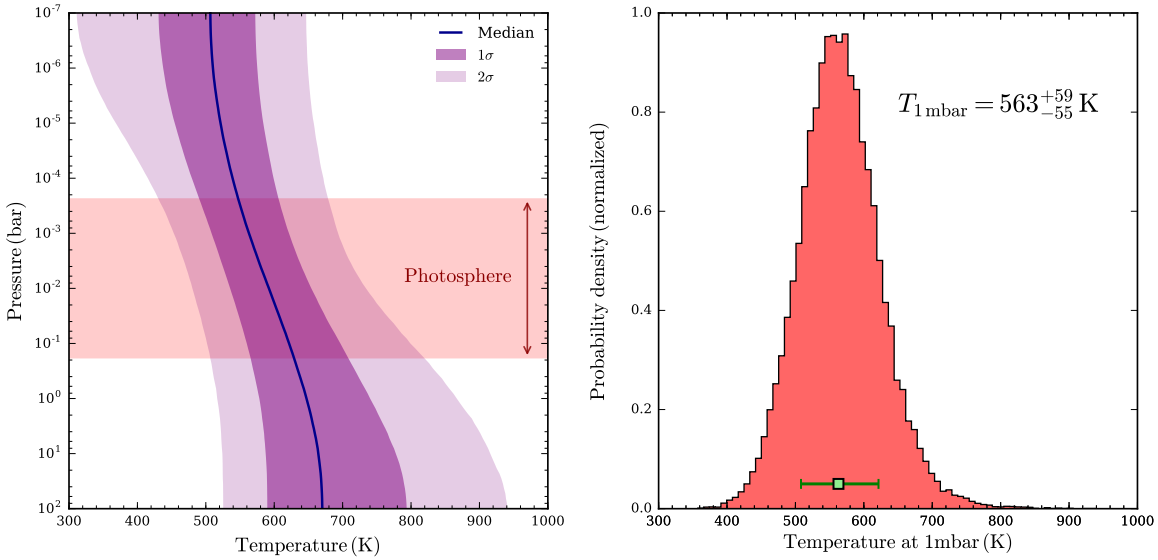


Fig. 6.7 Retrieved temperature structure at HAT-P-26b's terminator. Left: retrieved terminator P-T profile. The median profile is shown in blue, with corresponding 1σ and 2σ confidence regions in purple, derived from 30,000 random posterior samples. The $\tau \approx 1$ photosphere probed by the WFC3 observations is shaded in red. The pressures probed range from low pressures coinciding with maximal H_2O absorption (around the 1.432 µm datum), to deep regions away from H_2O absorption (around the 1.617 µm datum). Right: posterior for the temperature at 1 mbar, serving as a proxy for the typical temperature probed in the line-of-sight.

isothermal profiles are consistent with the retrieved 1σ profile. This relatively shallow gradient suggests purely isothermal models are compatible with present observations. This possibility will be assessed in section 6.3.5.

6.3.4 Clouds and hazes

The full retrieval obtained relatively loose constraints on the cloud and haze parameters. Specifically, I find a 2σ lower limit on the cloud top pressure of $\log(P_{\mathrm{cloud}}) > -4.28$, and a 2σ upper limit on the haze Rayleigh enhancement factor of $\log(a) < 4.72$. The haze slope is essentially unconstrained. The cloud fraction, $\bar{\phi}$, prefers cloud-free models. This arises from the joint posterior in the $\log(P_{\mathrm{cloud}}) - \bar{\phi}$ plane following an L-shape bordering an exclusion zone (see Figure 6.3, lower right). The exclusion zone is caused by high-altitude clouds with low cloud coverage and low-altitude clouds with high cloud coverage both producing similar spectra. If the cloud coverage fraction is forced to be uniform, the loss of this degree of freedom yields a more stringent constraint on the cloud top pressure ($\log(P_{\mathrm{cloud}}) \gtrsim -2.0$, not shown).

The retrieved cloud and haze properties can be understood as follows. The loose haze constraints arise from the lack of a slope in the visible data. Similarly, the substructure in the visible (for both the Hubble and Magellan datasets) is poorly fit by a constant transit depth, ruling out high-altitude uniform clouds. Models with high-altitude patchy clouds, which tend to broaden the wings of absorption features (e.g. Figure 2.4), are also not necessary to explain the widths of the near infrared H_2O features. I note that the non-inference of a cloud deck differs from the suggestion of clouds presented by Wakeford et al. (2017), as they did not consider visible sources of molecular opacity in their models. I have conducted additional retrievals without the ground-based observations (i.e. only using the data from Wakeford et al. (2017)) to verify that clouds and hazes are also not favoured by the space-based data alone. Taken together, these findings suggest that opacity due to clouds or hazes is not necessary to explain the present transmission spectrum of HAT-P-26b. I now turn to rigorously assess this possibility, along with the previous suggestion that the P-T profile is consistent with an isotherm, through a model complexity analysis.

6.3.5 Evaluating model complexity

I seek to establish the simplest atmospheric model capable of adequately explaining the transmission spectrum of HAT-P-26b. In the previous sections, I presented constraints on the atmospheric composition, temperature structure, and cloud properties which were fully marginalised over a wide range of potential model complexities. Here, I assess the extent to which each aspect of this complexity is warranted. As discussed in section 6.2.3, the guiding principle here is nested model comparisons. This exercise can be viewed as identifying models which maximise the Bayesian evidence (or those with a reduced chi-square closer to 1). Given that the 'full' retrieval had a reduced chi-square of 3.55 (Table 6.2), it is apparent that this 30-dimensional retrieval has many redundant parameters, which I will now endeavour to identify.

The optimal subset of chemical species is readily deduced. In section 6.3.2, I found that only H_2+He, H_2O, TiH, CrH, and ScH have Bayes factors clearly in excess of unity. A retrieval containing only this reduced set of chemical species, with all other parameters (P-T, clouds, and offset) remaining, was therefore computed. This reduced model, which I term 'P-T + Clouds', has 16 parameters, a Bayesian evidence of 356.06, and a much improved $\chi^2_{r,\mathrm{min}}$ of 2.21 – indicating the 14 removed chemical species are redundant parameters. The remaining set of molecules constitute the minimal set required to fit the observations.

Table 6.3 Model comparison: P-T structure and clouds on HAT-P-26b

Model	Evidence $\ln(\mathcal{Z}_i)$	Best-fit $\chi^2_{r,\,\min}$	Bayes Factor $\mathcal{B}_{0i}$	Significance of Ref.
P-T + Clouds	356.06	2.21	Ref.	Ref.
No Haze	356.30	2.10	0.79	N/A
Clear Skies	356.32	1.97	0.77	N/A
Iso + Clouds	356.24	1.94	Ref.	Ref.
No Haze	356.23	1.84	1.00	N/A
Clear Skies	356.22	1.75	1.02	N/A

Notes : Two reference models (bold) are considered: (i) 'P-T' models where $T = T(P)$, and (ii) 'Iso' models where $T(P) = \text{const}$. Both models include 2D clouds/hazes (section 2.2.4). All models include opacity due to H_2, He, H_2O, TiH, CrH, and ScH. $\chi^2_{r,\min}$ is the minimum reduced chi-square ($\chi^2/[N_{\text{data}} - N_{\text{params}}]$). The significance indicates the degree of preference for the reference model over each alternative model. N/A indicates no (or negative) evidence ($\mathcal{B}_{\text{ij}} \lesssim 1$) supporting hazes or clouds.

I assessed the role of temperature structure and clouds via further model comparisons. Two reference models are considered: (i) non-isothermal P-T profiles including clouds and hazes ('P-T + Clouds'), and (ii) isothermal atmospheres including clouds and hazes ('Iso + Clouds'). For each reference model, two additional retrievals were conducted: one without hazes, and one without clouds or hazes. These six retrievals range from 7-16 parameters. The model comparison is shown in Table 6.3. We see that models without clouds and hazes have a higher (or relatively unchanged) evidence and a lower $\chi^2_{r,\min}$. This reinforces the finding of section 6.3.4 that clouds and hazes minimally influence the retrieved spectra. The evidence is maximised for the non-isothermal model with a clear atmosphere, though the preference over isothermal models is marginal ($\mathcal{B}_{\text{ij}} = 1.1$ between the clear non-isothermal and clear isothermal models). The lowest $\chi^2_{r,\min}$ is for the isothermal clear model, for reasons I will now explain.

The best-fitting spectra from the 'P-T + Clouds' and 'Iso, Clear Skies' retrievals are compared in Figure 6.8. The fits are nearly indistinguishable, resulting in similar values of χ^2, despite the latter possessing 9 less free parameters. It is this reduced number of parameters that results in the isothermal clear model possessing the lowest reduced chi-square (1.75) of this model comparison. We thus see that the small temperature gradient inferred in section 6.3.3 results in only minimal differences in the fit quality when contrasted with purely isothermal models.

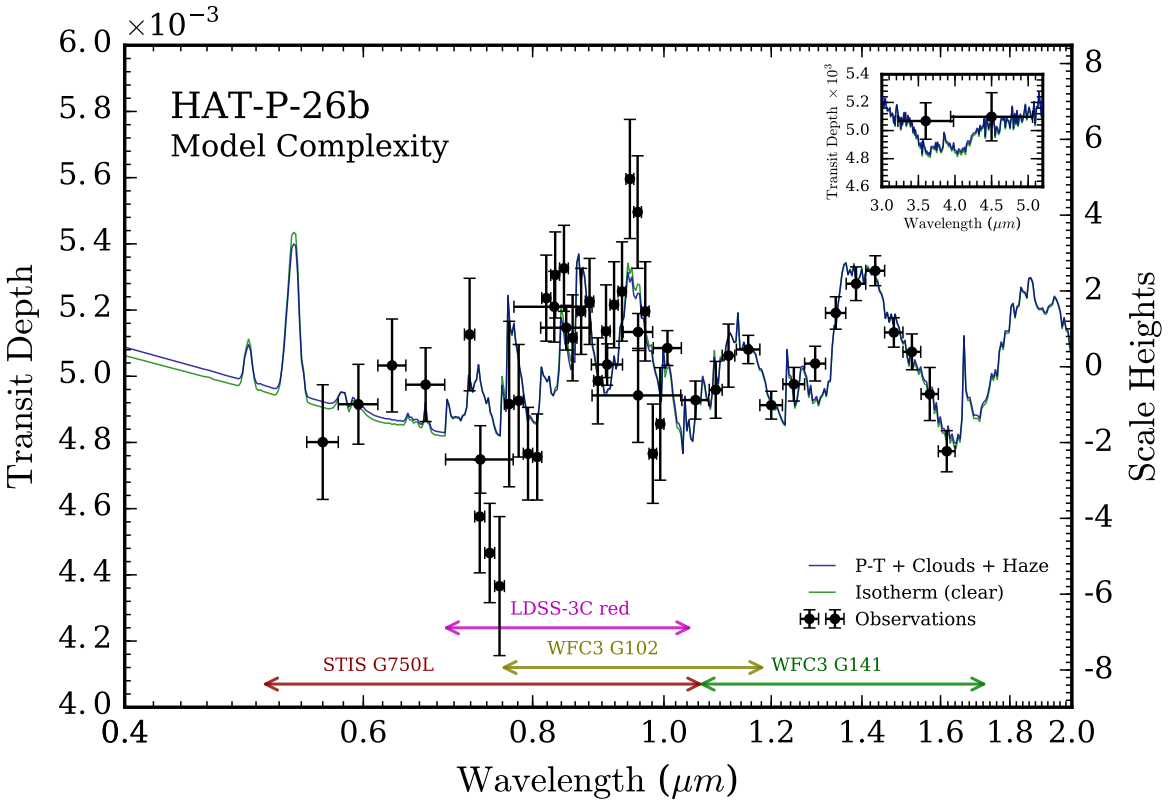
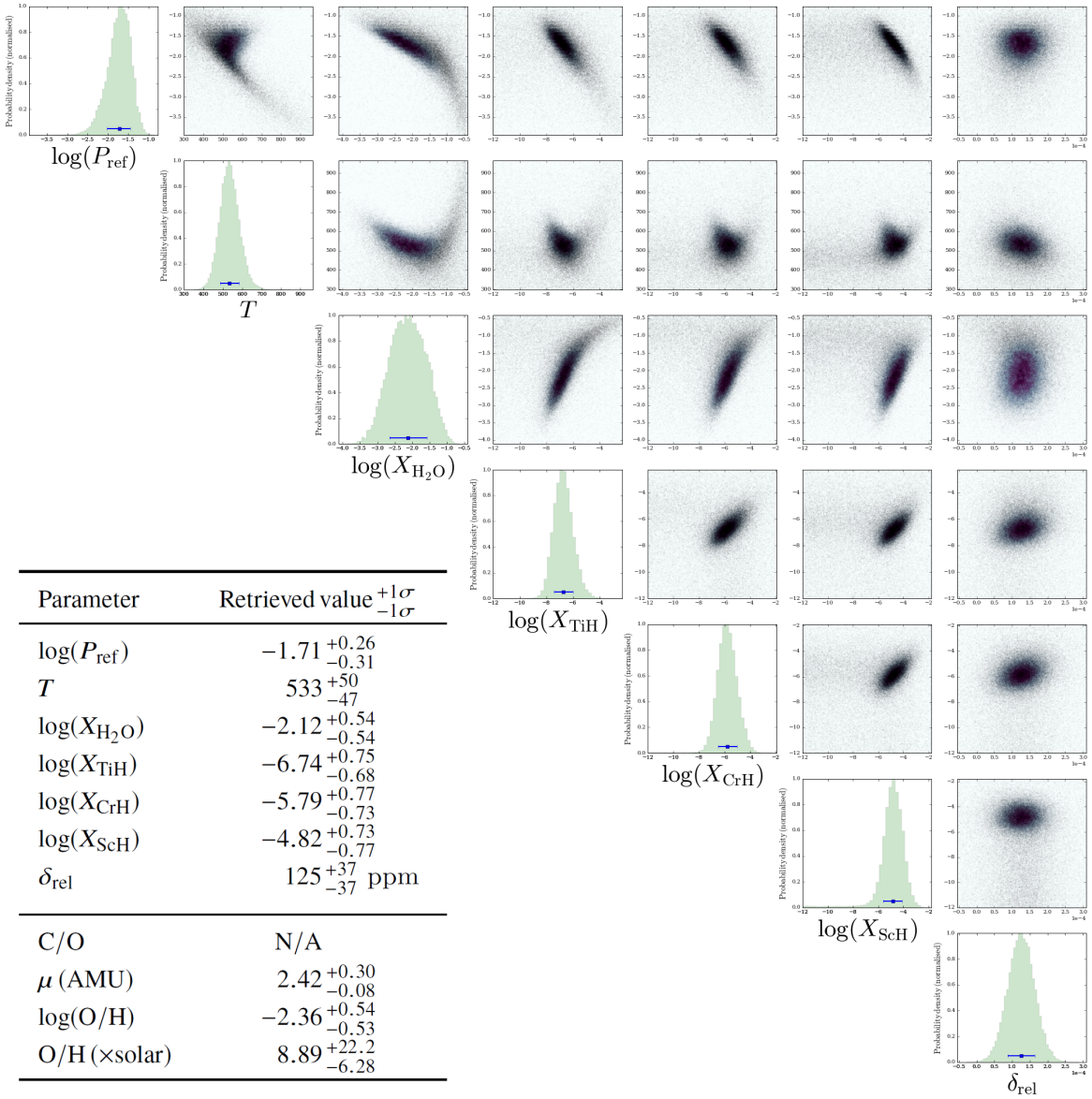


Fig. 6.8 Impact of P-T profiles & clouds on retrieved HAT-P-26b spectra. The maximum likelihood spectra from two retrieval models are shown: (i) non-isothermal P-T profiles including clouds and hazes (blue); and (ii) isothermal, clear, atmospheres (green). Both spectra are plotted at $R = 300$. The observed transit depths are shown by black circles, with corresponding instrument modes and spectral ranges shown. A relative offset of +125 ppm has been applied to the LDSS-3C red observations, based on the median retrieved $\delta_{\rm rel}$. The best-fitting spectra from the two models are nearly identical, indicating that the additional complexity of non-isothermal temperature structures and clouds / hazes are not required to explain the observations. Inset: the region surrounding the Spitzer band passes.

The posterior distribution from the isothermal, clear skies, retrieval is shown in Figure 6.9. This represents the 'minimal complexity' model that can explain the current observations. Despite having 23 less parameters than the 'full' model, the parameter inferences are consistent. Nevertheless, I caution against drawing inferences from such a simple model, due to the lack of marginalisation over other plausible atmospheric components. In totality, this analysis demonstrates that the additional complexity of non-isothermal temperature structures, clouds, and hazes are not necessary in order to explain current observations of HAT-P-26b.



Parameter	Retrieved value $^{+1\sigma}_{-1\sigma}$
$\log(P_{\rm ref})$	$-1.71^{+0.26}_{-0.31}$
T	533^{+50}_{-47}
$\log(X_{\rm H_2O})$	$-2.12^{+0.54}_{-0.54}$
$\log(X_{\rm TiH})$	$-6.74^{+0.75}_{-0.68}$
$\log(X_{\rm CrH})$	$-5.79^{+0.77}_{-0.73}$
$\log(X_{\rm ScH})$	$-4.82^{+0.73}_{-0.77}$
$\delta_{\rm rel}$	125^{+37}_{-37} ppm
C/O	N/A
μ (AMU)	$2.42^{+0.30}_{-0.08}$
log(O/H)	$-2.36^{+0.54}_{-0.53}$
O/H (×solar)	$8.89^{+22.2}_{-6.28}$

Fig. 6.9 Posterior from the 'minimal complexity' retrieval of HAT-P-26b. The corner plot depicts correlations between pairs of retrieved parameters and marginalised histograms for the values of each parameter extracted by POSEIDON. The median retrieved values and $\pm1\sigma$ constraints are shown on each histogram by blue error bars. The table inset summarises the statistical inferences. For each parameter, the median retrieved value and $\pm1\sigma$ confidence levels are given. Derived atmospheric properties are shown at the bottom of the table. C/O is given as N/A, due to it being formally zero in retrievals with no carbon-bearing species. The inferences are consistent with the 'full model' posterior given in Figure 6.3.

6.4 Implications

The retrieval analysis in this chapter has crystallised the question of HAT-P-26b's metallicity, whilst simultaneously revealing new surprises. I will now discuss the implications of these results. I first relate HAT-P-26b to the wider context of planets with well-determined metallicities in section 6.4.1. The derived composition is related to potential formation conditions in section 6.4.2. Finally, in section 6.4.3, I discuss the plausibility and potential origin of metal hydrides in the atmosphere of an exo-Neptune.

6.4.1 HAT-P-26b: an ice giant in context

The abundance constraints presented here revise the previous suggestion (Wakeford et al., 2017) of a low metallicity for HAT-P-26b. Though our values are consistent to within 1σ, the enhanced precision afforded by the inclusion of ground-based observations rules out a solar metallicity to $> 3\sigma$. There are, however, some key differences between our retrieval analyses. First, I consider possible visible and near-infrared opacity contributions from gas phase metal oxides and hydrides. The analysis of Wakeford et al. (2017) did not include these species, forcing their retrievals to include a cloud deck to provide visible opacity. This cloud deck, in turn, truncates the amplitude of the H_2O feature at 1.4 µm, reducing the quality of the fit and potentially resulting in an underestimation of the H_2O abundance. The present analysis, while including the possibility of clouds, does not find them necessary to explain the observations. Secondly, the metallicity of Wakeford et al. (2017) was derived by a model averaging procedure, whereby the inclusion of models assuming chemical equilibrium contributed disproportionately, despite not improving the overall fit, due to the lower number of free parameters. By contrast, I am able to explain the full peak-to-trough H_2O feature amplitude, without assuming chemical equilibrium (and with as few as 7 free parameters), and thereby obtain an accurate constraint on the H_2O abundance and other atmospheric properties.

Precise measurements of H_2O abundances in exo-Neptune atmospheres provide a fundamental component of ice giant compositions missing for the solar system ice giants, Uranus and Neptune. In particular, the H_2O abundance serves as a good indicator for the atmospheric oxygen abundance. At HAT-P-26b's inferred temperature (≈ 560 K), most of its oxygen is expected to be in H_2O, irrespective of the C/O ratio (Madhusudhan et al., 2016b). The new precise inference of O/H $= 18.1^{+25.9}_{-11.3}\times$ solar therefore places a strong constraint on the oxygen abundance of an ice giant atmosphere.

On the other hand, the H_2O abundance, and hence the oxygen abundance, is not known for the solar system ice giants. This is due to the low temperatures of Uranus and Neptune ($< 50\,K$), which causes all the H_2O to be sequestered in clouds sufficiently deep in the atmosphere to be inaccessible to observations (Atreya et al., 2018). Only their carbon abundances are known, via CH_4 abundance inferences, yielding C/H values of $80 \pm 20\times$ solar (Sromovsky et al., 2011; Karkoschka & Tomasko, 2011).

I compare the revised O/H ratio for HAT-P-26b to both the solar system C/H trend and O/H ratios from other exoplanets in Figure 6.10. Note that metallicities derived purely from WFC3 transmission spectra (e.g. Tsiaras et al., 2018; Fisher & Heng, 2018) are not shown here, as retrievals without visible data can be highly imprecise due to the 'normalisation degeneracy' (Heng & Kitzmann, 2017) and suffer from abundance biases (Pinhas et al., 2019; Welbanks & Madhusudhan, 2019). HAT-P-26b's metallicity lies slightly below the prediction from the solar system mass-metallicity trend, but is consistent to within 2σ. However, a degree of scatter about this trend is predicted from population synthesis models (Fortney et al., 2013). The precision of HAT-P-26b's metallicity (≈ 0.4 dex) is comparable to the most precisely determined hot Jupiter metallicities (e.g. ≈ 0.3 dex for HD 209458b in Chapter 3). It is this enhanced precision which robustly establishes the atmosphere of HAT-P-26b as metal enriched.

In contrast, most hot Jupiters appear to be consistent with either solar or sub-solar metallicities (Pinhas et al., 2019). However, one drawback of hot Jupiter metallicity estimates derived from H_2O abundances is the unknown O content in other molecules. For example, at the temperatures of hot Jupiters, around half the O is expected to reside in species other than H_2O, especially CO, for a solar C/O ratio (Madhusudhan, 2012). Many of the hot Jupiter metallicities shown in Figure 6.10 have attempted to account for this in their respective studies; either by multiplying the derived O/H by 2, or linking the H_2O abundance to a metallicity via the assumption of chemical equilibrium (Kreidberg et al., 2014b; Sedaghati et al., 2017; Pinhas et al., 2019). Such *ad hoc* corrections will not be necessary for analyses of future observations with long spectral baselines (especially with JWST, as we will see in Chapter 7), as the missing oxygen content can be directly constrained via the CO abundance. However, at the cooler temperature of HAT-P-26b the oxygen content is expected to mostly reside in H_2O (Madhusudhan et al., 2016b). The metallicity determination for HAT-P-26b presented here is therefore unique, representing the first empirically derived giant planet O abundance free from chemical equilibrium and oxygen partition assumptions.

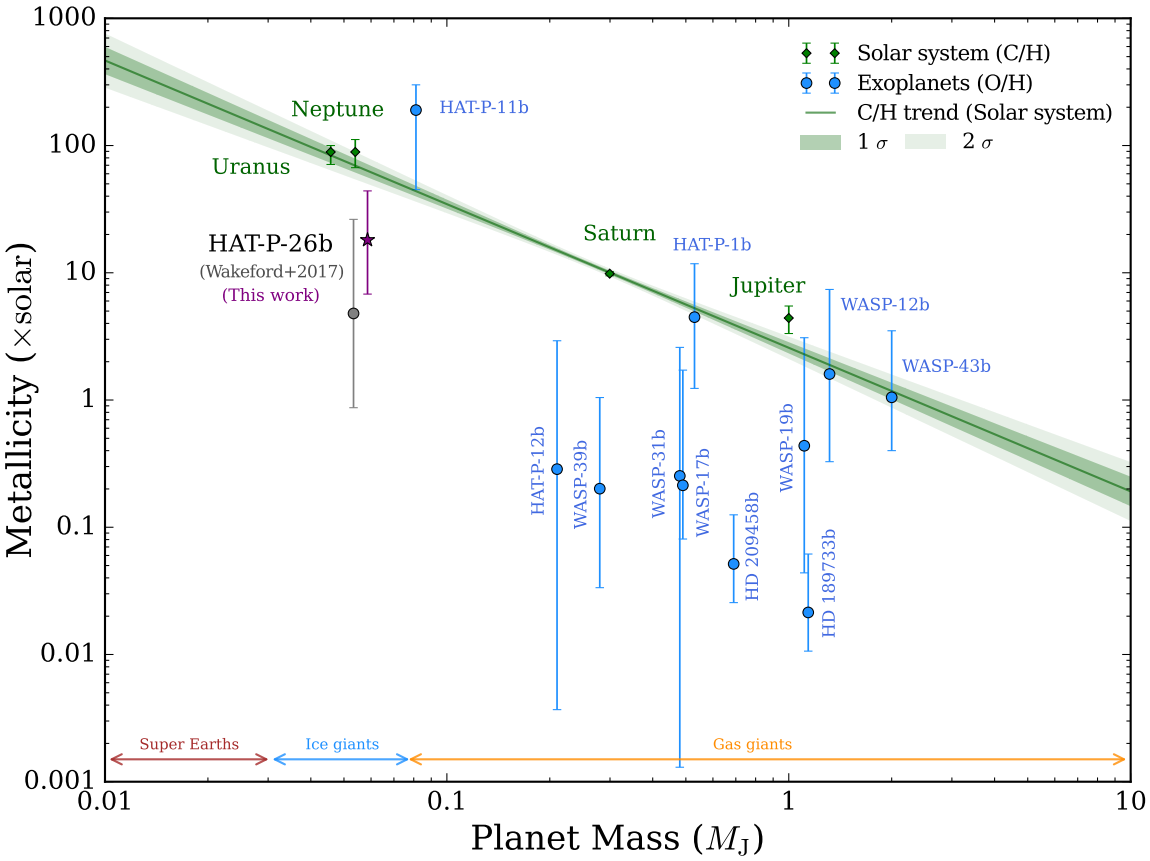


Fig. 6.10 Mass-metallicity diagram for exoplanets and solar system giants. Metallicity refers to $\frac{\mathrm{M/H}}{\mathrm{M/H}_\odot}$, where 'M' is any element heavier than He. M/H is taken as C/H for the solar system giants (determined via CH_4 abundances), and O/H for exoplanets (determined via H_2O abundances). The metallicities for Jupiter, Saturn, Neptune, and Uranus (Atreya et al., 2018; Karkoschka & Tomasko, 2011; Sromovsky et al., 2011) (green error bars), follow a linear trend in $\log(M_p) - \log(\mathrm{C/H})$. The median linear fit to the solar system metallicities is shown by the green line with corresponding 1σ and 2σ confidence regions. Exoplanets with metallicities derived from transmission spectra with detected H_2O features are over-plotted (Kreidberg et al., 2014b; Fraine et al., 2014; Sedaghati et al., 2017; Pinhas et al., 2019) (blue error bars). The previously reported metallicity of HAT-P-26b (Wakeford et al., 2017) (grey circle) is compared against the revised value from this analysis (purple star). The previous metallicity has been displaced leftwards in mass for clarity.

6.4.2 Formation conditions

The super-solar H_2O abundance I find for HAT-P-26b has important implications for understanding the formation of ice giants. Given that ice giant densities require significant solid content in the interior (see section 1.3.2), and that H_2O is the most dominant ice expected in planetesimals (Mousis et al., 2012; Johnson et al., 2012), its atmospheric abundance opens a window into the inventory of planetesimals accreted during planetary formation. In particular, close-in exo-Neptunes could arise from a number of different formation scenarios. On the one hand, they could have formed *in situ* in the inner part of the protoplanetary disk and accreted mainly primordial gas with little contamination from oxygen-rich planetesimals during their migration and growth (Rogers et al., 2011). This scenario would lead to atmospheric compositions with nearly solar O/H. On the other hand, they could have formed in the outer regions of the disk and accreted ice-rich planetesimals during their migration and grown (Madhusudhan et al., 2014c; Mordasini et al., 2016). This scenario would lead to super-solar O/H and C/H ratios for formation outside of the H_2O and CO / CO_2 snow lines, respectively, due to the accreted ices. This later scenario is thought to be the case for Uranus and Neptune due to their locations in the outer solar system (Helled & Bodenheimer, 2014).

The super-solar O/H of $18.1^{+25.9}_{-11.3}\times$ solar suggests HAT-P-26b formed beyond the H_2O snow line. The low C/O ratio (< 0.33), and hence low C/H, additionally suggests formation within the CO / CO_2 snow lines. The heavy-element content of the inferred metal hydrides could potentially have arisen from solid planetesimals dissolving into the H_2/He envelope, as discussed in the next section. Therefore, a consistent scenario is that HAT-P-26b formed between the H_2O and CO / CO_2 snow lines, accreted mainly oxygen-rich planetesimals, before migrating inwards to its present location.

6.4.3 Plausibility of inferred metal hydrides

The inference of metal hydrides at $> 4\sigma$ confidence is surprising, as gas phase metal hydrides are expected to condense at temperatures above those present at HAT-P-26b's terminator. In higher temperature atmospheres, such as L dwarfs, metal hydrides impart prominent absorption features at visible wavelengths (Kirkpatrick, 2005). Considering however the equilibrium temperature of HAT-P-26b ($T_{\rm eq} = 990$ K), the anticipated abundances of TiH and CrH are $\sim 10^{-17}$ and $\sim 10^{-9}$, respectively (assuming chemical equilibrium and solar elemental abundances, see Woitke et al., 2018) – many orders of magnitude lower than the retrieved abundances. The spectral features attributable to these species are thus at odds with equilibrium expectations.

One potential explanation is an unaccounted instrumental systematic affecting the visible transmission spectra. As I have utilised data taken from the ground (Stevenson et al., 2016) and space (Wakeford et al., 2017) at different epochs, differing data normalisations could arise due to stellar variability and/or differing reduction pipelines. Despite this, we saw in Figure 6.1 that the datasets agree relatively well in the regions of overlapping wavelength coverage. Nevertheless, I allowed a free offset between the datasets to account for this possibility. I also found that retrievals solely with the space-based observations still indicate the presence of metal hydrides (albeit at a reduced significance of 2.4σ). The agreement between ground-based and space-based observations suggests that the observed visible substructure is not of instrumental origin, and strengthens the interpretation of an atmospheric origin for these features.

The evidence for metal hydrides requires an extreme violation of local chemical equilibrium. One manner to achieve disequilibrium chemistry is vertical transport of material from the deep atmosphere, where the local conditions exceed the condensation temperature (see also the discussion in section 4.4.1). However, whilst vertical mixing can enhance some molecular abundances by $\sim$ 2-3 orders of magnitude over equilibrium expectations (Moses et al., 2011; Venot et al., 2012), such mechanisms would likely struggle to explain the retrieved metal hydride abundances (Figure 6.5).

I therefore propose an alternative scenario: secular contamination of the atmosphere by metal-rich planetesimals (e.g. Turrini et al., 2015; Pinhas et al., 2016). In this scenario, solid planetesimals rich in minerals containing heavy elements (Ti, Fe, Cr, etc.) impact the planet. The resulting high temperatures dissociate chemical bonds, enriching the atmosphere in these metals during the destructive breakup of the impactor. In the present case, high-temperature chemistry in the shocks and fireballs following an impact result in new molecules forming (Borunov et al., 1997), potentially producing transient signatures of metal hydrides.

The solar system provides many examples of this process. A well-known example is the impact of comet Shoemaker-Levy 9 into Jupiter, whereafter metallic species such as Fe and Cr were observed (Crovisier, 1995). Notable CO abundances in the stratospheres of Neptune and Saturn have also been attributed to a $\sim$ km-sized cometary impact within the last 200 years (e.g. Lellouch et al., 2005; Cavalié et al., 2010; Moses & Poppe, 2017). Finally, Mg+ ions in the upper atmosphere of Mars have recently been attributed to ablation of metallic meteorites (Crismani et al., 2017).

I examine the plausibility of this scenario via a simple calculation. The calculation is conducted for TiH and CrH, the metal hydrides with the highest significances (3.6σ and 2.1σ, respectively), with the plausibility of ScH to be discussed separately. Consider a planetesimal with a mass representative of solar system asteroids: $m_{\rm impact} \approx 10^{18}\,{\rm kg}$ (Carry, 2012). Taking this planetesimal as analogous to M-type asteroids, its composition may be estimated by noting that spectroscopic studies have suggested these objects may be the progenitors of iron meteorites (e.g. Fornasier et al., 2010). Such meteorites, in turn, tend to contain $> 90\%$ Fe by mass (Buchwald, 1977), along with minerals containing Ti (e.g. perovskite, $CaTiO_3$, ilmenite, $FeTiO_3$) and Cr (e.g. chromite, $FeCr_2O_4$) (Burbine, 2016). I estimate the Ti and Cr mass fraction of the asteroid by taking solar Ti/Fe and Cr/Fe ratios ($\sim 3 \times 10^{-3}$ and $\sim 10^{-2}$, respectively Asplund et al. (2009)). The masses contributed by a single impactor are thus $m_{\rm Ti,\,impact} \approx 3 \times 10^{15}\,{\rm kg}$ and $m_{\rm Cr,\,impact} \approx 9 \times 10^{15}\,{\rm kg}$.

Now consider the atmosphere of HAT-P-26b. I evaluate the masses of Ti and Cr present in the photosphere via $m_{\rm Ti,\,Cr,\,atm} = 4\pi \int_{r_1}^{r_2} \rho_{\rm Ti,\,Cr}\, r^2 dr$, where r_1 and r_2 are the radii where $P = 10^{-1}$ bar and $P = 10^{-4}$ bar (corresponding to the observed pressure range, see Figure 6.7). The Ti or Cr density, $\rho_{\rm Ti,Cr}$, follows from the 1σ abundance ranges (Figure 6.5). Evaluating this integral, I find $m_{\rm Ti,\,atm} \approx (1 - 30) \times 10^{14}\,{\rm kg}$ and $m_{\rm Cr,\,atm} \approx (1 - 160) \times 10^{14}\,{\rm kg}$. It is thus possible for a single such impactor to deliver sufficient Ti and Cr to be consistent with the observed TiH and CrH abundances.

However, metal hydrides are expected to eventually condense and sink into the deep atmosphere. A rough persistence timescale is given by the sedimentation timescale, $\tau_{\rm sed} \approx H/v_t$, where H is the vertical extent traversed and v_t is the terminal velocity of a condensate particle. Considering a condensate forming at $P = 10^{-2}$ bar, it must fall by $H \approx 2H_{\rm sc}$, where $H_{\rm sc}$ is the scale height, to leave the photosphere ($P \gtrsim 10^{-1}$ bar). I take $v_t = 2\beta r^2 g(\rho_c - \rho_{\rm atm})/9\eta$, assuming viscous flow (Ackerman & Marley, 2001). β is the Cunningham slip factor (here ≈ 1), r is the condensate radius, ρ_c the condensate density (e.g. $4000\,{\rm kg\ m^{-3}}$ for $CaTiO_3$), $\rho_{\rm atm}$ the atmospheric density ($\approx 5 \times 10^{-4}\,{\rm kg\ m^{-3}}$ at $P = 10^{-2}$ bar), and η the dynamic viscosity ($\approx 10^{-5}\,{\rm Pa\ s}$ for H_2 at $T \approx 560\,{\rm K}$, Ackerman & Marley (2001)). Taking a condensate radius of $r \approx 1\,\mu{\rm m}$, I estimate $v_t \approx 0.01\,{\rm ms^{-1}}$, and hence $\tau_{\rm sed} \approx 10^8\,{\rm s} \approx 4\,{\rm yr}$. If the observed signatures are to persist over time, replenishment rates of $\dot{m}_{\rm Ti,\,atm} \approx m_{\rm Ti,\,atm}/t_{\rm sed} \approx (3 - 80) \times 10^{13}\,{\rm kg\ yr^{-1}}$ and $\dot{m}_{\rm Cr,\,atm} \approx (3 - 400) \times 10^{13}\,{\rm kg\ yr^{-1}}$ would be required. Such rates could be achieved by a single impactor contributing $m_{\rm Ti,\,impact}$ and $m_{\rm Cr,\,impact}$ once every sedimentation timescale (4 yr), or by a greater impact rate of lower-mass asteroids.

The inferred ScH abundance (Figure 6.5) is likely inconsistent with a common mechanism (as the TiH and CrH abundances resemble solar Ti and Cr values, whilst ScH far exceeds the solar Sc abundance). However, the low statistical significance of ScH, partly due to a degeneracy with AlO, raises the possibility of an additional 'mystery absorber' not included in the present models. Such an absorber would possess infrared features around 1.06 µm, 1.25 µm, 1.6 µm, and 3.9 µm (see Figure 6.4). Ultimately, future observations with enhanced precision, resolution, and spectral baseline will be required to identify whether or not such a 'mystery absorber' is present.

6.5 Summary

In this chapter, I have conducted a detailed atmospheric retrieval analysis for the exo-Neptune HAT-P-26b. By simultaneously retrieving space and ground-based transmission spectra, this represents the most precise determination of atmospheric properties for an extrasolar ice-giant to date. The major results from this chapter are as follows:

1. I confirm the presence of H_2O in HAT-P-26b's atmosphere (7.2σ confidence). The H_2O abundance is precisely constrained to $1.5^{+2.1}_{-0.9}\%$ (a precision of 0.4 dex).

2. The metallicity is revised to $18.1^{+25.9}_{-11.3} \times$ solar. This is robustly super-solar ($> 3\sigma$), and consistent with the solar system mass-metallicity trend (to 2σ).

3. The visible transmission spectrum provides evidence for metal hydrides (4.1σ). Three candidate species are identified: TiH (3.6σ), CrH (2.1σ), and ScH (1.8σ). This represents the first inference of metal hydrides in an exoplanet atmosphere.

4. Non-detections of carbon-bearing species place a 2σ upper bound of C/O < 0.33.

5. The temperature at the terminator is $T_{1\,\mathrm{mbar}} = 563^{+59}_{-55}$ K, with a temperature gradient of ~ 80 K across the pressure range probed by the observations. This shallow gradient allows isothermal temperature profiles to explain current observations.

6. Clouds or hazes are not statistically supported by the present observations. Substructure in the visible and near-infrared is better explained by metal hydrides.

In the next chapter, I will demonstrate, using HAT-P-26b as a case study, how observations with the upcoming *James Webb Space Telescope* (JWST) will offer unprecedented insights into the nature of exo-Neptune atmospheres.

Chapter 7

Exploring an Alien Sky with the James Webb Space Telescope

7.1 Atmospheric retrieval in the era of JWST

In March 2021, the James Webb Space Telescope (JWST) is scheduled to launch. With a collecting area > 6 times that of Hubble, an extensive near-infrared wavelength range, superior spectral resolution, and continuous observing from an Earth-Sun L2 orbit, JWST will be a premiere observatory for exoplanet atmosphere studies. Compared to Hubble's spectral range of ~ 0.3–$1.7\,\mu$m at $R \lesssim 70$, JWST's instrument modes suitable for transiting exoplanet studies extend from ~ 0.6–$11\,\mu$m at $R \sim 100 - 3000$ (Beichman et al., 2014). A size comparison between JWST and Hubble is shown in Figure 7.1. The unprecedented sensitivity of JWST will enable detailed reconnaissance of exoplanetary atmospheres, ranging from hot Jupiters to potentially habitable terrestrial exoplanets. In recognition of this extraordinary promise, exoplanet studies have been awarded 133 hrs of Early Release Science (ERS) (Bean et al., 2018; Hinkley et al., 2018), along with $\sim 25\%$ of JWST's Cycle 1 Guaranteed Time Observations[1] (GTO).

A central theme of the earlier chapters of this thesis has been that precise long spectral baseline observations can constrain properties of exoplanet atmospheres. In Chapters 3 and 6, I demonstrated, for a hot Jupiter and exo-Neptune, that combined visible and near-infrared transmission spectra can yield precise H_2O abundances. We also saw in Chapter 4 (section 4.3) that the enhanced spectral resolution afforded by a ground-based transmission spectrum enabled a detection of heavy-element chemistry.

[1] A full list of approved JWST GTO programs is available on STScI's website

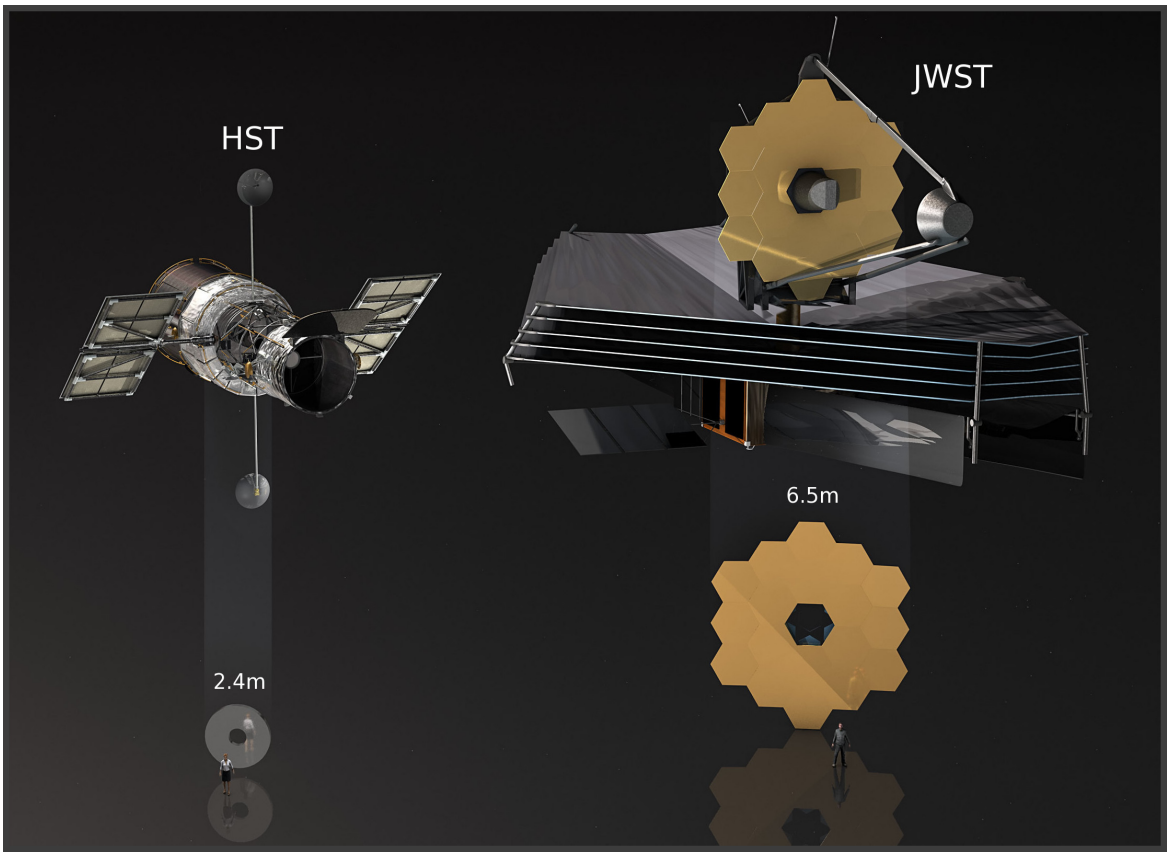


Fig. 7.1 The James Webb Space Telescope (JWST) compared to Hubble. JWST has a 6.5 m diameter segmented primary mirror, whilst Hubble's primary mirror measures 2.4 m. In terms of body dimensions, JWST's sunshield is $\approx 22\,\mathrm{m} \times 12\,\mathrm{m}$ whilst Hubble is $\approx 13.2\,\mathrm{m} \times 4.2\,\mathrm{m}$. Image credit: ESA/M. Kornmesser.

However, Chapter 4 (section 4.2) also revealed a key limitation of current observations; a lack of spectral coverage around particularly strong molecular absorption features, which prevented significant detections of NH_3 and HCN. The broad wavelength coverage promised by JWST will contribute enormously to these areas, resolve many outstanding questions, and open entirely new research directions.

In this chapter, I will demonstrate the potential for atmospheric retrieval of JWST observations to probe exoplanet atmospheres in unprecedented detail. As a case study, I consider simulated JWST transmission spectra of the exo-Neptune HAT-P-26b. This selection is made for three reasons: (i) HAT-P-26b is a guaranteed target in JWST's Cycle 1 GTO; (ii) warm exo-Neptunes ($\lesssim 1000\,\mathrm{K}$) are less favourable for transit spectra than hot Jupiters; and (iii) to compare constraints from cutting-edge observations, presented in Chapter 6, with those expected from JWST. In short, I seek to establish

the ability of JWST to characterise the atmosphere of a 'medium-difficulty' target (i.e. harder than a hot Jupiter, but less challenging than a super-Earth or terrestrial planet), comparing this to the state-of-the-art of what is currently possible.

HAT-P-26b will be observed by all four of JWST's instruments: NIRISS, NIRCam, NIRSpec, and MIRI (see Beichman et al., 2014, for a detailed overview of JWST's instruments in the context of exoplanet characterisation). Over a total of five transits, various spectroscopic modes across these instruments will produce a complete transmission spectrum from 0.6–11 µm. The goal of this chapter is to simulate these observations (for an assumed model atmosphere), retrieve the resultant synthetic transmission spectrum with POSEIDON, and compare the predicted atmospheric constraints with those found in Chapter 6. This serves to predict the constraining power of such an intensive observational campaign for a warm exo-Neptune.

A number of previous studies have employed retrievals of simulated JWST spectra to examine the precision of atmospheric constraints. Barstow et al. (2015) and Greene et al. (2016) considered transmission and emission spectra of archetypal exoplanets, ranging from hot Jupiters to super-Earths, generated synthetic JWST observations using various instrument mode combinations, and conducted retrievals to extract the atmospheric state of each planet. Greene et al. (2016) concluded that C/O and metallicity constraints of 0.2 dex and 0.5 dex, respectively, are anticipated from transmission spectra with a large spectral baseline. Information content analyses of JWST modes have similarly concluded that a wide spectral range, such as provided by NIRISS SOSS and NIRSpec G395H, have the greatest constraining power (Batalha & Line, 2017; Howe et al., 2017). One can therefore expect the 0.6–11 µm JWST transmission spectrum of HAT-P-26b to be rich in information. Most recently, Schlawin et al. (2018) considered transmission spectra retrievals of three representative JWST GTO targets, including HAT-P-26b, predicting C/O and metallicity constraints of ≈ 0.1 dex.

However, many transmission spectra retrieval studies simulating JWST observations have made a range of simplifying assumptions. The two most common assumptions are isothermal temperature profiles and chemical equilibrium, the validity of which cannot be expected in general. First, in the presence of a large temperature gradient across the pressure range probed by observations, the shapes of absorption features are altered compared to the isothermal case. This is due to absorption feature maxima probing higher altitudes (at cooler temperatures) than absorption minima at lower altitudes (at hotter temperatures), such that the changing temperature dependence of molecular cross sections distorts the overall feature shape. The assumption of an isothermal

temperature structure can therefore bias abundances retrieved from JWST spectra – by up to an order of magnitude for hot Jupiters (Rocchetto et al., 2016) – demonstrating that high-quality transmission spectra contain information on the shape of terminator temperature profiles (Barstow et al., 2013, 2015). Secondly, some retrieval analyses elect to impose chemical equilibrium (sometimes referred to as 'chemically consistent' retrieval) to retrieve C/O ratios and metallicities (e.g. Kreidberg et al., 2015; Schlawin et al., 2018). The motivation behind this approach is two-fold: (i) to reduce the number of parameters (from one parameter per retrieved species to just C/O and metallicity); and (ii) to specify uninformative priors on the C/O and metallicity parameters (see Line et al., 2013, for a discussion). The main disadvantage of this approach is its inability to explore disequilibrium chemistry, which can result from processes such as vertical mixing and photochemistry (see section 4.2). The analysis presented here relaxes these two assumptions.

In this chapter, I apply a more general retrieval approach, free from the assumptions of isothermal temperatures and chemical equilibrium, to simulated JWST data. Two unique findings are contributed: (i) the first investigation into the ability of JWST to detect and constrain metal hydride chemistry; and (ii) nested Bayesian model comparisons are utilised to predict detection significances for key molecules of interest (e.g. CH_4, CO_2). All reported constraints are marginalised over non-isothermal temperature structures, the influence of clouds, and disequilibrium chemistry. This minimal assumption approach, combined with realistic GTO observations, demonstrates the full potential of JWST to constrain exoplanetary atmospheres.

In what follows, I describe a reference atmospheric model for HAT-P-26b in section 7.2. The process of simulating JWST observations corresponding to this model, and the retrieval configuration applied to the synthetic data, are outlined in section 7.3. I present the atmospheric property constraints resulting from the retrieval analysis in section 7.4. Finally, in section 7.5, I discuss implications of these results for atmospheric retrievals in the era of JWST.

7.2 Constructing a reference exo-Neptune model

To simulate JWST observations of an exoplanet, a model atmosphere must first be created. A reference transmission spectrum is then constructed via radiative transfer through the atmosphere (as described in section 2.2.1). I first describe the selected reference model for HAT-P-26b's atmosphere in section 7.2.1. The corresponding transmission spectrum is then presented in section 7.2.2, followed by a discussion of how spectral signatures of particular molecules manifest in the fiducial spectrum.

7.2.1 Model atmosphere

The reference model atmosphere is constructed to be consistent with atmospheric inferences from current observations of HAT-P-26b (i.e. with Chapter 6) and expectations from chemical models of exo-Neptune atmospheres (see Madhusudhan et al., 2016b, for a review). For the chemical species inferred in section 6.3.2, I select mixing ratios representative of the retrieved posterior modal values (see Figure 6.5): $\log(X_{\mathrm{H_2O}}) = -1.83$, $\log(X_{\mathrm{TiH}}) = -6.2$, $\log(X_{\mathrm{CrH}}) = -5.7$, and $\log(X_{\mathrm{ScH}}) = -4.5$. For molecules only with upper limits (see Figure 6.3), I initially select representative equilibrium mixing ratios for a 900 K exo-Neptune from Moses et al. (2013b) (their Fig. 11), with a metallicity near to the derived value from Chapter 6 ($\sim 20\times$ solar, see Figure 6.6). The mixing ratios were then adjusted as necessary to ensure consistency with the derived upper limits for both the mixing ratios themselves and the C/O ratio. The resulting chosen values are: $\log(X_{\mathrm{CH_4}}) = -5.5$, $\log(X_{\mathrm{CO}}) = -2.7$, $\log(X_{\mathrm{CO_2}}) = -4.5$, $\log(X_{\mathrm{NH_3}}) = -6.0$, and $\log(X_{\mathrm{N_2}}) = -3.0$. It follows that this reference atmosphere has a metallicity of $20.1\times$ solar, C/O = 0.12, and $\mu = 2.62$.

Similarly, the temperature structure and cloud properties are chosen with previous inferences in mind. Considering that the model complexity analysis in section 6.3.5 identified a slight preference for models with non-isothermal P-T profiles, I adopt a P-T profile with $T_{1\,\mathrm{mbar}} = 560$ K and a gradient of ~ 80 K across the photosphere (consistent with the median retrieved P-T profile shown in Figure 6.7). P_{ref} is set to 10 mbar (consistent with Figures 6.3 and 6.9). As neither the retrieved posterior or model complexity analysis in Chapter 6 favoured clouds or hazes for HAT-P-26b, the reference model is chosen to have clear skies. This being said, clouds and hazes are included as parameters during the retrieval analysis, ensuring that all reported constraints include any degeneracies induced by the possibility of clouds or hazes.

7.2.2 Transmission spectrum & chemical signatures

The transmission spectrum corresponding to the reference model atmosphere is shown in Figure 7.2. Note that the model wavelength range has been extended (compared to Chapter 6), to 0.4–12 µm. The longer wavelengths beyond 5 µm are required to simulate MIRI observations. The reference spectrum is computed at $R = 10^4$, ensuring that the model spectral resolution exceeds the maximum resolving power of JWST's instrument modes ($R_{\max} \approx 3500$ for NIRSpec G395H). To allow a more direct comparison between the model spectrum and synthetic observations, binned model points, produced at $R = 100$, are also shown.

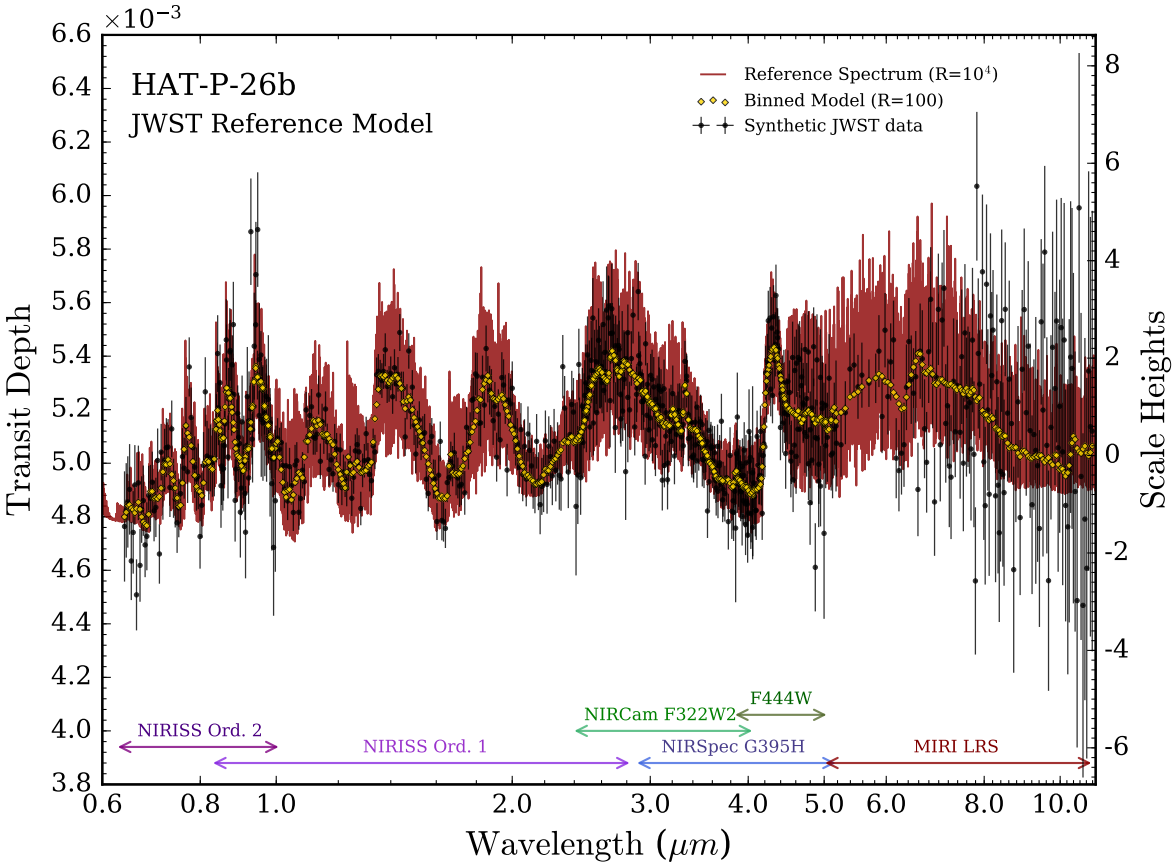


Fig. 7.2 Reference transmission spectrum of HAT-P-26b.
A reference model atmosphere, consistent with the findings in Chapter 6, results in the transmission spectrum (red). The model is computed and plotted at $R = 10^4$. Simulated JWST observations (black error bars) are generated using PandExo (Batalha et al., 2017) (see section 7.3.1). The instrument modes and spectral ranges, shown at the bottom, correspond to the GTOs for HAT-P-26b. Binned model points at the resolution of the synthetic observations ($R = 100$) are overlaid (gold diamonds).

The contributions of each molecule to the reference transmission spectrum are shown in Figure 7.3. By comparing the absorption feature amplitudes of each chemical species to typical noise levels in the synthetic observations (the creation of which will be described in section 7.3.1), one can gain qualitative intuition for which species will likely be detectable with JWST. I will compute rigorous detection significances via Bayesian model comparisons in section 7.4.3, but such an exercise is nevertheless instructive – especially given that the wavelength range of JWST observations extends beyond 5 µm, introducing new molecular signatures not discussed in previous chapters.

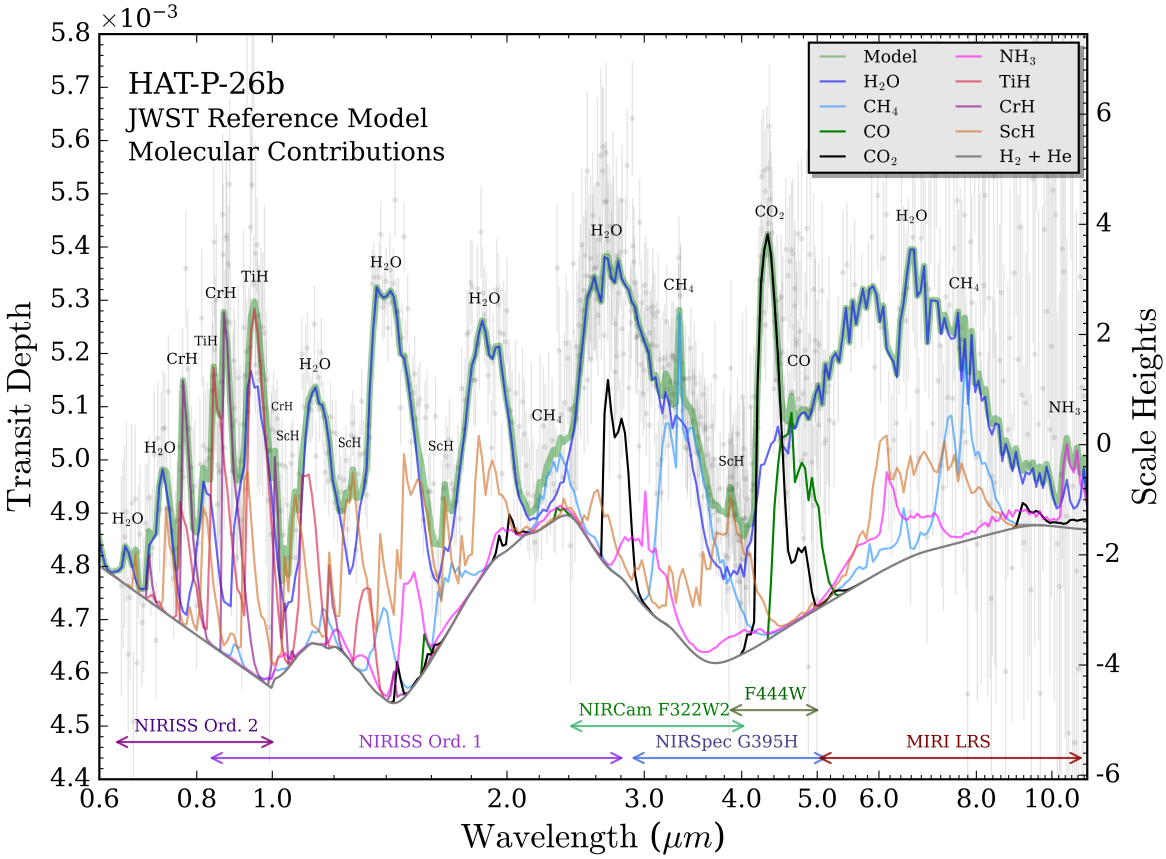


Fig. 7.3 Molecular signatures in the reference spectrum of HAT-P-26b. The green shading indicates the reference spectrum of HAT-P-26b. The grey line shows the continuum from H_2+He. Other lines shows contributions of each molecule above the H_2+He continuum. Prominent absorption features are labelled. N_2 has no notable absorption features (as it lacks an electric dipole moment), but indirectly influences the spectrum through the mean molecular weight and Rayleigh slope. All spectral contributions are plotted at $R = 100$ for comparison with the synthetic JWST data (transparent error bars) generated by PandExo (Batalha et al., 2017). The corresponding instrument modes and spectral ranges are shown at the bottom.

The infrared transmission spectrum beyond 1 µm is primarily shaped by H_2O absorption, with smaller contributions from other molecules containing carbon, oxygen, nitrogen, and hydrogen. H_2O is prominent across all instrument modes, suggesting an extremely strong detection ($> 10\sigma$). CO_2 possesses the strongest absorption feature in the spectrum, occurring around 4.3 µm. As this region is sampled by both the NIRCam F444W and NIRSpec G395H modes, I similarly expect a strong detection (e.g. > 10 simulated observations are $> 5\sigma$ deviant from the contribution of H_2O alone). CH_4 is notable around 2.3 µm and 3.4 µm, with a minor contribution around 8 µm. Even at the relatively low chosen CH_4 abundance ($10^{-5.5}$), I anticipate a detection of CH_4 to arise from combining NIRISS, NIRCam F322W2, and NIRSpec G395H observations. Detections of CO are expected to be challenging, as it only contributes around 4.6 µm. NH_3 is not expected to be detectable at the chosen abundance (10^{-6}), as its only prominent signature (around 10.5 µm) is buried in the high-noise MIRI data. However, higher NH_3 abundances than that considered here raises the possibility of NIRISS detections of NH_3 via the strong K-band feature at 2.2 µm (see section 4.2.2).

Signatures of metal hydrides largely dominate sub-micron wavelengths. Multiple sharp TiH and CrH features are probed by NIRISS SOSS, with the region from 0.84–1.0 µm, where the 1st and 2nd NIRISS orders overlap, coinciding with the strongest features of TiH and CrH. The additional short wavelength coverage of the 2nd order down to 0.64 µm further enhances the ability to assess the presence of TiH and CrH. Strong detections of these two metal hydrides are therefore expected. The utility of NIRISS to constrain the composition of exoplanetary atmospheres at $\lambda < 1$µm merits particular focus as, to my knowledge, this capability has not been studied in previous retrieval analyses of JWST spectra. However, I note that probing metal hydrides in this manner is more difficult for stars with J magnitude < 8.5 (as brighter targets saturate the SUBSTRIP256 subarray required for 2nd order observations). Absorption due to ScH, conversely, is spread more evenly over the infrared, occurring in three regions of minimal H_2O absorption observable by NIRISS, and around 4.0 µm accessible to both NIRCam filters and NIRSpec G395H. This latter signature, covered by three instrument modes, provides a powerful diagnostic for the presence (or absence) of ScH. With the key signatures of potentially observable molecules in mind, I now address the creation of synthetic observations for this spectrum.

7.3 Simulating and retrieving JWST observations

7.3.1 Generation of synthetic JWST observations

HAT-P-26b will be observed during three JWST GTO campaigns. Observations with NIRISS in the Single Object Slitless Spectroscopy (SOSS) mode and the NIRSpec G395H grism (GTO #1312[2], PI Nikole Lewis) will provide a nearly complete transmission spectrum from 0.6–5.2 µm. The NIRCam grism offers two modes, F322W2 and F444W (GTO #1185[3], PI Thomas Greene), providing additional coverage from 2.4–5.0 µm. Finally, MIRI's Low-Resolution Spectrometer (LRS) mode (GTO #1177[4], PI Thomas Greene) covers 5–11 µm. As simultaneous observations in different modes is not possible, a total of 5 transits will be observed. By combining these observations, a complete transmission spectrum from 0.6–11 µm will be obtained. These instrument modes and corresponding wavelength ranges are summarised in Table 7.1.

For each of these instrument modes, synthetic JWST observations are generated using the PandExo tool (Batalha et al., 2017). PandExo simulations require an input stellar spectrum, for which I use an interpolated model from the Phoenix Stellar Atlas (Husser et al., 2013). Specifically, I take HAT-P-26's spectrum to be specified by: $T_{\rm eff}$ = 5080 K, [Fe/H] = -0.04, $\log(g)$ = 4.56 (cgs), and normalised to J = 10.08. Given this stellar spectrum, and the reference planetary transmission spectrum from section 7.2.2, in transit and out of transit fluxes are computed by PandExo. Realistic noise levels are then calculated for each instrument mode, using exposure times and numbers of groups per integration as specified in each GTO proposal and reproduced in Table 7.1. A saturation limit of 80% full well is assumed throughout. Noise floors are set at 20 ppm, 30 ppm, 30 ppm, and 50 ppm for NIRISS, NIRCam, NIRSpec, and MIRI, respectively (Beichman et al., 2014; Greene et al., 2016). The resulting transit depths and uncertainties are binned to a common spectral resolution of R = 100, clipping wavelength ranges where falling sensitivity curves for a given mode dramatically increase transit depth uncertainties (e.g. MIRI LRS uncertainties beyond 11 µm exceed 1000 ppm). The final set of synthetic observations consists of 714 data points, with a mean precision of 120 ppm for 0.6–5.0 µm and 300 ppm beyond 5 µm. These synthetic observations can be seen in both Figures 7.2 and 7.3.

[2] https://www.stsci.edu/jwst/phase2-public/1312.pdf

[3] https://www.stsci.edu/jwst/phase2-public/1185.pdf

[4] https://www.stsci.edu/jwst/phase2-public/1177.pdf

Table 7.1 JWST Cycle 1 observations planned for HAT-P-26b

Mode	GTO #	λ (µm)	n_{groups}	t_{exposure}
NIRISS SOSS	#1312	$0.64 - 1.00$	5	6.94 hr
"	"	$0.84 - 2.81$	"	"
NIRCam F322W2	#1185	$2.41 - 4.05$	9	7.50 hr
NIRSpec G395H	#1312	$2.89 - 5.18$	37	6.94 hr
NIRCam F444W	#1185	$3.86 - 5.00$	20	7.51 hr
MIRI LRS	#1177	$5.02 - 11.00$	90	7.53 hr

Notes: NIRISS rows correspond to 2nd and 1st order spectra. Wavelength ranges are clipped where transit depth uncertainties exceed 1000 ppm. n_{groups} = number of groups per integration. Exposure times include the transit duration (2.455 hr) and out of transit baseline. The total number of transits across all modes is 5.

7.3.2 Retrieval configuration

A retrieval analysis can now be conducted on the synthetic JWST observations. The parametrisation is broadly as discussed in section 6.2.2. I prescribe free parameters to the mixing ratios of H_2O, CH_4, NH_3, CO, CO_2, AlO, TiH, CrH, and ScH. This set contains all the molecules used to generate the reference spectrum apart from N_2 (which has negligible spectral influence at the assumed abundance). AlO is included to ensure the degeneracy with ScH (see section 6.3.2) is factored into detection significances. Clouds and hazes are treated as before (see section 2.2.4). The P-T profile parameterisation has one minor change: P_{ref} is replaced by a new parameter, $R_{p,\,10\text{mbar}}$. This is because retrievals with this synthetic dataset are precise enough to encounter artefacts from the discrete pressure grid when P_{ref} is used as free parameter, which do not manifest using the continuous parameter $R_{p,\,10\text{mbar}}$. The assumed $P_{\text{ref}} = 10$ mbar for the reference model is equivalent to $R_{p,\,10\text{mbar}} = R_p = 0.544 R_J$. A uniform prior is prescribed for $R_{p,\,10\text{mbar}}$ between 80% and 120% of the true value. All other priors are as given in Table 6.1. In total, these retrievals have up to 20 parameters.

To investigate the ability of JWST to detect molecules and uncover non-isothermal P-T profiles, a total of 13 atmospheric retrievals were conducted. These consist of a full reference retrieval, 9 retrievals with one of each parameterised molecule removed in turn, a retrieval without H_2+He, a retrieval without metal hydrides, and a retrieval assuming an isothermal temperature structure. Spectra are generated from 0.6–11.0 µm on a model wavelength grid at $R = 5000$.

7.4 Characterising an exo-Neptune with JWST

With synthetic JWST observations in hand, I now demonstrate the potential of JWST to characterise an exo-Neptune atmosphere. The ability to recover the input transmission spectrum and temperature structure is discussed in sections 7.4.1 and 7.4.2, respectively. Predicted molecular detection significances and abundance constraints are presented in section 7.4.3, along with an elucidation of isothermal assumption biases. The section concludes with predictions for the ability of JWST to constrain the metallicity, mean molecular weight, and C/O of HAT-P-26b's atmosphere.

7.4.1 Retrieved transmission spectrum

Figure 7.4 compares the true reference transmission spectrum of HAT-P-26b to that retrieved by POSEIDON. It is seen that the median retrieved spectrum generally lies within 1σ of the true spectrum over the full 0.6–11 µm range. The one exception is CrH features, for which the retrieved spectrum overestimates the amplitude by $\sim 2\sigma$. This is due to an excess of elevated data points around the CrH feature at 0.88 µm for this particular noise instance. It is thus expected that the retrieved CrH abundance will be higher than the true value – a point I revisit in section 7.4.3. Despite this, the minimum reduced chi-square indicates an excellent fit: $\chi^2_{r,\mathrm{min}} = 1.04$ (for 694 degrees of freedom). Overall, the retrieval correctly identifies the molecules responsible for the main absorption features. Remarkably, there is even a hint of the NH_3 feature around 10.5 µm, though the 1σ contour remains consistent with models without NH_3 (matching the qualitative picture provided in section 7.2.2).

The posterior distribution corresponding to the retrieved transmission spectrum is shown in Figure 7.5. Most parameters are retrieved within 1σ of their true value (the one exception being the CrH abundance, as previously discussed). Molecules significantly contributing to the spectrum are well-constrained. In particular, no ScH-AlO degeneracy manifests, with the synthetic JWST data sufficient to correctly set an upper bound on AlO. The retrieval identifies two equivalent solutions to explain the clear atmosphere: (i) deep clouds with variable cloud coverage, or (ii) high-altitude clouds with zero cloud coverage (Figure 7.5, lower right). Contrasting this posterior with that from current observations (see Figure 6.3), new constraints on the P-T profile parameters are now evident (e.g. see the α_2 - $\log(P_1)$ plane). The implications of these enhanced constraints for the retrieved P-T profile is the next focus.

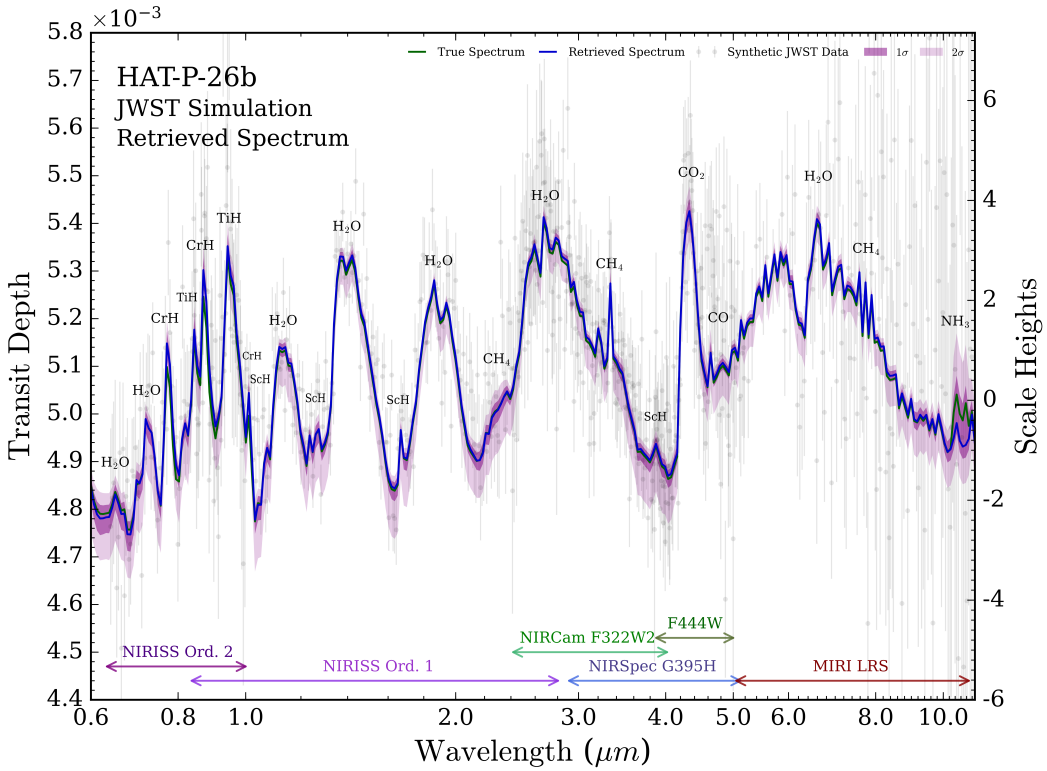
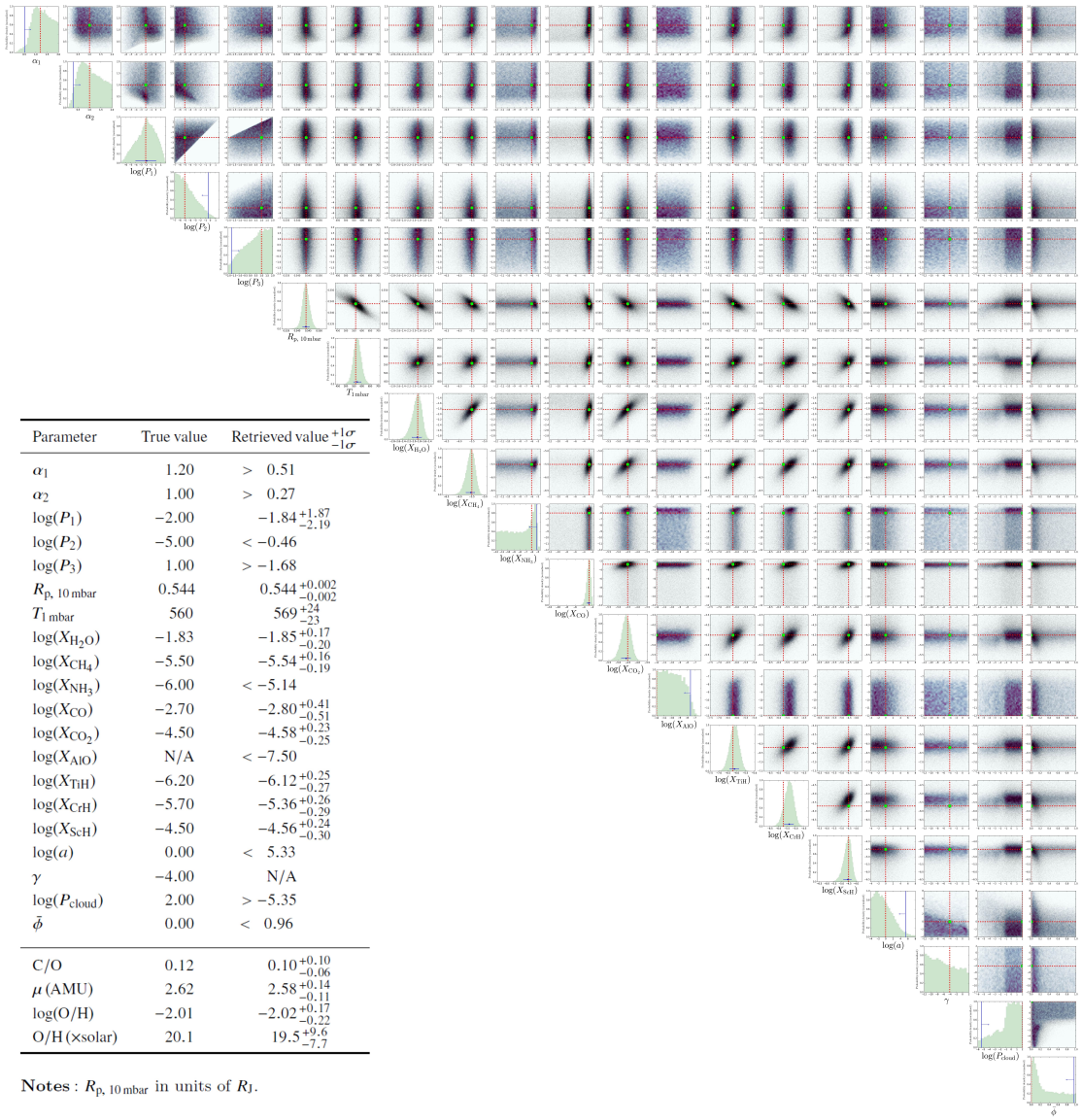


Fig. 7.4 Retrieved spectrum from simulated JWST data of HAT-P-26b. Simulated JWST observations (transparent error bars), with corresponding instrument modes and spectral ranges at the bottom, undergo a spectral retrieval by POSEIDON. The true input transmission spectrum (green) compares well with the median retrieved spectrum (blue), both plotted at the same spectral resolution as the synthetic observations ($R = 100$). 1σ and 2σ confidence regions (purple) are derived from 30,000 random posterior samples. Prominent molecular absorption features are labelled.

7.4.2 Temperature structure

I now assess the ability to retrieve terminator temperature structures from JWST observations of an exo-Neptune. Figure 7.6 compares the true P-T profile of the reference model atmosphere with profiles obtained from two retrievals: (i) the 'full' retrieval with the 6 parameter P-T profile usually used by POSEIDON (section 2.2.2); and (ii) a single parameter isothermal model with an otherwise identical atmosphere to the reference model. A Bayesian model comparison between these two models supports the interpretation favouring a non-isothermal profile, albeit marginally ($\mathcal{B}_{ij} = 1.3$). It is seen that the retrieval including a parametrised P-T profile correctly matches the true profile within 1σ over the entire atmosphere, whilst the isotherm is only consistent with the true profile from $\sim$ 2-30 mbar. The retrieved 1 mbar temperature, 569^{+24}_{-23} K,



Parameter	True value	Retrieved value $^{+1\sigma}_{-1\sigma}$
α_1	1.20	> 0.51
α_2	1.00	> 0.27
$\log(P_1)$	−2.00	$-1.84^{+1.87}_{-2.19}$
$\log(P_2)$	−5.00	< -0.46
$\log(P_3)$	1.00	> -1.68
$R_{\mathrm{p,\,10\,mbar}}$	0.544	$0.544^{+0.002}_{-0.002}$
$T_{\mathrm{1\,mbar}}$	560	569^{+24}_{-23}
$\log(X_{\mathrm{H_2O}})$	−1.83	$-1.85^{+0.17}_{-0.20}$
$\log(X_{\mathrm{CH_4}})$	−5.50	$-5.54^{+0.16}_{-0.19}$
$\log(X_{\mathrm{NH_3}})$	−6.00	< -5.14
$\log(X_{\mathrm{CO}})$	−2.70	$-2.80^{+0.41}_{-0.51}$
$\log(X_{\mathrm{CO_2}})$	−4.50	$-4.58^{+0.23}_{-0.25}$
$\log(X_{\mathrm{AlO}})$	N/A	< -7.50
$\log(X_{\mathrm{TiH}})$	−6.20	$-6.12^{+0.25}_{-0.27}$
$\log(X_{\mathrm{CrH}})$	−5.70	$-5.36^{+0.26}_{-0.29}$
$\log(X_{\mathrm{ScH}})$	−4.50	$-4.56^{+0.24}_{-0.30}$
$\log(a)$	0.00	< 5.33
γ	−4.00	N/A
$\log(P_{\mathrm{cloud}})$	2.00	> -5.35
$\tilde{\phi}$	0.00	< 0.96
C/O	0.12	$0.10^{+0.10}_{-0.06}$
μ (AMU)	2.62	$2.58^{+0.14}_{-0.11}$
log(O/H)	−2.01	$-2.02^{+0.17}_{-0.22}$
O/H (×solar)	20.1	$19.5^{+9.6}_{-7.7}$

Notes: $R_{\mathrm{p,\,10\,mbar}}$ in units of R_{J}.

Fig. 7.5 Posterior from retrieving synthetic JWST data of HAT-P-26b. The corner plot depicts correlations between pairs of retrieved parameters and marginalised histograms for each parameter extracted by POSEIDON. Red dashed lines and green squares indicate the true parameter values used to generate the data. The median retrieved values and $\pm1\sigma$ constraints are shown on each histogram by blue error bars. The table inset compares true values with retrieved inferences. Where parameters have clear upper and lower bounds, the median retrieved values and $\pm1\sigma$ confidence levels are given. Otherwise, parameter constraints are expressed as 2σ upper or lower bounds. N/A indicates redundant parameters (γ due to the unconstrained distribution). Derived atmospheric properties are shown at the bottom of the table.

agrees well with the true value of $T_{1\,\mathrm{mbar}} = 560\,\mathrm{K}$ (cf. $T = 584^{+24}_{-19}\,\mathrm{K}$ for the isothermal model). This constraint represents a 2.4 × improvement over the ∼60 K uncertainty obtained from current observations (see Figure 6.7). The temperature confidence intervals are tightest around 10 mbar – corresponding to the centre of the infrared photosphere – expanding significantly for pressures outside those directly probed by the observations ($P < 0.1$ mbar and $P > 100$ mbar). In totality, these results suggest that the photosphere is wide enough for JWST observations to be sensitive to a temperature gradient as small as ∼ 80 K. The effect of ignoring this temperature gradient on inferred chemical abundances is examined in the next section.

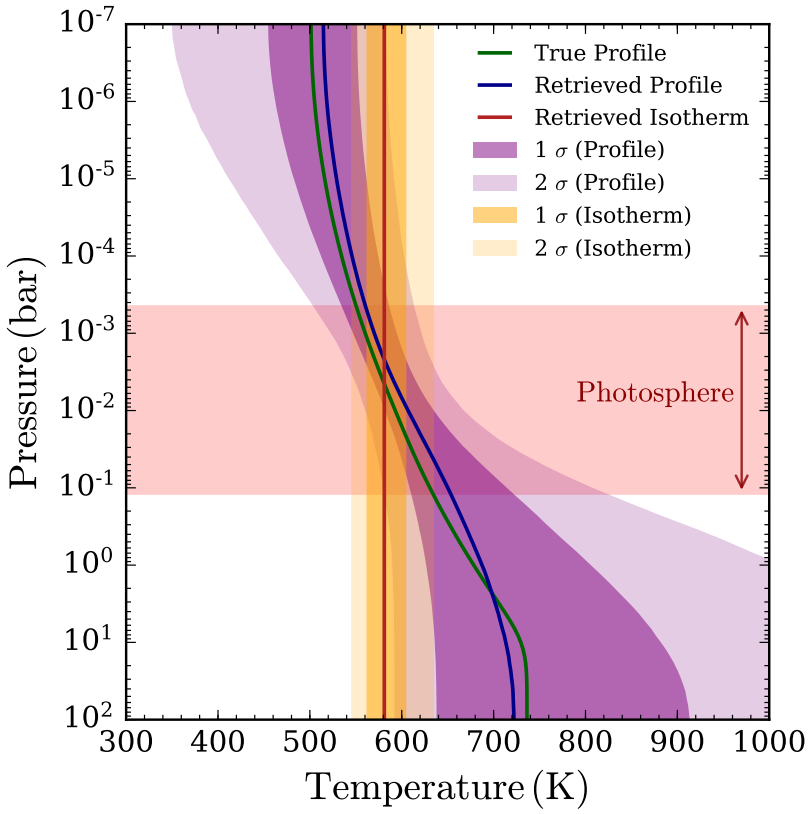


Fig. 7.6 Retrieved P-T profile from simulated JWST data of HAT-P-26b. The true input profile (green) is compared against median profiles from two independent retrievals: (i) a flexible P-T profile (blue), and (ii) an isotherm (red). Corresponding 1σ and 2σ confidence regions (purple for the P-T profile, orange for the isotherm) are derived from 30,000 random posterior samples. The isotherm agrees with the true profile across a decade in pressure around ∼ 10 mbar, whilst the P-T profile is consistent with the true profile over the entire atmosphere. The $\tau \leqslant 1$ infrared photosphere probed by the observations is shaded in red. The photosphere covers low pressures sampling absorption features to deep regions probing the continuum.

7.4.3 Chemical composition

The wavelength coverage and spectral resolution of JWST will significantly enhance constraints on the composition of exoplanetary atmospheres. This will manifest in two primary ways: (i) detections of new chemical species (along with increased significances for known species); and (ii) more stringent constraints on chemical abundances. Here I provide quantitative predictions for the performance of JWST in these two areas, considering the realistic case study of the HAT-P-26b GTO campaigns.

This analysis predicts new chemical species will be conclusively detected ($> 5\sigma$) by JWST. The model comparison arriving at detection significances for each molecular species is given in Table 7.2. From highest to lowest significance, I find detections of: CO_2 (13.1σ), TiH (8.3σ), CrH (8.0σ), CH_4 (6.2σ), and ScH (5.1σ). These detections join those of H_2+He and H_2O (already detected in section 6.3.2), at 25.8σ and 28.9σ, respectively. CO falls short of a conclusive detection (3.7σ), but would still be considered a 'strong' detection on the Jeffreys' scale (as $\mathcal{B}_{ij} > 150$). No evidence is found supporting the presence of NH_3 or AlO. The former finding reaffirms the qualitative expectation that the NH_3 abundance lies below the threshold to be extracted from the noise (see Figure 7.3). The latter finding indicates the degeneracy between ScH and AlO seen in the previous chapter should not preclude definitive identification of ScH with JWST. Finally, I find the combination of three metal hydrides inferred from current observations should be detectable at 13.7σ confidence by the planned JWST GTO campaigns. This provides a testable prediction to readily allow the metal hydride hypothesis to be critically assessed.

Chemical abundance constraints with JWST can reach precisions of < 0.3 dex for conclusively detected chemical species. Specifically, I find the following precisions: 0.72% (H_2+He), 0.19 dex (H_2O), 0.17 dex (CH_4), 0.46 dex (CO), 0.24 dex (CO_2), 0.26 dex (TiH), 0.28 dex (CrH), and 0.27 dex (ScH). This represents an improvement over current observations by a factor of > 2 for the H_2+He and H_2O abundances, and $\gtrsim 3$ for the metal hydrides. The posterior distributions for the abundances of these species are shown in Figure 7.7 – both for a retrieval assuming an isotherm and the full retrieval accounting for non-isothermal P-T profiles. Most abundances are correctly retrieved within 1σ. The lone exception is CrH, which lies just beyond 1σ of the true value (due to random scatter in the particular noise instance drawn for the synthetic NIRISS observations). I do not show NH_3 here, despite a suggestive posterior peak around the true abundance (see Figure 7.5), as the Bayesian model comparison established a non-detection. Assuming an isotherm does not overly affect abundance

Table 7.2 Predicted molecular detection significances from JWST spectra

Model	Evidence $\ln(\mathcal{Z}_i)$	Best-fit $\chi^2_{r,\,\min}$	Bayes Factor $\mathcal{B}_{0i}$	Significance of Ref.
Full Chemistry	5371.0	1.04	Ref.	Ref.
No H_2+He	5041.0	2.01	2×10^{143}	25.8σ
No H_2O	4957.3	2.17	5×10^{179}	28.9σ
No CH_4	5353.7	1.09	3.5×10^{7}	6.2σ
No NH_3	5371.2	1.03	0.80	N/A
No CO	5365.7	1.05	198	3.7σ
No CO_2	5287.4	1.27	2×10^{36}	13.1σ
No AlO	5371.8	1.03	0.40	N/A
No TiH	5338.7	1.13	1×10^{14}	8.3σ
No CrH	5341.0	1.13	1×10^{13}	8.0σ
No ScH	5360.1	1.06	6×10^{4}	5.1σ
No M-Hydrides	5280.2	1.26	3×10^{39}	13.7σ

Notes : The 'Full Chemistry' reference model includes chemical opacity due to H_2, He, H_2O, CH_4, NH_3, CO, CO_2, AlO, TiH, CrH, and ScH. The 'No M-Hydrides' model has TiH, CrH, and ScH removed. $\chi^2_{r,\,\min}$ is the minimum reduced chi-square ($\chi^2/(N_{\rm data} - N_{\rm params})$). The significance indicates the degree of preference for the reference model, highlighted in bold, over each alternative model. N/A indicates no (or negative) evidence ($\mathcal{B}_{\rm ij} \lesssim 1$) supporting a given molecule.

precisions, but does bias the abundances by $\sim$ 0.1-0.2 dex. This is sufficiently small to maintain consistency between the isothermal and non-isothermal posteriors (for the shallow temperature gradient assumed here). Note that, whilst trace abundances are biased to higher values, the bulk atmospheric component (here, H_2+He) is biased to lower abundances as mixing ratios must sum to unity.

Finally, we can examine the precision of derived atmospheric properties obtainable with JWST. In Figure 7.8, I compare the metallicity, mean molecular weight, and C/O constraints retrieved in Chapter 6 with those predicted from the simulated JWST retrievals (following the methodology of section 6.3.2 to compute these derived quantities). An improvement in metallicity and mean molecular weight determinations by a factor of $\sim$ 2 (0.4 $\rightarrow$ 0.2 dex and 0.26 $\rightarrow$ 0.13, respectively) is seen. The C/O posterior changes more dramatically, with the upper limit achieved from current observations transformed into a bounded constraint $\sim$ 0.35 dex wide (due to detections

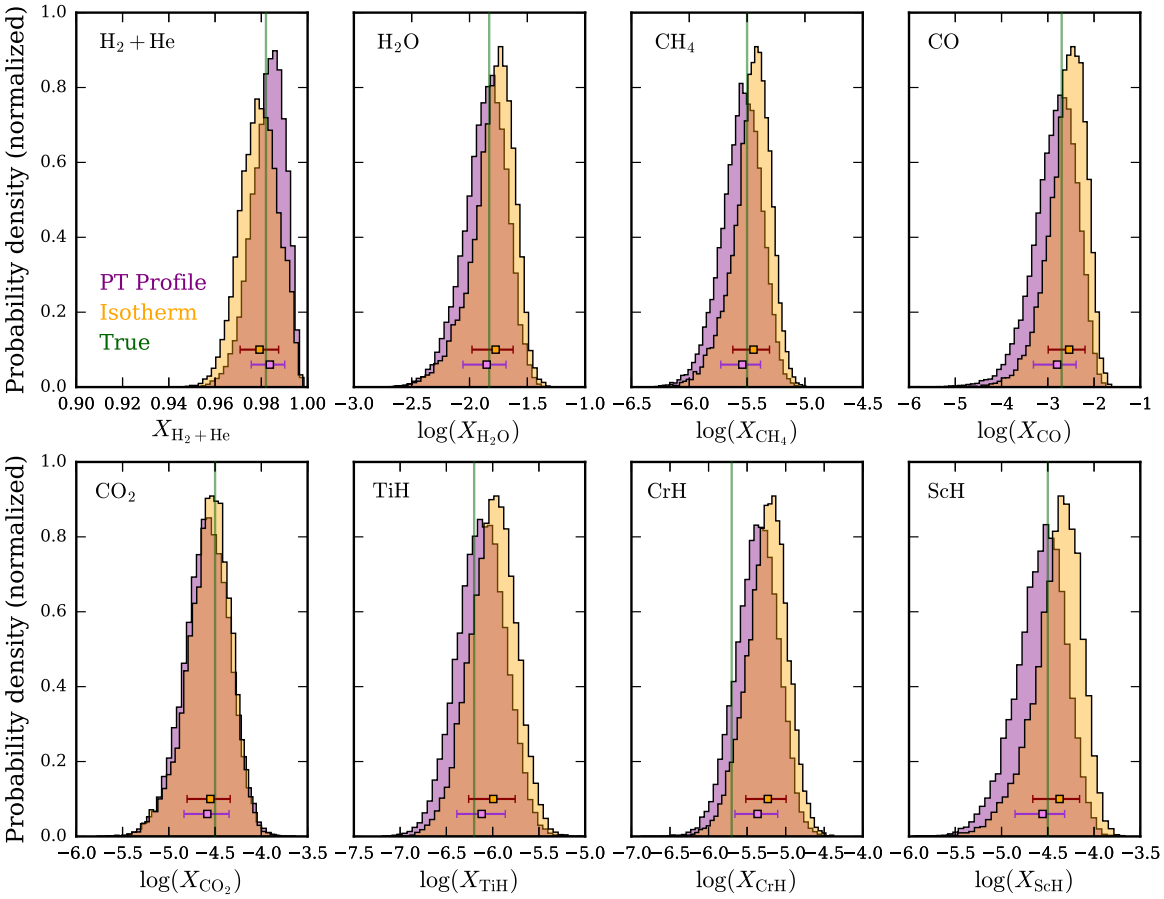


Fig. 7.7 Retrieved abundances from simulated JWST data of HAT-P-26b. Two posterior probability distributions are shown for the volume mixing ratios of each constrained chemical species: (i) a retrieval with a flexible P-T profile (purple), and (ii) a retrieval assuming an isothermal temperature structure (orange). The true value of each mixing ratio is shown by the green vertical lines. The median derived abundances and $\pm 1\sigma$ confidence levels from each retrieval are denoted by purple and orange error bars, for the P-T profile retrieval and isothermal retrieval, respectively. The abundance of the bulk component, H_2+He, is given in a linear scale, whilst the secondary component (H_2O) and trace gases are given on a log-scale.

of the carbon-bearing species CH_4, CO_2, and CO). All three derived properties match the values used to generate the synthetic observations to well within 1σ. Notably, this demonstrates that retrievals without the assumption of chemical equilibrium to guide them can produce unbiased measurements of atmospheric metallicities and C/O ratios.

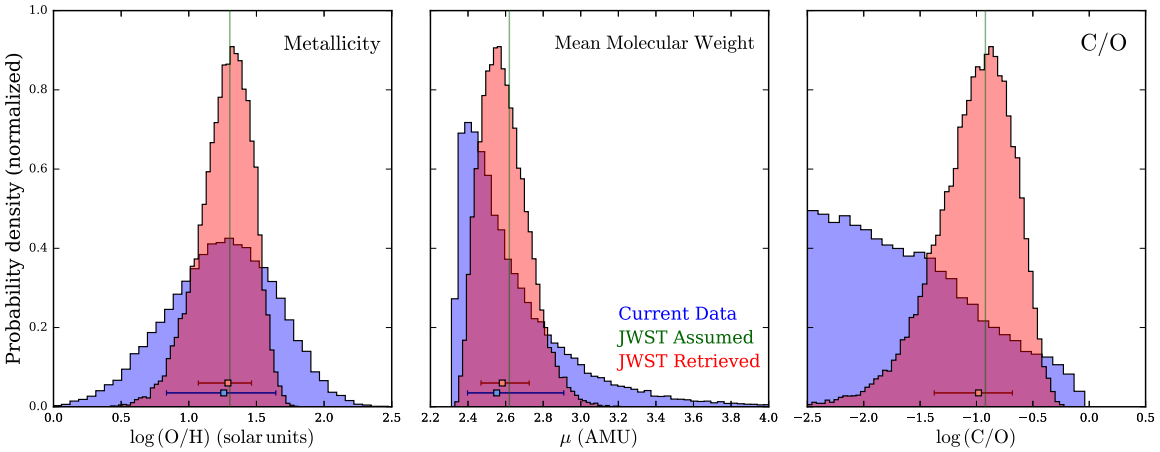


Fig. 7.8 Atmospheric properties of HAT-P-26b: current vs. JWST. Posteriors for the metallicity, mean molecular weight, and C/O ratio are shown. Inferences from current observations (blue) are compared against predicted constraints from retrieving synthetic JWST data of HAT-P-26b (red). The true values used to generate the synthetic JWST data are shown by green lines. Metallicity and C/O are given on a logarithmic scale. 'AMU' denotes atomic mass units. Median derived values and 1σ confidence levels are denoted by blue and red error bars from the current data and simulated JWST retrievals, respectively. No error bar is given for C/O for current observations, as only an upper limit can be placed (see Figure 6.6).

7.5 Implications

I will now briefly discuss implications of this JWST capability demonstration for extra-solar ice-giant characterisation. I first consider the improved compositional constraints promised by JWST in section 7.5.1. The importance of considering non-isothermal temperature structures in retrievals of JWST transmission spectra is addressed in section 7.5.2. Finally, in section 7.5.3, I examine the prospect of definitively detecting metal hydride chemistry in exoplanet atmospheres with JWST.

7.5.1 Constraining atmospheric compositions with JWST

I have demonstrated that JWST will be able to confidently detect (to $> 5\sigma$) a range of carbon-, hydrogen-, and oxygen-bearing molecules in an exo-Neptune atmosphere. Such detections, coupled with molecular abundance constraints at < 0.3 dex precision, will realise complete inventories of the main spectroscopic species contributing to O/H, C/H, and C/O ratios in transiting exoplanet atmospheres. A powerful demonstration of this is shown in Figure 7.8, whereby anticipated detections of carbon-bearing molecules

allow a bounded constraint on the C/O ratio of HAT-P-26b. Even for presently well-constrained abundances (e.g. the H_2O abundance), the additional wavelength coverage afforded by JWST's observing modes result in improved constraints by a factor of $\gtrsim 2$. In the case of HAT-P-26b's metallicity, for which I showed in Chapter 6 lies $> 3\sigma$ above solar metallicity, JWST will enable one to establish, to high confidence, whether or not it lies slightly below the solar-system mass-metallicity trend (see Figure 6.10). Though the present analysis does not predict an NH_3 detection at the assumed equilibrium abundance (NH_3 / H_2O $< 10^{-4}$), I note that a suggestive posterior peak near the correct abundance was deduced by the retrieval (see Figure 7.5). This suggests that disequilibrium nitrogen chemistry in exo-Neptune atmospheres can be probed by MIRI observations of the 10.5 µm NH_3 feature (complimenting the results of Chapter 4, which considered absorption features from 1–5 µm). Any such detections with JWST would then also yield constraints on N/H and N/O ratios in exo-Neptune atmospheres.

These predicted JWST constraints compliment those obtained by previous studies. My JWST metallicity precision of 0.2 dex is higher than the ≈ 0.5 dex found by Greene et al. (2016) – likely due to the inclusion of NIRSpec G395H and NIRISS SOSS 2nd order data in this study. My C/O precision of 0.35 dex is less than the ≈ 0.2 dex predicted by Greene et al. (2016). However, the reference C/O assumed here (0.12) is lower than the solar or super-solar values used in their retrievals, resulting in a lower achievable precision on C/O due to the weakened CH_4, CO_2, and CO absorption. Schlawin et al. (2018) predicted metallicity and C/O precisions for HAT-P-26b $\approx 3\times$ more precise than those presented here. Their narrower metallicity and C/O posteriors arise from assuming chemical equilibrium. I do not make this assumption, instead offering conservative constraints accounting for a wide range of possible disequilibrium chemistry. Line et al. (2013) posited that metallicities and C/O ratios derived from parameterised mixing ratios with log-uniform molecular priors are ultimately biased, which has been used as a justification for subsequent studies to impose chemical equilibrium. However, in Figure 7.8 I demonstrated that the approach used here successfully retrieves the metallicity and C/O without bias in the absence of equilibrium assumptions. Relaxing the assumption of chemical equilibrium is therefore the appropriate and recommended approach for atmospheric retrievals of JWST observations.

7.5.2 The importance of non-isothermal temperature profiles

In the context of hot Jupiters, Rocchetto et al. (2016) has shown that molecular abundances retrieved from synthetic JWST spectra can be biased by $\sim$ 1 dex if an isotherm is assumed. Here, I have demonstrated that this bias manifests even for much weaker photosphere temperature gradients ($\sim$ 80 K here, vs. $\gtrsim$ 800 K in Rocchetto et al. (2016)), resulting in abundances biased by $\sim$ 0.1-0.2 dex for HAT-P-26b. Whilst, in the present case, the bias from assuming an isotherm is not too severe (see Figure 7.7), a parameterised P-T profile generally produces more accurate results. Furthermore, the bias of trace abundances to higher values also biases the bulk atmospheric component to lower abundances. This, in turn, will bias the derived atmospheric mean molecular weight. Such an additional bias (though minor here due to the shallow temperature gradient), could prove more detrimental for exo-Neptune atmospheres with larger temperature gradients ($\gtrsim$ a few 100 K). In short, this study demonstrates that even shallow terminator temperature gradients have an important influence on compositional inferences. I therefore strongly caution against assuming isothermal temperature profiles when retrieving transmission spectra from JWST.

7.5.3 Heavy-element chemistry with JWST

One of the key results in Chapter 6 was strong evidence (at $> 4\sigma$) for metal hydrides in HAT-P-26b's atmosphere. Considering the extreme violation of chemical equilibrium this constitutes (see section 6.4.3), JWST observations will be timely to directly test this inference. In this chapter, I have demonstrated that NIRISS SOSS observations, especially its 2nd order, provide a powerful window to probe heavy-element chemistry in exoplanetary atmospheres. In particular, I have shown that a single transit with NIRISS will be sufficient to conclusively detect a combination of metal hydrides at $> 13\sigma$ confidence, with specific molecules (e.g. CrH) detected at $> 5\sigma$. There are two general takeaways from this finding: (i) despite the lack of data coverage for $\lambda < 0.6\,\mu$m, JWST transmission spectra can detect heavy-element molecules at high significance; and (ii) metal oxides and metal hydrides must be included as opacity sources for retrievals of JWST data (especially with NIRISS). The upcoming Cycle 1 JWST observations of HAT-P-26b will provide a crucial test of this hypothesis, enabling one to definitively confirm, or rule out, the presence of metal hydrides in HAT-P-26b's atmosphere. Should these metal hydrides be confirmed, this will potentially represent the first detection of external contamination in an exoplanet atmosphere.

7.6 Summary

In this chapter, I have demonstrated that the upcoming James Webb Space Telescope will enable one to characterise exoplanet atmospheres in exquisite detail; advancing what is possible today with existing ground and space-based instruments. I illustrated this through atmospheric retrievals of simulated JWST transmission spectra for the exo-Neptune HAT-P-26b. The major results from this chapter are as follows:

1. Carbon-bearing molecules will be detectable with JWST. Even at the low C/O established by present observations (< 0.33), CO_2, CH_4, and CO will be detectable at 13.1σ, 6.2σ, and 3.7σ, respectively.

2. A conclusive detection of metal hydrides (at $> 13\sigma$) is achievable with visible wavelength NIRISS observations. Several specific metal hydrides are also detectable at $> 5\sigma$ confidence.

3. Abundances for all species detected at $> 5\sigma$ confidence can be constrained to precisions < 0.3 dex. CO constraints are marginally weaker, at ≈ 0.5 dex.

4. The H_2O abundance, and hence metallicity, can be determined to 0.2 dex precision. Metallicity and mean molecular weight constraints will improve by a factor of $>$ 2 compared to current inferences.

5. C/O will be constrainable to < 0.4 dex with near-infrared JWST observations.

6. Temperature profile constraints will improve to $\pm$ 20 K.

7. Abundances are biased by $\sim$ 0.1-0.2 dex under the assumption of an isothermal atmosphere, when in reality a shallow ($\lesssim 100$ K) temperature gradient is present.

As the era of JWST dawns on the horizon, retrieval analyses with POSEIDON are well-placed to deliver new insights into the nature of exoplanet atmospheres.

Chapter 8

Concluding remarks

8.1 Revealing the nature of exoplanet atmospheres

Astronomers have always had a turbulent relationship with clouds. Long after we thought them vanquished by advanced space telescopes, it is perhaps the greatest irony that the very objects we are now studying, exoplanets, have cloudy skies. The complications caused by clouds in interpreting exoplanet spectra have led many to wonder if precise measurements of their atmospheric composition are even possible.

In this thesis, I have developed a new atmospheric retrieval approach capable of precisely measuring the atmospheric properties of cloudy exoplanets. The essential insight is that clouds are unlikely to entirely envelop the day-night terminator of exoplanet atmospheres. In other words, considering clouds as 2D entities with gaps in the cloud coverage allows one to peer through the atmosphere; *simultaneously* inferring chemical compositions, temperature structures, and cloud properties. The creation of a new retrieval code implementing this algorithm, POSEIDON, was described in Chapter 2. This work offers a new paradigm for the atmospheric retrieval of exoplanet transmission spectra, demonstrating that the promise of precision measurements for exoplanet atmospheres lies in the consideration of 2D atmospheric properties.

In Chapter 3, I demonstrated POSEIDON in an application to the high-quality transmission spectrum of the hot Jupiter HD 209458b. Despite the cloudy nature of the atmosphere, I obtained one of the most precise H_2O measurements in an exoplanet atmosphere to date, reported a $> 4\sigma$ detection of patchy clouds, and obtained the first evidence of nitrogen chemistry in a hot Jupiter atmosphere. This established, at high significance, that patchy clouds do indeed exist in hot Jupiter terminators. The central concept underlying POSEIDON, that patchy clouds allow cloud-composition degeneracies to be broken, was therefore vindicated in a real world setting.

Chapter 4 saw applications of POSEIDON to a population of hot Jupiters, yielding two major discoveries: (i) evidence of disequilibrium NH_3 and HCN chemistry for multiple exoplanets; and (ii) the first detection of TiO in an exoplanet atmosphere. The former finding illustrates that current observations already lie on the cusp of detecting nitrogen-bearing molecules in exoplanet atmospheres. Targeted observational campaigns focused on absorption features of NH_3 and HCN, ideally combining precise Hubble spectra with ground-based K-band observations, are already capable of detecting these species. The latter finding demonstrates that large ground-based telescopes can rival, and even exceed, the ability of space-based telescopes to characterise high-temperature exoplanet atmospheres at visible wavelengths. The new frontier of comparative planetology of heavy-element molecular chemistry opened by this work is an exciting direction now under investigation by multiple research groups.

In Chapter 5, I extended POSEIDON's opacity database to include new molecules and atoms which could reside in exoplanet atmospheres. In building this database, I also developed a new method for rapidly calculating molecular cross sections. These new opacities are the first step in enabling POSEIDON to model, and retrieve, spectra of exoplanets spanning the full range from ultra-hot Jupiters to terrestrial worlds.

In Chapter 6, I conducted atmospheric retrieval analyses for the lowest mass exoplanet with detected molecular absorption: the exo-Neptune HAT-P-26b. By obtaining the most precise metallicity for an extrasolar ice giant to date, a crucial clue has been obtained into the formation conditions of planets like Uranus and Neptune – for which oxygen abundances remain unknown. I further reported strong evidence ($> 4\sigma$) for the presence of metal hydrides - the first time such species have been inferred in an exoplanet atmosphere. The existence of metal hydrides in such a relatively cool atmosphere indicates significant disequilibrium chemistry, potentially arising from a recent planetesimal impact contaminating the upper atmosphere with heavy-elements.

Finally, in Chapter 7, I conducted a retrieval analysis on simulated James Webb Space Telescope (JWST) observations of HAT-P-26b, previewing the future potential of this research area. It was demonstrated that new chemical species will be detectable at high-significance, yielding precise constraints on elemental ratios such as C/O. Furthermore, the metal hydride hypothesis can be directly tested with JWST, as it will be able to detect specific metal hydrides at $> 5\sigma$ confidence. I also demonstrated that certain assumptions commonly applied to retrievals of current observations, chemical equilibrium and isothermal temperature structures, are neither necessary nor desirable when analysing JWST transmission spectra.

If there is one takeaway from these studies, it is this: *current observations are already capable of precisely characterising exoplanet atmospheres.* Given sufficient spectral resolution, data precision, and wavelength coverage, precise atmospheric property measurements can be obtained with presently-available instruments. This thesis has provided but a taste of the secrets waiting to be revealed in exoplanet atmospheres.

8.2 Future outlook

At the time of writing, over 4,000 exoplanets are known. Amongst this panoply of worlds, it appears that almost anything that can occur does exist somewhere in the diversity of worlds. In particular, there are already tantalising glimpses of a vast population of potentially habitable planets around M dwarf stars. The TESS mission is currently undertaking an all-sky survey for exoplanets around the nearest stars, poised to discover $\sim$ 20,000 new planets over the next few years (Sullivan et al., 2015). Complimenting this with upcoming space missions, such as WFIRST and PLATO, it is plausible that the exoplanet count will cross 100,000 by the 2030s.

With gaps in the exoplanet population rapidly filling in, our understanding of exoplanet atmospheres stands on the precipice of an even greater revolution. Successful atmospheric characterisation has already been achieved for tens of hot Jupiters and a few exo-Neptunes. In the next stage of this endeavour, the exploration of lower-mass planets, including super-Earths and terrestrial exoplanets, lies on the horizon. The upcoming launch of JWST in 2021 will offer the unprecedented sensitivity required to characterise these new classes of exoplanetary atmospheres.

Though this work has advanced atmospheric retrieval techniques, many areas still require attention to fully realise the promise of next generation observations. One area for further work is the inclusion of 2D (or even 3D) atmospheric properties in retrieval analyses. For example, in this thesis I have focused on inhomogenous clouds and have not examined the possibility of temperature or abundance variations around exoplanet terminators. If such variations can be retrieved, this offers the possibility of measuring horizontal atmospheric property gradients; directly testing predictions of 3D circulation models. As always, a balance between speed and complexity must be established for future retrieval studies considering such effects. Similarly, the complexity of cloud models will need to be critically addressed. Although current parametrisations have proven successful to date, additional refinements will be required to capture properties like the strong resonance features from cloud particles around 9–11 µm predicted by Mie theory (Wakeford & Sing, 2015; Pinhas & Madhusudhan, 2017).

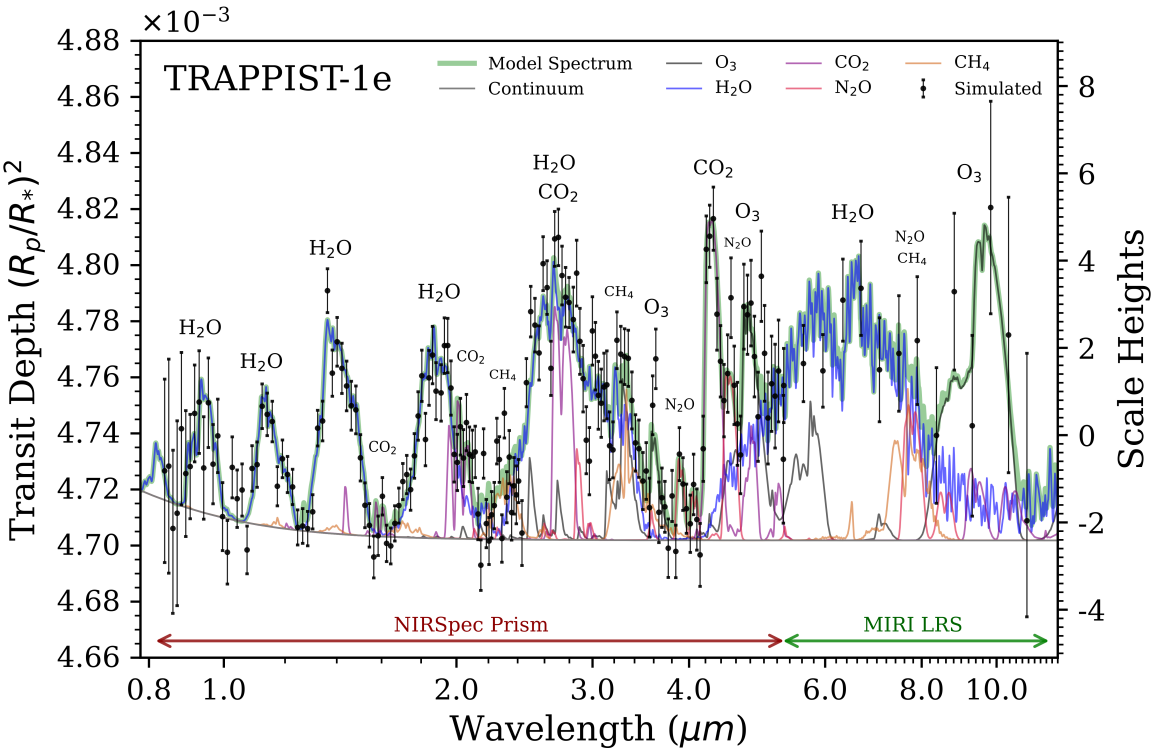


Fig. 8.1 Model transmission spectrum of a terrestrial exoplanet.
A model transmission spectrum for an Earth-like atmosphere around a habitable exoplanet is shaded in green. The planet and star properties correspond to the terrestrial exoplanet TRAPPIST-1e. Coloured lines depict contributions of notable gases to the spectrum. Synthetic JWST observations (black error bars) generated by PandExo for a 200 hr / 100 transit campaign are overlaid for comparison.

Finally, the atmospheric retrievals of giant planets presented in this work can be considered as a proving ground for one of the ultimate goals of exoplanet science: characterisation of habitable extrasolar planets. In Figure 8.1, I illustrate this enticing prospect with a model transmission spectrum of the terrestrial exoplanet TRAPPIST-1e (assuming an Earth-like atmosphere). With synthetic JWST data overlaid, it is seen that detections of gases such as H_2O, CO_2, CH_4, N_2O, and perhaps O_3, will soon be possible for temperate exoplanets. We stand at the threshold of realising one of the oldest scientific dreams: detecting life on another world. Nature always has a way of surprising us, whether that be here on Earth, or across the vastness of the Universe. In the distant skies of alien worlds, something incredible is waiting to be known.

Bibliography

Ackerman A. S., Marley M. S., 2001, ApJ, 556, 872

Adams J. C., 1846, MNRAS, 7, 149

Agol E., Fabrycky D. C., 2018, Transit-Timing and Duration Variations for the Discovery and Characterization of Exoplanets. p. 7, doi:10.1007/978-3-319-55333-7_7

Agúndez M., Venot O., Iro N., Selsis F., Hersant F., Hébrard E., Dobrijevic M., 2012, Astronomy and Astrophysics, 548, A73

Alessi M., Pudritz R. E., Cridland A. J., 2017, MNRAS, 464, 428

Amundsen D. S., Baraffe I., Tremblin P., Manners J., Hayek W., Mayne N. J., Acreman D. M., 2014, Astronomy and Astrophysics, 564, A59

Anglada-Escudé G., et al., 2016, Nature, 536, 437

Apai D., Radigan J., Buenzli E., Burrows A., Reid I. N., Jayawardhana R., 2013, ApJ, 768, 121

Arcangeli J., et al., 2018, ApJ, 855, L30

Armstrong D. J., de Mooij E., Barstow J., Osborn H. P., Blake J., Saniee N. F., 2016, Nature Astronomy, 1, 0004

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, Annual Review of Astronomy and Astrophysics, 47, 481

Atreya S. K., Crida A., Guillot T., Lunine J. I., Madhusudhan N., Mousis O., 2018, The Origin and Evolution of Saturn, with Exoplanet Perspective. Cambridge University Press, p. 5–43, doi:10.1017/9781316227220.002

Azzam A. A. A., Tennyson J., Yurchenko S. N., Naumenko O. V., 2016, MNRAS, 460, 4063

Baglin A., et al., 2006, in 36th COSPAR Scientific Assembly.

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

Barber R. J., Strange J. K., Hill C., Polyansky O. L., Mellau G. C., Yurchenko S. N., Tennyson J., 2014, MNRAS, 437, 1828

Barge P., et al., 2008, Astronomy and Astrophysics, 482, L17

Barklem P. S., Aspelund-Johansson J., 2005, Astron. and Astrophys., 435, 373

Barklem P. S., Collet R., 2016, Astronomy and Astrophysics, 588, A96

Barklem P. S., Piskunov N., O'Mara B. J., 2000, Astron. and Astrophys. Suppl. Ser., 142, 467

Barman T., 2007, ApJ, 661, L191

Barman T. S., Hauschildt P. H., Allard F., 2001, ApJ, 556, 885

Barman T. S., Macintosh B., Konopacky Q. M., Marois C., 2011, ApJ, 733, 65
Barman T. S., Konopacky Q. M., Macintosh B., Marois C., 2015, ApJ, 804, 61
Barstow J. K., Aigrain S., Irwin P. G. J., Bowles N., Fletcher L. N., Lee J. M., 2013, MNRAS, 430, 1188
Barstow J. K., Aigrain S., Irwin P. G. J., Kendrew S., Fletcher L. N., 2015, MNRAS, 448, 2546
Barstow J. K., Aigrain S., Irwin P. G. J., Sing D. K., 2017, ApJ, 834, 50
Barton E. J., Yurchenko S. N., Tennyson J., 2013, MNRAS, 434, 1469
Barton E. J., Hill C., Czurylo M., Li H. Y., Hyslop A., Yurchenko S. N., Tennyson J., 2017, Journal of Quantitative Spectroscopy and Radiative Transfer, 203, 490
Batalha N. E., Line M. R., 2017, AJ, 153, 151
Batalha N. E., et al., 2017, Publications of the Astronomical Society of the Pacific, 129, 064501
Bates D. R., Damgaard A., 1949, Philosophical Transactions of the Royal Society of London Series A, 242, 101
Baudino J. L., Bézard B., Boccaletti A., Bonnefoy M., Lagrange A. M., Galicher R., 2015, Astronomy and Astrophysics, 582, A83
Bean J. L., et al., 2018, Publications of the Astronomical Society of the Pacific, 130, 114402
Beaulieu J. P., et al., 2006, Nature, 439, 437
Beichman C., et al., 2014, Publications of the Astronomical Society of the Pacific, 126, 1134
Belorizky D., 1938, L'Astronomie, 52, 359
Benneke B., 2015, arXiv e-prints, p. arXiv:1504.07655
Benneke B., Seager S., 2012, ApJ, 753, 100
Benneke B., Seager S., 2013, ApJ, 778, 153
Bennett D. P., 2008, Detection of Extrasolar Planets by Gravitational Microlensing. p. 47, doi:10.1007/978-3-540-74008-7_3
Bento J., et al., 2014, MNRAS, 437, 1511
Benvenuto O. G., Fortier A., Brunini A., 2009, Icarus, 204, 752
Bessell M. S., 1991, AJ, 101, 662
Bétrémieux Y., Kaltenegger L., 2014, ApJ, 791, 7
Beuzit J.-L., et al., 2008, in Ground-based and Airborne Instrumentation for Astronomy II. p. 701418, doi:10.1117/12.790120
Biller B. A., Bonnefoy M., 2018, Exoplanet Atmosphere Measurements from Direct Imaging. p. 101, doi:10.1007/978-3-319-55333-7_101
Birkby J. L., 2018, Spectroscopic Direct Detection of Exoplanets. p. 16, doi:10.1007/978-3-319-55333-7_16
Birkby J. L., de Kok R. J., Brogi M., de Mooij E. J. W., Schwarz H., Albrecht S., Snellen I. A. G., 2013, MNRAS, 436, L35
Bodenheimer P., Hubickyj O., Lissauer J. J., 2000, Icarus, 143, 2
Bogaard M. P., Buckingham A. D., Pierens R. K., White A. H., 1978, J. Chem. Soc., Faraday Trans. 1, 74, 3008

Boley A. C., 2009, ApJL, 695, L53
Bolton S. J., et al., 2017, Space Sci. Rev., 213, 5
Bond I. A., et al., 2004, ApJ, 606, L155
Bonfils X., et al., 2013, Astronomy and Astrophysics, 549, A109
Bonfils X., et al., 2018, Astronomy and Astrophysics, 613, A25
Bonneau D., Josse M., Labyrie A., 1975, in de Jager C., Nieuwenhuijzen H., eds, Astrophysics and Space Science Library Vol. 54, Image Processing Techniques in Astronomy. p. 403, doi:10.1007/978-94-010-1881-4_52
Bonnefoy M., et al., 2016, Astronomy and Astrophysics, 587, A58
Borucki W. J., 2018, Space Missions for Exoplanet Science: Kepler/K2. p. 80, doi:10.1007/978-3-319-55333-7_80
Borucki W. J., Summers A. L., 1984, Icarus, 58, 121
Borucki W. J., et al., 2010, Science, 327, 977
Borucki W. J., et al., 2011, ApJ, 736, 19
Borunov S., Drossart P., Encrenaz T., Dorofeeva V., 1997, Icarus, 125, 121
Boss A. P., 1997, Science, 276, 1836
Boss A. P., 1998, Earth Moon and Planets, 81, 19
Boutle I. A., Mayne N. J., Drummond B., Manners J., Goyal J., Hugo Lambert F., Acreman D. M., Earnshaw P. D., 2017, Astronomy and Astrophysics, 601, A120
Bowler B. P., 2016, Publications of the Astronomical Society of the Pacific, 128, 102001
Bowler B. P., Nielsen E. L., 2018, Occurrence Rates from Direct Imaging Surveys. p. 155, doi:10.1007/978-3-319-55333-7_155
Bowler B. P., Liu M. C., Dupuy T. J., Cushing M. C., 2010, ApJ, 723, 850
Bracewell R. N., 1978, Nature, 274, 780
Brasseur G., De Rudder A., 1986, in Worrest R. C., Caldwell M. M., eds, Stratospheric Ozone Reduction, Solar Ultraviolet Radiation and Plant Life. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 1–28
Brewer J. M., Fischer D. A., Valenti J. A., Piskunov N., 2016, The Astrophysical Journal Supplement Series, 225, 32
Brogi M., Line M. R., 2019, AJ, 157, 114
Brogi M., Snellen I. A. G., de Kok R. J., Albrecht S., Birkby J., de Mooij E. J. W., 2012, Nature, 486, 502
Brogi M., de Kok R. J., Albrecht S., Snellen I. A. G., Birkby J. L., Schwarz H., 2016, ApJ, 817, 106
Brogi M., Line M., Bean J., Désert J. M., Schwarz H., 2017, ApJ, 839, L2
Brooke J. S. A., Bernath P. F., Western C. M., van Hemert M. C., Groenenboom G. C., 2014, Journal of Chemical Physics, 141, 054310
Brooke J. S. A., Bernath P. F., Western C. M., Sneden C., Afşar M., Li G., Gordon I. E., 2016, Journal of Quantitative Spectroscopy and Radiative Transfer, 168, 142
Brown T. M., 2001, ApJ, 553, 1006
Brown R. A., Burrows C. J., 1990, Icarus, 87, 484
Bruno G., 1584, De l'infinito universo et mondi (EN: On the infinite universe and worlds)

Buchhave L. A., et al., 2012, Nature, 486, 375
Buchner J., et al., 2014, Astronomy and Astrophysics, 564, A125
Buchwald V. F., 1977, Philosophical Transactions of the Royal Society of London Series A, 286, 453
Buenzli E., et al., 2012, ApJ, 760, L31
Burbine T. H., 2016, Asteroids: Astronomical and Geological Bodies. Cambridge University Press
Burningham B., Marley M. S., Line M. R., Lupu R., Visscher C., Morley C. V., Saumon D., Freedman R., 2017, MNRAS, 470, 1177
Burrows A., Sharp C. M., 1999, ApJ, 512, 843
Burrows A., Volobuyev M., 2003, ApJ, 583, 985
Burrows A., Marley M. S., Sharp C. M., 2000, ApJ, 531, 438
Burrows A., Ram R. S., Bernath P., Sharp C. M., Milsom J. A., 2002, ApJ, 577, 986
Burrows A., Dulick M., Bauschlicher C. W. J., Bernath P. F., Ram R. S., Sharp C. M., Milsom J. A., 2005, ApJ, 624, 988
Burrows A., Hubeny I., Budaj J., Knutson H. A., Charbonneau D., 2007, ApJ, 668, L171
Cabot S. H. C., Madhusudhan N., Hawker G. A., Gandhi S., 2019, MNRAS, 482, 4422
Cai K., Durisen R. H., Michael S., Boley A. C., Mejía A. C., Pickett M. K., D'Alessio P., 2006, ApJL, 636, L149
Cameron A. G. W., 1978, Moon and Planets, 18, 5
Campbell B., Walker G. A. H., 1979, PASP, 91, 540
Campbell B., Walker G. A. H., Yang S., 1988, ApJ, 331, 902
Carry B., 2012, Planetary and Space Science, 73, 98
Cassan A., et al., 2012, Nature, 481, 167
Cavalié T., et al., 2010, Astronomy and Astrophysics, 510, A88
Chabrier G., Johansen A., Janson M., Rafikov R., 2014, Protostars and Planets VI, pp 619–642
Chambers J. E., Wetherill G. W., 1998, Icarus, 136, 304
Chandrasekhar S., 1960, Radiative transfer
Charbonneau D., Brown T. M., Latham D. W., Mayor M., 2000, ApJL, 529, L45
Charbonneau D., Brown T. M., Noyes R. W., Gilliland R. L., 2002, ApJ, 568, 377
Charbonneau D., et al., 2005, ApJ, 626, 523
Charbonneau D., Knutson H. A., Barman T., Allen L. E., Mayor M., Megeath S. T., Queloz D., Udry S., 2008, ApJ, 686, 1341
Charpinet S., et al., 2011, Nature, 480, 496
Chauvin G., Lagrange A. M., Dumas C., Zuckerman B., Mouillet D., Song I., Beuzit J. L., Lowrance P., 2004, Astronomy and Astrophysics, 425, L29
Chen G., et al., 2018, Astronomy and Astrophysics, 616, A145
Cho J. Y. K., Menou K., Hansen B. M. S., Seager S., 2003, ApJ, 587, L117
Choi J., McCarthy C., Marcy G. W., Howard A. W., Fischer D. A., Johnson J. A., Isaacson H., Wright J. T., 2013, ApJ, 764, 131
Christiansen J. L., et al., 2010, ApJ, 710, 97

Clanton C., Gaudi B. S., 2017, ApJ, 834, 46
Cooper C. S., Showman A. P., 2006, ApJ, 649, 1048
Coppola C. M., Lodi L., Tennyson J., 2011, MNRAS, 415, 487
Cowan N. B., Agol E., 2008, ApJ, 678, L129
Cridland A. J., Pudritz R. E., Birnstiel T., Cleeves L. I., Bergin E. A., 2017, MNRAS, 469, 3910
Crismani M. M. J., et al., 2017, Nature Geoscience, 10, 401
Croll B., Jayawardhana R., Fortney J. J., Lafrenière D., Albert L., 2010, ApJ, 718, 920
Crossfield I. J. M., Hansen B. M. S., Harrington J., Cho J. Y. K., Deming D., Menou K., Seager S., 2010, ApJ, 723, 1436
Crossfield I. J. M., et al., 2014, Nature, 505, 654
Crovisier J., 1995, in European SL-9/Jupiter Workshop. pp 411–421
Cumming A., Butler R. P., Marcy G. W., Vogt S. S., Wright J. T., Fischer D. A., 2008, Publications of the Astronomical Society of the Pacific, 120, 531
Cushing M. C., et al., 2008, ApJ, 678, 1372
Cuthbertson C., Cuthbertson M., 1932, Proceedings of the Royal Society of London Series A, 135, 40
D'Angelo G., Lissauer J. J., 2018, Formation of Giant Planets. p. 140, doi:10.1007/978-3-319-55333-7_140
Dalgarno A., Williams D. A., 1962, ApJ, 136, 690
Dang L., et al., 2018, Nature Astronomy, 2, 220
Darby-Lewis D., Tennyson J., Lawson K. D., Yurchenko S. N., Stamp M. F., Shaw A., Brezinsek S., Contributors J., 2018, Journal of Physics B Atomic Molecular Physics, 51, 185701
Deleuil M., et al., 2018, Astronomy and Astrophysics, 619, A97
Delrez L., et al., 2018, MNRAS, 475, 3577
Deming D., Sheppard K., 2017, ApJ, 841, L3
Deming D., Seager S., Richardson L. J., Harrington J., 2005, Nature, 434, 740
Deming D., Harrington J., Laughlin G., Seager S., Navarro S. B., Bowman W. C., Horning K., 2007, ApJ, 667, L199
Deming D., et al., 2013, ApJ, 774, 95
Demory B.-O., Gillon M., Seager S., Benneke B., Deming D., Jackson B., 2012, ApJ, 751, L28
Demory B.-O., et al., 2013, ApJ, 776, L25
Demory B.-O., et al., 2016, Nature, 532, 207
Désert J. M., Vidal-Madjar A., Lecavelier Des Etangs A., Sing D., Ehrenreich D., Hébrard G., Ferlet R., 2008, Astronomy and Astrophysics, 492, 585
Diamond-Lowe H., Stevenson K. B., Bean J. L., Line M. R., Fortney J. J., 2014, ApJ, 796, 66
Dimitrijević M. S., Peach G., 1990, Astronomy and Astrophysics, 236, 261
Dodson-Robinson S. E., Veras D., Ford E. B., Beichman C. A., 2009, ApJ, 707, 79
Dominik C., Blum J., Cuzzi J. N., Wurm G., 2007, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V. p. 783 (`arXiv:astro-ph/0602617`)

Dong S., et al., 2006, ApJ, 642, 842
Dorn C., Mosegaard K., Grimm S. L., Alibert Y., 2018, ApJ, 865, 20
Dressing C. D., Charbonneau D., 2015, ApJ, 807, 45
Durisen R. H., Boss A. P., Mayer L., Nelson A. F., Quinn T., Rice W. K. M., 2007, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V. p. 607 (`arXiv:astro-ph/0603179`)
Ehrenreich D., et al., 2014, Astronomy and Astrophysics, 570, A89
Espinoza N., et al., 2019, MNRAS, 482, 2065
Esteves L. J., De Mooij E. J. W., Jayawardhana R., 2013, ApJ, 772, 51
Evans T. M., et al., 2016, ApJ, 822, L4
Evans T. M., et al., 2017, Nature, 548, 58
Evans T. M., et al., 2018, AJ, 156, 283
Feng Y. K., Robinson T. D., Fortney J. J., Lupu R. E., Marley M. S., Lewis N. K., Macintosh B., Line M. R., 2018, AJ, 155, 200
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 e-prints, p. arXiv:1306.2144
Fischer D. A., Valenti J., 2005, ApJ, 622, 1102
Fisher C., Heng K., 2018, MNRAS, 481, 4698
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, Publications of the Astronomical Society of the Pacific, 125, 306
Forgan D., Rice K., 2013, MNRAS, 432, 3168
Fornasier S., Clark B. E., Dotto E., Migliorini A., Ockert-Bell M., Barucci M. A., 2010, Icarus, 210, 655
Fortney J. J., 2005, MNRAS, 364, 649
Fortney J. J., Lodders K., Marley M. S., Freedman R. S., 2008, ApJ, 678, 1419
Fortney J. J., Shabram M., Showman A. P., Lian Y., Freedman R. S., Marley M. S., Lewis N. K., 2010, ApJ, 709, 1396
Fortney J. J., Mordasini C., Nettelmann N., Kempton E. M. R., Greene T. P., Zahnle K., 2013, ApJ, 775, 80
Fowler A., 1904, Proceedings of the Royal Society of London Series I, 73, 219
Fraine J., et al., 2014, Nature, 513, 526
Fressin F., et al., 2013, ApJ, 766, 81
Fujii Y., et al., 2018, Astrobiology, 18, 739
Gaidos E., Mann A. W., Kraus A. L., Ireland M., 2016, MNRAS, 457, 2877
Galicher R., Marois C., Macintosh B., Barman T., Konopacky Q., 2011, ApJ, 739, L41
Galicher R., et al., 2016, Astronomy and Astrophysics, 594, A63
Galle J. G., 1846, MNRAS, 7, 153
Gandhi S., Madhusudhan N., 2017, MNRAS, 472, 2334
Gandhi S., Madhusudhan N., 2018, MNRAS, 474, 271
Gatewood G., 1974, AJ, 79, 52
Gaudi B. S., 2010, arXiv e-prints, p. arXiv:1002.0332

Gharib-Nezhad E., Shayesteh A., Bernath P. F., 2013, MNRAS, 432, 2043
Gibson N. P., Aigrain S., Barstow J. K., Evans T. M., Fletcher L. N., Irwin P. G. J., 2013, MNRAS, 436, 2974
Gillon M., et al., 2017, Nature, 542, 456
Goldreich P., Tremaine S., 1979, ApJ, 233, 857
Goody R., West R., Chen L., Crisp D., 1989, JQSRT, 42, 539
Gordon I. E., et al., 2017, Journal of Quantitative Spectroscopy and Radiative Transfer, 203, 3
Goukenleuque C., Bézard B., Joguet B., Lellouch E., Freedman R., 2000, Icarus, 143, 308
Gould A., Loeb A., 1992, ApJ, 396, 104
Goyal J. M., Wakeford H. R., Mayne N. J., Lewis N. K., Drummond B., Sing D. K., 2019, MNRAS, 482, 4503
Gravity Collaboration et al., 2019, Astronomy and Astrophysics, 623, L11
Greenberg R., Wacker J. F., Hartmann W. K., Chapman C. R., 1978, Icarus, 35, 1
Greene T. P., Line M. R., Montero C., Fortney J. J., Lustig-Yaeger J., Luther K., 2016, ApJ, 817, 17
Griffith C. A., 2014, Philosophical Transactions of the Royal Society of London Series A, 372, 20130086
Grillmair C. J., et al., 2008, Nature, 456, 767
Grimm S. L., Heng K., 2015, ApJ, 808, 182
Grimm S. L., et al., 2018, Astronomy and Astrophysics, 613, A68
Guillot T., 2010, Astronomy and Astrophysics, 520, A27
Guillot T., Burrows A., Hubbard W. B., Lunine J. I., Saumon D., 1996, ApJ, 459, L35
Harrington J., Hansen B. M., Luszcz S. H., Seager S., Deming D., Menou K., Cho J. Y. K., Richardson L. J., 2006, Science, 314, 623
Hartman J. D., et al., 2011, ApJ, 728, 138
Hawker G. A., Parry I. R., 2019, MNRAS, 484, 4855
Hawker G. A., Madhusudhan N., Cabot S. H. C., Gandhi S., 2018, ApJ, 863, L11
Haynes K., Mandell A. M., Madhusudhan N., Deming D., Knutson H., 2015, ApJ, 806, 146
Hebb L., et al., 2010, ApJ, 708, 224
Hedges C., Madhusudhan N., 2016, MNRAS, 458, 1427
Heintz W. D., 1978, ApJ, 220, 931
Heinzel P., 1978, Bulletin of the Astronomical Institutes of Czechoslovakia, 29, 159
Helled R., Bodenheimer P., 2014, Astrophys. J., 789, 69
Helling C., et al., 2016, MNRAS, 460, 855
Heng K., Kitzmann D., 2017, MNRAS, 470, 2972
Heng K., Lyons J. R., 2016, ApJ, 817, 149
Heng K., Marley M. S., 2018, Radiative Transfer for Exoplanet Atmospheres. p. 102, doi:10.1007/978-3-319-55333-7_102
Heng K., Tsai S.-M., 2016, ApJ, 829, 104
Henry G. W., Marcy G. W., Butler R. P., Vogt S. S., 2000, ApJL, 529, L41

Hermes J. J., 2018, Timing by Stellar Pulsations as an Exoplanet Discovery Method. p. 6, doi:10.1007/978-3-319-55333-7_6
Hill R. J., Lawrence R. S., 1986, Infrared Physics, 26, 371
Hill C., Yurchenko S. N., Tennyson J., 2013, Icarus, 226, 1673
Hillenbrand L. A., 2008, Physica Scripta Volume T, 130, 014024
Hindmarsh W. R., Petford A. D., Smith G., 1967, Proceedings of the Royal Society of London Series A, 297, 296
Hinkley S., Carter A., Biller B., Skemer A., 2018, in European Planetary Science Congress. pp EPSC2018–923
Hinz P. M., Rodigas T. J., Kenworthy M. A., Sivanandam S., Heinze A. N., Mamajek E. E., Meyer M. R., 2010, ApJ, 716, 417
Hoeijmakers H. J., de Kok R. J., Snellen I. A. G., Brogi M., Birkby J. L., Schwarz H., 2015, Astronomy and Astrophysics, 575, A20
Hoeijmakers H. J., et al., 2018a, Nature, 560, 453
Hoeijmakers H. J., Schwarz H., Snellen I. A. G., de Kok R. J., Bonnefoy M., Chauvin G., Lagrange A. M., Girard J. H., 2018b, Astronomy and Astrophysics, 617, A144
Hoeijmakers H. J., et al., 2019, arXiv e-prints, p. arXiv:1905.02096
Hohm U., 1994, Chemical Physics, 179, 533
Holman M. J., et al., 2010, Science, 330, 51
Howard A. W., et al., 2012, The Astrophysical Journal Supplement Series, 201, 15
Howe A. R., Burrows A., Deming D., 2017, ApJ, 835, 96
Hsu D. C., Ford E. B., Ragozzine D., Ashby K., 2019, arXiv e-prints, p. arXiv:1902.01417
Hubeny I., 2017, MNRAS, 469, 841
Hubeny I., Burrows A., Sudarsky D., 2003, ApJ, 594, 1011
Huber K. F., Czesla S., Schmitt J. H. M. M., 2017, Astronomy and Astrophysics, 597, A113
Hubickyj O., Bodenheimer P., Lissauer J. J., 2005, Icarus, 179, 415
Huitson C. M., et al., 2013, MNRAS, 434, 3252
Husser T. O., Wende-von Berg S., Dreizler S., Homeier D., Reiners A., Barman T., Hauschildt P. H., 2013, Astronomy and Astrophysics, 553, A6
Ingraham P., et al., 2014, ApJ, 794, L15
Irwin P. G. J., et al., 2008, Journal of Quantitative Spectroscopy and Radiative Transfer, 109, 1136
Irwin P. G. J., Toledo D., Garland R., Teanby N. A., Fletcher L. N., Orton G. A., Bézard B., 2018, Nature Astronomy, 2, 420
Izidoro A., Raymond S. N., 2018, Formation of Terrestrial Planets. p. 142, doi:10.1007/978-3-319-55333-7_142
Jackson J. D., Fox R. F., 1999, American Journal of Physics, 67, 841
Jacob W. S., 1855, MNRAS, 15, 228
Janson M., Bergfors C., Goto M., Brandner W., Lafrenière D., 2010, in In the Spirit of Lyot 2010. p. E17
Jeffreys H., 1939, The Theory of Probability

Johnson T. V., Mousis O., Lunine J. I., Madhusudhan N., 2012, Astrophys. J., 757, 192
Jones M. I., et al., 2018, arXiv e-prints,
Joshi M., 2003, Astrobiology, 3, 415
Kaltenegger L., 2017, Annual Review of Astronomy and Astrophysics, 55, 433
Karalidi T., Apai D., Marley M. S., Buenzli E., 2016, ApJ, 825, 90
Karkoschka E., Tomasko M. G., 2011, Icarus, 211, 780
Karman T., et al., 2019, Icarus, 328, 160
Kataria T., Sing D. K., Lewis N. K., Visscher C., Showman A. P., Fortney J. J., Marley M. S., 2016, ApJ, 821, 9
Kilpatrick B. M., et al., 2018, AJ, 156, 103
King L. V., 1923, Proceedings of the Royal Society of London Series A, 104, 333
Kipping D. M., Tinetti G., 2010, MNRAS, 407, 2589
Kirkpatrick J. D., 2005, Annual Review of Astronomy and Astrophysics, 43, 195
Kitzmann D., Heng K., 2018, MNRAS, 475, 94
Kitzmann D., et al., 2018, ApJ, 863, 183
Knutson H. A., et al., 2007, Nature, 447, 183
Knutson H. A., Charbonneau D., Allen L. E., Burrows A., Megeath S. T., 2008, ApJ, 673, 526
Knutson H. A., et al., 2009, ApJ, 690, 822
Knutson H. A., et al., 2012, ApJ, 754, 22
Knutson H. A., Benneke B., Deming D., Homeier D., 2014a, Nature, 505, 66
Knutson H. A., et al., 2014b, ApJ, 794, 155
Komacek T. D., Showman A. P., 2016, ApJ, 821, 16
Konacki M., Torres G., Jha S., Sasselov D. D., 2003, Nature, 421, 507
Konopacky Q. M., Barman T. S., 2018, HR8799: Imaging a System of Exoplanets. p. 36, doi:10.1007/978-3-319-55333-7_36
Konopacky Q. M., Barman T. S., Macintosh B. A., Marois C., 2013, Science, 339, 1398
Kramer M., 2018, Pulsar Timing as an Exoplanet Discovery Method. p. 5, doi:10.1007/978-3-319-55333-7_5
Kramida A., Yu. Ralchenko Reader J., and NIST ASD Team 2018, NIST Atomic Spectra Database (ver. 5.6.1), [Online]. Available: `https://physics.nist.gov/asd` [2019, April 23]. National Institute of Standards and Technology, Gaithersburg, MD.
Kratter K., Lodato G., 2016, Ann. Rev. Astron. Astro., 54, 271
Kreidberg L., et al., 2014a, Nature, 505, 69
Kreidberg L., et al., 2014b, ApJ, 793, L27
Kreidberg L., et al., 2015, ApJ, 814, 66
Kreidberg L., et al., 2018, AJ, 156, 17
Kuiper G. P., 1951, Proceedings of the National Academy of Science, 37, 1
Kurucz R. L., 2014, Robert L. Kurucz on-line database of observed and predicted atomic transitions
Kurucz R. L., Peytremann E., 1975, SAO Special Report, 362, 1
Lacis A. A., Oinas V., 1991, Journal of Geophysical Research, 96, 9027

Lardner D., 1851, Handbook of Natural Philosophy and Astronomy. London : Taylor, Walton, and Maberly
Latham D. W., Mazeh T., Stefanik R. P., Mayor M., Burki G., 1989, Nature, 339, 38
Laughlin G., 2018, Mass-Radius Relations of Giant Planets: The Radius Anomaly and Interior Models. p. 1, doi:10.1007/978-3-319-55333-7_1
Lavie B., et al., 2017, AJ, 154, 91
Le Verrier U., 1846, Recherches sur les mouvements de la planète Herschel (EN: Research on the movements of Herschel's planet. Connaissance des temps, Bachelier
Lecar M., Podolak M., Sasselov D., Chiang E., 2006, ApJ, 640, 1115
Lecavelier Des Etangs A., Vidal-Madjar A., Désert J. M., Sing D., 2008, Astronomy and Astrophysics, 485, 865
Lee J. M., Fletcher L. N., Irwin P. G. J., 2012, MNRAS, 420, 170
Lee J.-M., Heng K., Irwin P. G. J., 2013, ApJ, 778, 97
Lee G., Dobbs-Dixon I., Helling C., Bognar K., Woitke P., 2016, Astronomy and Astrophysics, 594, A48
Lellouch E., Moreno R., Paubert G., 2005, Astronomy and Astrophysics, 430, L37
Li G., Harrison J. J., Ram R. S., Western C. M., Bernath P. F., 2012, Journal of Quantitative Spectroscopy and Radiative Transfer, 113, 67
Li G., Gordon I. E., Rothman L. S., Tan Y., Hu S.-M., Kassi S., Campargue A., Medvedev E. S., 2015, The Astrophysical Journal Supplement Series, 216, 15
Lin D. N. C., Papaloizou J., 1986, ApJ, 309, 846
Lin D. N. C., Bodenheimer P., Richardson D. C., 1996, Nature, 380, 606
Lindgård A., Nielson S. E., 1977, Atomic Data and Nuclear Data Tables, 19, 533
Line M. R., Parmentier V., 2016, ApJ, 820, 78
Line M. R., et al., 2013, ApJ, 775, 137
Line M. R., Teske J., Burningham B., Fortney J. J., Marley M. S., 2015, ApJ, 807, 183
Line M. R., et al., 2016, AJ, 152, 203
Lines S., et al., 2018, Astronomy and Astrophysics, 615, A97
Lissauer J. J., 1993, Ann. Rev. Astron. Astro., 31, 129
Lissauer J. J., Hubickyj O., D'Angelo G., Bodenheimer P., 2009, Icarus, 199, 338
Lockwood A. C., Johnson J. A., Bender C. F., Carr J. S., Barman T., Richert A. J. W., Blake G. A., 2014, ApJ, 783, L29
Lodders K., Fegley B., 2002, Icarus, 155, 393
Lodi L., Yurchenko S. N., Tennyson J., 2015, Molecular Physics, 113, 1998
Lovis C., et al., 2017, Astronomy and Astrophysics, 599, A16
Lubow S. H., Ida S., 2010, arXiv e-prints, p. arXiv:1004.4137
Lukashevskaya A. A., Lavrentieva N. N., Dudaryonok A. C., Perevalov V. I., 2016, Journal of Quantitative Spectroscopy and Radiative Transfer, 184, 205
Lupu R. E., Marley M. S., Lewis N., Line M., Traub W. A., Zahnle K., 2016, AJ, 152, 217
Lyulin O. M., Perevalov V. I., 2017, Journal of Quantitative Spectroscopy and Radiative Transfer, 201, 94
MacDonald R., 2018, New Scientist, 240, 38

MacDonald R. J., Madhusudhan N., 2017a, MNRAS, 469, 1979
MacDonald R. J., Madhusudhan N., 2017b, ApJ, 850, L15
MacDonald R. J., Madhusudhan N., 2019, MNRAS, 486, 1292
MacDonald R. J., Marley M. S., Fortney J. J., Lewis N. K., 2018, ApJ, 858, 69
Macintosh B. A., et al., 2008, in Adaptive Optics Systems. p. 701518, doi:10.1117/12.788083
Macintosh B., et al., 2015, Science, 350, 64
Madhusudhan N., 2012, ApJ, 758, 36
Madhusudhan N., 2019, arXiv e-prints, p. arXiv:1904.03190
Madhusudhan N., Seager S., 2009, ApJ, 707, 24
Madhusudhan N., Seager S., 2011, ApJ, 729, 41
Madhusudhan N., et al., 2011, Nature, 469, 64
Madhusudhan N., Knutson H., Fortney J. J., Barman T., 2014a, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. p. 739 (arXiv:1402.1169), doi:10.2458/azu_uapress_9780816531240-ch032
Madhusudhan N., Crouzet N., McCullough P. R., Deming D., Hedges C., 2014b, ApJ, 791, L9
Madhusudhan N., Amin M. A., Kennedy G. M., 2014c, ApJ, 794, L12
Madhusudhan N., Apai D., Gandhi S., 2016a, arXiv e-prints, p. arXiv:1612.03174
Madhusudhan N., Agúndez M., Moses J. I., Hu Y., 2016b, Space Sci. Rev., 205, 285
Madhusudhan N., Bitsch B., Johansen A., Eriksson L., 2017, MNRAS, 469, 4102
Malbet F., Sozzetti A., 2018, Astrometry as an Exoplanet Discovery Method. p. 196, doi:10.1007/978-3-319-55333-7_196
Mandel K., Agol E., 2002, ApJ, 580, L171
Mandell A. M., Haynes K., Sinukoff E., Madhusudhan N., Burrows A., Deming D., 2013, ApJ, 779, 128
Mansfield C. R., Peck E. R., 1969, Journal of the Optical Society of America (1917-1983), 59, 199
Mao S., Paczynski B., 1991, ApJL, 374, L37
Marley M. S., Robinson T. D., 2015, Annual Review of Astronomy and Astrophysics, 53, 279
Marley M. S., Gelino C., Stephens D., Lunine J. I., Freedman R., 1999, ApJ, 513, 879
Marley M. S., Saumon D., Goldblatt C., 2010, ApJ, 723, L117
Marley M. S., Ackerman A. S., Cuzzi J. N., Kitzmann D., 2013, Clouds and Hazes in Exoplanet Atmospheres. p. 367, doi:10.2458/azu_uapress_9780816530595-ch15
Marois C., Macintosh B., Barman T., Zuckerman B., Song I., Patience J., Lafrenière D., Doyon R., 2008, Science, 322, 1348
Marois C., Zuckerman B., Konopacky Q. M., Macintosh B., Barman T., 2010, Nature, 468, 1080
Márquez-Neila P., Fisher C., Sznitman R., Heng K., 2018, Nature Astronomy, 2, 719
Martin R. G., Livio M., 2012, MNRAS, 425, L6
Martin G., Fuhr J., Wiese W., 1988, J. Phys. Chem. Ref. Data Suppl., 17
Masseron T., et al., 2014, Astronomy and Astrophysics, 571, A47

Mayor M., Queloz D., 1995, Nature, 378, 355
Mayor M., et al., 2009, Astronomy and Astrophysics, 493, 639
Mayor M., et al., 2011, arXiv e-prints, p. arXiv:1109.2497
McKemmish L. K., Yurchenko S. N., Tennyson J., 2016, MNRAS, 463, 771
McKemmish L. K., Masseron T., Hoeijmakers H. J., Perez-Mesa V., Grimm S. L., Yurchenko S. N., Tennyson J., 2019, arXiv e-prints, p. arXiv:1905.04587
Miller-Ricci Kempton E., Rauscher E., 2012, ApJ, 751, 117
Millholland S., Laughlin G., 2017, AJ, 154, 83
Min M., 2017, Astronomy and Astrophysics, 607, A9
Mizuno H., Nakazawa K., Hayashi C., 1978, Progress of Theoretical Physics, 60, 699
Mizus I. I., Alijah A., Zobov N. F., Lodi L., Kyuberis A. A., Yurchenko S. N., Tennyson J., Polyansky O. L., 2017, MNRAS, 468, 1717
Mollière P., Snellen I. A. G., 2019, Astronomy and Astrophysics, 622, A139
Mollière P., van Boekel R., Dullemond C., Henning T., Mordasini C., 2015, ApJ, 813, 47
Mordasini C., van Boekel R., Mollière P., Henning T., Benneke B., 2016, ApJ, 832, 41
Morgan W. W., Keenan P. C., Kellman E., 1943, An atlas of stellar spectra, with an outline of spectral classification
Morley C. V., Fortney J. J., Kempton E. M. R., Marley M. S., Visscher C., Zahnle K., 2013, ApJ, 775, 33
Morley C. V., Knutson H., Line M., Fortney J. J., Thorngren D., Marley M. S., Teal D., Lupu R., 2017, AJ, 153, 86
Moses J. I., 2014, Philosophical Transactions of the Royal Society of London Series A, 372, 20130073
Moses J. I., Poppe A. R., 2017, Icarus, 297, 33
Moses J. I., et al., 2011, ApJ, 737, 15
Moses J. I., Madhusudhan N., Visscher C., Freedman R. S., 2013a, ApJ, 763, 25
Moses J. I., et al., 2013b, ApJ, 777, 34
Moulton F. R., 1899, AJ, 20, 33
Mousis O., Lunine J. I., Madhusudhan N., Johnson T. V., 2012, ApJ, 751, L7
Movshovitz N., Bodenheimer P., Podolak M., Lissauer J. J., 2010, Icarus, 209, 616
Mróz P., et al., 2017, Nature, 548, 183
Mróz P., et al., 2018, AJ, 155, 121
Mulders G. D., 2018, Planet Populations as a Function of Stellar Properties. p. 153, doi:10.1007/978-3-319-55333-7_153
Mulders G. D., Pascucci I., Apai D., 2015, ApJ, 814, 130
Murphy S. J., Bedding T. R., Shibahashi H., 2016, ApJL, 827, L17
Murray C. D., Correia A. C. M., 2010, Keplerian Orbits and Dynamics of Exoplanets. pp 15–23
Nitz D. E., Wickliffe M. E., Lawler J. E., 1998, Astrophys. J. Suppl. Ser., 117, 313
Nugroho S. K., Kawahara H., Masuda K., Hirano T., Kotani T., Tajitsu A., 2017, AJ, 154, 221
OGLE Collaboration et al., 2019, Astronomy and Astrophysics, 622, A201

Öberg K. I., Murray-Clay R., Bergin E. A., 2011, ApJ, 743, L16
Olivero J., 1977, Journal of Quantitative Spectroscopy and Radiative Transfer, 17, 233
Oppenheimer B. R., et al., 2013, ApJ, 768, 24
Oreshenko M., Heng K., Demory B.-O., 2016, MNRAS, 457, 3420
Parmentier V., Crossfield I. J. M., 2018, Exoplanet Phase Curves: Observations and Theory. p. 116, doi:10.1007/978-3-319-55333-7_116
Parmentier V., Showman A. P., Lian Y., 2013, Astronomy and Astrophysics, 558, A91
Parmentier V., Fortney J. J., Showman A. P., Morley C., Marley M. S., 2016, ApJ, 828, 22
Parmentier V., et al., 2018, Astronomy and Astrophysics, 617, A110
Patrascu A. T., Yurchenko S. N., Tennyson J., 2015, MNRAS, 449, 3613
Peach G., 1981, Advances in Physics, 30, 367
Petigura E. A., Howard A. W., Marcy G. W., 2013, Proceedings of the National Academy of Science, 110, 19273
Petigura E. A., et al., 2018, AJ, 155, 89
Pickering J. C., Thorne A. P., Perez R., 2001, Astrophys. J. Suppl. Ser., 132, 403
Pinhas A., Madhusudhan N., 2017, MNRAS, 471, 4355
Pinhas A., Madhusudhan N., Clarke C., 2016, MNRAS, 463, 4516
Pinhas A., Madhusudhan N., Gandhi S., MacDonald R., 2019, MNRAS, 482, 1485
Pino L., et al., 2018, Astronomy and Astrophysics, 612, A53
Piskorz D., et al., 2016, ApJ, 832, 131
Piskorz D., et al., 2017, AJ, 154, 78
Piso A.-M. A., Pegues J., Öberg K. I., 2016, ApJ, 833, 203
Pollacco D. L., et al., 2006, PASP, 118, 1407
Pollack J. B., Hubickyj O., Bodenheimer P., Lissauer J. J., Podolak M., Greenzweig Y., 1996, Icarus, 124, 62
Polyansky O. L., Kyuberis A. A., Zobov N. F., Tennyson J., Yurchenko S. N., Lodi L., 2018, MNRAS, 480, 2597
Pont F., Sing D. K., Gibson N. P., Aigrain S., Henry G., Husnoo N., 2013, MNRAS, 432, 2917
Pudritz R. E., Cridland A. J., Alessi M., 2018, Connecting Planetary Composition with Formation. p. 144, doi:10.1007/978-3-319-55333-7_144
Pueyo L., 2018, Direct Imaging as a Detection Technique for Exoplanets. p. 10, doi:10.1007/978-3-319-55333-7_10
Raassen A. J. J., Uylings P. H. M., 1998, Astronomy and Astrophysics, 340, 300
Radigan J., 2014, ApJ, 797, 120
Radziemski L. J., Engleman Jr. R., Brault J. W., 1995, Phys. Rev. A, 52, 4462
Rafikov R. R., 2005, ApJL, 621, L69
Ralchenko Y., Kramida A., Reader J., NIST ASD Team 2010, NIST Atomic Spectra Database (ver. 4.0.0), [Online].
Rauscher E., Kempton E. M. R., 2014, ApJ, 790, 79
Raymond S. N., O'Brien D. P., Morbidelli A., Kaib N. A., 2009, Icarus, 203, 644
Redfield S., Endl M., Cochran W. D., Koesterke L., 2008, ApJ, 673, L87

Reuyl D., Holmberg E., 1943, ApJ, 97, 41
Riaud P., Schneider J., 2007, Astronomy and Astrophysics, 469, 355
Ribas I., et al., 2018, Nature, 563, 365
Richard C., et al., 2012, JQSRT, 113, 1276
Ricker G. R., et al., 2014, in Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave. p. 914320 (arXiv:1406.0151), doi:10.1117/12.2063489
Rivlin T., Lodi L., Yurchenko S. N., Tennyson J., Le Roy R. J., 2015, MNRAS, 451, 634
Roberge A., Kamp I., 2010, Protoplanetary and Debris Disks. pp 269–295
Robinson T. D., 2017, ApJ, 836, 236
Robinson T. D., Fortney J. J., Hubbard W. B., 2017, ApJ, 850, 128
Rocchetto M., Waldmann I. P., Venot O., Lagage P. O., Tinetti G., 2016, ApJ, 833, 120
Rodgers C. D., 2000, Inverse Methods for Atmospheric Sounding: Theory and Practice, doi:10.1142/3171.
Roellig T. L., et al., 2004, The Astrophysical Journal Supplement Series, 154, 418
Rogers T. M., 2017, Nature Astronomy, 1, 0131
Rogers L. A., Bodenheimer P., Lissauer J. J., Seager S., 2011, ApJ, 738, 59
Rosenblatt F., 1971, Icarus, 14, 71
Rothman L. S., et al., 1998, Journal of Quantitative Spectroscopy and Radiative Transfer, 60, 665
Rothman L. S., et al., 2010, Journal of Quantitative Spectroscopy and Radiative Transfer, 111, 2139
Rowan D., et al., 2016, ApJ, 817, 104
Rumble J. R., 2018, CRC Handbook of Chemistry and Physics (99th edition)
Ryabchikova T., Piskunov N., Kurucz R. L., Stempels H. C., Heiter U., Pakhomov Y., Barklem P. S., 2015, Physica Scripta, 90, 054005
Safronov V. S., 1972, Evolution of the protoplanetary cloud and formation of the earth and planets.
Sahlmann J., Lazorenko P. F., Ségransan D., Martín E. L., Queloz D., Mayor M., Udry S., 2013, Astronomy and Astrophysics, 556, A133
Sahu K. C., et al., 2006, Nature, 443, 534
Santerne A., et al., 2016, Astronomy and Astrophysics, 587, A64
Sato M., Hansen J. E., 1979, Journal of Atmospheric Sciences, 36, 1133
Saumon D., Marley M. S., Cushing M. C., Leggett S. K., Roellig T. L., Lodders K., Freedman R. S., 2006, ApJ, 647, 552
Schäfer U., Yang C.-C., Johansen A., 2017, Astronomy and Astrophysics, 597, A69
Schlawin E., Greene T. P., Line M., Fortney J. J., Rieke M., 2018, AJ, 156, 40
Schneider J., Dedieu C., Le Sidaner P., Savalle R., Zolotukhin I., 2011, Astronomy and Astrophysics, 532, A79
Schreier F., 1992, Journal of Quantitative Spectroscopy and Radiative Transfer, 48, 743
Schwartz J. C., Cowan N. B., 2015, MNRAS, 449, 4192

Schwarz H., Brogi M., de Kok R., Birkby J., Snellen I., 2015, Astronomy and Astrophysics, 576, A111
Schwenke D. W., 1998, Faraday Discussions, 109, 321
Seager S., 2010, Exoplanet Atmospheres: Physical Processes
Seager S., Sasselov D. D., 1998, ApJ, 502, L157
Seager S., Sasselov D. D., 2000, ApJ, 537, 916
Seager S., Kuchner M., Hier-Majumder C. A., Militzer B., 2007, ApJ, 669, 1279
Sedaghati E., et al., 2017, Nature, 549, 238
See T. J. J., 1895, AJ, 15, 180
Sellke T., Bayarri M. J., Berger J. O., 2001, The American Statistician, 55, 62
Sengupta S., 1975, JQSRT, 15, 159
Shabram M., Fortney J. J., Greene T. P., Freedman R. S., 2011, ApJ, 727, 65
Shallue C. J., Vanderburg A., 2018, AJ, 155, 94
Sharp C. M., Burrows A., 2007, The Astrophysical Journal Supplement Series, 168, 140
Sheppard K. B., Mandell A. M., Tamburo P., Gand hi S., Pinhas A., Madhusudhan N., Deming D., 2017, ApJ, 850, L32
Showman A. P., Guillot T., 2002, Astronomy and Astrophysics, 385, 166
Showman A. P., Polvani L. M., 2011, ApJ, 738, 71
Showman A. P., Cooper C. S., Fortney J. J., Marley M. S., 2008, ApJ, 682, 559
Showman A. P., Fortney J. J., Lian Y., Marley M. S., Freedman R. S., Knutson H. A., Charbonneau D., 2009, ApJ, 699, 564
Shporer A., Hu R., 2015, AJ, 150, 112
Silvotti R., et al., 2007, Nature, 449, 189
Simos T. E., et al., 2006, Chemical Modelling. SPR - Chemical Modelling Vol. 4, The Royal Society of Chemistry, doi:10.1039/9781847555267
Sing D. K., López-Morales M., 2009, Astronomy and Astrophysics, 493, L31
Sing D. K., Vidal-Madjar A., Désert J. M., Lecavelier des Etangs A., Ballester G., 2008, ApJ, 686, 658
Sing D. K., et al., 2011, Astronomy and Astrophysics, 527, A73
Sing D. K., et al., 2013, MNRAS, 436, 2956
Sing D. K., et al., 2015, MNRAS, 446, 2428
Sing D. K., et al., 2016, Nature, 529, 59
Skemer A. J., et al., 2012, ApJ, 753, 14
Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, Vol. 735, American Institute of Physics Conference Series. pp 395–405, doi:10.1063/1.1835238
Sneep M., Ubachs W., 2005, Journal of Quantitative Spectroscopy and Radiative Transfer, 92, 293
Snellen I. A. G., Brown A. G. A., 2018, Nature Astronomy, 2, 883
Snellen I. A. G., Albrecht S., de Mooij E. J. W., Le Poole R. S., 2008, Astronomy and Astrophysics, 487, 357
Snellen I. A. G., de Mooij E. J. W., Albrecht S., 2009, Nature, 459, 543
Snellen I. A. G., de Kok R. J., de Mooij E. J. W., Albrecht S., 2010, Nature, 465, 1049

Snellen I. A. G., Brandl B. R., de Kok R. J., Brogi M., Birkby J., Schwarz H., 2014, Nature, 509, 63
Snellen I., et al., 2015, Astronomy and Astrophysics, 576, A59
Soboczenski F., et al., 2018, arXiv e-prints, p. arXiv:1811.03390
Song I., Schneider G., Zuckerman B., Farihi J., Becklin E. E., Bessell M. S., Lowrance P., Macintosh B. A., 2006, ApJ, 652, 724
Sousa-Silva C., Al-Refaie A. F., Tennyson J., Yurchenko S. N., 2015, MNRAS, 446, 2337
Sousa S. G., et al., 2008, Astronomy and Astrophysics, 487, 373
Sousa S. G., Santos N. C., Israelian G., Mayor M., Udry S., 2011, Astronomy and Astrophysics, 533, A141
Southworth J., Wheatley P. J., Sams G., 2007, MNRAS, 379, L11
Spake J. J., et al., 2018, Nature, 557, 68
Sparks W. B., Ford H. C., 2002, ApJ, 578, 543
Speagle J. S., 2019, arXiv e-prints, p. arXiv:1904.02180
Spergel D., et al., 2015, arXiv e-prints, p. arXiv:1503.03757
Spiegel D. S., Silverio K., Burrows A., 2009, ApJ, 699, 1487
Sromovsky L. A., Fry P. M., Kim J. H., 2011, Icarus, 215, 292
Stevens D. J., Gaudi B. S., 2013, Publications of the Astronomical Society of the Pacific, 125, 933
Stevenson K. B., et al., 2014, Science, 346, 838
Stevenson K. B., Bean J. L., Seifahrt A., Gilbert G. J., Line M. R., Désert J.-M., Fortney J. J., 2016, ApJ, 817, 141
Stevenson K. B., et al., 2017, AJ, 153, 68
Strand K. A., 1943, PASP, 55, 29
Struve O., 1952, The Observatory, 72, 199
Sudarsky D., Burrows A., Pinto P., 2000, ApJ, 538, 885
Sudarsky D., Burrows A., Hubeny I., 2003, ApJ, 588, 1121
Suissa G., Chen J., Kipping D., 2018, MNRAS, 476, 2613
Sullivan P. W., et al., 2015, ApJ, 809, 77
Sumi T., MOA OGLE Collaboration 2011, in AAS/Division for Extreme Solar Systems Abstracts. p. 1.03
Swain M. R., et al., 2009, ApJ, 704, 1616
Tashkun S. A., Perevalov V. I., 2011, Journal of Quantitative Spectroscopy and Radiative Transfer, 112, 1403
Tashkun S. A., Perevalov V. I., Lavrentieva N. N., 2016, Journal of Quantitative Spectroscopy and Radiative Transfer, 177, 43
Tennyson J., Yurchenko S. N., 2017, Molecular Astrophysics, 8, 1
Tennyson J., Yurchenko S., 2018, Atoms, 6, 26
Tennyson J., et al., 2016, Journal of Molecular Spectroscopy, 327, 73
Thiabaud A., Marboeuf U., Alibert Y., Leya I., Mezger K., 2015, Astronomy and Astrophysics, 574, A138
Tobin J. J., et al., 2015, ApJ, 805, 125

Todorov K. O., Line M. R., Pineda J. E., Meyer M. R., Quanz S. P., Hinkley S., Fortney J. J., 2016, ApJ, 823, 14
Torres G., Winn J. N., Holman M. J., 2008, ApJ, 677, 1324
Trotta R., 2008, Contemporary Physics, 49, 71
Trotta R., 2017, arXiv e-prints, p. arXiv:1701.01467
Tsiaras A., et al., 2016a, ApJ, 820, 99
Tsiaras A., Waldmann I. P., Rocchetto M., Varley R., Morello G., Damiano M., Tinetti G., 2016b, ApJ, 832, 202
Tsiaras A., et al., 2018, AJ, 155, 156
Tuomi M., Anglada-Escudé G., Gerlach E., Jones H. R. A., Reiners A., Rivera E. J., Vogt S. S., Butler R. P., 2013, Astronomy and Astrophysics, 549, A48
Turrini D., Nelson R. P., Barbieri M., 2015, Experimental Astronomy, 40, 501
Udalski A., Szymanski M., Kaluzny J., Kubiak M., Mateo M., 1992, Acta Astronomica, 42, 253
Underwood D. S., Tennyson J., Yurchenko S. N., Huang X., Schwenke D. W., Lee T. J., Clausen S., Fateev A., 2016, MNRAS, 459, 3890
Unsöld A., 1955, Physik der Sternatmospharen, MIT besonderer Berucksichtigung der Sonne.
Vanderspek R., et al., 2018, arXiv e-prints,
Venot O., Hébrard E., Agúndez M., Dobrijevic M., Selsis F., Hersant F., Iro N., Bounaceur R., 2012, Astronomy and Astrophysics, 546, A43
Venot O., Hébrard E., Agúndez M., Dobrijevic M., Selsis F., Hersant F., Iro N., Bounaceur R., 2013, in Trigo-Rodriguez J. M., Raulin F., Muller C., Nixon C., eds, Vol. 35, The Early Evolution of the Atmospheres of Terrestrial Planets. p. 67, doi:10.1007/978-1-4614-5191-4_6
Vidal-Madjar A., Lecavelier des Etangs A., Désert J. M., Ballester G. E., Ferlet R., Hébrard G., Mayor M., 2003a, Nature, 422, 143
Vidal-Madjar A., Lecavelier des Etangs A., Désert J. M., Ballester G. E., Ferlet R., Hébrard G., Mayor M., 2003b, Nature, 422, 143
Vigan A., et al., 2017, Astronomy and Astrophysics, 603, A3
Wahl S. M., et al., 2017, Geophysical Research Letters, 44, 4649
Wakeford H. R., Sing D. K., 2015, Astronomy and Astrophysics, 573, A122
Wakeford H. R., Sing D. K., Evans T., Deming D., Mandell A., 2016, ApJ, 819, 10
Wakeford H. R., et al., 2017, Science, 356, 628
Wakeford H. R., Wilson T. J., Stevenson K. B., Lewis N. K., 2019, Research Notes of the American Astronomical Society, 3, 7
Waldmann I. P., 2016, ApJ, 820, 107
Waldmann I. P., Tinetti G., Rocchetto M., Barton E. J., Yurchenko S. N., Tennyson J., 2015a, ApJ, 802, 107
Waldmann I. P., Rocchetto M., Tinetti G., Barton E. J., Yurchenko S. N., Tennyson J., 2015b, ApJ, 813, 13
Ward W. R., 1997, Icarus, 126, 261
Warner B., 1968, MNRAS, 139, 115

Weidenschilling S. J., Marzari F., 1996, Nature, 384, 619
Welbanks L., Madhusudhan N., 2019, arXiv e-prints, p. arXiv:1904.05356
Wende S., Reiners A., Seifahrt A., Bernath P. F., 2010, Astronomy and Astrophysics, 523, A58
Wiese W. L., Smith M. W., Glennon B. M., 1966, Atomic transition probabilities. Vol.: Hydrogen through Neon. A critical data compilation. US Government Printing Office
Wilson P. A., et al., 2015, MNRAS, 450, 192
Winn J. N., 2010, arXiv e-prints, p. arXiv:1001.2010
Winn J. N., 2018, Planet Occurrence: Doppler and Transit Surveys. p. 195, doi:10.1007/978-3-319-55333-7_195
Winn J. N., Fabrycky D. C., 2015, Annual Review of Astronomy and Astrophysics, 53, 409
Wittenmyer R. A., et al., 2016, ApJ, 819, 28
Woitke P., Helling C., Hunter G. H., Millard J. D., Turner G. E., Worters M., Blecic J., Stock J. W., 2018, Astronomy and Astrophysics, 614, A1
Wolszczan A., Frail D. A., 1992, Nature, 355, 145
Wong I., et al., 2016, ApJ, 823, 122
Wong A., Yurchenko S. N., Bernath P., Müller H. S. P., McConkey S., Tennyson J., 2017, MNRAS, 470, 882
Wood M. P., Lawler J. E., Sneden C., Cowan J. J., 2013, ApJs, 208, 27
Wright J. T., Marcy G. W., Howard A. W., Johnson J. A., Morton T. D., Fischer D. A., 2012, ApJ, 753, 160
Yadin B., Veness T., Conti P., Hill C., Yurchenko S. N., Tennyson J., 2012, MNRAS, 425, 34
Yan Z.-C., Drake G. W. F., 1995, Phys. Rev. A, 52, 4316
Yan F., Fosbury R. A. E., Petr-Gotzens M. G., Zhao G., Wang W., Wang L., Liu Y., Pallé E., 2015, International Journal of Astrobiology, 14, 255
Yang J., Cowan N. B., Abbot D. S., 2013, ApJ, 771, L45
Youdin A. N., Goodman J., 2005, ApJ, 620, 459
Yurchenko S. N., Barber R. J., Tennyson J., 2011, MNRAS, 413, 1828
Yurchenko S. N., Blissett A., Asari U., Vasilios M., Hill C., Tennyson J., 2016, MNRAS, 456, 4524
Yurchenko S. N., Amundsen D. S., Tennyson J., Waldmann I. P., 2017, Astronomy and Astrophysics, 605, A95
Yurchenko S. N., Sinden F., Lodi L., Hill C., Gorman M. N., Tennyson J., 2018a, MNRAS, 473, 5324
Yurchenko S. N., Bond W., Gorman M. N., Lodi L., McKemmish L. K., Nunn W., Shah R., Tennyson J., 2018b, MNRAS, 478, 270
Yurchenko S. N., Williams H., Leyland P. C., Lodi L., Tennyson J., 2018c, MNRAS, 479, 1401
Yurchenko S. N., Al-Refaie A. F., Tennyson J., 2018d, Astronomy and Astrophysics, 614, A131
Zahnle K. J., Marley M. S., 2014, ApJ, 797, 41

Zahnle K., Marley M. S., Fortney J. J., 2009, arXiv e-prints, p. arXiv:0911.0728
Zeng L., Jacobsen S. B., 2017, ApJ, 837, 164
Zhang J., Kempton E. M. R., Rauscher E., 2017, ApJ, 851, 84
Zhu W., Petrovich C., Wu Y., Dong S., Xie J., 2018, ApJ, 860, 101
Zingales T., Waldmann I. P., 2018, AJ, 156, 268
Zurlo A., et al., 2016, Astronomy and Astrophysics, 587, A57
de Kok R. J., Brogi M., Snellen I. A. G., Birkby J., Albrecht S., de Mooij E. J. W., 2013, Astronomy and Astrophysics, 554, A82
de Kok R. J., Birkby J., Brogi M., Schwarz H., Albrecht S., de Mooij E. J. W., Snellen I. A. G., 2014, Astronomy and Astrophysics, 561, A150
de Pater I., Lissauer J. J., 2001, Planetary Sciences
de Wit J., Seager S., 2013, Science, 342, 1473
van de Kamp P., 1963, AJ, 68, 295
van de Kamp P., Lippincott S. L., 1951, AJ, 56, 49
von der Goltz D., Hansen W., Richter J., 1984, Physica Scripta, 30, 244