Title
Advancing Retrievals of Exoplanetary Spectra in the Era of Large Space-Based Telescopes

Author
Feng, Ying

Publication Date
2020

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

ADVANCING RETRIEVALS OF EXOPLANETARY SPECTRA IN THE ERA OF LARGE SPACE-BASED TELESCOPES

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

ASTRONOMY AND ASTROPHYSICS

by

Ying Feng

June 2020

The Dissertation of Ying Feng is approved:

Jonathan J. Fortney, Chair

Michael R. Line

Andrew Skemer

Xi Zhang

Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Advancing Retrievals of Exoplanetary Spectra in the Era of Large Space-Based Telescopes

by

Ying Feng

Exoplanet atmospheres tell the story of diverse worlds: what they are made of and how they came to be. We use theoretical models to make sense of the narrative encoded in each atmospheric data set. In particular, we extract information with retrievals, which couple statistical tools with high signal-to-noise spectral data to derive and estimate atmospheric properties. The inferred atmospheric structure and molecular abundances then influence our understanding of the Solar System in the context of exoplanets.

Retrievals have become an essential tool in both understanding existing data and quantitatively informing the needs of future missions. Consequently, this thesis focuses on the interplay between data and models, the core components of a retrieval.

First, I demonstrate the need to evaluate model assumptions in order to extract meaningful constraints from spectra. While varying in sophistication, most model-spectra comparisons fundamentally assume "1D" model physics. However, we know that planetary atmospheres are inherently "3D" in their structure and composition. Within a Bayesian retrieval framework, I show how the assumption of a single 1D thermal profile can bias our interpretation of the thermal emission spectrum of an unresolved hot Jupiter's atmosphere that is composed of two thermal profiles. I extend

the application to full spectroscopic phase curves. For modern data, I reveal that the constraint for water vapor abundance is robust independent of model setup; however, methane is artificially well-constrained to incorrect values. Furthermore, I find that the 1D setup is insufficient at fitting data that the *James Webb Space Telescope* may measure, causing many abundance biases.

Next, I develop an inverse modeling framework to estimate the science return of proposed missions that aim to perform reflected light spectroscopy of rocky exoplanets around Sun-like stars. By combining an albedo model, an instrument noise model, and a Bayesian inference tool, I explore retrievals of atmospheric and bulk properties as a function of data signal-to-noise ratio (SNR) and resolution (R). I present the recommended R and SNR combinations to achieve detection or constraint of key indicators of habitability on rocky planets such as water vapor, oxygen, and ozone.

Acknowledgments

At some point along the way, I heard the saying, "a PhD is the greatest gift you can give yourself." In some respects, that has been true – I never could have anticipated the experiences and growth (sometimes more of the personal than the scientific) from the last six years. Looking back on all the difficult and joyous moments of grad school, I am profoundly grateful for the gifts of support and love from those in my life.

Firstly, I thank my parents. I had the best teacher growing up – my dad, who taught me to see the beauty in knowledge and curiosity – and nothing quite compares to coming home to my mom's cooking. Even continents apart, I feel their care and encouragement. I also thank my sister, without whom I would not have seen *Contact* or eventually become interested in astronomy. She has always been there for me.

I was lucky to have received an outstanding astronomy education in undergrad at Penn State, which set me up for success at UCSC, and I am indebted to teachers like Don Schneider and Mike Eracleous. In addition, I am grateful for Jason Wright, my undergraduate adviser. To my dear friends Richa, Rohana, Gloria, Sarah T, and Tesia: I deeply cherish our bonds that remain strong in spite of distance and time.

Next, I thank my adviser, Jonathan Fortney, for his mentorship, compassion, and patience. He gave me the space to explore career options and always listened when I was struggling. He provided me with the opportunities to promote my work and shared the human side of being an academic – I know I will be there at the next group barbecue. Without Jonathan, I would not have been able to work with two of my other phenomenal mentors, Mike Line and Ty Robinson. Mike taught me everything I know about retrievals and supported me as I grew to harness my own expertise. With Ty, I got to spend time thinking about habitability, and he was always kind and patient as I asked question after question. I'd also like to thank my thesis committee members Andy Skemer, who taught me so much about astrostatistics, and Xi Zhang, for his friendliness and the enthusiastic energy he brought to group meetings. I also thank Ruth Murray-Clay for being on my qualifying committee. Furthermore, I thank Natasha Batalha for the long chats in our offices about life and science. Special thanks to Claire Dorman and Megan Shabram for their time and wisdom while I tried to figure out the next steps in my career.

I am also fortunate to have found community among the redwoods at UCSC. I would first like to thank a special group of friends, many of whom wore multiple hats (from cohort to housemate to officemate): Tiffany, Eric, Alexa, Asher, Sarah M, Brittany, Ari, Enia, Viraj, and Layla. Even as I became a more senior grad student, I still leaned on the wonderful grads who came before me: Angie, Rosalie, Claire, Jenn, Katie, Jieun, Emily, Camille, Anna, Eric L, Katie, Marie, Anna, and Caitlin. I am proud to have witnessed the growth of the UCSC astronomy department, and I am thankful to the members of our department who made the space welcoming, especially my fellow grad students, for our shared adventures, outings, and consumption of delicious baked goods. Thanks to the fun and supportive Fortney group: Caroline, Chris, Daniel, Diana, Maggie, Callie, Molly, Naor, and Adam. To Writing Together: getting my thesis done would have been that much harder without our gatherings.

I would especially like to thank Tiffany Hsyu, who has been in the majority of my memories made during grad school. My experience would have been completely different without her. From tea time to camping, Europe to PIE, from doing homework to putting together our theses, I am glad to have shared the journey with such a thoughtful and fun confidant.

Even when grad school did become isolating, I was blessed to have had a strong and amazing group of women that lifted me up. Here's to The Best Group Ever, for the vulnerability, empathy, and laughs we shared over the last several years. I thank Audrey for bringing us together. I also thank Sharon and Digby for their understanding and support as I got to know myself better.

Finally, I want to thank Felipe Ardila, for his belief in me, for helping me grow, and for bringing me into his loving family. Every day, I am the happiest as we wander and wonder through the vastness of space, enjoying the birds, movies, and snacks along the way.

Published Material

The text of this dissertation includes reprints of the following previously published material led by Ying Feng, with the permission of the listed co-authors.

Chapter 2 was published in the literature as Feng et al. 2016, ApJ, 829, 52. I was responsible for developing the models, performing the analysis, preparing the figures, and writing the manuscript. I thank my co-authors Michael R. Line, Jonathan J. Fortney, and Vivien Parmentier for their guidance and feedback. I thank my co-authors Kevin B. Stevenson, Jacob Bean, and Laura Kreidberg for providing the observations made with the NASA/ESA *Hubble Space Telescope* and the NASA *Spitzer Space Tele-*

scope.

Chapter 4 was published in the literature as Feng et al. 2018, AJ, 155, 200. I was responsible for developing the models, performing the analysis, preparing the figures, and writing the manuscript. I thank my co-author Tyler D. Robinson for his generous guidance throughout the course of this research. I thank my co-authors Jonathan J. Fortney, Roxana E. Lupu, and Mark S. Marley for their help with model development. I thank my co-authors Nikole K. Lewis, Bruce Macintosh, and Michael R. Line for their feedback and support.

Scientific Acknowledgments

I gratefully acknowledge research support from the National Science Foundation Graduate Research Fellowship under grant DGE1339067 and the ARCS Foundation Scholarship. The computation for my dissertation research was performed by the UCSC Hyades supercomputer, which was supported by the National Science Foundation (award number AST-1229745) and the University of California, Santa Cruz. I thank Brant Robertson and Shawfeng Dong for their maintenance and support of Hyades over the years. My work was further made possible by support from the UCSC Other Worlds Laboratory, the WFIRST Science Investigation Team program, and the 2016 Kavli Summer Program in Astrophysics.

For my sister

Life has been easier to navigate

with her as my role model

Chapter 1

Introduction

The end of the 20th Century saw an expansion in humanity's exploration of the universe: ours was no longer the only known solar system in the Galaxy. Thousands of extrasolar planets, or exoplanets, have since been discovered. As we enter the 2020s and beyond, we are launching into in-depth characterization of these distant worlds.

From super Earths to hot Jupiters, tightly-packed multi-planet systems to massive, self-luminous young planets, exoplanets continually present conditions both similar and unfamiliar to our own planetary system. Without the ability to send probes like we can in the Solar System, we turn to the power of remote sensing. We rely on discovery methods to infer bulk properties such as mass and radius. We rely on the atmospheres to deduce chemical composition, atmospheric circulation, and the presence of aerosols. But in any flavor of exoplanet characterization, we rely foremost on the synergy between telescope observations and theoretical models.

As described in this chapter, theory and data have mutually pushed each other to improve. A bountiful trove of unprecedented data will follow the launch of future large space-based telescopes. All the proposed missions considered by the forthcoming astronomy decadal survey highlight exoplanet characterization as a key area. While it is difficult to anticipate the phenomena future data will capture, we can still prepare by improving existing models.

Through this thesis, I present several frameworks that advance statistical modeling of exoplanet atmosphere observations. I also show how these models quantitatively drive the decisions in the design of upcoming missions. First, in this chapter, I present relevant characterization methods, modeling approaches, and findings followed by a summary of questions sparked by future space telescopes and how this thesis intends to address them.

1.1 Fantastic Atmospheres and How to Study Them

The first planet discovered around another solar-type star exceeded expectations in many ways. One surprise was its position relative to its host star. 51 Pegasi b (Mayor & Queloz 1995), at 0.46 Jupiter masses, has an orbital semimajor axis of ~ 0.05 au.

Core accretion, a commonly proposed formation mechanism for the Solar System, fosters the idea that gas giants like Jupiter were able to reach runaway accretion and so retain a thick hydrogen and helium (H/He)-dominated atmosphere. Effective accretion is possible beyond the ice line, where volatile compounds such as water or carbon dioxide can condense into ices and provide more accretion material. When the

Solar System was forming, for instance, the water ice line was approximately at 3 au (e.g., Qi et al. 2013), well within Jupiter's current orbital distance. A massive planet like 51 Pegasi b existing so near its host star, where it would be too hot for volatiles to remain solid, challenged this theory, invoking new theories of migration (Lin et al. 1996). This has also earned planets of this class the title of "hot Jupiter." Hot Jupiters can be tidally locked; the intense irradiation on the permanent dayside can lead to large day-night temperature contrasts.

Novel results from the discovery of exoplanets have refreshed our perspective on planetary systems in the Galaxy time after time. Two of these discovery methods are now our main pathways to exoplanet characterization: transits and direct imaging.

1.1.1 Transit Spectroscopy

The transit method detects planets which cross in front of their host star from the observer's point of view. Planets with short-periods are prime targets for transit surveys. Transiting planet atmospheres can be studied through two primary techniques: transmission spectroscopy and emission spectroscopy. Figure 1.1 illustrates when these techniques are applied during the orbit of a transiting planet.

Transmission spectroscopy occurs during transit, as we observe the stellar flux filtering through the planet's atmosphere. Opacity sources (absorbing or scattering atoms and molecules, cloud/haze particles) impede the transmission of starlight, causing the planet to appear "larger" at certain wavelengths. The relative transit depth as a function of wavelength corresponds to a transmission spectrum. Transmission spectroscopy probes the day/night limbs of the atmosphere, giving us access to pressures

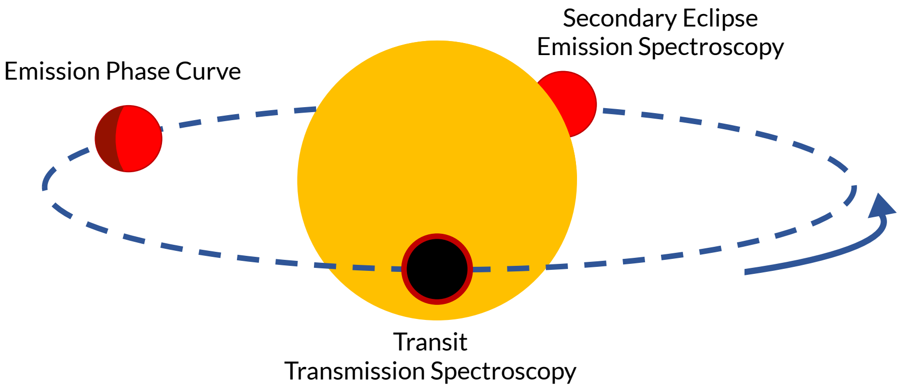


Figure 1.1: A schematic of a transiting exoplanet (in black) and its atmosphere (in red). At transit, we perform transmission spectroscopy. When the planet is behind the host star, we perform emission spectroscopy at secondary eclipse. Measuring the emission signature over the orbit constructs an emission phase curve.

higher up in the atmosphere. While transmission spectra are not investigated in this thesis, they have yielded impactful findings. In particular, they have revealed the prominence of hazes and clouds in exoplanet atmospheres (Kreidberg et al. 2014b; Sing et al. 2016).

Emission spectroscopy, featured in Chapters 2 and 3, measures the thermal emission from an exoplanet. For a transiting planet, this means monitoring the flux from the planet-and-star system over its orbit. We can obtain the planet-to-star flux ratio,

$$\frac{F_p}{F_s} = \frac{R_p^2\, I(\lambda, T_p)}{R_s^2\, I(\lambda, T_s)}, \tag{1.1}$$

where $I(\lambda, T)$ corresponds to the intensity (e.g., blackbody) as a function of wavelength λ and temperature T. When the planet is behind the star, we observe the secondary eclipse and can infer the flux contribution of the planet's day side. Emission spectroscopy probe deeper into the atmosphere than transmission spectroscopy. Importantly, these spectra are sensitive to both the atmospheric composition and thermal structure. Furthermore, when we have the flux ratios as a function of wavelength for the full orbit, we have an emission spectroscopic phase curve. Phase curves provide us information about the atmosphere as a function of longitude.

Space-based observatories such as the *Hubble Space Telescope*(HST) and the *Spitzer Space Telescope* (Spitzer) pioneered atmospheric observations in the optical, UV, and infrared wavelengths. As Figure 1.2 shows, almost a hundred exoplanets have had their atmospheres detected by at least one technique (emission or transmission) in at least one spectral band. Many do not have more than photometric observations, but tens have high-precision spectra with data typically spanning $\sim 0.5-2\mu$m, depending on the instrument. Modeling efforts, described in more detail in Section 1.2, are providing initial constraints on molecular and atomic species in the atmosphere, thermal profiles, cloud/haze properties, and atmospheric dynamics.

Unlike on Solar System gas giants, where the temperature is too low, water (H_2O) on hot Jupiters is in its gas phase. We can link its constraint to ratios of elemental abundances, such as the carbon-to-oxygen ratio (C/O). Beyond the iceline, gas becomes oxygen-poor once water, the dominant carrier of oxygen, becomes solid (Öberg et al. 2011) – the solids are thus oxygen-rich. The composition of a planet depends on the relative accretion of gas and solids. An estimate of the atmospheric C/O ratio, then, has implications for the planet's formation location.

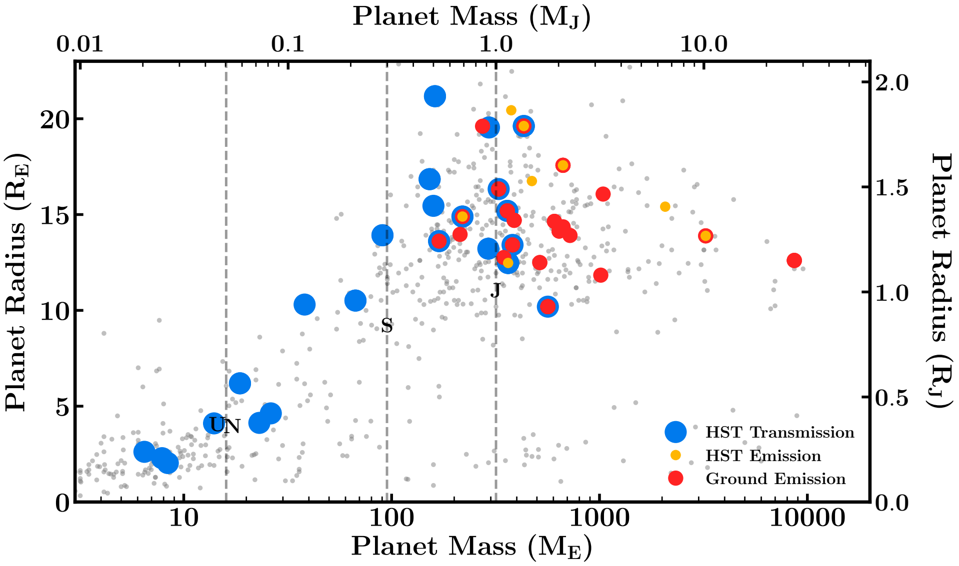


Figure 1.2: Masses vs. radii for known exoplanets. A subset has been highlighted to indicate transiting planets with observed atmospheres. Blue/yellow dots: planets with transmission/emission data with the *Hubble Space Telescope* (HST). Red dots: planets with observations from ground-based facilities. Saturn, Jupiter, Uranus, and Neptune are marked in black with their initials and vertical dashed lines. Credit: Madhusudhan (2019), with data from the NASA Exoplanet Archive at `https://exoplanetarchive.ipac.caltech.edu/`.

A complicating factor to measuring molecular abundances is disequilibrium chemistry. When chemical reaction rates are slower compared to dynamic processes (e.g., vertical mixing), we should not expect chemical equilibrium to hold. For example, Jupiter's upper atmosphere appears to have higher carbon monoxide (CO) abundance than predicted by chemical equilibrium. One explanation is the upwelling of CO from deeper in the atmosphere due to convection, effectively becoming quenched when chemical reactions slow at lower temperatures (Bézard et al. 2002; Cooper & Showman 2006). The accurate inference of gas abundances in hot Jupiters is key to identifying such processes, if present.

1.1.2 Direct Imaging Spectroscopy

Direct imaging involves suppressing the light from a host star to image orbiting planets and measure their spectra directly. Starlight suppression is immensely challenging. The contrast at visible wavelengths between Jupiter and the Sun is $\sim 10^{-9}$, and that of the Earth-Sun system is $\sim 10^{-10}$. Depending on the host star's own spectrum, going to redder wavelengths helps decrease the contrast. Ground-based instruments have successfully imaged massive, young exoplanets at wide orbital separations (> 10 au), with contrast levels going down to $\sim 10^{-6}$. These young planets are still cooling off from formation and are self-luminous; their characterization occurs typically occurs in the infrared (e.g., Macintosh et al. 2014; Rajan et al. 2017).

While this is the science of today, this thesis focuses on the science of the future: reflected light spectra from terrestrial planets in visible wavelengths. The goal is to eventually characterize Earth-like planets around Sun-like stars to better understand our Earth in the broader context of potentially habitable systems. Reflected light spectroscopy of these objects will be likely only be done in space, as ground-based facilities may not reach the necessary contrast ratios at orbital separations as close in as such a star's Habitable Zone (HZ). The HZ (Kasting et al. 1993; Kopparapu et al. 2013) is the region around a star where a terrestrial planet can sustain liquid water on its surface. Planets in the HZ around Sun-like stars will be difficult to study with transit spectroscopy because transits are infrequent, and the transit probability (based on stellar radius and planet semimajor axis as $\frac{R_{\rm star}}{a}$) and transit depth (area of the star blocked by planet) are low.

Reflected-light spectra of planets, like emission spectra, can probe deeper into atmospheres; for rocky planets, this may even be down to the surface. These spectra represent light directly from the atmosphere (or starlight that has entered and scattered back out), unlike transmission spectra which consist of starlight traversing through longer optical paths through the upper atmosphere. In addition, water vapor, needed to establish habitable conditions, is found lower in the atmosphere on Earth than a transmission observation may probe. A prominent water band sits at $0.87 - 1.05\mu$m. Beyond this feature, the visible light regime ($0.4 - 1\mu$m) hosts a wealth of information regarding a terrestrial planet's atmosphere, as discussed in Chapter 4.

A few challenges to interpreting reflected light spectra include the lack of knowledge about planetary bulk properties (orbit, radius, mass). However, follow-up observations (e.g., proper motion studies) can complement spectroscopic studies in providing constraints on bulk parameters.

1.2 Interpreting Atmospheric Data

Once exoplanet spectra have been measured, the next challenge lies in their interpretation. A spectrum, or even photometric points, encodes the intricate processes occurring within a distant exoplanet's atmosphere. Modeling an atmosphere is, in a sense, like judging a book by its cover. However, astronomers are well-equipped to gain insight through self-consistent or inverse models.

1.2.1 Self-Consistent Models

A self-consistent model computes (i.e., predicts) spectra of atmospheres based on assumptions about macroscopic properties like gravity, irradiation, and elemental abundances. Self-consistent models range in complexity, from one-dimensional (1D) to three-dimensional (3D).

A 1D model is typically plane-parallel and maintains thermochemical and radiative-convective equilibrium (net flux is conserved). Radiative transfer determines how radiation interacts with matter, like stellar radiation entering a planet's atmosphere, or the conversion of internal heat flux from the interior to radiation which exits the atmosphere. By solving radiative transfer for a pressure grid, the resulting thermal profile and chemical composition then yield a spectrum. We can compare the predicted spectra from 1D models with different initial assumptions to observed data and get an intuition of a planet's atmosphere. Chemical abundances, for instance, are set initially depending on the metallicity or C/O ratio assumed (e.g., Mollière et al. (2015)). Opacity sources (gas species or clouds and hazes in the atmosphere), which play a key role in the interaction with radiation, depend on these abundances.

A 3D model is also known as a general circulation model (GCM). GCMs take as input bulk planetary properties and the irradiation field to solve equations of fluid motion for the full 3D atmospheric structure, from chemistry to dynamics. Modelers have coupled the results of the GCMs, which include the thermal profiles as a function of planetary longitude and latitude, with a radiative transfer scheme to produce spectra. These models are well-suited for examining phase-resolved spectra and thermal phase curves. In fact, GCMs predicted strong equatorial jets in hot Jupiter atmospheres which can cause the hot spot of the irradiated day side to shift from the sub-stellar point (Showman & Guillot 2002). This has been observed in several planets, such as HD 189733b (Knutson et al. 2008). Based on GCM studies, we anticipate diverse temperature structures and abundances in exoplanet atmospheres.

1.2.2 Inverse Models

Atmospheric retrieval is the process of inferring the atmospheric properties of an exoplanet from its observed spectrum. Since they back out conditions from data, they are also called inverse models. Figure 1.3 shows the key components of a retrieval in relation to one another: a parametric forward model, the observed data, and an optimization algorithm. The optimization scheme – a statistical inference tool – is a means of estimating the model parameters by numerous evaluations of the corresponding generated forward model spectra fitted to the observed data. Parameter estimation performed through Bayesian inference (Section 1.2.3) constructs probability distributions for each parameter.

Parameter estimation also maps out degeneracies between variables. Such a process requires at least an order of 10^5 - 10^6 model evaluations against the data. Computational efficiency becomes crucial in this effort. Retrievals are thus designed to produce spectra far quicker than self-consistent models, enabling wider exploration of parameter space. Self-consistent models, meanwhile, assume to know the atmospheric physics and chemistry conditions in order to reach converged steady state solutions; these assumptions can miss domains that exoplanet atmospheres occupy given their

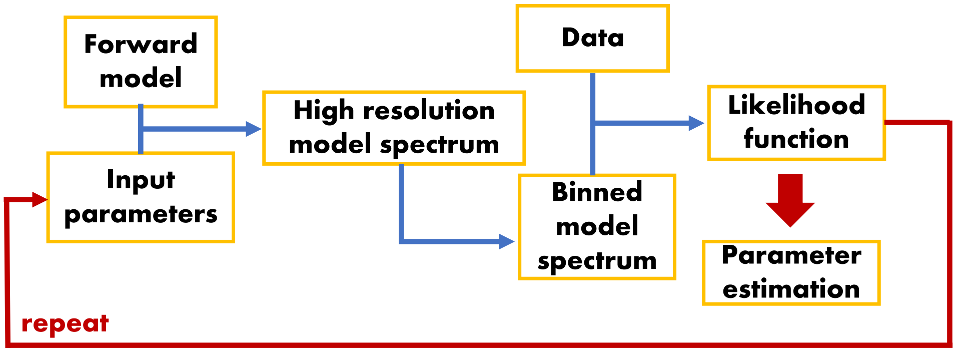


Figure 1.3: Schematic of a retrieval model. The forward model takes input values for its parameters to generate a high resolution spectrum. This spectrum is binned to match the resolution of the observed data. Next, the binned model and the data are given to an optimization scheme. In a Bayesian framework (Section 1.2.3), we repeatedly sample the parameter space and calculate the likelihood function to construct the posterior probability distribution function for the parameters.

diversity (Madhusudhan 2018).

In general, a retrieval model first defines the atmospheric structure (e.g., how temperature, gas concentrations, or cloud/haze presence vary with pressure) and then performs a radiative transfer calculation based on its properties, generating a spectrum. The parameterization of the forward model depends on the observational context. Chapters 2 and 3, for instance, describe the forward modeling of cloud-free hot Jupiter emission spectra. The thermal profile (temperature as a function of pressure) is a key property to model, along with chemical composition (gas mixing ratios). Chapter 4 features reflected light spectra from rocky planets. A different set of molecules and opacities would be in play for terrestrial planets, and modeling these spectra invokes scattering effects.

Early-on exoplanet atmosphere retrieval relied on a grid-search method. In

Madhusudhan & Seager (2009), this entailed defining the parameter space for 10 variables and generating a grid of 10^7 models. A goodness-of-fit contour map can be calculated and provide confidence intervals for each parameter. Such an approach is limited by grid resolution and computational power, especially when trying to explore high-dimensional parameter spaces with strong degeneracies. Furthermore, the confidence intervals do not correspond to actual probability distributions for a variable. A Bayesian framework addresses these issues, and Bayesian inference has since become a prominent method in exoplanet atmosphere inverse modeling (e.g., Madhusudhan et al. 2011; Lee et al. 2012; Line et al. 2013; Benneke 2015). In the following sections, we will examine the setup and implementation of Bayesian retrievals.

1.2.3 Bayesian Inference

Bayesian inference is a widely-adopted approach for parameter estimation. In this context, Bayes' theorem is specifically

$$P(\theta|d, M) = \frac{P(d|\theta, M)P(\theta|M)}{P(d|M)}, \tag{1.2}$$

where θ is the set of parameters for a model M and d is the set of observed data. $P(\theta|d, M)$ describes the posterior probability distribution for θ, i.e., our degree of belief about the value of θ given the measurements of d. $P(d|\theta, M)$ is the likelihood function, or how likely it is that we would obtain the data under the parameters θ for model M. Assuming that the data $(d_1, ..., d_N)$ are measurements with Gaussian-distributed errors $(\sigma_1, ..., \sigma_N)$, as I do in this thesis, the likelihood becomes

$$\mathcal{L}(\theta, M) \equiv P(\theta|d, M) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(d_i - M_i(\theta))^2}{2\sigma_i^2}\right). \tag{1.3}$$

$P(\theta|M)$ represents the prior probability distribution, quantifying our knowledge of θ before taking any data. The $P(d|M)$ term, also called the “evidence”, normalizes the posterior to unity. It is defined as

$$Z \equiv P(d|M) = \int \mathcal{L}(\theta, M) P(\theta|M) d\theta, \tag{1.4}$$

and Z is relevant for comparing competing models. Model comparison is further discussed in Section 1.2.3.

In this thesis, I will employ two implementations of Bayesian inference to numerically obtain posterior distributions: Markov Chain Monte Carlo and Nested Sampling.

Markov Chain Monte Carlo

A Markov Chain Monte Carlo (MCMC) algorithm starts from an initial guess and then chains together a sequence of points in parameter space through random sampling. The probability of the $(n + 1)$-th element in a Markov chain only depends on the probability of the n-th element. Markov chains can converge to a steady state in which the values no longer change with n. The samples drawn at this stage are considered to be from the target distribution, i.e., the joint posterior of all parameters. Typically, a transition probability defines the chance of moving from point $\theta^{(n)}$ to $\theta^{(n+1)}$ within the parameter space, thus adding elements to the chain (or not).

Several techniques exist to guide the exploration of parameter space. One variant of MCMC is known as affine-invariant ensemble sampling (Goodman & Weare 2010), which has been developed by Foreman-Mackey et al. (2013) into the python sampler `emcee`. In Chapter 4, I use `emcee` to perform my retrievals.

Instead of one chain of samples, Goodman & Weare (2010) proposed ensemble samplers. Here, the Markov chain is constructed from the state space of ensembles, where each ensemble, $\vec{X}$ consists of L walkers $\vec{X}_k$. Generally speaking, affine-invariant transformation (e.g., $y = ax + b$) maintains collinearity (e.g., parallel lines are transformed to parallel lines). An affine-invariant algorithm means that the ease of sampling is independent of how isotropic or anisotropic a probability density function is, and this characteristic is implemented when updating the position of the walkers in an ensemble within `emcee`:

$$X_k(n) \to Y = X_j + Z(X_k(n) - X_j), \tag{1.5}$$

where walker X_k is using a complementary walker also from the ensemble, X_j. Z is a random scaling variable from a distribution $g(Z = z)$ which satisfies

$$g(\frac{1}{z}) = zg(z). \tag{1.6}$$

For an N-dimensional parameter space, the chain accepts the proposed jump if r drawn from $[0, 1]$ is less than or equal to $q = \min(1, Z^{N-1}\frac{P(Y)}{P(X_k(n))})$ (the second term contains the ratio of the probabilities of the current and proposed positions). The position of X_k remains as is if the jump is rejected. Foreman-Mackey et al. (2013) lay out the algorithm in more detail, including how `emcee` can easily be parallelized to speed up computation, which is one of its advantages for atmospheric retrievals. Briefly, this jump procedure repeats for all walkers, which concludes one iteration, and the ensemble sampling continues until convergence.

With the posterior distribution constructed, we can obtain the one-dimensional marginal posterior for, e.g., θ_1 among N parameters θ by integrating

$$P(\theta_1|d, M) = \int P(\theta|d, M) d\theta_2 ... d\theta_N. \tag{1.7}$$

In practice, the Markov chains enable us to count the number of samples within bins defined for the range of θ_1 while ignoring the coordinates valued $\theta_2, ..., \theta_N$ to get the one-dimensional posterior. The posterior distribution then provides intervals of constraints (e.g., the 1σ range represents 68% of the samples) for θ_1.

Nested Sampling

Models are not one-size-fits-all. Depending on the data, multiple models may be pertinent. This necessitates the process of model selection. As described in Equation 1.4, the Bayesian evidence Z is used for this purpose. For model M, we can use Bayes' theorem to show that

$$P(M|d) \propto P(M) P(d|M). \tag{1.8}$$

The quantity on the lefthand side is the "model posterior". By taking the ratio of the posterior probability for models M_0 and M_1, we determine

$$\frac{P(M_0|d)}{P(M_1|d)} = B_{01}\frac{P(M_0)}{P(M_1)}, \tag{1.9}$$

where B_{01}, the Bayes factor, compares the models' evidences:

$$B_{01} \equiv \frac{P(d|M_0)}{P(d|M_1)} = \frac{Z_0}{Z_1}. \tag{1.10}$$

We interpret $B_{01} > (<) 1$ as an increase (decrease) of the support for M_0 in favor of M_1, given the data.

However, computing the evidence, a multi-dimensional integral of the likelihood of the data over the parameter space, is complex. Skilling (2004) proposed nested sampling to efficiently calculate the evidence. Nested sampling is also an algorithm that constructs the posterior probability distribution. Its efficiency lies in the transformation of Equation 1.4 into a one-dimensional integral:

$$Z = \int_0^1 \mathcal{L}(X)dX. \tag{1.11}$$

Here, by first defining a prior volume X from $dX = P(\theta|M)d\theta$, we get

$$X(\lambda) = \int_{\mathcal{L}(\theta,M)>\lambda} P(\theta|M)d\theta, \tag{1.12}$$

meaning $X(\lambda)$ represents the volume of parameter space above a level of likelihood as designated by λ. In other words, $\mathcal{L}(\theta, M) = \lambda$ establishes an iso-likelihood contour ("nested" contours) to carve out this volume. As λ increases, the enclosed volume X

decrease from 1 to 0. $\mathcal{L}(X)$ is then the inverse of Equation 1.12, and Equation 1.11 is an integral of a function that is positive and decreasing.

Supposing we evaluate the likelihoods $\mathcal{L}_i = \mathcal{L}(X_i)$ for a sequence of decreasing values X_i (i.e., $0 < X_m < ... < X_1 < X_0 = 1$). We can then numerically approximate Z as a weighted sum with weights w_i ($\frac{1}{2}(X_{i-1} - X_{i+1})$ for the trapezoidal rule):

$$Z = \sum_{i=1}^{m} w_i \mathcal{L}_i. \tag{1.13}$$

Skilling (2004) outlines the algorithm in detail. In a nutshell, we start with a set of N points ("live points") $\theta_1, ..., \theta_N$ from the prior. For each step i in j iterations, we set $\mathcal{L}_i$ to be the lowest of the current likelihood values and increment Z by $w_i \mathcal{L}_i$. The lowest likelihood point is replaced by a new one drawn from $\mathcal{L}(\theta, M) > \mathcal{L}_i$. This particular constraint of how samples are drawn is a challenge to implement in the nested sampling approach.

Randomly drawing from the prior results in a decrease of the acceptance of new samples with decreasing prior volume. Other algorithms were created to address this. MultiNest, an algorithm developed by Feroz & Hobson (2008) and Feroz et al. (2009), is one of them. MultiNest follows from the idea of ellipsoidal nested sampling (Mukherjee et al. 2006). Instead of points, ellipsoidal sampling approximates an iso-likelihood contour with an ellipsoid defined from the covariance matrix of a set of live points (wherein the lowest-likelihood point has been rejected). Within this ellipsoid, new points are selected from the prior until obtaining one with a likelihood greater than the removed likelihood point.

While ellipsoidal nested sampling is efficient for unimodal posterior distributions, Shaw et al. (2007) proposed the use of clusters of active points, each with its own ellipsoidal bound, to better deal with multimodal distributions. MultiNest, short for multimodal nested sampling, improves upon this idea. MultiNest later on became adapted into python (PyMultiNest) by Buchner et al. (2014). Lupu et al. (2016) demonstrated that `emcee` and PyMultiNest perform similarly. Given its ability to determine Z and usefulness for model comparison, PyMultiNest is selected for the retrieval analyses in Chapters 2 and 3.

1.3 Retrievals: State-of-the-Art and the Future

Retrievals have only just begun revealing the nature of exoplanet atmospheres. Many studies characterizing transiting planets make use of the *Hubble Space Telescope* (HST) and the *Spitzer Space Telescope* (Spitzer). Spitzer operated in the infrared, and, after losing its cryogenic cooling, could still perform photometric measurements for exoplanets with its IRAC (Infrared Array Camera) instrument. The bands at 3.6μm and 4.5μm in particular were prolific for atmospheric observations up until Spitzer's decommissioning in early 2020. The phase curves obtained with Spitzer were studied with self-consistent 3D GCMs (e.g., Showman et al. 2009; Dobbs-Dixon & Agol 2013; Wong et al. 2016). However, performing retrievals on only two photometric points would overfit the data and render the interpretation prone to many degeneracies.

The rise of exoplanet atmosphere retrievals coincided with the installation of the Wide Field Camera 3 (WFC3) instrument on HST in 2009. The WFC3 G141 grism, offering spectroscopy at $R \sim 100$ between $1.1 - 1.4\mu$m, has transformed the data landscape for transiting planets. Combined with IRAC data, we have even been able to obtain phase-resolved spectra for several planets.

Based on these data sets, retrievals on emission spectra have successfully constrained H_2O abundance in atmospheres. The H_2O abundance has in turn been used to infer C/O ratios assuming thermal chemical equilibrium (e.g., Line et al. 2016), suggesting solar or sub-solar quantities. Estimates for other molecules, such as CO_2 or CH_4 remain unconstrained under the available data.

Most planets studied with retrievals do not show evidence of thermal inversion. These include HD 209458b (Line et al. 2016) and WASP-43b (Stevenson et al. 2014). Yet, for a class of extremely irradiated planets, called ultra-hot Jupiters (with equilibrium temperature > 2000 K), retrieval studies have found thermal inversions; WASP-121b (Evans et al. 2017) is an example. Because the thermal profile is intricately linked with abundance inference, its accurate reconstruction is important.

Data with more complete and wider wavelength coverage will allow more precise inference of atmospheres. The highly anticipated *James Webb Space Telescope* (JWST), with its 6.5m primary mirror, will have numerous spectroscopic modes that span $0.6 - 28\mu$m. Numerous studies have employed retrievals to examine the potential of JWST observations for transiting planets (Greene et al. 2016; Taylor et al. 2020; Venot et al. 2020). In addition to JWST, the ARIEL mission is the first selected mission by the European Space Agency that will be dedicated to the characterization of exoplanet atmospheres (Eccleston et al. 2016; Tinetti et al. 2016). ARIEL expects to perform photometry and spectroscopy throughout the $0.5 - 7.8\mu$m range for up to 1000 transiting planets (Tinetti et al. 2018).

While the performance of any retrieval depends on the data, in this thesis, I highlight the need to carefully weigh the choice of models used in a retrieval. Recently, the accuracy of the 1D forward model in retrievals has been challenged. Studies examining both transmission and emission spectra are beginning to incorporate 3D effects to see if the inference changes, for a given data set (Line & Parmentier 2016; Feng et al. 2016; Blecic et al. 2017). I show that, as higher-quality data become available, the 3D nature of the planet cannot be ignored.

Farther in the future, the astronomical community strives to perform space-based direct imaging characterization of exoplanets. The Wide Field Infrared Surveying Telescope (WFIRST) will act as a technology demonstration with its coronagraph instrument, which enables high-contrast imaging and visible-light spectroscopy of nearby exoplanets. The targets are massive, cool planets at much wider orbital separations than a hot Jupiter, more like our own Jupiter. Lupu et al. (2016) and Nayak et al. (2017) built a retrieval framework to model gas giants in reflected light and study the tradeoff between data resolution and signal-to-noise ratio for constraining atmospheric properties. This work takes advantage of Bayesian inference to make quantitative estimates of the science expected from a mission like WFIRST.

Two other missions, the Habitable Exoplanet Observatory (HabEx) and the Large Ultraviolet Optical Infrared Surveyor (LUVOIR), have been proposed to tackle the feat of reflected light characterization of rocky exoplanets around Sun-like stars

(Gaudi et al. 2020; The LUVOIR Team 2019). Reflected light spectroscopy is an exciting next chapter in the study of exoplanets, and further retrieval work is needed to prepare for its prospects.

1.4 Structure of This Work

Chapters 2 and 3 feature emission spectroscopy data from transiting gas giants. Chapter 2 tests the assumption of 1D hemispheric averaging in retrieval models. For a canonical hot Jupiter – WASP-43b – with a day-night temperature contrast, I consider both the typical 1D model, which assumes one TP profile, and a "2D" model that averages the fluxes from two TP profiles, one hot and one cool. These two models are applied to simulated plus actual modern observations and simulated JWST data. Chapter 3 extends the previous study to spectroscopic phase curve data sets. I investigate the behavior of the 1D vs. 2D models as a function of orbital phase. This study presents several ways to portray two TP profiles. I highlight once more the differences in inference between simulated vs. observed data, and modern vs. future data.

Chapter 4 shifts the focus from the importance of data in evaluating the performance of models to the utility of models in identifying the quality of data needed to meet science goals in mission design. In chapter 4, I introduce a retrieval framework for rocky terrestrial planets seen with reflected light spectroscopy. I simulated albedo spectra and data sets of varying wavelength coverage and signal-to-noise ratio. The retrievals of these data sets then reveal the robustness of detection or constraint for planetary properties with implications for habitability.

Chapter 2

The Impact of Non-Uniform Thermal Structure on the Interpretation of Exoplanet Emission Spectra

2.1 Introduction

Even a cursory view of images of solar system planets shows us that these planets have complex atmospheres. It is readily appreciated that not all latitudes and longitudes look alike. A view of Jupiter at 5 μm shows bright bands and spots, where, due to locally optically thin clouds, thermal emission can be seen from deeper, hotter atmospheric layers. Looking at Mars in visible light, we can often see locations obscured by thin cirrus clouds in the atmosphere, and at other locations we can see down to the surface. These different locations not only appear different to our eyes; the spectra of light that they reflect and emit also differ. When it is possible to resolve the disk of the planets under study, quite detailed levels of information can be determined: for instance, changing cloud properties with latitude, different atmospheric temperature-pressure (TP) profiles with solar zenith angle, and compositional differences in updrafts vs. downdrafts.

However, if a planet is tens of parsecs distant, there is no path to spatially resolving the visible hemisphere (with current technology). Observers probe the spectra reflected or emitted by the visible hemisphere, but there is generally little hope of assessing how diverse or uniform the visible hemisphere is. Typically, when comparing observations to the spectra from either self-consistent radiative-convective forward models (e.g. Burrows et al. 2007; Fortney et al. 2008; Marley et al. 2012; Barman et al. 2011), or from data-driven retrievals (e.g. Madhusudhan & Seager 2010; Line et al. 2014), the spectrum, or set of spectra, are generated and aim to represent hemispheric average conditions. However, while the calculation of such a spectrum, and its comparison to data, is relatively straightforward, it has been unclear how dependent our inferences are for TP profile structure, cloud optical depth, and chemical abundances from this important initial assumption.

Recent work on matching the spectra of some brown dwarfs and directly imaged planets points to problems with the homogeneous atmosphere assumption, with best-fit radiative-convective forward models coming from spectra generated from linear combinations of “cloudy” and “clear” atmospheres, or atmospheres with weighted areas of “thick” and “thin” clouds (Skemer et al. 2014; Buenzli et al. 2014). The variable nature of brown dwarf thermal emission, now well documented over several years via photometry (e.g., Enoch et al. 2003; Artigau et al. 2009; Radigan et al. 2014) and spectroscopy (Buenzli et al. 2014, 2015), also indicates inhomogeneity in the visible hemisphere, with emission that changes due to rotation and/or atmospheric dynamics (Robinson & Marley 2014; Zhang & Showman 2014; Morley et al. 2014; Zhou et al. 2016).

In the realm of retrievals, could a search through phase space for a best-fit to a measured spectrum lead to well-constrained yet biased or incorrect constraints on atmospheric properties when we assume planet-wide average conditions? This seems like a real possibility, and one well worth investigating in a systematic way. With the advent of higher signal-to-noise spectroscopy from the ground (Konopacky et al. 2013) and the coming launch of the *James Webb Space Telescope* (*JWST*), which will deliver excellent spectra for many planets over a wide wavelength range, we aim to test the 1D planet-wide average assumption systematically. We want to furthermore determine, when the data quality is high enough, if we can justify a more complex inhomogeneous model.

Recently, Line & Parmentier (2016) investigated for transmission spectra how the signal of high atmospheric metallicity inferred under planet-wide average conditions can be mimicked by a uniform lower metallicity together with a high cloud over a part of the planet’s terminator. Our work here is on thermal emission and is entirely complementary. We take the first step in characterizing how a diverse visible hemisphere may impact atmospheric retrievals.

Our paper is organized as follows: Section 2.2 describes the setup, retrieval approach, and methodology. In Section 2.3, we describe our findings. In Section 2.4, we present the application to WASP-43b. We discuss our results in Section 2.5 and conclude with future expansions.

2.2 Methodology

2.2.1 Setup

We present a simple case to illustrate the impact of a planet's spatially varying thermal structure on retrievals. The model setup features two different TP profiles, equally weighted in surface area, in a cloud-free atmosphere with planet-wide uniform abundances. This case is relevant to two simple kinds of atmospheres. One is a "checkerboard" atmosphere of equal-area hotter and colder areas, with applicability to brown dwarfs and imaged planets. Another is a transiting planet with a hot day side and cold night side, as viewed at one-quarter or three-quarter phase, meaning half-day and half-night. The equal-weighting average allows for symmetry in viewing geometry and limb-darkening effects. Each TP profile generates emitted fluxes at the top of the atmosphere. The observed spectra result from the average of the fluxes. From these averaged spectra, we generate data as observed with typical *Hubble Space Telescope* (*HST*) Wide Field Camera 3 (WFC3)+*Spitzer* Infrared Array Camera (IRAC) and *JWST* modes. We then perform atmospheric retrievals on these synthetic data assuming either a single profile (1TP) or two profiles (2TP). Figure 2.1 shows the setup.

As an initial investigation, we primarily explore the role that temperature contrast has in biasing the retrieval results, specifically on hot Jupiters. The TP profiles are offset at the top of the atmosphere from each other by a factor (i.e. contrast) defined as

$$1 - \frac{T_{\mathrm{TOA,c}}}{T_{\mathrm{TOA,h}}} \tag{2.1}$$

where $T_{\mathrm{TOA,c}}$ and $T_{\mathrm{TOA,h}}$ are the top-of-atmosphere (TOA) temperatures for the cold ("night") and hot ("day") TP profiles, respectively. Under different observational setups, we (1) determine the biases in the atmospheric abundances when one global TP profile is assumed for a planet that is actually composed of two TP profiles and (2) quantitatively determine the justification for the inclusion of a second TP profile within a nested model hypothesis testing framework (e.g., Trotta 2008; Cornish & Littenberg 2007). In what follows, we describe the necessary tools to accomplish these tasks.

2.2.2 Modeling Tools

The thermal infrared radiative transfer model we use is described in detail in Line et al. (2013). It numerically solves the thermal infrared radiation problem for a plane-parallel atmosphere with absorption, emission, and no scattering given a TP profile and uniform-with-altitude gas abundances. We consider absorption due to CH_4, CO_2, CO, H_2O, NH_3, He, and H_2. The molecular abundances for generating the synthetic spectra are chosen to be in rough agreement with solar elemental abundances in thermochemical equilibrium at a representative photospheric pressure (~100 mbars)

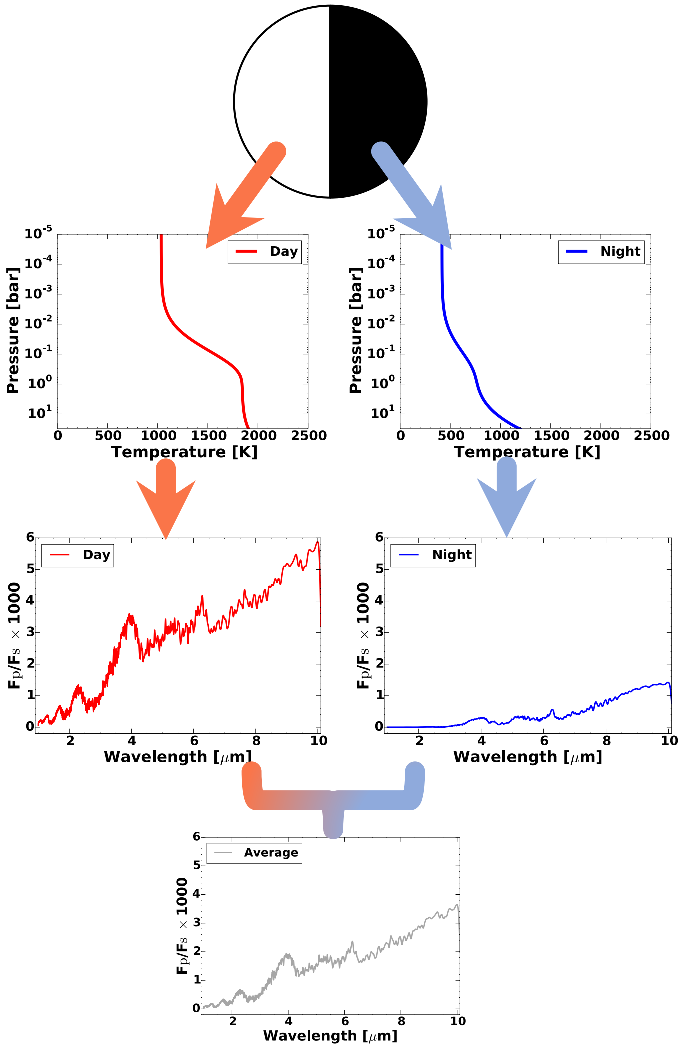


Figure 2.1: Schematic demonstration of our set up. We assume a planet with two equally weighted thermal structures with a cloud-free atmosphere of uniform composition. The fluxes from both thermal profiles are then averaged to create the disk integrated spectrum upon which we perform the retrievals.

along the prescribed thermal profile. When computing the spectra for two TP profiles, we assume the dayside abundances for both, consistent with expectations from horizontal mixing (e.g., Cooper & Showman 2006; Agúndez et al. 2014). The opacity database is described in Freedman et al. (2014).

We set and retrieve for the temperature profiles using the Parmentier & Guillot (2014) 5-parameter prescription (two visible opacity parameters ($\log \gamma_1$, $\log \gamma_2$), partitioning between the two visible streams (α), infrared opacity ($\log \kappa_{IR}$), and the fraction of absorbed incident flux (β); see Equations 13, 14 in Line et al. (2013) and Table 2.1). The internal temperature, T_{int}, is an additional parameter we specify, but it is not one of the retrieved quantities. We fix T_{int} to 200K (Guillot 2010) which prevents the TP profiles from ever reaching 0 K. Given the molecular abundances and thermal structures, we use four point Gaussian quadrature to compute the full disk-integrated spectrum for the day and night profiles separately. By taking the average of the "hemispheres" or "checkerboard patches" and dividing by a stellar spectrum, we generate the planet-to-star flux ratios. Taking the average of the disk-integrated fluxes is equivalent to weighting each profile by the same area, thus invoking the same limb-darkening effects. Note that this need not be true in the case of hot-spot or "crescent phase" models, where there is asymmetry in limb darkening, which we will investigate in a later publication.

The high-resolution model spectra are then appropriately convolved and interpolated to the "observational" wavelength grid. Poisson noise (no systematic noise is included) is then added to each data point. For the *HST* WFC3+*Spitzer* IRAC setup, we assume error bars representative of current observations (e.g., 35 ppm error bars at 0.035 μm resolution, R$\sim$40, for WFC3, and 70 ppm error bars for the *Spitzer* IRAC 3.6 and 4.5 μm channels). For the *JWST* observational setup, we use the noise model described in Greene et al. (2016), covering 1-11 μm and combining modes from the

NIRISS, NIRCam, and MIRI instruments. This noise model simulates the uncertainties obtainable with the observation of a single transit or secondary eclipse for a hot Jupiter. We adopt planet and stellar parameters for the HD 189733 system (Table 2.1).

We use an atmospheric retrieval approach to explore the biases introduced in assuming a single TP profile for a spectrum composed of two separate TP profiles and the detectability of multiple profiles. Much of the thermal infrared retrieval machinery is based on the CHIMERA retrieval suite already described in Line et al. (2013) and Line et al. (2014) and subsequently applied in Kreidberg et al. (2014a) and Stevenson et al. (2014). However, for the Bayesian inference problem, rather than using the differential evolution Monte Carlo approach, we use the `multinest` algorithm (Feroz et al. 2009) as implemented with the `pymultinest` routine (Buchner et al. 2014) because it not only has the ability to produce posterior samples, but it also efficiently computes the Bayesian evidence, or the integral of the posterior over the parameter volume. The Bayesian evidence is required for model comparison and selection, and it is a numerical encapsulation of the balance between goodness-of-fit and a model's simplicity. It can be thought of as the more rigorous computation of the commonly used Bayesian Information Criterion (BIC). For application of the `multinest` algorithm and model selection to exoplanet spectra, we refer the reader to Benneke & Seager (2013), Waldmann et al. (2015b), Waldmann et al. (2015a), and Line & Parmentier (2016). For a summary of Bayesian model selection and evidence computation, we refer the reader to Trotta (2008) and Cornish & Littenberg (2007).

From the synthetic model spectra, we aim to determine the constraints on the

Table 2.1: Model parameter values

Parameter	Value	TP Parameter	Value
R_p (R_{J})	1.138	$\log \gamma_1$	-1
R_* ($R_\odot$)	0.756	$\log \gamma_2$	-1
T_* (K)	5040	$\log \kappa_{\mathrm{IR}}$	-1
a (AU)	0.031	α	0.5
T_{int} (K)	200	β_{day}	1
$\log(g)$ (cm s^{-2})	3.34	β_{night}	0.2
$\log f_{\mathrm{H_2O}}$	-3.37		0.4
$\log f_{\mathrm{CH_4}}$	-9		0.6
$\log f_{\mathrm{CO}}$	-3.7		0.8
$\log f_{\mathrm{CO_2}}$	-9		
$\log f_{\mathrm{NH_3}}$	-9		

Note. — Nominal system and TP shape parameters used to generate our synthetic spectra. Stellar and planetary parameters are based on the HD 189733 system. For definitions of the TP parameters, see Line et al. (2013). Solar proportion Hydrogen and Helium are assumed to make up the remaining gas abundance.

uniform-with-altitude abundances for CH_4, CO_2, CO, H_2O, NH_3, and both of the TP profiles. We assume the same 4 "shape" parameters (two visible opacity parameters, partitioning between the two visible streams, and the infrared opacity) for both TP profiles but allow for a different ratio of the absorbed-to-incident flux (e.g., some combination of albedo and redistribution), represented by parameter β in Table 2.1. β also acts as a multiplicative factor between the two contrasting profiles, as illustrated by our definition of contrast in Equation 2.1. This leads to a total of 11 free parameters for the 2TP model and 10 for the 1TP model. We assume uniform-in-$\log_{10}$ priors for the 5 gas volume mixing ratios ranging from -12 to 0 and top-of-atmosphere temperatures ranging from zero to twice the irradiation temperature.

Table 2.2: Retrieval Results and Bayesian Model Evidence for 2nd TP profile

	WFC3		*JWST*	
Contrast term	$\ln(B)$[a]	σ[b] 2nd TP	$\ln(B)$	σ 2nd TP
0.2	-1.06±0.68	<0.1	12.93	5.44[c]
0.4	1.12±0.56	2.05±0.37	274.8	>20
0.6	2.49±2.10	2.54±1.00	967.9	>20
0.8	1.77±0.75	2.41±0.34	1836	>20
1TP	-1.26	< 0.1	-2.92	< 0.1

Note. — The last row, "1TP", reports the case for which we generated the spectrum with one TP profile and retrieved for two. For both observational setups, in this scenario, a 2nd TP profile is not favored. Contrast term is $1 - \beta_{\rm night}$ (see Table 2.1), and "σ 2nd TP" is the detection significance of the 2nd TP profile.

[a] Bayes factor, calculated as the difference in the natural log of the evidence between the larger model (2TP) and the smaller model (1TP).

[b] We consider a $> 3.6\sigma$ detection to be strong (Trotta 2008, Table 2)

[c] Using a different noise instance, we find a 4.2σ detection of the 2nd profile. While $\ln(B)$ changed, the 2nd TP is still detected robustly. We also calculate the BIC for the noise instance with 5.4σ detection significance. Our Δ BIC = 23, which is above the threshold (Δ BIC > 10) for strong evidence against the model with the larger BIC (in our case, the 1TP scenario). It also corresponds to 5σ detection significance, consistent with the Bayesian evidence result. Small differences in χ^2 are magnified if there are many points, as with *JWST* data.

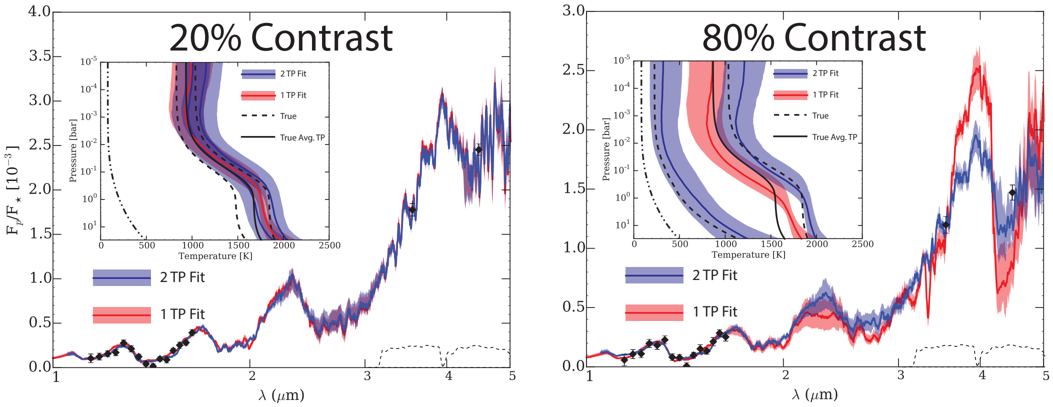


Figure 2.2: *HST* WFC3 + *Spitzer* IRAC 1TP vs. 2TP fit and temperature profiles (insets) retrieval summary. The left panel shows the results for the low (20%) contrast while the right shows the results for high (80%) contrast. The data simulated with 2 TP profiles are shown as the black diamonds with error bars (WFC3 between 1 and 2 μm and the *Spitzer* IRAC points at 3.6 and 4.5 μm). The fits and temperature profiles are summarized with a median (solid line) and 68% confidence interval (spread) generated from 1000 randomly drawn parameter vectors from the posterior. Red corresponds to the fits/temperature profiles resulting from a single profile fit, while blue represents the result of including two temperature profiles in the retrieval. The black dashed lines in the temperature profile insets are the two TP profiles used to generate the simulated data (i.e., the "true" TP profiles). For comparison, we also include the flux-averaged TP profile ($T^4_{\rm avg} = \frac{1}{2}(T^4_{\rm day} + T^4_{\rm night})$), shown as the solid black line in the insets. The dot-dashed TP profile is the coldest profile permitted by the model: a non-irradiated cooling profile governed by the 200K internal temperature. By eye, the 1TP vs. 2TP performances at 20% contrast are comparable. Based on the Bayesian evidence, the detection of the second profile is not significant ($< 0.1\sigma$). At 80% contrast, the two retrieved spectra are visibly different. The second profile is detected to 2.4σ significance.

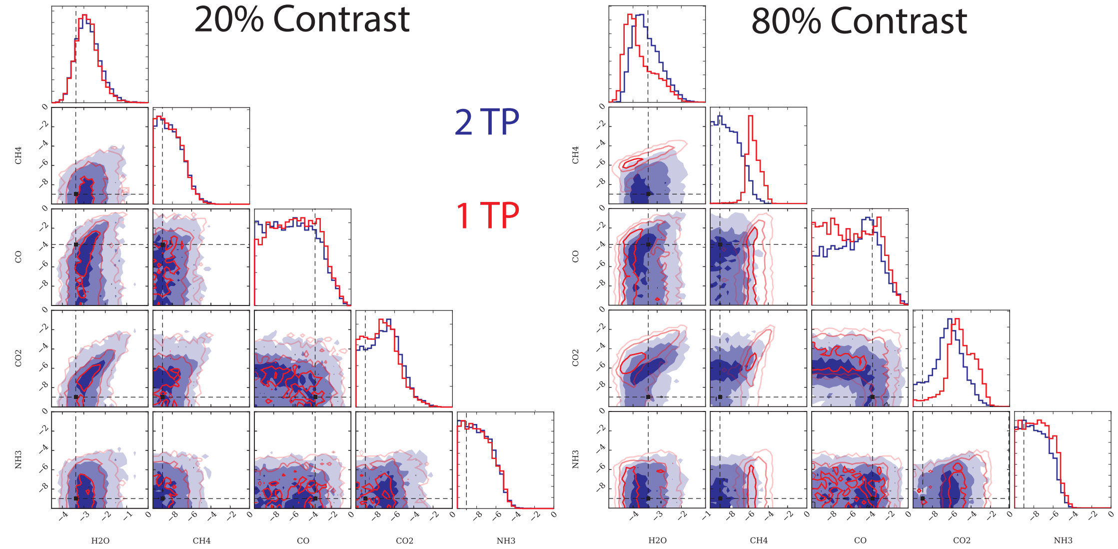


Figure 2.3: Summary of the posterior probability distributions of the molecular abundances for the low (20%, left) and high (80%, right) contrast cases under the *HST* WFC3+*Spitzer* IRAC observational scenario. The red and blue 1- and 2-D histograms correspond to 1TP and 2TP scenarios. The dashed lines in the 1-D histograms and intersection of the dashed lines in the 2-D histograms are the true molecular abundances used to generate the synthetic data. The detection significance of the second profile from the 2TP retrieval is $< 0.1\sigma$ at 20% contrast, and the posterior distributions show that invoking a second profile did not improve our abundance estimation. At 80% contrast, where the detection significance is 2.4σ, we still note the similarities in the posterior distributions for most species. However, in the case of CH_4, the 1TP approach, bound by the radiative transfer properties of one profile, overestimates both its abundance and the precision. When we include a second profile, we are able to recover a more realistic and representative distribution for the CH_4 abundance.

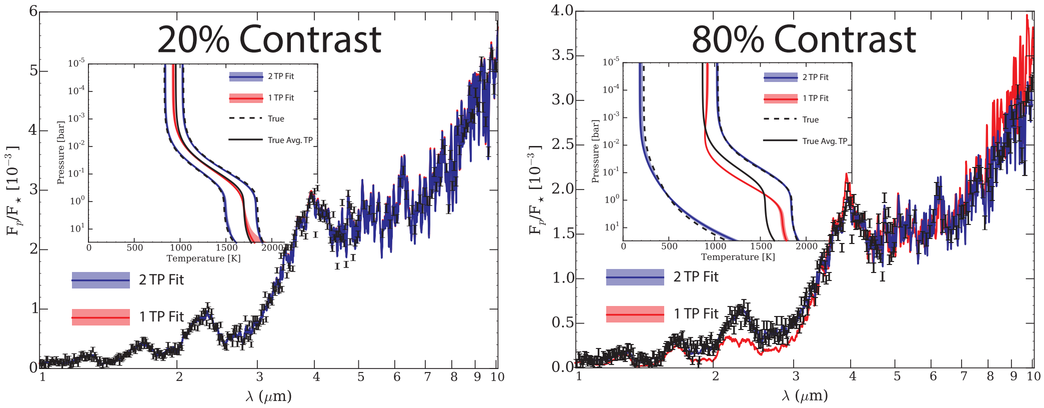


Figure 2.4: *JWST* 1TP vs. 2TP fit and temperature profiles (insets) retrieval summary. The left shows the results for the low (20%) contrast while the right shows the results for high (80%) contrast. The data simulated with 2 TP profiles are shown as the black error bars. The fits and temperature profiles are summarized with a median (solid line) and 68% confidence interval (spread) generated from 1000 randomly drawn parameter vectors from the posterior. Red corresponds to the fits/temperature profiles resulting from a single TP profile fit, while blue represents the result of including two temperature profiles in the retrieval. The black dashed lines in the temperature profile insets are the two TP profiles used to generate the simulated data (e.g., the "true" TP profiles). For comparison, we also include the flux-averaged TP profile ($T^4_{\rm avg} = \frac{1}{2}(T^4_{\rm day} + T^4_{\rm night})$), shown as the solid black line in the insets. At 20% contrast, while the retrieved fits appear similar, we find that the second TP profile is detected to $\sim 5\sigma$ significance. At 80% contrast, the 1TP retrieved spectra poorly fit the data, especially at $2-3$ μm and at longer wavelengths.

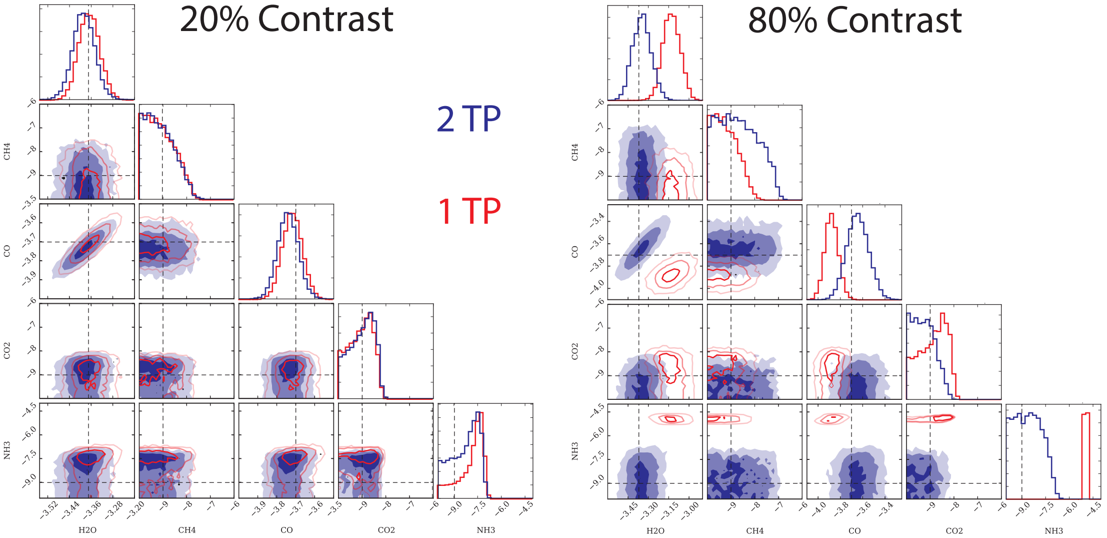


Figure 2.5: Summary of the posterior probability distributions of the molecular abundances for the low (20%, left) and high (80%, right) contrast cases under the *JWST* observational scenario. The red and blue 1- and 2-D histograms correspond to 1TP and 2TP scenarios. The dashed lines in the 1-D histograms and intersection of the dashed lines in the 2-D histograms are the true molecular abundances used to generate the synthetic data. When the contrast is 20%, the second profile is detected to $\sim 5\sigma$. When the contrast is 80%, the second profile is detected to $> 20\sigma$. We see that, at higher contrasts, the 1TP retrieval case is a poor representation of the abundances. We also note the over-constraint of NH_3 under the 1TP prescription. This behavior is analogous to the CH_4 abundance inference using one profile that we saw with WFC3+IRAC data. Once a second profile is included, we recover the true abundance of NH_3.

2.3 Results

We present our retrieval results on the synthetic spectra simulated with *HST* WFC3+*Spitzer* IRAC and *JWST*. For each telescope combination, we produce spectra for four levels of contrast between the two TP profiles (see Equation 2.1) of 0.2, 0.4, 0.6, and 0.8. For each spectrum, we perform a 1TP and a 2TP retrieval. We also test the inclusion of two TP profiles in the retrieval when only one profile was used to generate the spectrum and synthetic data. We compare the performance of the two models by the Bayes factor, summarized in Table 2.2. Based on the retrievals, we can explore the biases resulting from retrieving for a single TP profile when the spectrum is generated with two. We also quantify the detectability of a second TP profile as contrast changes.

We summarize the retrieval results comparing the 1TP and the 2TP retrievals for only the extreme contrasts, 0.2 and 0.8. In our figures, we also include a flux weighted (averaging T^4) profile for each contrast to guide the eye when interpreting the 1TP retrievals. One would expect a single representative TP profile to closely match the flux weighted profile. The gas abundance retrievals are summarized with a pairs-plot showing both the 1D and 2D marginalized posteriors (Figure 2.3 for *HST*+*Spitzer*, and Figure 2.5 for JWST). The TP profiles and spectra are summarized with a median and 1-sigma spread reconstructed from 1000 randomly drawn posterior samples (Figure 2.2 for *HST*+*Spitzer*, Figure 2.4 for JWST).

We note that, because we knew *a priori* (from test simulations) that the detection significance would be marginal for a 2nd TP profile within the *HST*+*Spitzer* setup, we tested their robustness by performing the retrievals and nested model comparison on six noise instances per contrast setup. At low detection values (less than $\sim 3\sigma$), the exact detection significance is very sensitive to a particular random noise instance. Thus, in Table 2.2, we show a mean value and error on the results for the *HST*+*Spitzer* observational setup. This is not an issue for the *JWST* observational setup as the detection significances are always above a significant threshold.

2.3.1 Findings for Simulated *HST* WFC3+*Spitzer* IRAC Observations

Figure 2.2 and 2.3 summarize the results for the 0.2 and 0.8 contrast cases. The left panel of Figure 2.2 shows the retrieved TP profiles and model spectra for the low contrast (0.2) scenario. The spectra are nearly indistinguishable. This results in our inability to robustly distinguish a 2nd TP profile as the 68% confidence envelopes for each of the two TP profiles strongly overlap with each other and with the error envelope for the single TP profile. The retrieved molecular abundance posteriors (Figure 2.3, left panel) are also nearly indistinguishable between the 1 and 2TP cases. Unsurprisingly, the nested model comparison results in a non-detection for a 2nd TP profile in the 0.2 contrast scenario (Table 2.2). In fact the Bayes factor, B, is less than 1 ($\ln B < 0$) suggesting that the improvement in the spectral fit is outweighed by the increased model complexity.

At higher contrast (0.8), there is a greater deviation in the retrieved model spectra at longer wavelengths ($> 2\mu$m). The 1TP spectra have to contort themselves with a strong peak-to-trough "N"-shaped feature between 3 and 5 μm in order to fit the two IRAC points. The broadband integration over wavelength does not allow us to tell the difference between the two scenarios. The day and night TP profiles, in contrast to the 0.2 scenario, are widely separated outside of their 68% confidence intervals. However, the 68% confidence envelope of the 1TP profile largely encompasses the flux-weighted TP profile, especially over the range where the observations probe (between 1 and 0.01 bars, Stevenson et al. 2014). We note that the fixed internal temperature of 200 K sets a lower limit for the night side profile, while the retrieved profiles serve as an upper limit. We saw this by examining the histogram of retrieved temperatures at a certain pressure (4 mbar); the distribution is unbounded but consistent with the coldest permitted temperature. While the detection significance for the 2nd TP profile in Table 2.2 is higher than in the 0.2 contrast case, it is still not considered significant. However, it makes sense that an increase in contrast should result in higher evidence for a 2nd TP profile.

Perhaps the most striking find in this high-contrast scenario is the strong differences in the molecular abundance constraints, in particular that of CH_4. While the 2TP scenario (the "true" model) results in an upper limit on the methane abundance, as expected given the low (non-detectable) input value used, the 1TP profile scenario results in a strong methane constraint. This strong constraint, however, is several standard deviations away from the true input abundance. In essence, assuming only one TP profile results in an artificial constraint on the methane abundance. This is a key result that we would like to highlight. The narrow constraint is due to the high sensitivity of the fit (due to the topology of the hyper-dimensional likelihood volume) to small changes in the CH_4 abundance within the 1TP setup, very much like what would happen if one were to fit a constant to linear data. This is largely driven by the IRAC 3.6 μm point. In terms of the other species, we find that the distribution for CO_2 is sensitive to the noise instance of the data points (especially 4.5 μm), and performs more closely to the true value under a 2TP retrieval depending on the noise instance.

The dramatic change in emission from $3 - 5\mu$m in the 1TP-retrieved spectra for WFC3+IRAC (Figure 2b), showing strong emission and then absorption, merited additional modeling to investigate its cause. To investigate these prominent absorption features present across the IRAC bands we performed a 1TP retrieval where all the abundances were fixed to their true values, to better understand the role of TP profile shape in generating the spectrum. With this reduced parameter space, it was more readily apparent that the retrieved TP profiles featured a significant temperature gradient – 1000 K – that spanned $\sim 700 - 1700$ K over a relatively narrow pressure range ($0.01 - 1$ bar). These large differences in temperatures probed naturally leads to the striking features (strong emission and absorption) seen in the spectrum, while a more isothermal profile would yield more muted contrasts in emission.

Finally, from Table 2.2, we find that the detection of a 2nd TP profile is below what is commonly considered significant (3-4σ), especially when considering the uncertainties, but that all contrasts greater than 0.2 are more justified in including the 2nd TP profile. The marginal detections are a result of the complex interplay between the intrinsic temperature contrast, wavelength coverage, and feature signals-to-noise. Furthermore, as a sanity check, we find evidence *against* ($\ln B < 0$) the inclusion of a 2nd TP profile when only one is used to create the spectra.

2.3.2 Findings for Simulated JWST data

Figure 2.4 and 2.5 summarize the findings for 0.2 and 0.8 contrast cases. We find that for low contrast (0.2) there is not a significant bias in the retrieved molecular abundance when using one TP profile, and that the retrieved single TP profile matches the flux-averaged TP profile quite well. However, we still find a significant detection ($> 5\sigma$) of a 2nd TP profile. This suggests that fit with the single TP profile is not quite as good as the fit with two, though apparently indistinguishable by eye, even when taking into account the Occam's penalty (Table 2.2; Gregory 2005, page 49).

The situation changes, however, for large contrasts (0.8). The 1TP fit is noticeably worse between ~ 1.6 and 3.3μm, and then again at the longest wavelengths (Figure 2.4). The shape of the spectrum is different enough with two TP profiles that a model with single TP profile simply cannot accommodate. Because WFC3+IRAC do not cover $2-3$ μm, we would not have known that this range is sensitive to the large TP contrasts in our particular toy atmosphere. Thanks to *JWST*s wavelength coverage, we see that, at large contrasts, a second profile is needed, and this profile is detected to $> 20\sigma$. Furthermore, the 1TP model results in significant abundance biases. The H_2O abundance is much higher (relative to the uncertainty) than the truth, CO is slightly underestimated, and ammonia off by ~ 4 orders of magnitude, with an artificially small uncertainty. This is a cautionary note that small uncertainties on parameter values should be taken with a grain of salt if a model is inadequately fitting the data. This behavior is analogous to the CH_4 abundance inference using one profile that we saw with WFC3+IRAC data. Once a second profile is included, we recover a distribution representative of the true abundance of NH_3.

For the remaining contrast cases (0.4 and 0.6), we find overwhelming evidence ($> 20\sigma$) for the presence of a 2nd TP profile. We also find, as expected, that there is little evidence for a 2nd TP profile from an object with only one TP profile. All of this taken together suggests that *JWST* observations of thermal emission spectra will be *extremely* sensitive to the presence of multiple TP profiles (given reasonable observational assumptions).

2.4 Application to WASP-43 b

As an application to real observations, we test our two TP profile assumption on the well-characterized hot Jupiter WASP-43b. WASP-43b was observed as part of a large *HST* Legacy program (PI Jacob Bean) with WFC3 providing 3 primary transits, 2 secondary eclipses, and 3 full spectroscopic phase curves (Kreidberg et al. 2014a; Stevenson et al. 2014). Such phase curve observations provide a glimpse into the 3-dimensional structure of a planet as different wavelengths probe different atmospheric pressures and the different phases probe different planetary longitudes. These published results were interpreted (using CHIMERA, the same model used here) assuming a single TP profile representation for each spectrum at every phase. We now know from our synthetic tests above that, for objects with strong day-night contrasts (as WASP-43b possesses), assuming a single TP profile for a single disk-integrated spectrum may result in biased abundances. Motivated by recent full phase IRAC 3.6 and 4.5 μm observations of WASP-43b (Stevenson et al., in prep), we decided to revisit interpretation of the spectral energy distribution of non-secondary or primary eclipse phases within our newly developed two-TP profile framework. For an initial exploration, we focus on the first quarter *HST* WFC3+*Spitzer* IRAC data (eastern hemisphere). This phase represents exactly the geometry explored in above examples: half "day", half "night". We utilize the same forward model and retrieve for the identical set of molecules as on our simulated data. Figure 2.6 summarizes the relevant results. In addition to the first quarter, we examined the third quarter (western hemisphere) as well as the day side emission data.

For the first quarter, as in our synthetic WFC3+IRAC example, we find evidence for a bias in the CH_4 abundance. Assuming a single TP profile forces a solution that results in an overly well constrained methane abundance, an abundance that is a few sigma larger than anticipated from solar composition gas in thermochemical equilibrium at dayside photospheric conditions (1700K, 400 mbars). Once again, one would not expect such a good constraint given that these particular observations only provide a single measurement, the 3.6 μm band, on a methane absorption feature. However, as in the above synthetic examples, we find that the water abundance is robust against the 1 vs. 2 TP profile assumption. This is because water is primarily constrained at shorter wavelengths where the impact of including a second TP profile is minimized.

We determine the justification for the inclusion of a second TP profile by comparing the Bayesian evidence for a model with and without the second profile. Upon doing so, we determine that the second TP profile is justified at the 3.3 σ level (just below what would be considered "strong" on a Jeffery's scale (Trotta 2008)). The Bayes factor is $\ln(B) = 3.99$. While this is not the strongest of detections, when combined with the CH_4 bias, it warrants the inclusion of the second TP profile. We also find that the two retrieved profiles match remarkably well with the hemispheric TP profiles retrieved for the day side and night side spectra presented in Stevenson et al. (2014) as well as the hottest day side profile from the General Circulation Model (GCM) in Kataria et al. (2015). The projected spectra between 3.8 and 5 μm show the strongest divergence between the 1- and 2-TP profile fits, followed by wavelengths between 2.2 and 3.5 μm . Future higher resolution observations should focus on these spectral regions to boost the detection level of a “2nd TP profile”.

When we investigated the day side emission data, the one and two TP profile scenarios yielded similar results, consistent with what we saw when the contrast is low between two profiles: on the day side, a second, cooler profile is not needed to explain the data.

We then examined the role of multiple TP profiles for the the third quarter. We found that the second TP profile is not justified by the data (2.7σ). Like the first quarter single TP profile fit, we find a well-constrained methane abundance using one profile. However, after including a second profile, the methane posterior remained constrained unlike in the case of the first quarter.

While the first and third quarters seem in conflict with regards to the impact of a second TP profile, the full phase curve for WASP-43b shows asymmetry, suggesting that the third quarter is not the exact “opposite” to the first quarter in the sense of contrast. When it comes to seeking trends or consistency throughout the phase, we have to be wary and investigate more thoroughly to differentiate what is truly representative of the atmosphere and what is the artificial manifestation of e.g., the sensitivity to the slope between band-integrated data points.

The full phase curve data of WASP-43b continue to serve as a benchmark data set in the context of the 3D nature of planets and push us to better our model interpretations, which are especially important for future exoplanet characterization observatories.

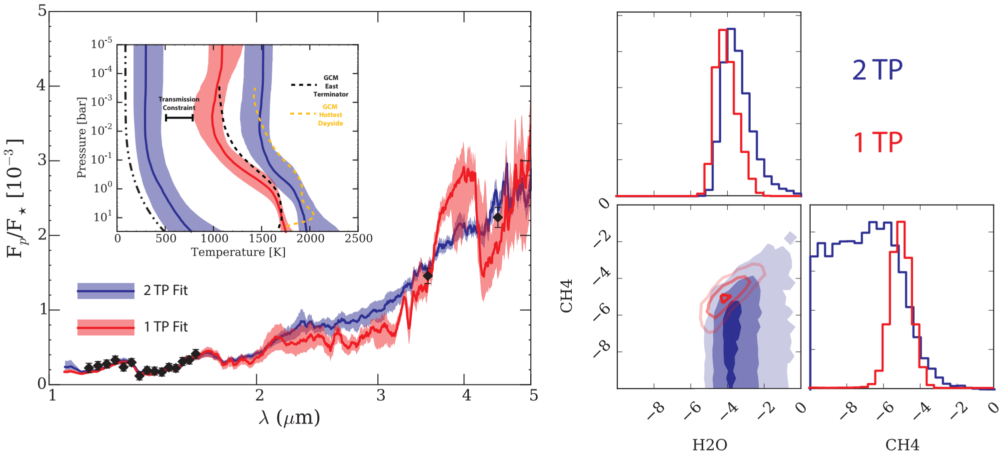


Figure 2.6: Summary of the 1 TP vs. 2 TP retrievals on the *HST* WFC3 + *Spitzer* IRAC observations of WASP-43b. In the left panel, the data are shown as the black diamonds with error bars (WFC3 between 1 and 2 μm and the *Spitzer* IRAC points at 3.6 and 4.5 μm). The fits and temperature profiles (inset) are summarized with a median (solid line) and 68% confidence interval (spread) generated from 1000 randomly drawn parameter vectors from the posterior. Red corresponds to the fits/temperature profiles resulting from a single TP profile fit, while blue are a result of including two temperature profiles in the retrieval. The dot-dashed TP profile is the coldest profile permitted by the model: a non-irradiated cooling profile governed by the 200K internal temperature. At two sigma, the retrieved night-side TP profile is consistent with the coldest permitted profile, suggesting that the retrieved night-side temperatures are an upper limit. We also show GCM derived TP profiles for the east terminator (black dashed) and dayside (purple dashed). The single TP profile fit matches the east terminator GCM profile well, while the dayside TP in the 2TP fit matches the GCM derived dayside TP profile reasonably well. The "scale height" temperature retrieved from the WASP-43b transmission spectra (Kreidberg et al. 2014a) is shown as the horizontal error bar. This temperature assumes an isothermal profile seen in transmission. Finally, the water and methane abundance posteriors are shown in the right panel. For simplicity, we do not show the posteriors of the other molecules whose abundances were retrieved (NH_3, CO, CO_2). Note the water abundance here seems invariant under the 1-(red) or 2-(blue) TP assumptions, but the methane abundance is artificially well-constrained when assuming only 1 TP. Approximate thermochemical equilibrium molecular abundances at 1700K, 0.4 bars (dayside photospheric conditions) with solar elemental composition are shown with the dashed lines.

2.5 Discussion and Summary

The interpretation of exoplanet spectra is complex; the conclusions we draw about the composition, thermal structures, and other properties of exoplanet atmospheres strongly depend on our model assumptions. In this pilot study, we explored the biases in thermal structure and molecular abundances as a result of the commonly used assumption of "1D" on the interpretation of transiting exoplanet emission spectra. We generated spectra from a simple "2D" setup of a planetary hemisphere composed of two thermal profiles, representative of either a "checkerboard" hemisphere, which may physically correspond to a planet peppered with various convective cells, or a "half-and-half" planet, similar to simultaneously observing a hot day side and cooler night side. We then applied commonly used atmospheric retrieval tools under the assumption of a single 1D homogeneous hemisphere to one that is inherently "2D".

Within this setup, we explored how the biases in the abundances and 1D thermal profile are influenced by varying degrees of "contrast" between the two TP profiles for two different observational situations. We found that, for current observational setups, *HST* WFC3+*Spitzer* IRAC, while the inclusion for a 2nd thermal profile is largely unjustified within a nested Bayesian hypothesis testing framework (i.e., the fits do not improve enough to justify the additional parameter), significant biases in the abundance may exist at large contrasts. In particular we found that an artificially precise constraint on the methane abundance can be obtained when assuming a hemisphere composed of a single 1D thermal profile. For a representative *JWST* observational scenario (1-11 microns requiring the NIRISS, NIRCam, and MIRI instruments), we found strong evidence of a 2nd profile in all contrast cases. While little molecular abundance biases appeared to exist for the lowest contrast (0.2), significant biases exist in the water, carbon monoxide, and ammonia abundances for high contrast (0.8). We also found that the retrieval was able to accurately recover both TP profiles when included in the model.

Conceptually, we can understand why the 1TP retrieval performs poorly in the case of large contrast by considering just the blackbody spectra of the day and the night sides. Because the night side flux is much lower, the averaged flux we observe is essentially half of the day side flux. This averaged spectrum is then not of a blackbody form. The 1TP approach can be thought of the attempt to fit one blackbody to the averaged spectrum – it cannot simultaneously fit for both the peak location and the amplitude. An alternative way to fit for the lowered flux, and allowing the fitting of the peak, is to change the emitting area. In our case, that area is fixed, making that not applicable. The 1TP retrieval has to rely on the flexibility provided by tweaking the thermal profile and abundances. With a 2TP approach, we are able to halve one of the blackbodies in the same way the data are generated, and we can better characterize this simple day-night atmosphere.

As a practical real-world example, we tested the 1 TP vs. 2 TP profile on the first quarter phase, third quarter phase, and day side emission spectra of the hot Jupiter WASP-43b as observed with *HST* WFC3 (Stevenson et al. 2014) and *Spitzer* IRAC (Stevenson et al., in prep). For the dayside, the results are analogous to the low contrast synthetic cases. For the first quarter, we found, much like in our high contrast synthetic model scenarios, that a strong methane bias appears when assuming only a single 1D profile, but that the retrieved water abundance remains robust under the different assumptions. The artificially strong methane constraint is driven by the requirement to fit the IRAC 3.6μm point given only a single TP profile to work with, whereas the water abundance constraint is driven primarily by the WFC3 data of which is less impacted by the assumption of one or two TP profiles. The inclusion of a 2nd TP profile in this particular scenario is justified at the moderate to strong 3.3 σ level.

It is prudent for us to note, however, that for WASP-43b vertical mixing could potentially reproduce our single TP scenario retrieved methane abundance ($\sim 10^{-5}$). The abundance of methane near the base of the single TP profile at typical CH_4-CO quench pressures of ~10 bars (e.g., Moses et al. 2011; Line et al. 2011, 1600 K) is a few $\times 10^{-5}$. So, in a sense, if we assume a single TP profile, we would arrive to the conclusion that the measured methane abundance is indicative of disequilibrium chemistry to a high degree of constraint (i.e., solar composition thermochemical equilibrium would have been ruled out at several sigma in this scenario). Instead, if we assume two TP profiles, the methane upper limit would be consistent with both pure thermochemical equilibrium at solar composition or solar composition with quenched methane. We are inclined to believe the latter scenario (two profiles) given our synthetic test cases and the fairly strong detection thresh-hold for the 2nd TP profile. The broad methane upper limit permits both chemical situations. Furthermore, Kreidberg et al. (2014a) found only an upper limit on the methane abundance from the day side emission and transmission spectra of WASP-43b. Had disequilibrium methane been as present as it appeared so here, under the single TP assumption, we would have expected a similar, if not higher, degree of constraint on the methane abundance due to the slightly higher signal-to-noise of the feature during occultation. This WASP-43b example clearly points out a degeneracy in the interpretation of the spectrum, non-equilibrium chemistry or not, which can only be lifted with a robust determination of additional TP profiles that comes from higher S/N spectra over a wider wavelength range.

For the third quarter, the posterior for methane remains constrained regardless of the retrieval set-up. Instead of a statement on the chemical processes present at this phase, we take this result to highlight future work that should be done to examine the effects of utilizing broadband photometric points and the consistency of retrievals for a full phase curve.

2.6 Future Work

As we continue to push the envelope in exoplanet atmosphere observations, at the cutting edge we will always be trying to make initial inferences about planetary climate and atmospheric abundances from data with limited wavelength ranges and less than ideal signal-to-noise. Here we have shown that with sparse data, and even with outstanding data over a wide wavelength range, that modeling choices can dramatically impact our view of an atmosphere's retrieved parameters. In addition to considering and defending choices made for observational strategies and data reduction methods, it would be wise for us to also consider choices made in the construction of our model retrievals.

This manuscript serves as an initial investigation of the impact of spatial inhomogeneities on our interpretation of emission spectra. Much work remains to be explored, including, but not limited to the impacts of: spatially non-homogeneous molecular abundances driven by disequilibrium processes or instantaneous equilibrium, day side single or multiple "hot-spots", optically thick non-uniform clouds (like in brown dwarfs), and a more thorough sweep of the observational parameter space (wavelength coverage, signal-to-noise, resolution).

The exploration of observational set-ups is especially important in the coming years. The current wavelength coverage provided by WFC3+IRAC does not offer the information necessary to differentiate between potentially contrasting profiles. The *JWST* results show the potential wealth of information at wavelengths not currently probed by space-based observations. For our explored case, with its prescribed abundances and parameters, $2-3\mu$m are essential in highlighting thermal contrast. It will be worthwhile to explore this behavior under different conditions. We also emphasize this characteristic because it demonstrates our ability to determine diagnostic wavelengths indicative of key features in an atmosphere with future observations in mind. Moving forward, we aim to explore how we can minimize the observational coverage needed while maximizing our inference.

Our investigation, along with the recent exploration of non-uniform terminator cloud cover by Line & Parmentier (2016), serves to demonstrate that there is a strong need to consider the non-homogeneous nature of transiting exoplanets when interpreting their spectra. While there has been a push to develop ever more sophisticated and complicated 1D models, we have shown that even the simplest of 2D assumptions can strongly impact the models, and may even potentially dwarf the impact of the more sophisticate physics being explored in the 1D models. Moving forward, we suggest a balanced approach between complicated 1D models and simple 2D models when interpreting transit (both emission and transmission) spectra. Starting from simple models and working toward more complicated models permits us to better understand the importance of the inclusion of additional model physics (e.g., Showman & Polvani 2011).

Chapter 3

2D Retrieval Frameworks for Hot Jupiter Phase Curves

3.1 Introduction

Hot Jupiters have complex atmospheres; they are expected to be tidally-locked and experience large day-night temperature contrasts, along with significant variations in abundances and cloud properties (Parmentier & Crossfield 2018). Phase curve observations of tidally locked exoplanets probe the longitudinal variations in temperature, composition, and cloud properties, acting as a powerful diagnostic of energy and chemical transport (e.g., Agúndez et al. 2012; Komacek et al. 2017; Drummond et al. 2018; Steinrueck et al. 2019). Furthermore, precision abundance ratios can potentially be tied back to models of planet formation (e.g., Öberg et al. 2011; Madhusudhan et al. 2014; Espinoza et al. 2017). With the promise of better precision on the horizon from future observatories such as the *James Webb Space Telescope*, we will be able to deepen the level of our characterization of exoplanet atmospheres. It is critical to our understanding of these worlds to assess the accuracy with which this information can be constrained by leveraging the synergies between observations and modeling efforts.

Atmospheric retrievals have emerged as a powerful tool for determining atmospheric properties such as molecular/elemental abundances, cloud properties, and thermal structures from exoplanet spectra (Madhusudhan & Seager 2009; Line et al. 2012; Lee et al. 2012; Benneke 2015; Madhusudhan 2018). Inverse modeling is driven by the data set, wavelength coverage, and observation uncertainty; just as much, retrieval-based atmospheric inference is highly model dependent, a realization that has recently received well-deserved attention. As more inverse models are developed and data diversity continues to increase, we face a growing suite of choices regarding radiative transfer, chemistry, and aerosol treatment. (Changeat et al. 2019; Mollière et al. 2019; Mai & Line 2019; Iyer & Line 2020; Barstow et al. 2020). The specifics are ever evolving as it can be complex to pinpoint what a model may be lacking within the context of a specific data set.

A challenging aspect of retrievals is maintaining a computationally efficient forward model within common Bayesian frameworks while balancing adequately sophisticated implementation of the necessary atmospheric physics. Given the disk-integrated nature of the observed spectra, retrieval models have typically assumed 1D treatment of the temperature-pressure (TP) profile and chemistry. Yet, for instance in the case of a planet observed at quadrature, where half the dayside and half the nightside are visible, the hemispherically averaged spectrum would include contribution from contrasting hot and cool temperature-pressure (TP) profiles. Consequently, we used this case in our previous work (Feng et al. 2016) to demonstrate that a 1D retrieval model assumption affects atmospheric inference and can introduce unwanted biases. The significance of the impact depends on the type of data set and temperature contrast between the day and the night. This work found that methane is mischaracterized for simulated *Hubble Space Telescope* Wide Field Camera 3 (WFC3) and *Spitzer Space Telescope* Infrared Array Camera (IRAC) data (hereafter HST+Spitzer) - biased to a precise but inaccurate posterior distribution. Furthermore, for simulated JWST data, even water is mischaracterized.

In addition to Feng et al. (2016), numerous other works have explored the impacts of the 1D treatment for inherently-3D exoplanet atmospheres. Line & Parmentier (2016) demonstrated how nonuniform terminatory cloud cover can mimic high mean molecular weight atmospheres in transmission spectra observations. Caldas et al. (2019) identified biases in interpreting transmission spectra associated with the day-night temperature gradient through the limb of the atmosphere. Specifically, the 1D retrieval models perfectly fitted 3D generated spectra, but resulted in substantial biases in retrieved abundance. Blecic et al. (2017) used 3D general circulation model (GCM) outputs to generate synthetic emission spectra at secondary eclipse on which to test a 1D retrieval, finding that the 1D TP profile resembles the arithmetic average over the profiles within the 3D model. These results were dependent on the data quality, from the wavelength coverage to instrument resolution.

One promising avenue for elucidating the 3D structure of an exoplanet's atmosphere is the spectroscopic phase curve. With a different hemispheric average observed at each phase, we use phase curves to investigate energy transport, atmospheric dynamics, chemistry, and cloud distribution and composition.

Typically, sophisticated 3D general circulation models have been the preferred approach for interpreting spectrophotometric phase curve observations (e.g., Showman et al. 2009; Kataria et al. 2015; Mendonça et al. 2018b). Stevenson et al. (2014, 2017) provide the first spectrophotometric phase curve data set of a hot Jupiter (WASP-43b, Hellier et al. 2011) of which at the time was interpreted with the simplistic 1D model. This likely has resulted in strong abundance biases, as shown in Feng et al. (2016). Retrievals that can accommodate the inherent 3D nature of phase data are relatively new due to the challenge of adding the necessary complexity in a computationally efficient manner as well as higher fidelity data (as for WASP-43b).

Recently, Irwin et al. (2020) adapted the NEMESIS optimal estimation retrieval code to perform "2.5-D" retrievals on the spectroscopic phase curve of WASP-43b. By using a parameterized prescription of assigning temperature and composition as a function of longitude and latitude, Irwin et al. (2020) are able to retrieve thermal structures for WASP-43b that are more consistent with GCM predictions than the simplified 1D models. However, the optimal estimation method is limited in its ability to provide a wide enough sampling of parameter space or a means of model comparison based on data quality.

While state-of-the-art phase curves obtained with HST+Spitzer only exist for a handful of exoplanets (e.g., Maxted et al. 2013; Stevenson et al. 2017; Kreidberg et al. 2018; Arcangeli et al. 2019), JWST will enhance this technique and measure phase curves across longer wavelength ranges (Bean et al. 2018). Irwin et al. (2020) identify JWST as an important avenue to explore in the context of spectroscopic phase curves. Venot et al. (2020) simulated JWST MIRI ($5-12\,\mu$m) observations of the phase curve of WASP-43b; while thorough, their retrieval study uses the typical 1D approach. Potential pitfalls resulting from a 1D approach need clarifying. Recently, Taylor et al. (2020) consider the possibility of identifying non-uniform thermal structure from emission spectra of WASP-43b given different observing modes of JWST NIRSpec. Taylor et al. (2020) also find evidence of biased molecule abundance estimates described in Feng et al. (2016), detailing the dependence on model choice and wavelength range.

As such, we seize a unique opportunity in this study to complement previous work by incorporating phase geometry within a Bayesian retrieval framework such that we can robustly explore the following questions: What inferences remain consistent as a function of phase for a planet with 2 contrasting TP profiles? What will we gain when we use JWST? Is there any advantage in leveraging the full set of phase curve data together? With our nested sampling retrieval, we are able to illustrate the posterior distribution of parameters as a function of phase. In this way, we build on previous work by Feng et al. (2016) to systematically explore biases resulting from retrieval assumptions by introducing additional model complexity.

In this paper, we present a new framework that uses spherical trigonometry to properly model the phase geometry within the CHIMERA retrieval suite (Line et al.

2013). Section 3.2 describes the relevant adaptations and our investigation setup. We simulated HST+Spitzer phase curves to anchor our understanding of thermal inhomogeneity based on the 1TP and 2TP models. We then retrieve on Stevenson et al. (2017) WASP-43b data as well as simulated JWST phase curves of a model planet. Section 3.3 compiles these results. As in Feng et al. (2016) and Taylor et al. (2020), we focus on model comparison in the this study; we provide a guide as to which phases need the appropriate 2TP modeling of the large day-night temperature contrast to accurately interpret the atmosphere. We conclude with Section 3.4 through a discussion of our findings and future work.

3.2 Methodology

In Feng et al. (2016), we tackled thermal inhomogeneity with a simplified experiment: A fully symmetric scenario where half the emitting area is attributed to a hotter TP profile, while the other half is from a colder profile, permitting a simple averaging of the resultant spectra. Such a scenario would be applicable to the quadrature phases, assuming the day and night side can each be well-represented by a single thermal profile, as well as a planet "checkered" with hot and cold patches. For non-symmetric cases, we need a more sophisticated geometry to account for the differing contributions as well as the appropriate limb darkening. In the subsections that follow, we provide an overview of the radiative transfer and retrieval analysis. Next, we describe the modifications to the model used in Feng et al. (2016) to generate spectra at arbitrary orbital phases. We also detail our validation and lay out the investigation plan.

3.2.1 Overview of modeling tools

The core radiative transfer routines remain identical to those in Feng et al. (2016). Line et al. (2013) provide a detailed description. Given a TP profile and vertically uniform gas mixing ratios, we solve for the outgoing thermal radiation in a cloud-free, plane parallel atmosphere. We retrieve for H_2O, CH_4, CO, CO_2, and NH_3, and assume solar composition H_2/He as the background filler. As in Feng et al. (2016), opacities for these gases are drawn from the database described in Freedman et al. (2014) and Lupu et al. (2014, Table 2).

As in our previous work, we consider the distinction between 1TP and 2TP. We focus on a large temperature contrast case (80%) to see what effects may be dominant across phase. Table 3.1 lists the parameters we retrieve for in the 2TP model. We calculate the TP profile using the approach from Parmentier & Guillot (2014) (see also Equations 13 and 14 in Line et al. (2013)). Each profile is defined by five parameters: two visible-to-infrared mean opacity ratios ($\log \gamma_1$ and $\log \gamma_2$), the partitioning between the two visible streams (α), the infrared opacity ($\log \kappa_{\mathrm{IR}}$), and the fraction of absorbed incident flux (β_{TP}). We specify β_{day} for the day profile and β_{night} for the night profile while letting the two profiles share the other parameters. For values $\beta_{\mathrm{day}} = 1$ and $\beta_{\mathrm{night}} = 0.2$, we establish the contrast of 80% between the day and night sides. The contrast factor can also be thought of as $1 - \frac{T_{\mathrm{TOA,c}}}{T_{\mathrm{TOA,h}}}$ (Feng et al. 2016). The second term is a ratio between the temperature at the top of the atmosphere (TOA) from the cool profile (c) and the TOA temperature from the hot profile (h).

In other words, the 2TP model retrieves for 12 parameters (6 molecules + 6

TP parameters) while the 1TP model retrieves for 11 (6 molecules + 5 TP parameters). All scenarios assume an internal temperature, $T_{\mathrm{int}} = 200\mathrm{K}$ (although higher values can be expected in hot Jupiters; see Thorngren et al. (2019)). We assume constant-with-altitude and constant-with-longitude (e.g., Cooper & Showman 2006; Mendonça et al. 2018b) mixing ratios loosely consistent with thermochemical equilibrium and solar composition abundances. We adopt WASP-43b planetary properties (Hellier et al. 2011) in our model (see also Table 3.1).

Typical radiative transfer uses an "N-point" (4 in Feng et al. (2016)) Gaussian quadrature to compute the TOA outgoing fluxes whereby the observed disk can be divided up into concentric "annuli" (Figure 3.1) with intensities computed independently at each annulus (given the quadrature μ). An example crescent phase in Figure 3.1 shows how each annulus contains different contributions of "dayside" and "nightside" regions (and properties thereof). In this work, we adopt the same Gaussian quadrature scheme but have to apply geometric corrections to account for the varying viewing geometry as a function of phase (e.g., for uneven day-night temperature variations). Rather than "pixelating" the planet (e.g., Fortney et al. 2006; Showman et al. 2009; Cahoy et al. 2010), we divide it up in annuli (a natural radiative transfer coordinate system) whereby we can assign individual atmospheric properties (temperature, composition, etc.) that will dictate the upwelling intensity beam. Summing over these beams will produce the appropriate disk-integrated flux accounting for atmospheric inhomogeneity. Appendix A describes in detail the adjustments needed to accommodate arbitrary phase angles within the Gaussian quadrature/concentric annuli radiative transfer framework. We

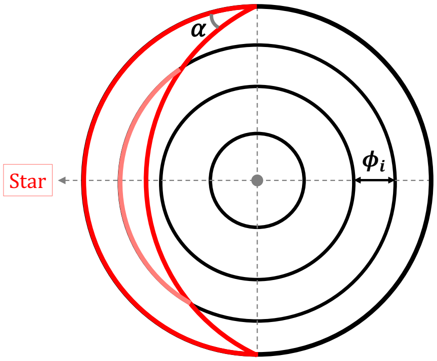


Figure 3.1: Diagram of hemisphere visible to the observer at phase angle α. The visible hemisphere is divided into annuli based on Gaussian quadrature angles, which are used to define the annulus width ϕ_i. The "dayside" region (in red) intersects the annuli at different points. By determining the areas of these segments within each annulus, we can determine the fractional contribution of "day" and "night" for the annulus, as described more in Appendix A. Within the "2TP" approach, all annuli areas encompassed by red will be assigned a "dayside" TP, and in black, a "nightside" TP.

will refer to this updated model as 2TP-Crescent.

Table 3.2 lists the orbit fraction and phase angle we consider, consistent with the phase curve data set presented in Stevenson et al. (2014, 2017). Hereafter, we will refer to orbital phases by a corresponding number as listed in Table 3.2. The new geometric implementation also allows for the retrieval of day-side hot spot properties (modeled as inner annuli with higher temperature than remaining annuli, Figure 3.1), a thorough exploration of which is beyond the scope of our current study.

We pair our modified forward model with `pymultinest` (Buchner et al. 2014), the python implementation of the `multinest` algorithm (Feroz et al. 2009), to perform Bayesian parameter estimation and model selection, following standard guidelines (e.g., Trotta 2008; Cornish & Littenberg 2007).

Table 3.1: Model parameter values

Parameter	Value	TP Parameter	Value
R_p (R_J)	0.93	$\log \gamma_1$	-1
R_* ($R_\odot$)	0.598	$\log \gamma_2$	-1
T_* (K)	4400	$\log \kappa_\mathrm{IR}$	-1
a (AU)	0.01424	α	0.5
T_int (K)	200	β_day	1
$\log(g)$ (cm s^{-2})	3.34	β_night	0.2
$\log f_{\mathrm{H_2O}}$	-3.37	$\mathrm{f_{day}}^1$	see eq. A.1
$\log f_{\mathrm{CH_4}}$	-9		
$\log f_{\mathrm{CO}}$	-3.7		
$\log f_{\mathrm{CO_2}}$	-9		
$\log f_{\mathrm{NH_3}}$	-9		

Note. — Nominal system and TP shape parameters used to generate our synthetic spectra. Stellar and planetary parameters are based on the WASP-43 system (Hellier et al. 2011). For definitions of the TP parameters, see Line et al. (2013). Solar proportion Hydrogen and Helium are assumed to make up the remaining gas abundance.

3.2.2 Investigation set-up

Within our new framework, we explore three phase curve observational setups under four different retrieval model assumptions. The three observational scenarios are:

- Simulated (of which we know the "truth" values) HST+Spitzer observations based on the Stevenson et al. (2017) WASP-43b data set
- The actual Stevenson et al. (2017) WASP-43b HST+Spitzer phase curve data set
- Simulated JWST phase curve observations of WASP-43b

For the simulated data, we only consider phases between transit and secondary eclipse due to symmetry over the orbit (i.e., we do not assume hot spot offsets or other asymmetries). We assume cloud-free atmospheres in order to not confuse any degeneracy

Table 3.2: Reference for phase angles

Phase	Angle (°)	Fraction
0	22.5	0.0625
1	45.0	0.125
2	67.5	0.1875
3	90.0	0.25 (quadrature)
4	112.5	0.3125
5	135.0	0.375
6	157.5	0.4375
7	180.0	0.5 (secondary eclipse)
8	202.5	0.5625
9	225.0	0.625
10	247.5	0.6875
11	270.0	0.75 (quadrature)
12	292.5	0.8125
13	315.0	0.875
14	337.5	0.9375

Note. — We assume the full orbit (360°) is divided into 15 phases. Phase 0 is right after transit; phase 7 is secondary eclipse; and phase 14 is right before transit. Orbital fraction is the phase value between 0 and 1. The phase angle (between the sub-observer point and sub-stellar point) is defined from transit; see Figure A.2(b).

Table 3.3: Data sets and relevant model scenarios

Scenario	HST+ Spitzer	WASP-43b	JWST	Joint Phases
1TP	×	×	×	–
2TP-Crescent	×	×	×	×
2TP-Free	–	×	–	–
2TP-Fixed	–	–	–	×

Note. — Tabulating the results presented in Section 3.3. "×" indicates which models were used to perform retrievals on which data sets; "–" indicates the results are not included in Section 3.3.

arising between the geometric/model assumptions and basic atmospheric properties, like abundances and thermal profiles. Future work looking into the nature of inhomogeneous clouds is most certainly a next step.

The synthetic HST+Spitzer uncertainties are pulled phase-by-phase from the Stevenson et al. (2017) WASP-43b data set. For the simulated JWST data, we use the same data setup as described in Feng et al. (2016) (based on the Greene et al. (2016) noise model), which assumes a single transit each in NIRISS, NIRCam, and MIRI LRS, covering 1-10 μm at a resolution (R) of 100. For the synthetic data, we do not apply random noise to the instrument-resolution points. As in past works (Feng et al. 2018; Krissansen-Totton et al. 2018; Mai & Line 2019; Changeat et al. 2019), we opt not to randomize the simulated data points as to mitigate random bias due to outlier noise instance draws.

Table 3.3 lists the data sets considered in our study along with corresponding model scenarios whose results are presented in Section 3.3. In total, we examine four retrieval model assumptions:

- 1TP: assumes a single TP profile and and one set of gas mixing ratios regardless of the observed phase, like in Feng et al. (2016) and applied in Stevenson et al. (2014, 2017).

- 2TP-Crescent: retrieves a hot day profile and cool night profile. Hot and cool fluxes are distributed according to phase and combined with annuli (Appendix A), such that any limb darkening differences between hot and cold are included.

- 2TP-Fixed: retrieves two profiles. Hot and cool fluxes are combined via linear combination using $F_{\text{total}} = F_{\text{hot}} * f_{\text{day}} + (1 - f_{\text{day}}) * F_{\text{cool}}$. Equation A.1 determines phase-dependent day-side contribution f_{day}

- 2TP-Free: retrieves two profiles. Fluxes calculated in the same way as 2TP-Fixed, but the retrieval treats the dayside area fraction f_{day} as an additional free parameter (as in Taylor et al. 2020).

We first retrieve on each "phase" independently (phase-by-phase). In this situation, the "day" and "night" TP profiles are allowed to vary from phase to phase. We then use the 2TP-Crescent and 2TP-Fixed schemes to perform a joint retrieval on all phases simultaneously, assuming the same day and night TP profiles at each phase. The likelihood used in the retrieval is the sum of individual likelihoods from each phase. This latter investigation aims to determine if improved abundance constraints are achievable if we assume a priori that the "day" and "night" profiles remain the same throughout the orbit.

3.3 Results

Sections 3.3.1 through 3.3.4 provide detailed retrieval results for each of the data sets we explored. We show the posteriors of the abundances as a function of phase, retrieved pressure-temperature profiles at several points in the orbit, and the spectral fits to the data for select phases.

The first data set is a simulated HST+Spitzer phase curve, where all input parameters are known. Next, we perform a similar analysis on real WASP-43 HST+Spitzer phase curve data and examine the differences. We expand next to a simulated JWST data set with a higher signal-to-noise and wider wavelength range. Finally, we explore the concept of "joint retrieval" where we use the full suite of phase curve data to seek tightened error bars on atmospheric quantities.

The focus of our study is on the difference between 1TP and 2TP models as a function of phase and determining the phases at which we are justified in employing a more complex 2D model. We thus synthesize our different scenarios by presenting an overview comparing the use of a homogeneous 1D model and a more complex model (2TP-Crescent and 2TP-Free specifically) on phase-resolved spectra. Figure 3.2 summarizes the justification of the 2TP model over the 1TP model as a function of phase for each of our cases. We will return to the figure throughout the paper. Following Trotta (2008) and Gordon & Trotta (2007), we convert the Bayes factor into detection significance of the 2TP model. For each set of data we consider, we illustrate the degree to which it is justified to use the more complex 2TP model (Crescent or Free) to interpret the data. For example, for HST simulated data, the leftmost panel of Figure 3.2, phases

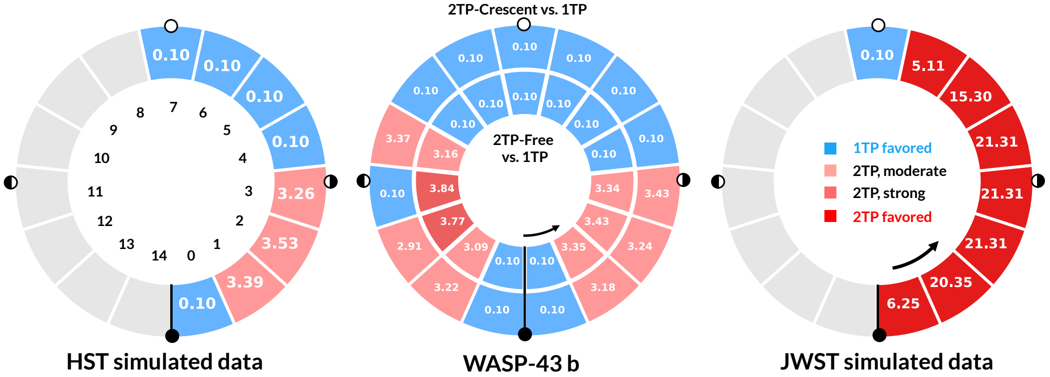


Figure 3.2: Model comparisons as a function of phase. See text for how detection significances are computed. The significance values represented here follow: $\sigma_{2\mathrm{TP}} \leq 0.1\sigma$: insignificant (blue), $2.7\sigma < \sigma_{2\mathrm{TP}} < 3.6\sigma$: moderate evidence (light pink), $3.6\sigma \leq \sigma_{2\mathrm{TP}} < 5\sigma$: strong evidence (deep pink), and $\sigma_{2\mathrm{TP}} \geq 5\sigma$ - significant evidence (red). Left: Detection significance of the 2TP-Crescent model compared to the 1TP model for the simulated HST+Spitzer data. Due to symmetry, we only simulated half the orbit. Phase are labeled by their numbers (Table 3.2). Middle: Detection significance for observed WASP-43b HST+Spitzer data. Outer ring compares 2TP-Crescent to the 1TP model. Inner ring compares the 2TP-Free and 1TP models. Full orbit is considered. Right: 2TP-Crescent vs 1TP comparison on simulated JWST data. Due to symmetry, only half the orbit is considered.

1-3 moderately favor the addition of the night profile (indicated by red color). For the rest of the half orbit from transit to secondary eclipse, there is insufficient evidence to suggest the 2TP model should be used for the data (indicated by blue color). However, as we will see in the posteriors of the abundances, phases 4 and 5 also reveal biased methane constraints if we use the 1TP model.

3.3.1 Control Case: HST and Spitzer Synthetic Data

The results from our synthetic data set of HST+Spitzer observations serve as a guide for our intuition, allowing us to identify trends and biases before considering measured spectroscopic phase curves. Figure 3.3 presents the 1TP and 2TP posteriors

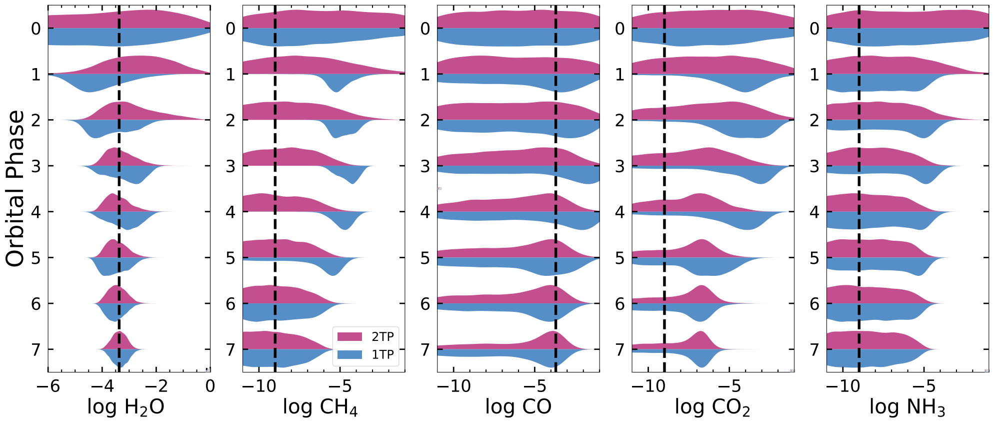


Figure 3.3: Abundance vs. phase results from HST+Spitzer simulated data for H_2O, CH_4, CO, CO_2, NH_3 for the 1TP model (blue) and the 2TP-Crescent model (dark pink). For each panel, we plot the marginalized posterior probability distribution of the log of the molecule's mixing ratio as a function of orbital phase. For simulated data, we only consider half an orbit (transit to secondary eclipse), or eight orbital steps. For each molecule, we indicate the input abundance value with the vertical dashed line. This simulated data set is only able to accurately constrain H_2O abundance; both 1TP and 2TP-Crescent models provide consistent posteriors for H_2O. The other molecules have only upper limit estimates with the 2TP-Crescent model. For most of the phases, the 1TP model produces biased CH_4 abundances (constrained at values orders of magnitude above the input). CO_2 is biased toward higher values under the 1TP model for half the phases.

of the molecular abundances as a function of phase for the simulated HST+Spitzer data. Phase 0, just after transit, exhibits no constraining power for the mixing ratio of any molecule under either model due to the low feature signal-to-noise. The posteriors are identical for phase 7 (secondary eclipse), as expected, given that there is no visible night-side flux.

H_2O and CO are the only two molecules with high enough abundances in the input model such that we should expect detection. We find no significant bias in the H_2O abundance when using the overly simplistic 1D TP profile. The “truth” falls within the 1σ range of the retrieved distribution at most phases (2σ, at worst). Phases for which there is more viewable "dayside" (hotter), the constraints are more precise (about 0.5 dex), and gradually decline towards the cooler phases (phase 0), simply due to the reduced feature signal-to-noise. CO is largely unconstrained at all phases in both models as only the Spitzer 4.5μm photometric point is sensitive to this molecule .

The original input mixing ratios for CH_4, CO_2, and NH_3 are all well below typical detectable amounts (10^{-9}), such that only upper limits would be anticipated. As such, biases become obvious when the retrieved distributions for some of these molecules are tightly constrained at elevated abundances. For example, we find that in Figure 3.3 between phases 1 and 5, the 1TP model retrieves a tight constraint on CH_4 a few orders of magnitude higher than the input (median of $\sim 3.5 \times 10^{-5}$ vs. 10^{-9}), as seen in our 2016 paper. However, using the correct 2TP-Crescent model (which was used to generate the simulated data), we retrieve only an upper limit at all phases, as expected. NH_3 shows similar behavior as CH_4: The 1TP model appears to produce more of a constraint at higher values (while still an upper limit) than the 2TP-Crescent model, as seen in phases 2 - 5 in Figure 3.3. This suggests that the 1TP model results in an NH_3 bias as well, though not as extreme as in the case for CH_4.

CO and CO_2 show similar trends in Figure 3.3 due to their overlapping spectral features over single 4.5μm Spitzer point (Line et al. 2016). In fact, CO_2 presents more bias when using the incorrect 1TP model. This apparent bias appears even at full phase (phase 7) in the correct 2TP-Crescent model because a large sample of models accumulate at high CO and CO_2 abundances simply due to the overlapping degeneracy.

Figure 3.4 shows the progression of the retrieved pressure-temperature profiles from the two models as a function of phase. For 1TP, as the phase gets closer to secondary eclipse, the retrieved profile matches more closely with the input dayside profile. The 1TP retrieved profiles are biased towards hotter temperatures at phases near primary transit ($<$ phase 3). This is because the 1TP profile is attempting to strike a balance between the nightside and dayside fluxes, as discussed in Feng et al. (2016). When implementing the 2TP-Crescent model, the retrieved day and night profiles more-or-less retrieve the input profiles, with mild bias for phases before phase 3, where more of the nightside TP profile is present. At secondary eclipse, there is no emission signal from the night, so the TP profile for the night at this phase is completely unconstrained, effectively filling out the prior.

Posterior-representative spectra are shown in Figure 3.5 for both the 1TP and 2TP-Crescent scenarios. The 2TP profile properly fits the data but the 1TP case struggles for phases between transit and first quarter (0 - 2). Given the sparse data coverage with HST and Spitzer, model differences are most noticeable in unmeasured spectral regions, with resultant 1TP spectra presenting deeper absorption features owing to the steeper temperature gradients retrieved in the 1TP model. At phases closer to secondary eclipse, the discrepancy wanes as the dayside TP/spectra more prominently represent the total.

3.3.2 Application to the Observed WASP-43b Data Set

Here, we consider the retrieval outcome using actual observations of WASP-43b from Stevenson et al. (2014); Kreidberg et al. (2014a); Stevenson et al. (2017) using HST+Spitzer. We consider results from the 1TP, 2TP-Crescent, and 2TP-Free models.

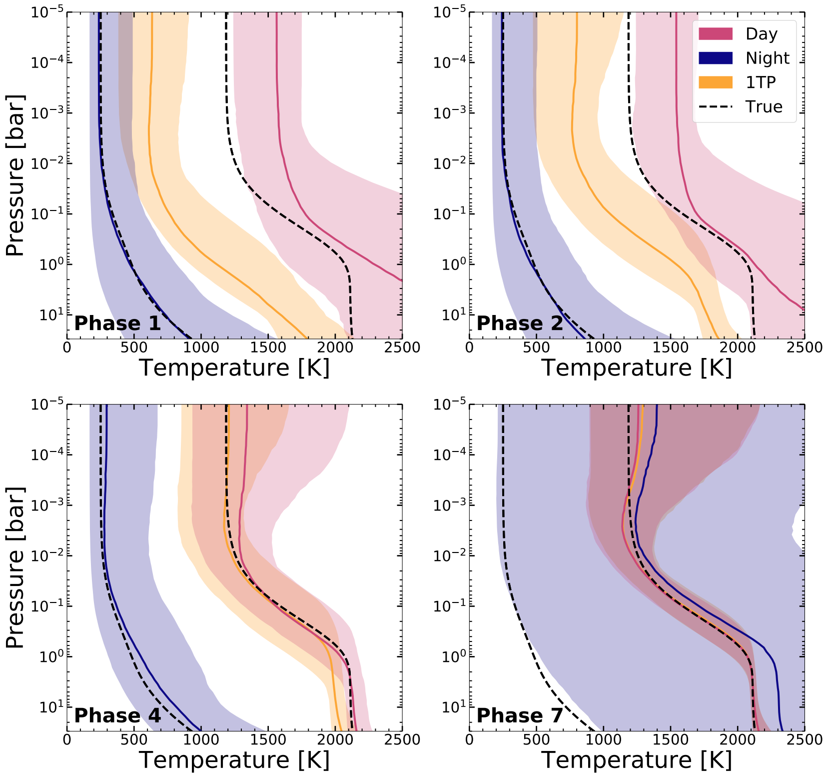


Figure 3.4: Temperature-pressure (TP) profiles for simulated HST+Spitzer data. We selected phases 1, 2, 4, and 7 to show the change in TP profile constraint as a function of phase. In each panel, the dashed line represent the true input profiles for the day and night sides. The retrieved 2TP-Crescent profiles are in blue (night) and red (day). The retrieved 1TP profiles are in yellow. For each distribution, we show the median profile in a solid line, surrounded by the 2σ spread in profiles. 1TP-retrieved profiles fall in between the true day and true night profiles, shifting toward hotter temperatures until reproducing the true day profile at secondary eclipse. The 2TP-Crescent model provides constraints on the night profiles for most of the orbit, until secondary eclipse where there is negligible night side emission. There is a preference for hotter temperatures for the day side at phases closer to transit (more of the night side visible), but once we reach quarter phase and above, the day side profile is accurately constrained.

The two different 2TP retrieval implementations allow for comparison to simulated data results and exploration of asymmetry in the atmosphere as a function of phase. The 2TP-Free model enables the latter because the dayside contribution, as parameterized by $f_{\rm day}$, is a retrieved quantity.

As Figure 3.2 (middle panel, outer ring) shows, phases 1-3, 10, 12, and 13 are moderately in favor of the 2TP-Crescent model. The inner ring in the same panel compares the 1TP model and the 2TP-Free model. For this case, phases 1-3, 10, and 13 are moderately in favor of the Free model; phases 11 and 12 are strongly in favor.

Figure 3.6, like Figure 3.3, summarizes the molecular abundance constraints for the 1TP and 2TP-Crescent scenarios, in this case for all 15 phases in the Stevenson et al. (2017) WASP-43b data set.

We find that the retrieved distributions for all gases mostly resemble the trends seen from the simulated data set (Figure 3.3). There are no substantial biases in the constraints on H_2O, CO, CO_2, or NH_3, though the water abundance appears to increase at phases between first and third quarter (e.g., around secondary eclipse)

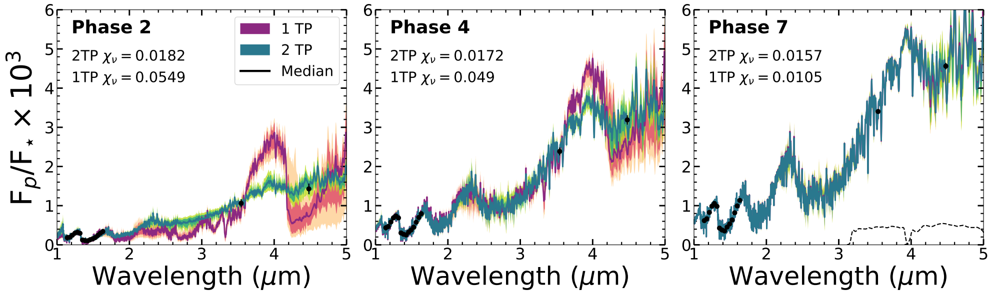


Figure 3.5: Simulated HST+Spitzer data and resultant representative fits drawn from the posterior for phases 2 (just before first quarter), 4 (just after first quarter) and 7 (secondary eclipse). We include the Spitzer 3.6μm and 4.5μm filter profiles in the phase 7 panel. 1TP spectra are in magenta while 2TP-Crescent spectra are in green. For each set of model spectra, we plot the median (solid line), 1σ, and 2σ contour. We include corresponding χ_ν values for the 1TP and 2TP(-Crescent) models, which can be small because random noise is not included. The spectra from the two models differ the most at phases close to transit; they become more similar as the phases advance to secondary eclipse, where they overlap. The biased CH_4 and CO_2 abundances result in more spectral contrast between 3 and 5μm for the 1TP profile scenario. With such distinct spectra at phases showing more night side emission, data filling the gaps between HST and Spitzer observations would be able to differentiate between these two models.

H_2O is the only well-constrained (i.e., bounded) species. Overall, there is no significant difference in the posteriors between the 1TP and the 2TP-Crescent models for H_2O, NH_3, CO, or CO_2. Although, at phase 10 (almost at third quarter), the 2TP H_2O posterior indicates slightly elevated values than the rest of the orbit.

We see the same constrained distributions for CH_4 that are only present under the 1TP model as seen in the simulated data set. Once again, the affected phases are when the visible hemisphere is dominated by the night side. The only phase that does not have an upper limit distribution under the 2TP model is phase 11. In this case, we see a well-constrained posterior consistent with what the 1TP finds.

Figure 3.7 summarizes the retrieved TP profiles for select phases, as in Figure 3.4. The behavior of the retrieved profiles over the orbit resemble what is seen from simulated data. We note that the overlap between the 1TP model profiles and the day side profiles from the 2TP-Crescent model at phase 4 is smaller than in the simulated case. The day side profiles also appear more isothermal, with a smaller temperature gradient through photospheric pressures.

In Figure 3.8, we show representative model fits for both the 1TP and 2TP-Crescent scenarios. The fits of these two models are statistically similar (e.g., at phase 7, 1TP model's $\chi_\nu = 1.71$, while 2TP-Crescent's $\chi_\nu = 2.00$) for phases around secondary eclipse (6 - 8). There is little "nightside" contribution in these cases, permitting an adequate representation with a single TP. At most phases shown, the 1TP model requires significantly larger abundances (sometimes for NH_3 or CH_4 depending on the phase) and a steeper temperature gradient than the 2TP-Crescent model, resulting in much

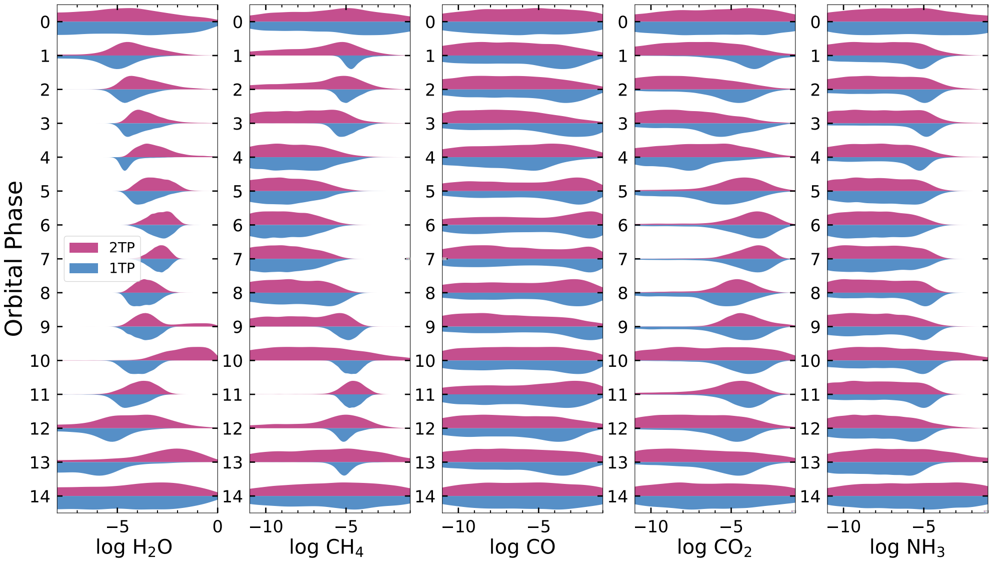


Figure 3.6: Abundance vs. phase results from WASP-43b data for H_2O, CH_4, CO, CO_2, and NH_3 for the 1TP model (blue) and the 2TP-Crescent model (dark pink). For each panel, we plot the posterior probability distribution of the log of the molecule's mixing ratio as a function of orbital phase. We see artificially tight constraints of CH_4 at several phases when using the 1TP model. With 2TP-Crescent, CH_4 at phase 11 is also constrained. However, considering the constraints (of lack thereof) of all the phases can help identify potential outlier distributions. H_2O constraints from the two models are consistent, with similar increases in estimates from transit to secondary eclipse. There is no constraining power within the data sets for CO or NH_3. CO_2 is largely unconstrained with the exception of phases near secondary eclipse. In some cases the 1TP model results in overconstrained abundances relative to the 2TP-Crescent model. CO_2 constraints are challenging to interpret due to the 1-to-1 degeneracy with CO as a result of the overlapping bands over the 4.5μm Spitzer point.

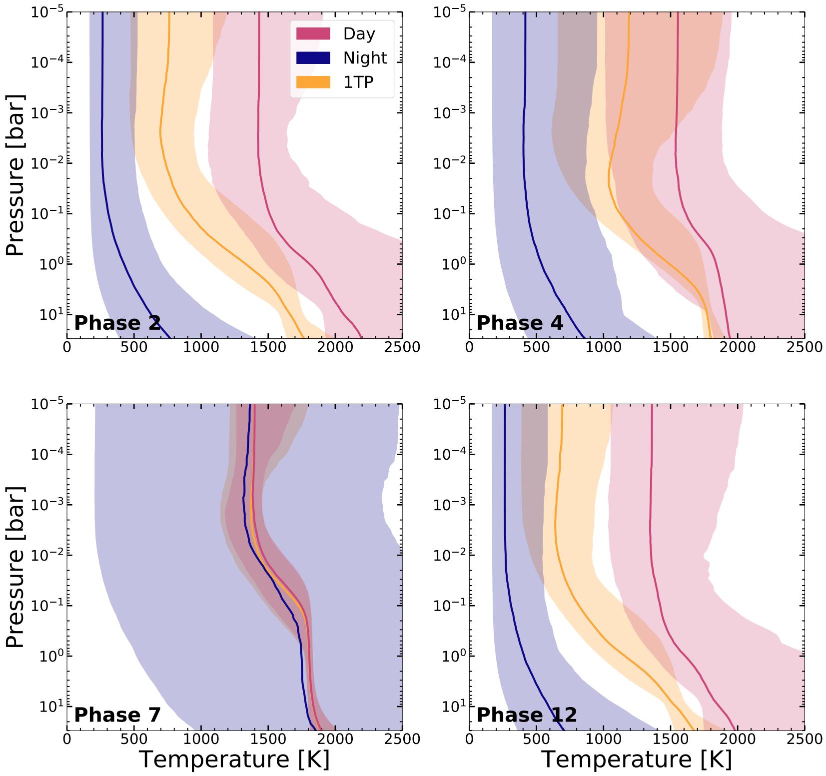


Figure 3.7: Temperature-pressure (TP) profiles for HST+Spitzer data of WASP-43b. Phases 2, 4, 7, and 12 are shown to illustrate the constraint behavior with phase. The retrieved 2TP-Crescent profiles are in blue (night) and red (day). The retrieved 1TP profiles are in yellow. For each distribution, we show the median profile in a solid line, surrounded by the 2σ spread in profiles based on reconstructed from random posterior parameter draws. Phases 2 and 12 are symmetric in the orbit (just after and just before transit, respectively), resulting in similar retrieved profiles under both models. The retrieved 1TP profile overlaps perfectly with the 2TP-Crescent dayside profile at secondary eclipse (phase 7) as there is no contributing flux from the nightside. Retrieved day and night profiles from the 2TP-Crescent model are relatively similar from phase to phase, further evidence of a large day-night temperature contrast in the atmosphere.

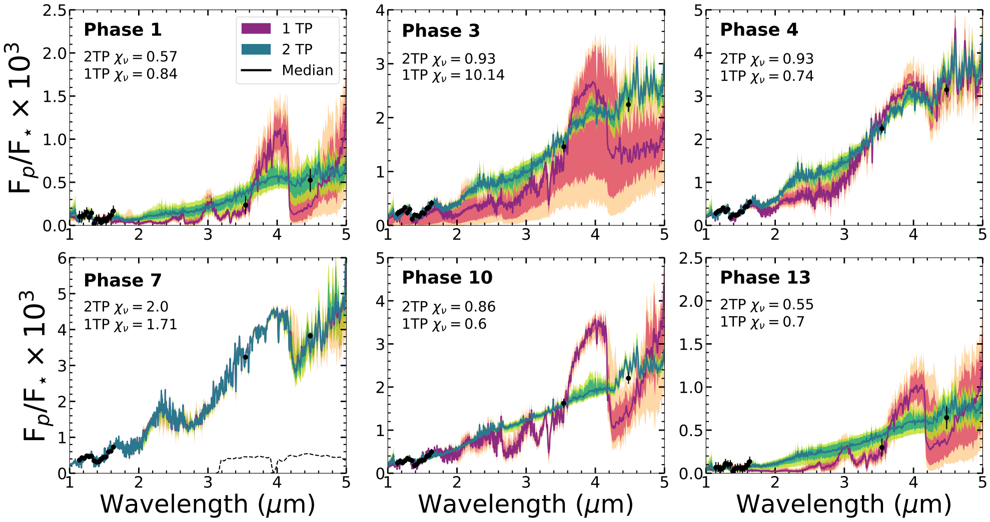


Figure 3.8: WASP-43b data (HST+Spizter) and high-resolution spectra generated with random posterior draws from the retrieval. Shown here are the spectra for phases 1, 3, 4, 7, 10, and 13. We include corresponding χ_ν values for the 1TP and 2TP(-Crescent) models. In the panel of phase 7, we overplot the Spitzer 3.6μm and 4.5μm filter transmission. 1TP spectra are in magenta while 2TP-Crescent spectra are in green. For each set of model spectra, we plot the median, 1σ, and 2σ contour. The 1TP model struggles to fit the 4.5μm Spitzer point at more crescent phases (dominated by night side). The 2TP-Crescent model spectra look more featureless in comparison at these phases, reflecting the corresponding unconstrained posteriors of the atmospheric gases. At phases closer to secondary eclipse, data between 2 and 3μm are needed to separate the two models.

more spectral contrast, which yields deep absorption features, to fit the spectra and photometry. The 2TP fitted spectra, in comparison, have less spectral contrast overall. Interestingly, while Phases 4 and 10 are in theory geometrically symmetric, the overall flux at phase 10 is lower and the difference between the two models is more noticeable at that phase. This is possibly due to the presence of an offset hotspot, leading to asymmetry between phases before and after secondary eclipse and a difference in fitted model parameters.

2TP-Crescent vs. 2TP-Free

As introduced in Section 3.2.2, we consider several ways to retrieve an atmosphere with two contrasting thermal profiles. For observed data in particular, we are interested in identifying longitudinal asymmetry in the atmosphere (e.g., due to a hot spot offset). Using the 2TP-Free model can inform us about potential inhomogeneity, and we may study how the interpretation of an atmosphere changes when there is more flexibility in geometry. Here we highlight the differences between the 2TP-Crescent model and the 2TP-Free model on the WASP-43b data set.

As described in Section 3.2.2, the 2TP-Crescent model uses the geometry described in Figure A.2. The 2TP-Free model uses a linear combination of fluxes from the day side and fluxes from the night side, parameterized with f_{day}, which accounts for the fraction contributed by the day side. There is no assumption of geometry or symmetry in the free model. The f_{day} parameter determines how much of the final spectrum is contributed by the hot TP profile, while the remaining flux is then attributed to the cool TP profile. We have the ability to see whether the retrieved f_{day} from a certain phase's data set is different from the corresponding value based on Equation A.1

In Figure 3.9, we show the posterior distributions as a function of phase for the parameter f_{day} from the 2TP-Free model. In particular, we note the lower-than-expected value at phase 2, 9, 11, and 12. This suggests a preference for lower flux from the day side or hotter temperatures and more flux from the night side or lower temperatures. Furthermore, this preference is seen mostly past secondary eclipse. WASP-43b has a known hot spot offset such that the maximum flux occurs before secondary eclipse

(Stevenson et al. 2017); the behavior in $f_{\rm day}$ as a function of phase is thus evidence for the offset. We can also examine phases 5 through 7 (right before to during secondary eclipse) and see that the $f_{\rm day}$ posteriors appear similar rather than finding higher values at secondary eclipse.

Next, we consider how the additional geometric flexibility impacts atmospheric inference. Figure 3.10 shows the posterior distributions for H_2O and CH_4 for these two models. CO, CO_2, and NH_3 are virtually identical and these comparisons have been left off the figure. The most significant differences for H_2O and CH_4 occur on the night side as seen after secondary eclipse, in particular at phases 11 and 12. These correspond to some of the phases that returned lower-than-expected day side flux contribution in the 2TP-Free model. The 2TP-Crescent model constrains CH_4 at these two phases. The 2TP-Free model, on the other hand, shows no detection of the molecule.

In terms of H_2O, the 2TP-Free model shows similar posterior distributions from phases 10 through 13. These values are consistent with the findings at phases 10 and 13 under the 2TP-Crescent model. However, the 2TP-Free values are elevated compared to the posteriors for phases 11 and 12 with 2TP-Crescent. These are the two phases that became non-detection under 2TP-Free for CH_4.

We find that only phase 11 has significance greater than 0.1σ when justifying the 2TP-Free model, which has one additional free parameter $f_{\rm day}$, over the 2TP-Crescent model. The detection significance at phase 11 is 3σ, or moderately in favor of 2TP-Free (see Figure 3.2). As a result, the more justified 2TP model appears to be the 2TP-Crescent model for modern observations. The 2TP-Free model, however, is able

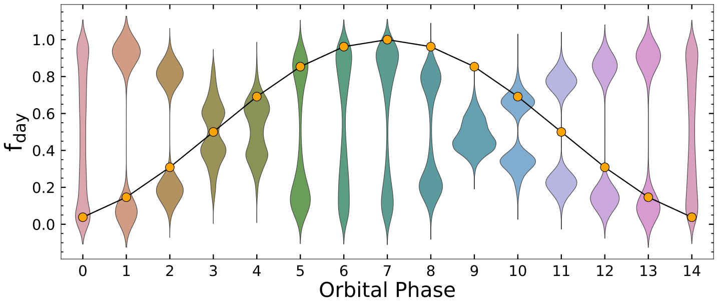


Figure 3.9: Posterior distribution of f_{day} from the 2TP-Free model using WASP-43b data. The distributions are bimodal due to the fact that we imposed no geometric information, thus showing the symmetric nature of each phase. Overplotted (orange circles connected with black line) is the expected emitting fraction for each phased based on Equation A.1. These expected values correspond to the total contribution from the day side in the 2TP-Crescent model. Phases 2, 9, 11, and 12 have posteriors constraining lower values than the expected, suggesting a preference for lower temperatures and less contribution from the day profile.

to account for existing asymmetries in the phase curve. In this regard, we note that the 2TP-Crescent model can be modified in the future to retrieve for an arbitrary phase angle such that the dayside contribution is not pre-determined based on geometry.

3.3.3 Simulated JWST data

From the above analysis on both simulated and true HST+Spitzer data, we found that substantial abundance and temperature biases exist when applying a single TP profile to several planetary phases. A challenging aspect emerged: standard Bayesian nested modeling tools seemed unable to rule out the simpler model which resulted in biased results. We now present the effects of the 1TP vs 2TP-Crescent assumption on simulated JWST data with the anticipation that the more suitable model will be more obvious.

A look back at the rightmost panel of Figure 3.2 shows that all phases (with

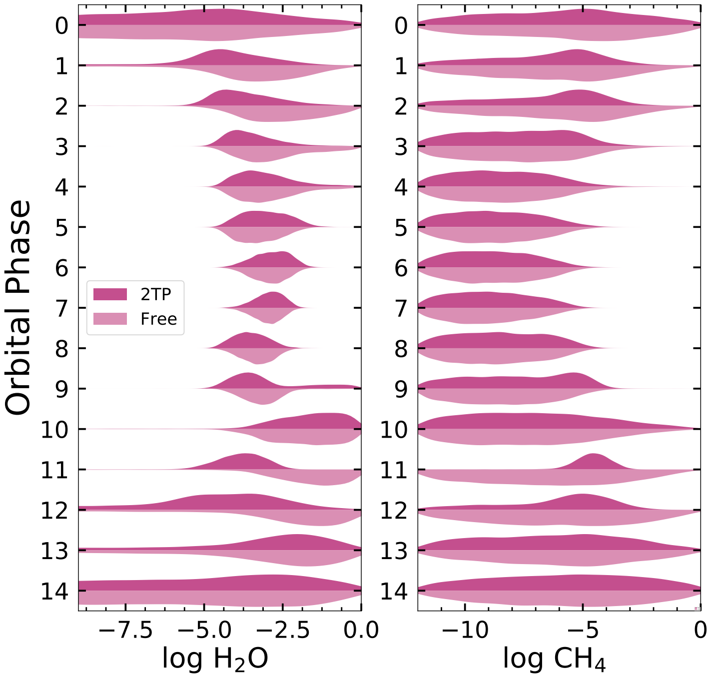


Figure 3.10: Abundance vs. phase results from WASP-43b data for H_2O (left) and CH_4 (right) for the 2TP-Crescent model (dark pink) and the 2TP-Free model (light pink). For each panel, we plot the posterior probability distribution of the log of the molecule's mixing ratio as a function of orbital phase. A noticeable difference is how 2TP-Crescent's constraint of CH_4 at phase 11 becomes a non-detection with 2TP-Free. The H_2O distributions from phases 10 - 13 with the 2TP-Free model look more similar to one another but these values are larger than estimates from the rest of the orbit.

the exception of secondary eclipse due to the complete lack of nightside contribution) demonstrate a clear preference for the more complex 2TP-Crescent model. We describe this result here.

The broad (1-10μm) and higher resolution ($R = 100$) simulated JWST data provides ultra-precise constraints on the gas mixing ratios when using the correct model. This high quality data permits us to readily disprove the 1TP hypothesis at all phases ($> 5\sigma$), at least in this scenario of a phase curve from a planet with large day-night temperature contrast. Figure 3.11 shows the retrieved molecular abundance constraints as a function of phase. Phase 7, or secondary eclipse, is the only phase where the 1TP model does not produce a bias, as expected. Phase 0 is not informative for abundances for either model, owing to the overall low night-side flux. For many phases, strong abundance biases across all the gases persist under the incorrect 1TP profile model.

While for the simulated HST data there is negligible difference in H_2O inference between the two models, all of the 1TP posteriors here for H_2O miss the input value until phase 6. The problem is worsened by the precision in the biased posterior distributions. For example, the 1TP H_2O constraint in phase 1 is $\log \mathrm{H_2O} = -5.28^{+0.14}_{-0.15}$ when in fact the true mixing ratio is $\log \mathrm{H_2O} = -3.37$ (or a 1TP bias of 12.7σ). The 2TP-Crescent (true model) at phase 1, however, results in a less precise ($^{+}_{-}0.35$), but more accurate (unbiased) constraint. Unsurprisingly, the water precision improves (up to $^{+}_{-}0.5$ dex at secondary eclipse) as more dayside is visible due to the higher emission feature signal to noise. We see up to a factor of 5 improvement in precision over the phase 7 posterior with 2TP-Crescent model using simulated HST data.

The 1TP CO constraints present less of a bias than H_2O, as seen in Figure 3.11. CO begins to largely deviate from the truth between phases 0 and 4. Although, phase 2 provides an unbiased yet over-constrained abundance compared to the 2TP-Crescent constraint. By phase 7, we can constrain CO using the 2TP-Crescent model to a 1σ of $\sim$ 90ppm ($\log \mathrm{CO} = -3.69^{+0.1}_{-0.09}$).

The poor performance of the 1TP model is further evident in CH_4, CO_2, NH_3 estimates. We find biased posteriors that are well constrained to 1/10th of an order of magnitude but their median values are several orders of magnitude away from the truth. CH_4, for instance, suffers from bias at phases 2-4. CO_2 is biased at phases 1 through 5, and NH_3 from phase 2 to 5. Meanwhile, the 2TP-Crescent model only detects upper limits for these molecules.

Figure 3.12 summarizes the retrieved TP profiles based on the two models. While overall trends are similar to what we see in simulated HST data, we find much more precisely constrained profiles across the orbit, by a factor of several better than with current data. Consequently, we note the presence of artificial temperature inversions at phases 2 and 4 under the 1TP model between 10^{-1} bar and 10^{-4} bar. At other phases, such as phase 3 and 5 (not shown in Figure 3.12), we do not find this phenomenon. This is yet another example of false conclusions that could arise from the overly simplistic 1TP profile assumption.

Figure 3.13 shows the simulated data and fitted spectra from the two models. As anticipated based on the abundance inference, the 1TP model produces poor fits. At phase 2, we see that the model spectra miss the data points between $1.4 - 3.2\mu$m,

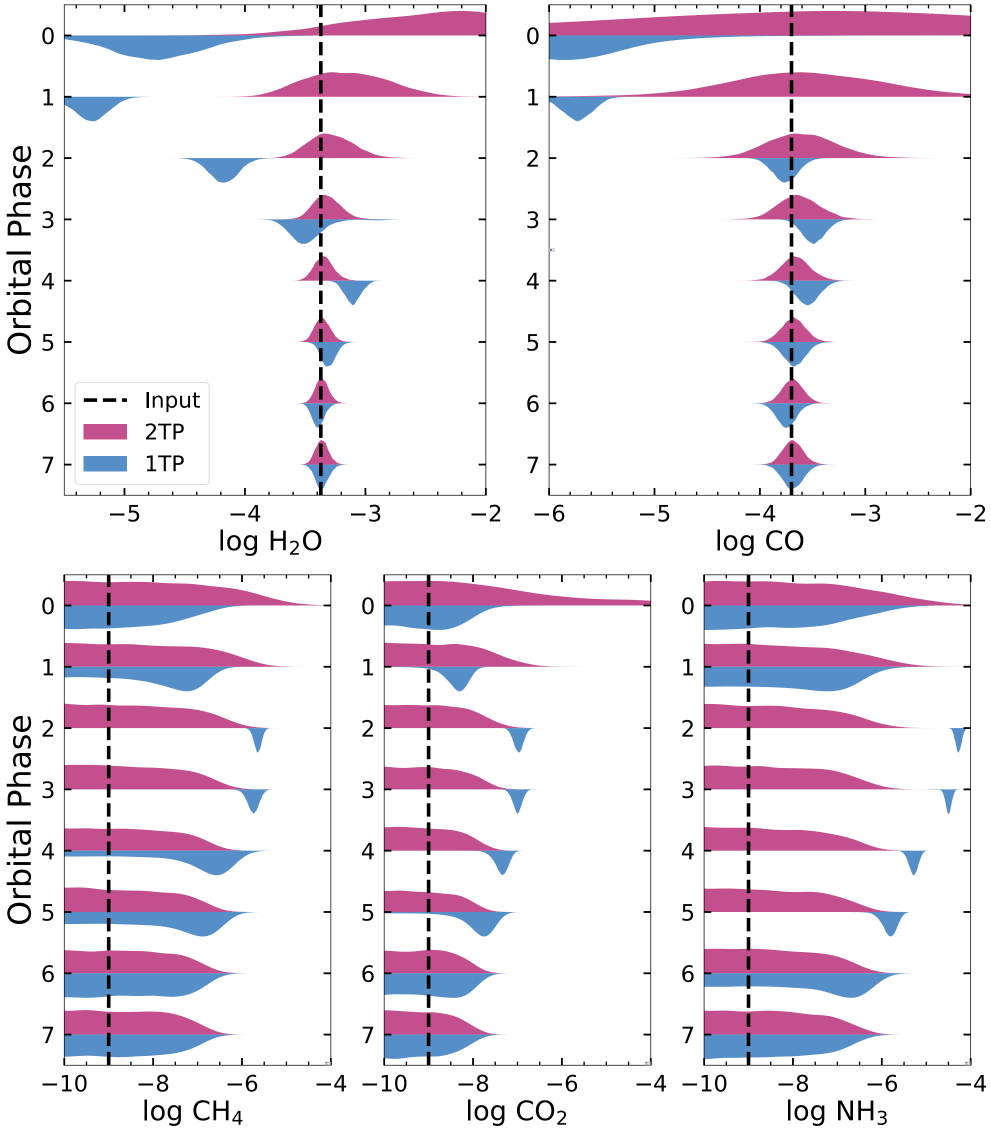


Figure 3.11: Abundance vs. phase results from *JWST* simulated data for H_2O, CH_4, CO, CO_2, NH_3 for the 1TP model (blue) and the 2TP-Crescent model (dark pink). For each panel, we plot the posterior probability distribution of the log of the molecule's mixing ratio as a function of orbital phase. The distributions are set to show the same total height at each phase and thus do not show the relative probability. For simulated data, we only consider half an orbit (transit to secondary eclipse), or eight orbital steps. For each molecule, we indicate its input abundance value with the vertical dashed line. The 1TP model produces constrained but bias posteriors for all molecules at multiple phases. Most of them have incorrect estimates for half the orbit. With the 2TP-Crescent model, we can get well-constrained and accurate estimates of H_2O and CO. We have upper limits for the remaining molecules, which do not have large input values to begin with.

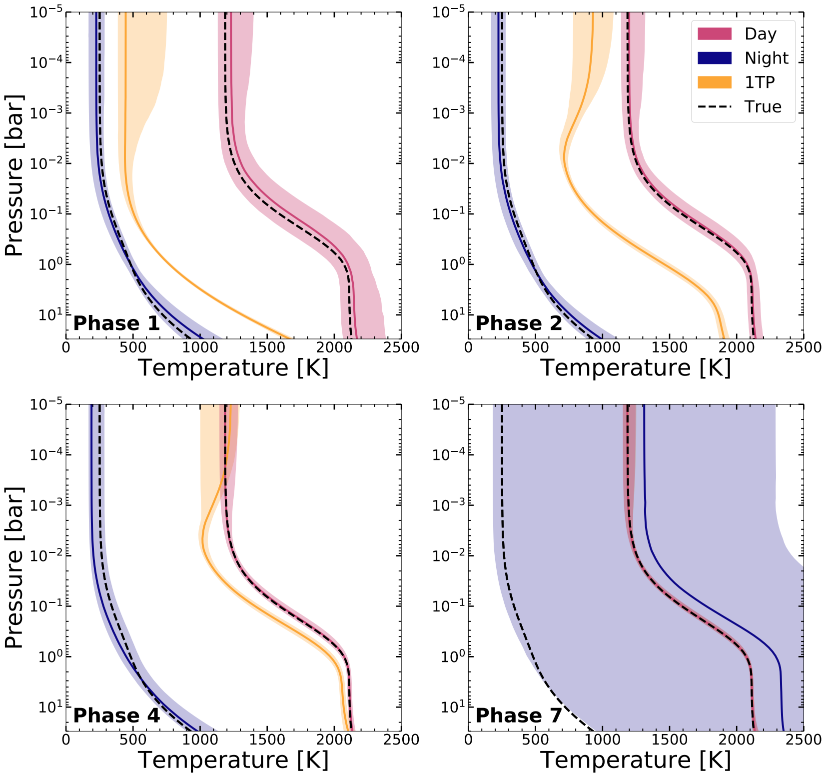


Figure 3.12: Temperature-pressure (TP) profiles for simulated JWST data. We selected phases 1, 2, 4, and 7 to show the change in TP profile constraint as a function of phase. In each panel, the dashed line represent the true input profiles for the day and night sides. The retrieved 2TP-Crescent profiles are in blue (night) and red (day). The retrieved 1TP profiles are in yellow. For each distribution, we show the median profile in a solid line, surrounded by the 2σ spread in profiles based on reconstructed random posterior draws. For certain phases, the 1TP profiles appear to have a temperature inversion. The 1TP profiles are close to the day-side profile as early as phase 4 (half day, half night). The 2TP profiles for day and night are accurate and precise.

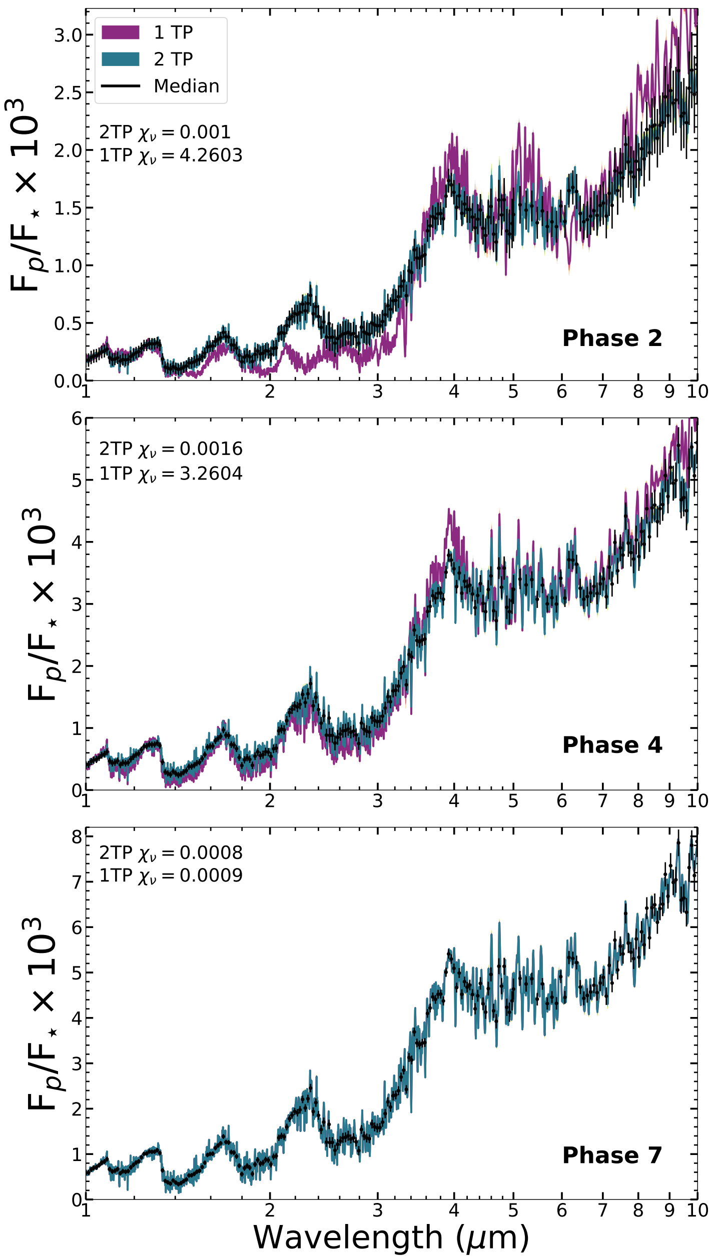


Figure 3.13: Simulated *JWST* data and high-resolution spectra generated with random posterior draws from the retrieval. Shown here are the spectra for phases 2, 4, and 7. 1TP spectra are in magenta while 2TP-Crescent spectra are in green. For each set of model spectra, we plot the median, 1σ, and 2σ contour. We include corresponding χ_ν values for the 1TP and 2TP(-Crescent) models, which can be small because random noise is not included. The *JWST* results are so precise that the contours are difficult to see. The 1TP spectra do not fit the majority of the data points at phases besides secondary eclipse.

with a reduction in flux with respect to the data. Then, between $3.6 - 10\mu$m, we see elevated flux. The quality of fit from the two models is easily distinguishable due to the precise 1σ and 2σ contours, and the 1TP model is not able to properly fit the data until just before secondary eclipse.

3.3.4 Joint Phase Retrieval

The previous phase-by-phase analyses are useful for determine phase-dependent properties like abundance variations. They also demonstrated the possibility of strong biases when using too simplistic of a model. Here we explore the feasibility of a joint phase curve retrieval whereby we simultaneously retrieve upon all phases, locking certain properties at each phase. The goal is to determine if improved precision can be obtained on atmospheric properties that are expected to be uniform with phase. For instance, while we expect temperatures and the abundances of some species (and clouds) to vary with longitude, the overall atmospheric metallicity, or carbon-to-oxygen ratios should not (provided the appropriate chemistry is taken into account). This in a sense would be an intermediate step before employing full "3D GCM-retrievals".

We explore the joint phase curve retrieval on three cases: both the simulated and Stevenson et al. (2017) HST+Spitzer WASP-43b phase curve data as well as simulated JWST phase curve data. In all three observational scenarios, we explore two separate 2TP modeling assumptions: the first is our new geometric method where limb darkening is appropriately accounted for per Gaussian quadrature annulus (2TP-Crescent), and the second is 2TP-Fixed, where we apply parameter f_{day}, a set value based on Equation A.1, to account for phase-dependent fluxes.

In these examples, we assume that the gas mixing ratios are the phase-independent quantities with only the contribution of the temperature profiles changing with phase. Figure 3.14 summarizes the abundance constraints under these conditions. We also include the constraints from Sections 3.3.2 and 3.3.3 at phase 7 (secondary eclipse) for comparison.

For most molecules, the joint retrieval shows similar results to the phase 7 retrieval in the case of simulated data. This is expected as the abundances are designed to remain constant with phase. The only case where the two joint methods differ significantly is for CH_4. While the joint 2TP-Crescent retrieval does not retrieve phantom elevated amounts of CH_4, the joint 2TP-Fixed retrieval suffers from this bias. CO is detected at the input value while having a large tail to the distributions that does not rule out lower values. Both models also detect CO_2 for the simulated data; however, we attribute these to the correlation seen between CO and CO_2 based on Spitzer data points.

For WASP-43b phase curve data, we see larger discrepancies between the two joint approaches as well as the results from the individual phase 7 retrieval. In contrast to the simulated case, it appears that the secondary eclipse conditions are not representative of the planet over the orbit. We do not see any CH_4 bias for either model. The CO distributions peak at different values, unlike in the simulated data results. CO_2, which is constrained again, exhibits similar results, with secondary eclipse peaking at the higest value, followed by joint 2TP-Crescent and then joint 2TP-Fixed; this is demonstration of the CO-CO_2 correlation.

We find a well-constrained distribution for water from the observed data, $\log \mathrm{H_2O} = -3.71^{+0.31}_{-0.24}$, when using the joint 2TP-Crescent retrieval. This is consistent with the 2TP-Fixed model finding within 1σ but about 3σ away from the constraint at secondary eclipse.

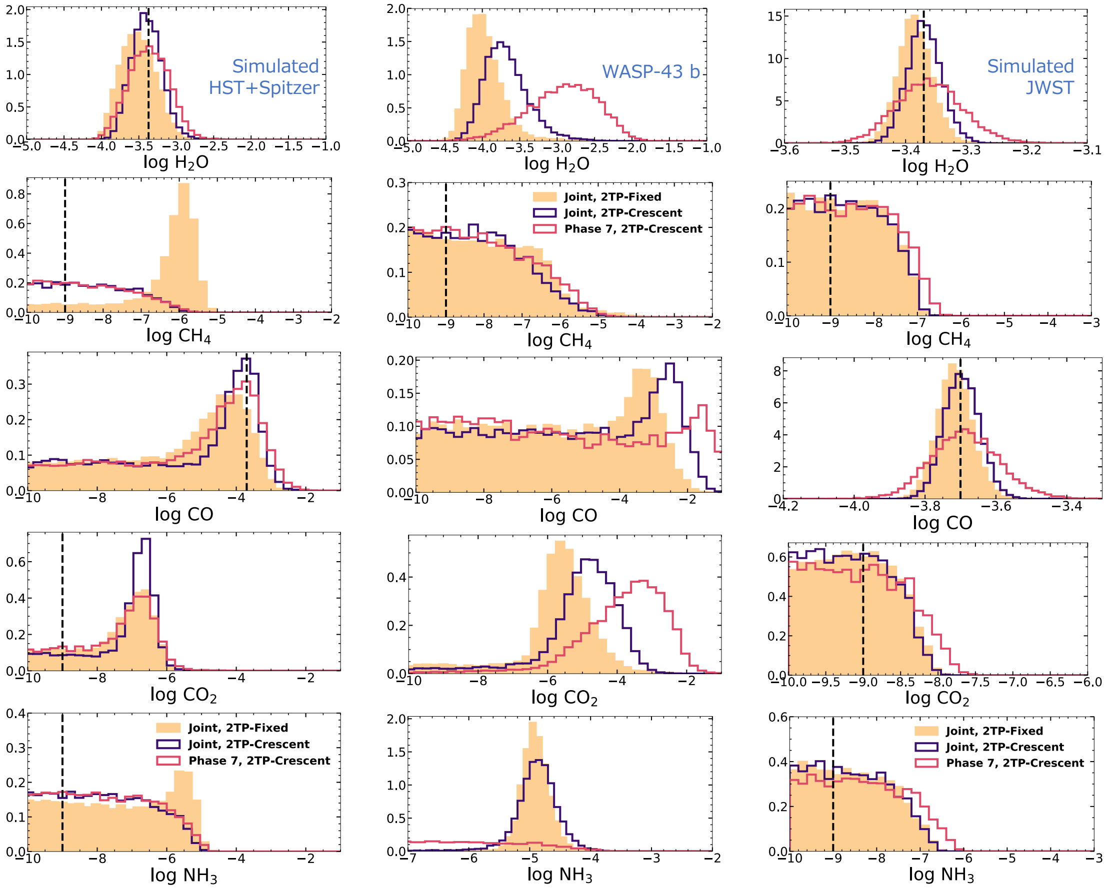
Kreidberg et al. (2014a) demonstrated the power of combining posteriors from multiple HST WFC3 data sets (secondary eclipse and transit) to precisely estimate the H_2O abundance for WASP-43b. Stevenson et al. (2017) added Spitzer data in the atmospheric analysis of WASP-43 b. Using a 1TP retrieval model, Stevenson et al. (2017) grouped dayside-dominant and nightside-dominant phases together, providing a H_2O estimate for each as there appeared to be lower H_2O on the nightside. Figure 3.15 compares the H_2O distribution from the joint retrieval of WASP-43b spectroscopic phase curve data to the estimates from Kreidberg et al. (2014a) and Stevenson et al. (2017). Our approach differs in that we did not combine posteriors from the phase-by-phase retrievals after the fact but instead utilized the sum of log-likelihoods at each phase to derive a "self-consistent" posterior. The joint phase retrieval places the mixing ratio of H_2O to be between a 1σ range of $1.1 \times 10^{-4} - 3.9 \times 10^{-4}$, more consistent with the Stevenson et al. (2017) estimate using dayside phases (1σ of $1.4 \times 10^{-4} - 6.1 \times 10^{-4}$). We see less variation as a function of phase in H_2O abundance using the 2TP-Crescent model, thus we did not differentiate between day and night phases. As seen in Figure 3.15(a), our H_2O estimate is not only consistent with Kreidberg et al. (2014a) but is also more precise. We note that we agree with the constraint determined by Irwin et al. (2020) as well, where H_2O is $(2 - 10) \times 10^{-4}$ for WASP-43b based on Stevenson et al.

(2017) data.

Another noteworthy result from the joint WASP-43b retrievals is the constraint of NH_3. Both 2TP-Fixed and 2TP-Crescent approaches agree on $\log \mathrm{NH_3} = -4.89^{+0.29}_{-0.34}$. Strong constraints should always be met with skepticism. However, we note that the retrieved constraint is consistent with expectations from solar-composition disequlibirum chemical models of similar hot Jupiters ($\sim 1 \times 10^{-7} - 1 \times 10^{-4}$ over the atmospheric pressures probed in emission; e.g., Moses et al. (2011), although this depends on the value of $T_{\rm int}$ (Thorngren et al. 2019)).

Figure 3.16 shows the spectra fits from the joint retrieval of the observed WASP-43b data compared to the 2TP-Crescent phase-by-phase retrievals. The fits from the joint retrieval are poorer fits to the data than the phase-by-phase case. We expect this result because the joint retrieval is restricted by needing to fit the same set of abundances and profiles to different phases at once. The only change from phase to phase in the joint retrieval is the relative area between the hotter and the cooler profile, and thus the shape of the overall spectra from the joint retrieval looks the same at all phases, just at varying levels of total flux. Yet, the use of the full phase curve data set is worth further development since the data from one phase is not independent to the next. Together they paint a holistic image of a planet's atmosphere, and that relationship between phases should be reflected in a retrieval framework for phase curves (e.g., Irwin et al. 2020).

For completeness, in Figure 3.17, we show TP profile constraints under the joint retrieval for all three data sets. The joint retrievals place tighter constraints on

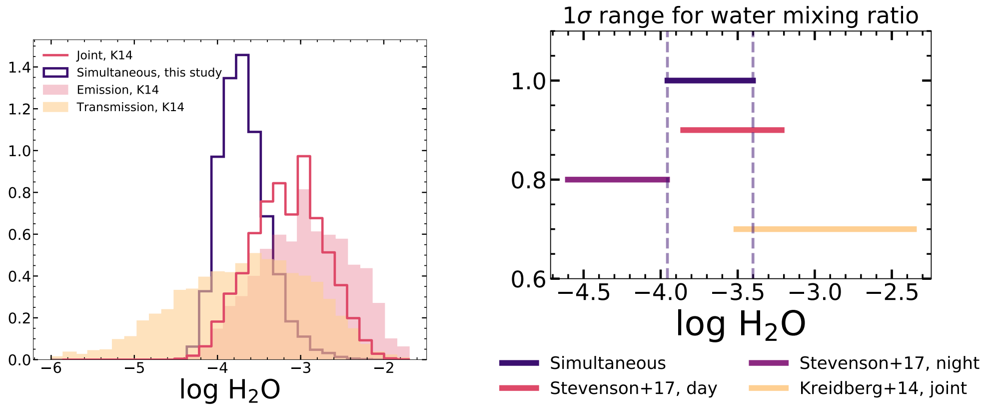


(a) Simulated HST+Spitzer: Constraints from joint retrievals

(b) WASP-43b: Constraints from joint retrievals

(c) JWST: Constraints from joint retrievals

Figure 3.14: Left: Constraints distributions of the five retrieved molecules from simulated HST+Spitzer data with: (1) averaged posterior from the phase-by-phase retrievals from Section 3.3.1, (2) joint retrieval of all phases using the 2TP-fixed model,(3) joint retrieval of all phases using the 2TP-Crescent model. Dashed line indicates the input value for each molecule. Middle: Posterior distributions of the same cases using WASP-43b data. The joint retrieval is able to return more precise distributions; in some cases, however, the advantage of combining multiple data sets also enhances bias in the result. Right: Posterior distributions of the same cases using simulated JWST data. Distributions from jointly-done retrievals indicate stronger, more precise detection. 2TP-Fixed and 2TP-Crescent approaches yield similar results.



(a) Comparing H_2O constraints: joint vs. Kreidberg et al. (2014a) (K14) transmission, emission, and joint

(b) Comparing 1σ for H_2O estimates

Figure 3.15: Comparing existing H_2O estimates and the estimate from this study using the joint retrieval for WASP-43b. Fig. 3.15(a): We plot the posterior distribution for H_2O from the joint retrieval along with the distributions from Kreidberg et al. (2014a). These include the posterior based on secondary eclipse only, transmission only, and the joint distribution (multiplication of the two posteriors) from the two sets of observations. Fig. 3.15(b): Illustration of the 1σ range of H_2O estimates from the Kreidberg et al. (2014a) joint distribution, Stevenson et al. (2017), and this study. The Stevenson et al. (2017) results are based on multiplying the posteriors from phases grouped as day (first to third quarter) and night and determine corresponding joint H_2O posteriors. Vertical dashed lines are placed to guide the eye during comparison. The joint retrieval constraint of H_2O is lower than the Kreidberg et al. (2014a) 1σ range, but it is overall consistent with their joint distribution and with the dayside estimate from Stevenson et al. (2017).

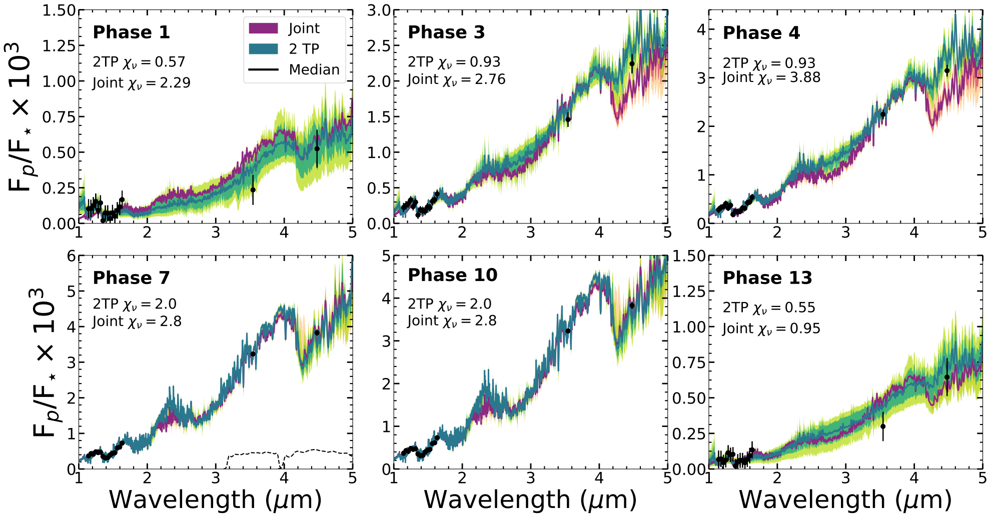
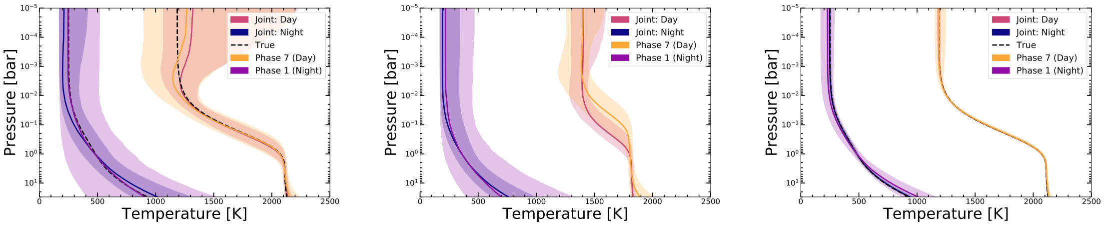


Figure 3.16: WASP-43b data (HST+Spizter) and high-resolution spectra generated with random posterior draws from the joint retrieval. Shown here are the spectra for phases 1, 3, 4, 7, 10, and 13. Overplotted are the spectral fits from the phase-by-phase 2TP-Crescent retrievals (see Figure 3.8). We include corresponding χ_ν values for the Joint and 2TP(-Crescent) cases. In the panel of phase 7, we overplot the Spitzer 3.6μm and 4.5μm filter transmission. Jointly-fit spectra are in magenta while 2TP-Crescent spectra are in green. For each set of model spectra, we plot the median, 1σ, and 2σ contour. Although the constraints are more precise with the joint retrievals, the goodness-of-fit is worse compared to the phase-by-phase scenario. This is expected given only one set of parameters (abundances, TP profiles) were allowed in order to fit all the phases.

the TP profiles for all data sets; based on the simulated cases, we can see that the joint retrieval is able to accurately reconstruct the true profiles. That is to say, if an atmosphere is indeed dominated by two contrasting profiles, the joint approach is able to identify them.

Finally, Figure 3.14(c) summarizes the abundance constraints resulting from the joint-phase fits for the simulated JWST data. The major advantage seen in simulated retrieval results of JWST data is once again the precision along with the accuracy; here, the precision improvement is almost a factor of two. We also find that 2TP-Fixed and 2TP-Crescent provide similar constraints. This suggests that 2TP-Fixed imposes



(a) TP constraints from joint retrievals, simulated HST/Spitzer

(b) TP constraints from joint retrievals, WASP-43b

(c) TP constraints from joint retrievals, simulated JWST

Figure 3.17: Constraints of day and night temperature-pressure profiles from joint retrievals of (left) simulated HST/Spitzer data, (middle) observed WASP-43b data, and (right) simulated JWST data. The profiles shown are the median fit and 2σ envelope of fits from the retrievals. Also included are the phase-by-phase retrieval results of the day profile from phase 7 (secondary eclipse) and night profile from phase 1 (right after transit); see Section 3.3.1. Dashed lines are the input profiles for the simulated data. The joint retrievals are able to accurately model the true profiles in simulated cases, and provide more precise constraints on the TP profiles than the phase-by-phase retrievals for all data sets.

little bias in the context of our simulated JWST data.

3.4 Discussion and Conclusions

Spectroscopic phase curves offer insight into planetary climate and chemistry by providing a measure of the 2D nature of species abundances and temperature. We can maximize our leverage of these powerful data sets with atmospheric retrievals.

We have generalized our previous non-homogeneous temperature retrieval methodology (Feng et al. 2016) to arbitrary phases using a new geometry scheme. We investigated several TP modeling scenarios: 1TP, 2TP-Crescent, 2TP-Free, and 2TP-Fixed. A 2TP setup uses two profiles to explain the variation in flux as a function of phase while a 1TP setup relies on a changing profile throughout the orbit. We combined these scenarios with different observational setups: simulated HST WFC3 spectroscopy with Spitzer IRAC photometry, actual HST+Spitzer data for WASP-43b from Stevenson et al. (2017), and simulated JWST data (NIRISS + NIRCam + MIRI LRS). By both using Bayesian model selection and examining the posteriors with respect to the input values for simulated data, we were able to determine which phases need the use of a 2TP model to accurately interpret the atmosphere.

Our simulated HST+Spitzer data setups provided the following insights:

- Even phases closer to transit (i.e., more nightside) can constrain the dayside profile because the hotter profile provides more flux.
- We are justified in using a 2TP model for three out of eight phases (which span half an orbit).
- For HST+Spitzer, several of the phases that favored the 1TP model in model-

comparison context returned biased abundance posterior distributions. In particular, CH_4 appeared artificially constrained at higher values than the input for phases 1 through 5 of the simulated HST+Spitzer data.

- H_2O constraints are robust regardless of model choice.
- Upper limits are placed on CO but CO_2 has biased detection; this results from the correlation between the two molecules given overlapping features within the Spitzer photometric bands.

In the case of simulated JWST data:

- Every phase strongly justified the use of two profiles except for secondary eclipse (as there is no second TP contribution to the flux). The 1TP model was severely biased in terms of abundance retrievals for all molecules.
- Certain instances of the 1TP model also showed signs of temperature inversion in the profile, adding another layer to its inaccuracy.
- With the 2TP-Crescent model, JWST provides precise constraints on the mixing ratios of H_2O and CO, offering up to a factor of five improvement over HST+Spitzer results.
- We found that the wavelengths of 1.4-3.2 and 3.6-10 micron best differentiated the 1TP and the 2TP models.

The important distinguishing wavelength ranges we identified are missing in modern observations. Taylor et al. (2020) examined the information content for different

JWST observing modes for a planet with inhomogeneous temperature structure, establishing NIRSpec's observing range (~1-5 μm) as especially effective. By considering four different amounts of hot profile contribution, Taylor et al. (2020) also investigated the impact of phase angle, finding prominent abundance biases when relying on the 1D model as we have. When we combine longitudinal information from different phases, we can leverage the retrieval technique to better optimize JWST observations.

Meanwhile, our application of the different model approaches to the Stevenson et al. (2017) WASP-43b phase curve data reveals:

- A little under half of the phases have moderate evidence in favor of a 2D model, mostly for phases with more night side visible.
- The 2TP-Crescent model retrieved consistent profiles for the day and night sides over the orbit.
- The 1TP model finds peaked CH_4 distributions at 8 out of the 15 orbital phases that are reminiscent of the biased distributions seen in the simulated data set.
- There is no evidence of CO, while CO_2 is constrained. This is likely the same artificial behavior seen in the simulated data caused by the CO-CO_2 correlation.
- H_2O is mostly consistent from phase to phase. However, the 2TP-Free model prefers higher values at phases 10-13 than 2TP-Crescent (and 1TP). Thus, the implementation of a more complex model (thermal inhomogeneity in our case) can affect interpretation.

- We identified the evidence of the hot-spot offset on WASP-43b based on the 2TP-Free model results.

We also found that we can combine observed data and our phase-dependent retrieval approach to identify interesting phenomena associated with asymmetry over an orbit. The 2TP-Crescent model retrieves upper limits for CH_4 except at phase 11; with the Free model, that phase is an upper limit instead. The relaxed assumption about day-side emission fraction in the 2TP-Free model allows the model to fit for the slope between the Spitzer points with less day side flux contribution while the 2TP-Crescent model needed significant absorption due to CH_4 to do so. It is helpful to look at the full orbit to spot consistency or outliers.

Recently, Mendonça et al. (2018a) and Morello et al. (2019) reanalyzed the Spitzer points from Stevenson et al. (2017) - several of the "anomalous" phases (e.g., phase 11, or 0.75) where we detected CH_4 in the 2TP-Crescent model have had their points shifted upwards, particularly the 3.6 μm band point. While we did not retrieve on this reanalysis to be consistent with the approach in Irwin et al. (2020), this could yield different results of CH_4, CO, and CO_2 abundances. Further investigation is warranted; however, this is another reason why retrieving on the full phase curve can be beneficial to provide a more holistic picture of an atmosphere.

Finally, we introduced the concept of a joint phase curve retrieval and applied that to simulated HST+Spitzer data, the Stevenson et al. (2017) WASP-43b data set, and simulated JWST observations and found:

- The 2TP-Crescent and 2TP-Fixed models were consistent in performance, although the 2TP-Fixed model was more prone to biased detection.

- NH_3 is tightly constrained (about half a dex) to $\sim 10^{-5}$ for WASP-43b by the 2D models. This is an interesting constraint that is plausibly consistent with expectations from disequilibrium chemistry models of similar temperature hot Jupiters. JWST will show if this is real or another bias.
- We can place a constraint on H_2O for WASP-43b at 1σ range of 1.1×10^{-4} – 3.9×10^{-4}, increasing precision while remaining consistent with previous studies.
- For simulated JWST data, the increase in precision of the constraints compared to the phase-by-phase approach is notable for H_2O and CO, approximately by a factor of two.

Based on these results, it would be best to combine both the phase-by-phase and joint retrieval approaches. Phase-by-phase retrievals identify outliers, providing more insight to accuracy, while the joint can improve precision for molecules with accurate inference. We should also be strategic about applying different models to different phases in order to probe the different temperature structures (e.g., 1D for secondary eclipse).

3.4.1 Future Work

The value of phase curve retrievals is clear. The promise of richer data sets will necessarily demand the advancement of 3D retrieval techniques, in turn improving our understanding of the 3D-nature of planets. In this study, we used our forward model to generate the data. Onwards, we plan to use 3D GCM models (e.g., Blecic et al. 2017;

Irwin et al. 2020) to provide the spectra that we retrieve to better identify degeneracies and inform retrieval forward model expansions, including:

1. Non-uniform chemistry. As our model currently assumes the same chemical composition and distribution for the day and night profiles, an important next upgrade should be to allow differences in the mixing ratios. Bayesian model comparison will once again be important in determining if the data quality justifies the inclusion of the extra parameters associated with more complex chemical profiles. Three-dimensional modelling has shown evidence of species transport (Drummond et al. 2020), so retrievals should study its detectability and differentiate between transport and biased constraints. Another important consideration is that we are not seeing the constraint capabilities of observations for molecules such as CO_2, CH_4, or NH_3, which were chosen to be low in abundance in our study.

2. Various temperature contrasts. We note that unlike in Feng et al. (2016), where we vary the temperature contrast between the day and night side, we maintained a fixed, large contrast to better isolate differences. This contrast also models the observed WASP-43b data well. In follow-up work, we suggest implementing multiple contrasts to better model other hot Jupiters. It may be worthwhile to include a means of modeling a more gradual temperature change from the day to night side, as done in Irwin et al. (2020); this could be done in conjunction with abundance variation.

3. Clouds. Our scenarios remain cloud-free, so the implementation of clouds in our retrieval framework can extend our study of existing and future data sets. Irwin et al. (2020)'s 2.5D retrieval does not assume clouds either, and infer the presence of

clouds on the night side of WASP-43b based on the retrieved low temperatures, which are evident in our retrievals as well. Venot et al. (2020) perform cloudy and cloud-free retrievals on simulated JWST MIRI phase curves of WASP-43b. This important groundwork finds that the $5 - 12\,\mu$m range can confirm or rule out the presence of clouds, although this is dependent on cloud bulk properties and composition. With a variety of aerosol implementation available, there is much to be done in cloudy retrievals (Barstow et al. 2020). It is, however, worth noting that strong evidence for relatively cloud-free daysides.

4. More specialized geometry. We can apply our annulus framework to geometries beyond crescent phases. For instance, we can explore hot spots on the dayside prominent near secondary eclipse, complementing future eclipse mapping studies. Ohno & Zhang (2019) presented the impact of planetary obliquity on phase curves, and our retrieval model can be adapted to study the detection of obliquity.

Chapter 4

Characterizing Earth Analogs in Reflected Light: Atmospheric Retrieval Studies for Future Space Telescopes

4.1 Introduction

The scientific field of exoplanets has been rapidly advancing since the hallmark discovery of the first planet orbiting a Sun-like star (Mayor & Queloz 1995). Following the launch of NASA's *Kepler* mission (Borucki et al. 2003, 2011), the field has seen the discovery of thousands of transiting exoplanets and the exciting result that planets with radii between 0.75–2.5 $R_{\oplus}$ are common around solar-type stars (Burke et al. 2015). Only within the last decade have observational studies for exoplanet atmospheric characterization seen substantial development, starting with the first detection of an exoplanet's atmosphere by Charbonneau et al. (2002).

To date, the majority of exoplanet atmospheric characterization investigations have focused on transiting worlds. Hot Jupiters, owing to their large sizes and short orbital periods, are typically emphasized as targets for these studies. Characterization of small, potentially rocky exoplanets is limited to worlds with cool stellar hosts (K and M dwarfs), which offer favorable planet-to-star size ratios. Recently, de Wit et al. (2016) studied the combined transmission spectra of two transiting Earth-sized planets orbiting the ultracool dwarf TRAPPIST-1 using the *Hubble Space Telescope*. While no gas absorption features were detected by de Wit et al. (2016), this work highlights the improvements in signal size when terrestrial-sized transiting planets are studied around low-mass stars. Additionally, since the Habitable Zone (Kasting et al. 1993) around a low-mass star is relatively close-in, characterization studies of potentially habitable exoplanets around cool stars can benefit from the frequency of transit events. However, for Sun-like hosts, the planet-to-star size ratio is much less favorable and the Habitable Zone is located far from the star, thus severely limiting the potential for atmospheric characterization.

Direct, high-contrast imaging has now emerged as an essential technique for studying the atmospheres of planets at larger orbital separations from their host star (i.e., at orbital distances $\gtrsim 1$ au). Thus far, high-contrast imaging has been proven successful in studying atmospheres of young, self-luminous gas giants in the near-infrared and mid-infrared (e.g., Barman et al. 2011; Skemer et al. 2014; Macintosh et al. 2015). These worlds, owing to their intrinsic brightness, have typical contrast ratios of 10^{-4} with respect to their hosts. A true Jupiter analog at visible wavelengths, by comparison, would have a contrast ratio of 10^{-9}, while an Earth analog would have a contrast ratio of order 10^{-10}. Reflected light in the visible probes to atmospheric depths of up to ~ 10 bar for giant planets (Marley et al. 2014), which is complimentary to the relatively low pressures probed in transit observations (typically less than 10–100 mbar). Additionally, the wavelength range of 0.4–1.0 μm holds rich information about a planet's atmosphere, including signatures of methane, water vapor, and haze (Marley et al. 2014; Burrows 2014).

In spite of the incredible technological challenges, there are multiple planned or in-development space-based missions that would be capable of high-contrast imaging of exoplanets in reflected light. First among these will be NASA's Wide-Field InfraRed Survey Telescope (*WFIRST*, Spergel et al. 2013), which was identified as the top priority space mission in the 2010 National Academy of Sciences Decadal Survey of Astronomy and Astrophysics[2]. The *WFIRST* mission will carry a Coronagraphic Instrument (CGI) with imaging capability and a visible-light integral field spectrograph of wavelength resolution ~ 50 (Noecker et al. 2016; Trauger et al. 2016; Seo et al. 2016; Cady et al. 2016; Balasubramanian et al. 2016; Groff et al. 2018). Although envisioned primarily as a technology demonstrator, it may study the atmospheres of relatively cool gas giant exoplanets that have been previously detected using the radial velocity technique (Traub et al. 2016).

---

[2] `http://sites.nationalacademies.org/bpa/bpa_049810`

While *WFIRST* could also have some capability to survey stars in the solar neighborhood for lower-mass planetary companions (Burrows 2014; Greco & Burrows 2015; Spergel et al. 2015; Savransky & Garrett 2016; Robinson et al. 2016), it is anticipated that the core optical throughput of the *WFIRST* CGI will be low for planetary signals. This stems primarily from the complexities of accommodating for *WFIRST*'s on-axis secondary mirror and support structures within the high-contrast instruments (Traub et al. 2016; Krist et al. 2016). Low throughput drives long requisite integration times, thereby likely making spectroscopic observations of smaller, less-bright worlds (such as super-Earth exoplanets) unfeasible except around the very closest stars (Robinson et al. 2016). However, if the *WFIRST* spacecraft were to be paired with an external starshade (Cash 2006; Kasdin et al. 2012), the CGI can be operated in a direct mode without coronagraphic masks, substantially increasing throughput. High-contrast imaging of sub-Neptune and terrestrial-sized exoplanets may then become possible. The feasibility of a starshade "rendezvous" with the *WFIRST* spacecraft is under active investigation (Seager et al. 2015; Crill & Siegler 2017).

In advance of the 2020 astronomy and astrophysics decadal survey, several large-scale space-based mission concepts are being studied[3]. Of these, two have a strong focus on the characterization of rocky exoplanets with direct imaging: the Habitable Exoplanet Imaging Mission (HabEx; Mennesson et al. 2016) and the Large UltraViolet/Optical/InfraRed Surveyor (LUVOIR; Peterson et al. 2017). HabEx and LUVOIR are incorporating aspects of design that would allow the detection of water vapor and biosignatures on planets in the Habitable Zones of nearby Sun-like stars. It is

---

[3] `https://science.nasa.gov/astrophysics/2020-decadal-survey-planning`

therefore timely and critical that we explore observational approaches that maximize science yield during the development of these large-scale mission concepts as well as the *WFIRST* rendezvous concept. To accomplish this, we must perform atmospheric and instrument modeling to simulate the types of spectra we can expect to measure, and we must develop tools to infer planetary properties from these simulated observations.

Traditionally, the comparison to a limited range of forward models has been used to infer atmospheric properties (such as temperature structure and gas abundances) from spectral observations. This involves iterating to a radiative-convective solution for a given set of planetary parameters (e.g., gravity, metallicity, equilibrium abundances, incident flux), and can include detailed treatment of aerosols, chemistry, and dynamics within the model atmosphere (Marley & Robinson 2015). The goal is to generate a spectrum that matches available data and, thus, offers one potential explanation for the world's atmospheric state (e.g., Konopacky et al. 2013; Macintosh et al. 2015; Barman et al. 2015). A more data-driven interpretation of atmospheric observations is accomplished through inverse modeling, or retrievals. Developed for Solar System studies and remote sensing (e.g., Rodgers 1976; Irwin et al. 2008), retrievals have become a valuable tool in constraining our understanding of the atmospheres of transiting exoplanets. Early exoplanet retrieval work invoked grid-based optimization schemes (Madhusudhan & Seager 2009), while subsequent works have taken advantage of Bayesian inference with methods such as optimal estimation and Markov chain Monte Carlo (MCMC) (e.g., Lee et al. 2012; Benneke & Seager 2012; Line et al. 2013).

Several studies have examined the hypothetical yield from characterizing giant exoplanets observed with a space-based coronagraph (such as *WFIRST*) with retrieval techniques. Marley et al. (2014), for example, modeled spectra we could expect from known radial velocity gas giants if observed by the *WFIRST* CGI. Given the diversity of cool giant planets, the model spectra have a variety of input assumptions for clouds, surface gravity, and atmospheric metallicity. Marley et al. (2014) then applied retrieval methods to these synthetic spectra, enabling the exploration of how well atmospheric parameters are constrained under varying quality of data. Lupu et al. (2016) further investigated the feasibility of characterizing cool giant planet atmospheres through retrieval, focusing on the ability to constrain the CH_4 abundance and cloud properties. The systematic study of the impact of conditions like signal-to-noise ratios or wavelength resolution is essential to quantifying the scientific return of these reflected-light observations. Nayak et al. (2017) considered the impact of an unknown phase angle on the inference of properties such as planet radius and gravity. In all of these studies, the signal-to-noise ratio (SNR) of the data has a significant influence on the constraints of atmospheric properties.

Previous work on smaller planets in the context of possible future space missions includes von Paris et al. (2013), who synthesized infrared emission observations of a cloud-free, directly-imaged Earth-twin, and employed a least-squares approach and χ^2 maps to perform retrievals and explore parameter space (considering the effects of instrument resolution and SNRs). A collection of recent studies (Wang et al. 2017b; Mawet et al. 2017; Wang et al. 2017a) examined atmospheric species detection using "High Dispersion Coronagraphy", which couples starlight suppression technologies with high resolution spectroscopy. In these studies, simulated observations (typically at spectral resolutions, $R = \lambda/\Delta\lambda$, of many hundreds to tens of thousands) are cross-correlated with template molecular opacity spectra to explore the feasibility of species detection. While this novel approach can yield detections of key atmospheric constituents, the abundance of these these atmospheric species cannot be robustly constrained.

To date, there still does not exist a systematic study of atmospheric characterization of small exoplanets using retrieval techniques on reflected light observations at spectral resolutions relevant to *WFIRST* rendezvous, HabEx, and LUVOIR. Motivated by this need, we present here our extension of Bayesian retrieval techniques into the terrestrial regime. We construct a forward model suitable for simulating reflectance spectra of Earth-like planets in the visible wavelength range of 0.4 μm to 1.0 μm. We explore retrievals of planetary and atmospheric properties from simulated data sets at varying spectral resolutions and SNRs. A retrieval framework such as this allows us to quantify uncertainties we expect for key planetary parameters given certain observing scenarios. Thus, our approach enables us to search for the minimal observing conditions that achieve the scientific goal of identifying traits associated with habitability and life. In particular, we are interested in our ability to detect and constrain abundances of molecules such as water, oxygen, and ozone, characterize basic properties of a cloud layer, and measure bulk parameters such as radius.

In section 4.2, we describe our forward model and construction of simulated data. In section 4.3, we validate our forward model by building up retrieval complexity (i.e., number of retrieved parameters). We perform a study of retrieval performance with respect to spectral resolution and SNR in section 4.4, with implications for HabEx/LUVOIR. We also study the retrieval performance for data sets expected from a *WFIRST* rendezvous scenario, where the CGI would provide modest-resolution spectroscopy in the red (600–970 nm) and photometry in the blue (480–600 nm). We present our discussion and conclusions in sections 4.5 and 4.6, respectively.

4.2 Methods

The observed quantity for a directly imaged exoplanet in reflected light at a given phase (i.e., planet-star-observer) angle, α, is the wavelength dependent planet-to-star flux ratio,

$$\frac{F_p}{F_s} = A_g \Phi(\alpha) \left(\frac{R_p}{r} \right)^2 , \tag{4.1}$$

where A_g is the geometric albedo, $\Phi(\alpha)$ is the phase function, R_p is the radius of the planet, and r is the orbital separation. The phase function (which depends on wavelength) translates the planetary brightness at full phase (i.e., where $\alpha = 0°$) to its brightness at different phase angles. The wavelength dependent geometric albedo is the ratio between the measured flux from the planet at full phase to that from a perfectly reflecting Lambert (i.e., isotropically-reflecting) disk with the size of the planet. We denote the product of the geometric albedo and the phase function as the phase dependent "reflectance" of the planet. In general, the geometric albedo encodes information about the composition and structure (i.e., "state") of an atmosphere, while the phase function is strongly related to the scattering properties of an atmosphere (e.g., Marley et al. 1999; Burrows 2014).

To understand the information contained in direct imaging observations of exoplanets in reflected light, we employ a retrieval (or inverse analysis) framework that consists of several linked simulation tools and models. Of central importance is a well-tested three-dimensional albedo model—described in greater depth below—that computes a reflectance spectrum at high resolution for a planet given a description of its atmospheric state (McKay et al. 1989; Marley et al. 1999; Cahoy et al. 2010; Lupu et al. 2016; Nayak et al. 2017). When coupled with a simulator for degrading a high resolution spectrum to match the resolution of an instrument, we refer to these two tools as the "forward model." By adding simulated noise to forward model spectra, we generate faux "observations" of worlds as would be studied by future high-contrast imaging missions. To create "observed" spectra, we adopt a widely-used direct imaging instrument simulator (Robinson et al. 2016) that generates synthetic observations given an input, noise-free spectrum.

Given an "observed" planet-to-star flux ratio spectrum, our inverse analyses use a Bayesian inference tool that compares the observation to forward model outputs to sample the posterior probability distributions for a collection of atmospheric state parameters. In other words, our inverse analyses indicate what range of atmospheric state parameters (e.g., gas abundances) adequately describe a direct imaging observation. Our Bayesian parameter estimations use an open-source affine invariant Markov-Chain Monte Carlo (MCMC) ensemble sampler—`emcee` (Goodman & Weare 2010; Foreman-Mackey et al. 2013).

In this work, retrieval analyses generally proceed by first simulating a noise-free spectrum of a world with a known atmospheric state (e.g., Earth). We then add simulated observational noise to this spectrum to create a synthetic observation. Following Bayesian parameter estimation on this synthetic observation, we can compare a retrieved atmospheric state to the original, known atmospheric state, thereby allowing us to understand how observational noise affects our ability to deduce the true nature of an exoplanetary atmosphere.

4.2.1 Albedo Model

Our three-dimensional albedo model (see also Cahoy et al. 2010) divides a world into a number of plane-parallel facets with coordinates of longitude (ζ) and co-latitude (η), with the former referenced from the sub-observer location and the latter ranging from 0 at the northern pole to π at the southern pole. A single facet has downwelling incident stellar radiation from a zenith angle $\mu_{\rm s} = \cos\theta_{\rm s} = \sin\eta\cos\left(\zeta - \alpha\right)$, where, as earlier, α is the phase angle. The facet reflects to the observer in a direction whose zenith angle is given by $\mu_{\rm o} = \cos\theta_{\rm o} = \sin\eta\cos\zeta$. Note that, at full phase (where the geometric albedo is defined) the observer and the source are colinear such that $\mu_{\rm o} = \mu_{\rm s}$ for all facets.

The atmosphere above each facet is divided into a set of pressure levels, and we perform a radiative transfer calculation to determine the emergent intensity. With the intensities calculated for an entire visible hemisphere, we follow the methods outlined by Horak (1950) and Horak & Little (1965) to perform integration using Chebychev-Gauss quadrature, thus producing the reflectance value at a given wavelength. We repeat this procedure at each of the wavelength points within a specified range to build up a reflectance spectrum.

Taking $I(\tau,\mu,\phi)$ to be the wavelength-dependent intensity at optical depth τ in a direction determined by the zenith and azimuth angles μ and ϕ, we ultimately need to determine the emergent intensity from each facet in the direction of the observer, $I(\tau=0,\mu_\mathrm{o},\phi_\mathrm{o})$. Thus, for each facet we must solve the one-dimensional, plane-parallel radiative transfer equation,

$$\mu\frac{dI}{d\tau} = I(\tau,\mu,\phi) - S(\tau,\mu,\phi), \tag{4.2}$$

where S is the wavelength dependent source function. Following Meador & Weaver (1980), Toon et al. (1989), and Marley & Robinson (2015), the source function is

$$\begin{aligned} S(\tau,\mu,\phi) = {} & \frac{\bar{\omega}}{4\pi}F_\mathrm{s}\cdot p(\tau,\mu,\phi,-\mu_\mathrm{s},\phi_\mathrm{s})\cdot \mathrm{e}^{-\tau/\mu_\odot} \\ & + \bar{\omega}\int_0^{2\pi} d\phi' \int_{-1}^{1}\frac{d\mu'}{4\pi}\cdot I(\tau,\mu',\phi')\cdot p(\tau,\mu,\phi,\mu',\phi'), \end{aligned} \tag{4.3}$$

where $\bar{\omega}$ is the single scattering albedo, F_s is the incoming stellar flux at the top of the atmosphere (which we normalize to unity so that emergent intensities correspond to reflectivities), ϕ_s is the stellar azimuth angle, and p is the scattering phase function. Note that our source function does not include an emission term since we are not computing thermal spectra. Recall that the first term in Equation 4.3 describes directly scattered radiation from the direct solar beam while the final term describes diffusely scattered radiation from the (μ',ϕ') direction scattering into the (μ,ϕ) direction.

Like most standard tools for solving the radiative transfer equation, we separate treatments of directly-scattered radiation from diffusely-scattered radiation, and, for both, it is convenient to express the scattering phase function in terms of a unique scattering angle, Θ. As single scattered radiation typically has more distinct forward and backward scattering features, we choose to represent the scattering phase function for the direct beam with a two-term Henyey-Greenstein (TTHG) phase function (Kattawar 1975),

$$p_{\mathrm{TTHG}}(\Theta) = f p_{\mathrm{HG}}(g_{\mathrm{f}}, \Theta) + (1 - f) p_{\mathrm{HG}}(g_{\mathrm{b}}, \Theta) \tag{4.4}$$

where p_{HG} is the Henyey-Greenstein (HG) phase function with,

$$p_{\mathrm{HG}} = \frac{1}{4\pi} \frac{1 - \bar{g}^2}{(1 + \bar{g}^2 - 2\bar{g}\cos\Theta)^{3/2}} \ . \tag{4.5}$$

Recall that a TTHG phase function can represent both forward and backward scattering peaks, while the (one term) Henyey-Greenstein phase function only has one peak (typically in the forward direction). In the previous expressions, $\bar{g}$ is the asymmetry parameter, f is the forward/backward scatter fraction, g_{f} is the asymmetry parameter for the forward-scattered portion of the TTHG, and g_{b} is the asymmetry parameter for the backward-scattered portion of the TTHG. For the forward and backward scattering portions of the TTHG phase function, we use $g_{\mathrm{f}} = \bar{g}$, $g_{\mathrm{b}} = -\bar{g}/2$, and $f = 1 - g_{\mathrm{b}}^2$. Substituting these into the TTHG phase function expression yields,

$$p_{\mathrm{TTHG}} = \frac{\bar{g}^2}{4} p_{\mathrm{HG}}(-\bar{g}/2, \Theta) + (1 - \frac{\bar{g}^2}{4}) p_{\mathrm{HG}}(\bar{g}, \Theta). \tag{4.6}$$

For radiation that is single-scattered from the solar beam to the observer, the scattering geometry is fixed by the planetary phase angle such that $\Theta = \pi - \alpha$. Our choice of parameters, and their relation to $\bar{g}$, in the TTHG was designed by Cahoy et al. (2010) to roughly reproduce the phase function of liquid water clouds. This parameterization, however, is different from that proposed by Kattawar (1975). We do not expect our results to be sensitive to the details of a particular phase function treatment as Lupu et al. (2016) showed that scattered-light retrievals struggle to constrain phase function parameters.

We adopt a standard two-stream approach to solving the radiative transfer equation (Meador & Weaver 1980). In this case, the diffusely-scattered component of the source function is azimuthally averaged. Combined with our representation of the directly-scattered component, we have,

$$\begin{aligned} S(\tau, \mu, \mu_{\mathrm{s}}, \alpha) = & \frac{\bar{\omega}}{4\pi} F_{\mathrm{s}} \cdot p_{\mathrm{TTHG}}(\mu, -\mu_{\mathrm{s}}) \cdot \mathrm{e}^{-\tau/\mu_{\mathrm{s}}} \\ & + \frac{\bar{\omega}}{4\pi} \int_{-1}^{1} I(\tau, \mu') p(\mu, \mu') d\mu', \end{aligned} \tag{4.7}$$

where the azimuth-averaged phase functions are given by,

$$p(\mu, \mu') = \frac{1}{2\pi} \int_{0}^{2\pi} p(\mu, \phi, \mu', \phi) d\phi \ . \tag{4.8}$$

We represent the azimuth-averaged scattering phase functions as a series of Legendre polynomials, $P_l(\mu)$, expanded to order M with,

$$p(\mu, \mu') = \sum_{l=0}^{M} g_l P_l(\mu) P_l(\mu') \ , \tag{4.9}$$

where the phase function moments, g_l, are defined according to,

$$g_l = \frac{2l+1}{2} \int_{-1}^{1} p(\cos\Theta) P_l(\cos\Theta) d\cos\Theta \ . \tag{4.10}$$

The first moment of the phase function is related to the asymmetry parameter, with $\bar{g} = g_1/3$. We use a second order expansion of the phase function, giving,

$$p(\mu, \mu') = 1 + 3\bar{g}\mu\mu' + \frac{g_2}{2}(3\mu^2 - 1)(3\mu'^2 - 1) \ . \tag{4.11}$$

In a given atmospheric layer of our albedo model, the optical depth is the sum of the scattering optical depth and the absorption optical depth, $\tau = \tau_{\rm scat} + \tau_{\rm abs}$. The scattering optical depth has contributions from Rayleigh scattering and clouds, so that $\tau_{\rm scat} = \tau_{\rm Ray} + \bar{\omega}_{\rm cld}\tau_{\rm cld}$, where $\bar{\omega}_{\rm cld}$ is the cloud single scattering albedo. The single scattering albedo for a layer is then $\bar{\omega} = \tau_{\rm scat}/\tau$. We determine the asymmetry parameter, $\bar{g}$, with an optical depth weighting on the Rayleigh scattering asymmetry parameter (which is zero) and the cloud scattering asymmetry parameter, yielding $\bar{g} = \bar{g}_{\rm cld}(\tau_{\rm cld}/\tau_{\rm scat})$. When representing the second moment of the phase function, we use $g_2 = \frac{1}{2}(\tau_{\rm Ray}/\tau_{\rm scat})$ so that g_2 tends towards the appropriate value for Rayleigh scattering (i.e., 1/2; Hansen & Travis 1974) when the Rayleigh scattering optical depth dominates the scattering optical depth.

4.2.2 Model Upgrades

As compared to prior investigations that have used the Cahoy et al. (2010) albedo tool (e.g., Lupu et al. 2016; Nayak et al. 2017), we have updated the model to include an optional isotropically-reflecting (Lambertian) lower boundary (mimicking a planetary surface), and have added pressure-dependent absorption due to H_2O, O_3, O_2, and CO_2. CH_4 remains a radiatively active species in the model, as in previous studies. We also include Rayleigh scattering from H_2O, O_2, CO_2, and N_2 (in addition to H_2 and He from previous studies). As in Lupu et al. (2016), we allow for an extended gray-scattering cloud in our atmospheres.

In Lupu et al. (2016), their two-layer cloud model atmosphere includes a deeper, optically thick cloud deck that essentially acts as a reflective surface. Unlike the gas giants within that study, though, terrestrial planets have a solid surface we can probe. We characterize our isotropically-reflecting lower boundary using a spherical albedo for the planetary surface, A_{s}, which represents the specific power in scattered, outgoing radiation compared to that in incident radiation. For this study, we simply adopt gray surface albedo values, which reduces complexity and computation time. For the inhomogeneous surface of a realistic Earth, featuring oceans and continents, the surface albedo is wavelength-dependent, and we hope to investigate the significance of such surfaces in future work.

We undertook a test to check our reflective lower boundary condition in the limit of a transparent atmosphere. Without atmospheric absorption or scattering, our assumption of a Lambertian surface would imply that the reflectivity (or phase function) determined by our albedo code should follow the analytic Lambert phase function,

$$\Phi_{\mathrm{L}}(\alpha) = \frac{\sin\alpha + (\pi - \alpha)\cos\alpha}{\pi}. \tag{4.12}$$

Figure 4.1 compares the model phase function with the analytic phase function and shows complete agreement, confirming that our treatment of the surface is correct.

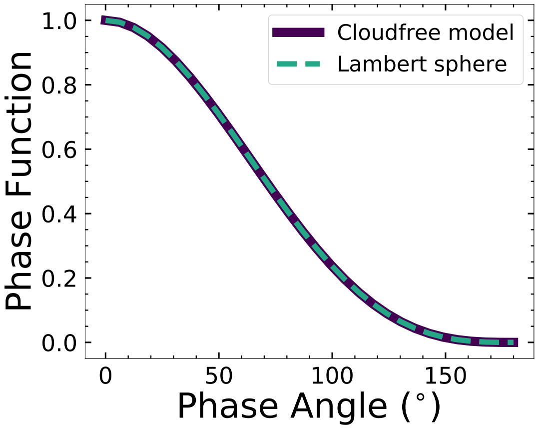


Figure 4.1: Comparing our model phase function to the analytic Lambertian phase function (Equation 4.12). No atmospheric absorption or scattering is present in the forward model.

Previous work featuring the albedo model adopted here used a pre-defined atmospheric pressure grid. To accommodate the finite surface pressures of rocky planets as well as the various combinations of cloud parameters our retrievals will explore, we instead establish an adaptive method of determining the pressure grid. Here, we divide the atmosphere into a pressure grid of N_{level}, bounded by $P = P_{\mathrm{top}}$ at the top of the atmosphere and $P = P_0$ at the surface. In a cloud-free scenario, we simply divide the atmosphere evenly in log-P space.

For our simulations that include a single cloud deck, we adaptively determine the pressure value at each level depending on the location, thickness, and optical depth of the cloud. The quantities that define the cloud deck are p_{t}, the cloud-top pressure, dp, the atmospheric pressure across the cloud, and τ, the cloud optical depth. We begin by assigning a number of layers to the cloud, imposing two conditions: (1) there should be at least three model pressure layers to each atmospheric pressure scale height ($\texttt{perH} = 3$), and (2) the cloud optical depth in a layer must remain below at most 5 ($\texttt{maxtau} = 5$). This allows us to avoid any one layer from spanning a large extent within the atmosphere, and also avoids cloud layers that have extremely large scattering optical depths.

When beginning our gridding process, we propose an initial number of cloud layers, $N_{\mathrm{c}} = \texttt{perH} \times \texttt{numH}$, where $\texttt{numH} = \ln \frac{p_{\mathrm{t}}+dp}{p_{\mathrm{t}}}$ is the number of e-folding distances through the cloud (serving as a proxy for scale height). The aerosol optical depth for each pressure layer within the cloud would then simply be $\Delta\tau = \frac{\tau}{N_{\mathrm{c}}}$. However, if $\Delta\tau > \texttt{maxtau}$, we adjust the cloud resolution by increasing N_{c} by a factor of $\frac{\Delta\tau}{\texttt{maxtau}}$ and then round up to the nearest integer. In other words, we increase the resolution of the pressure grid through the cloud until the layer optical depth is under `maxtau`. We determine successive pressure level values through the cloud with $p[i] = p[i-1] + \Delta \ln p$, where $\Delta \ln p = \frac{\ln{(p_{\mathrm{t}}+dp)} - \ln p_{\mathrm{t}}}{N_{\mathrm{c}}}$, starting from the top of the cloud. We divide the remaining $N_{\mathrm{level}} - N_{\mathrm{c}}$ levels in uniform $\ln p$ space on either side of the cloud, weighted by the number of pressure scale heights above (N_{t}) and below (N_{b}) the cloud. Figure 4.2 visualizes the three portions of the atmosphere.

For simplicity, we assume an isothermal atmosphere (at $T = 250$ K), as temperature has little effect on the reflected-light spectrum (Robinson 2017). Pressure, however, has a strong impact on molecular opacities, as seen in Figure 4.3. We incorporated high-resolution pressure-dependent opacities for all molecules in our atmosphere. The absorption opacities are generated line-by-line from the HITRAN2012 line list (Rothman et al. 2013) for seven orders of magnitude in pressure ($10^{-5} - 10^{2}$ bar) at T, spanning our entire wavelength range at < 1 cm^{-1} resolution. Figure 4.3 also illustrates how absorption features of H_2O, O_2, and O_3 change when in an atmosphere of 1 bar versus one of 10 bar.

We interpolate our high-resolution opacity tables to the slightly lower resolution of the forward model in order to maintain short model runtimes while not affecting the accuracy of the output spectra. For each model layer, we interpolate over the opacities from our table given the pressure. The chemical abundances in our forward model atmosphere are constant as a function of pressure, and we also adopt a uniform acceleration due to gravity.

We have also added an option to include partial cloudiness across a planetary disk, whose fractional coverage is described by f_c. To mimic partial cloudiness as we see on Earth, we call the forward model twice. We use the same set of atmospheric and planetary parameters for both calls, except for the cloud optical depth. “Cloudy” is the call that has a non-zero cloud optical depth, while “cloud-free” is the call where we set cloud optical depth to zero. Each call returns a geometric albedo spectrum, and we combine the two sets with the fractional cloudiness parameter such that the combined spectrum follows $f_c \times \text{cloudy} + (1 - f_c)\times$ cloud-free.

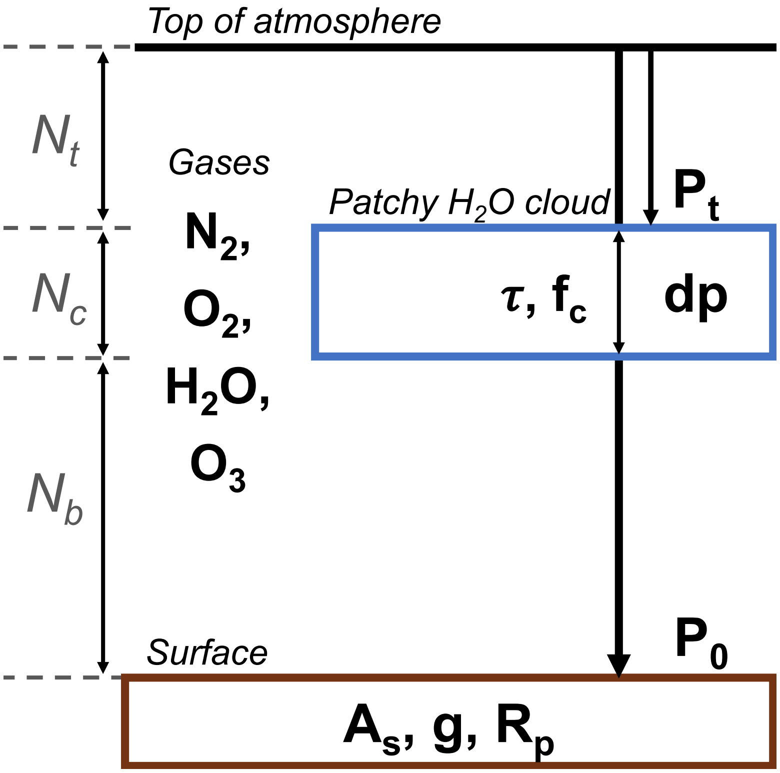


Figure 4.2: Illustrative schematic of our model atmosphere's structure. The atmosphere has $N_t + N_c + N_b$ layers. Table 4.1 lists the definitions, fiducial values, and priors of the presented parameters.

4.2.3 Albedo Model Fiducial Values and Validation

The generalized three-dimensional albedo model described above can simulate reflected-light spectra of a large diversity of planet types, spanning solid-surfaced worlds to gas giants with a variety of prescribed atmospheric compositions. For the present study, however, we choose to focus on Earth-like worlds, which are described in detail below. Thus, we define a set of fiducial model input parameters that are designed to mimic Earth and thereby enable us to generate simulated observational datasets for an Earth twin.

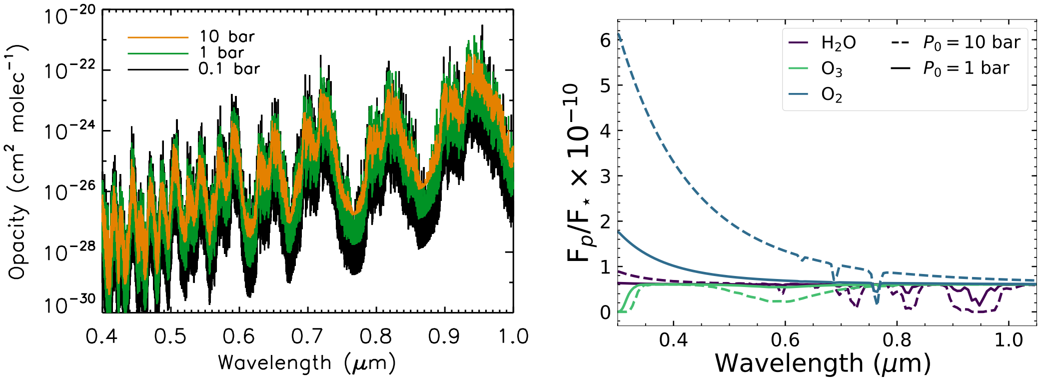


Figure 4.3: *Left:* High resolution (1 cm^{-1}) H_2O opacities from 0.4-1.0 μm at three different pressures: 0.1 bar, 1 bar, and 10 bars. *Right:* Absorption features in a $R = 140$ spectrum from 0.3 - 1.05 μm of H_2O, O_2, and O_3 at fiducial mixing ratios listed in Table 4.1 at $P = 1$ bar and $P = 10$ bar. For each spectrum here, the atmosphere only contains the stated molecule and a radiatively inactive filler gas to match the pressure.

Table 4.1 summarizes the fiducial model parameter values adopted for our Earth twin. Also shown are the priors for these parameters, which we use when performing retrieval analyses. For an Earth-like setup, the surface atmospheric pressure is $P_0 = 1$ bar and we adopt a surface albedo of $A_\mathrm{s} = 0.05$, which is representative of mostly ocean-covered surface. We adopt a uniform acceleration due to gravity of $g = 9.8$ m s^{-2} and set the planetary radius to $R_\oplus$. For convenience, we sometime refer to these four variables (P_0, A_s, g, and R_p) as the bulk planetary and atmospheric parameters.

We focus on molecular absorption due to H_2O, O_3, and O_2. While our albedo model includes opacities from CH_4 and CO_2 as well, we omit these two species as the reflected-light spectrum of Earth in the visible contains no strong features for these molecules. The input values for the molecular abundances (or volume mixing ratios) are $H_2O = 3 \times 10^{-3}$, $O_3 = 7 \times 10^{-7}$, and O_2= 0.21. These abundance values are based on column weighted averages from a standard Earth model atmosphere with vertically-

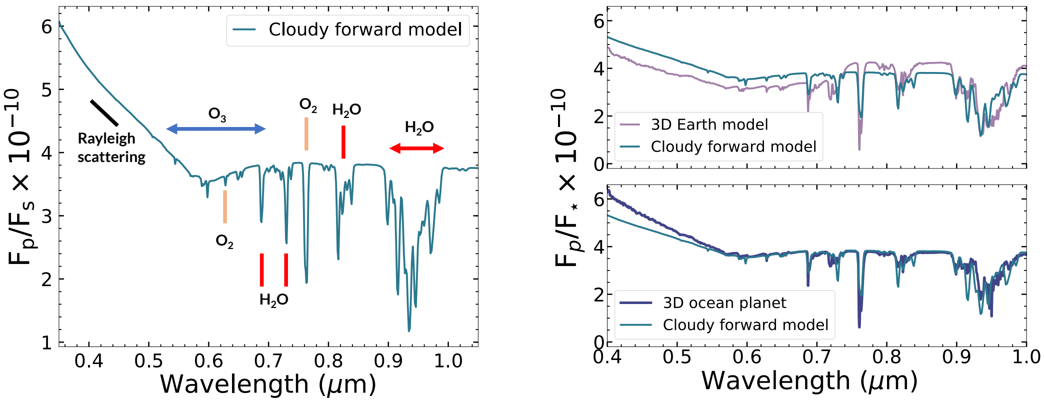


Figure 4.4: *Left:* The spectrum generated with the forward model in this study using fiducial values from Table 4.1. Key spectral features from the atmospheric species in our model are labeled. *Right, top:* Comparison of the cloudy forward model in this study using fiducial values from Table 4.1 to a spectrum from a more computationally complex three-dimensional (3D) forward model of Earth at full phase described in Robinson et al. (2011). *Right, bottom:* Comparison of the cloudy forward model to a spectrum of a planet generated using the 3D model from Robinson et al. (2011) that is like Earth except it only has ocean coverage.

varying gas mixing ratios (McClatchey et al. 1972). The primary Rayleigh scatterer and background gas in our fiducial model is N_2, whose abundance makes up the remainder of the atmosphere after all other gases are accounted for (i.e., roughly 0.79). Rayleigh scattering is treated according to Hansen & Travis (1974) with constants to describe the scattering properties of N_2, O_2, and H_2O from Allen & Cox (2000). We do not include polarization or Raman scattering effects.

Our cloud model was designed to be minimally parametric while still enabling us to sufficiently reproduce realistic spectra of Earth. Our single-layer gray H_2O cloud has $\bar{\omega} = 1$ and $\bar{g} = 0.85$, which are characteristic of water clouds across the visible range. These two parameters were fixed to minimize retrieval model complexity, as we believe that water is the most likely condensate for worlds in the Habitable

Zone. Nevertheless, future studies may not wish to assume values of $\bar{\omega}$ and $\bar{g}$ *a priori*. Cloud top pressure (p_{t}) and fractional coverage (f_c) are set at 0.6 bar and 50%, respectively, which are roughly consistent with observations of optically thick cloud coverage on Earth (Stubenrauch et al. 2013). Cloud thickness (dp) and optical depth (τ) were set to 0.1 bar and 10, respectively, based on results from the MODIS instrument (http://modis-atmos.gsfc.nasa.gov) used in Robinson et al. (2011).

With fiducial values chosen, we validate our forward model against a simulated high-resolution disk-integrated spectrum of Earth at full phase, as shown in Figure 4.4. The comparison spectrum is produced by the NASA Astrobiology Institute's Virtual Planetary Laboratory (VPL) sophisticated 3D line-by-line, multiple scattering spectral Earth model (Robinson et al. 2011). The Robinson et al. (2011) tool can simulate images and disk-integrated spectra of Earth from the ultraviolet to the infrared. It has been validated against observations at visible wavelengths taken by NASA's *EPOXI* mission (Robinson et al. 2011) and NASA's *LCROSS* mission (Robinson et al. 2014).

Features of the Robinson et al. (2011) model include Rayleigh scattering due to air molecules, realistic patchy clouds, and gas absorption from a variety of molecules, including H_2O, CO_2, O_2, O_3, and CH_4. Surface coverage of different land types (e.g., forest, desert) is informed by satellite data, and water surfaces incorporate specular reflectance of sunlight. A grid of thousands of surface pixels are nested beneath a grid of 48 independent atmospheric pixels, all of equal area. For each surface pixel, properties from the overlying atmospheric pixels are used as inputs to a full-physics, plane-parallel radiative transfer solver— the Spectral Mapping Atmospheric Radiative

Transfer (SMART) model (Meadows & Crisp 1996). Intensities from this solver are integrated over the pixels with respect to solid angle, thereby returning a disk-integrated spectrum.

The sophistication of the Robinson et al. (2011) model makes it unsuitable to retrieval studies, however, as model runtimes are measured in weeks for the highest-complexity scenarios. This, in part, justifies our adoption of a minimally-parameteric albedo model (whose runtime is measured in seconds). Furthermore, as in Figure 4.4, our efficient albedo model reproduces all of the key features of the Robinson et al. (2011) model. The most notable differences—that the efficient model, as compared to the Robinson et al. (2011) model, is more reflective in the blue and less reflective in the red—are simply due to our adoption of a gray surface albedo. Land and plants, which cover roughly 29% of Earth's surface, are generally more reflective in the red than in the blue. Figure 4.4 also compares a spectrum from our forward model against a spectrum of a partially clouded ocean planet generated with the Robinson et al. (2011) model. This ocean world is identical to Earth except for the fact that its surface is covered entirely by an ocean, with no land present. The surface albedo in the ocean model is gray beyond 500 nm; shortward of this the reflectivity increases, likely leading to the discrepancy in our comparison at the bluest wavelengths. Still, with a more accurate match to a planet that has a nearly gray albedo through the visible, we consider our assumption of gray surface albedo to be the main reason for the discrepancies when compared to the Robinson et al. (2011) realistic model.

Finally, in our albedo model we set 100 facets for the visible hemisphere and

Table 4.1: List of the 11 retrieved parameters in the complete cloudy model, their descriptions, fiducial input values, and corresponding priors.

Parameter	Description	Input	Prior
$\log P_0$ (bar)	Surface pressure	$\log(1)$	[-2,2]
$\log \mathrm{H_2O}$	Water vapor mixing ratio	$\log(3 \times 10^{-3})$	[-8,-1]
$\log \mathrm{O_3}$	Ozone mixing ratio	$\log(7 \times 10^{-7})$	[-10,-1]
$\log \mathrm{O_2}$	Molecular oxygen mixing ratio	$\log(0.21)$	[-10,0]
R_p ($\mathrm{R}_\oplus$)	Planet radius	1	[0.5, 12]
$\log g$ (m s^{-2})	Surface gravity	$\log(9.8)$	[0,2]
$\log A_s$	Surface albedo	$\log(0.05)$	[-2, 0]
$\log p_t$ (bar)	Cloud top pressure	$\log(0.6)$	[-2,2]
$\log dp$ (bar)	Cloud thickness	$\log(0.1)$	[-3,2]
$\log \tau$	Cloud optical depth	$\log(10)$	[-2,2]
$\log f_c$	Cloudiness fraction	$\log(0.5)$	[-3,0]

calculate a high-resolution geometric albedo spectrum at 1000 wavelength points between 0.35μm and 1.05μm. Like Lupu et al. (2016), we only consider a planet at full phase ($\alpha = 0°$). While direct imaging missions will not obtain observations of exoplanets at full phase, this assumption makes little difference for our results as we are not computing integration times and only work in SNR space. Additionally, as Nayak et al. (2017) followed up Lupu et al. (2016) by retrieving phase information from giant planets in reflected light, we anticipate performing a similar expansion in the future. Our forward model has 61 pressure levels in an isothermal atmosphere of 250 K, bounded below by a reflective surface. The top of the atmosphere is set at $P_{\mathrm{top}} = 10^{-4}$ bar.

4.2.4 Retrieval Setup and Noise Model

We convert a high resolution geometric albedo spectrum to a synthetic planet-to-star flux ratio spectrum given the resolution of an instrument and a noise model. We

Table 4.2: Simulated data sets.

	$R = 70,\ R = 140$	*WFIRST* Rendezvous [a]
Wavelength (μm)	0.4 – 1.0	0.506, 0.575;[b] $R = 50$: 0.6 – 0.97[c]
Data quality	$\mathrm{SNR}_{550\mathrm{nm}} = 5,\ 10,\ 15,\ 20$	$\mathrm{SNR}_{600\mathrm{nm}} = 5,\ 10,\ 15,\ 20$

Note. — We do not randomize the noise for any of the data sets.

[a] Using *WFIRST* Design Cycle 7 values from `https://wfirst.ipac.caltech.edu/sims/Param_db.html`

[b] The first photometric band is centered on 0.506 μm and covers 0.48–0.532 μm. The second photometric band is centered on 0.575 μm and covers 0.546–0.6 μm. We assume 100% transmission.

[c] We combine three integral field spectrograph bands into one at $R = 50$ from 0.6 μm to 0.97 μm. Separated, they are 0.6–0.72 μm, 0.7–0.84 μm, and 0.81-0.97 μm.

Table 4.3: Four cumulative models for retrieval validation, as described in Section 4.3.

Model Variant		Retrieved Parameters	$\mathrm{N}_{\mathrm{param}}$
I	Surface conditions	P_0, A_s	2
II	+ Bulk properties	P_0, A_s, g, R_p	4
III	+ Gas mixing ratios	P_0, A_s, g, R_p H_2O, O_2, O_3	7
IV	+ Cloud properties	P_0, A_s, g, R_p H_2O, O_2, O_3 p_t, dp, τ, f_c	11

Note. — See Table 4.1 for the corresponding definition and prior of each parameter. Model IV represents the full suite of parameters and can serve as a reference for the fixed parameters in Models I through III.

then apply a Bayesian inference tool on the synthetic data set to sample the posterior probability distributions of the forward model input parameters. To perform Bayesian parameter estimation, we utilize the open-source affine invariant Markov-Chain Monte Carlo (MCMC) ensemble sampler `emcee` (Goodman & Weare 2010; Foreman-Mackey et al. 2013). Ensemble refers to the use of many chains, or walkers, to traverse parameter space; as a massively parallelized algorithm, it is computationally efficient. Affine-invariance refers to the invariant performance under linear transformations of parameter space, enabling the algorithm to be insensitive to parameter covariances (Foreman-Mackey et al. 2013). With a cloudy retrieval, we can expect complex correlations that a sampler should be able to reveal. As it is more agnostic to the shape of the posterior, we choose `emcee` following Nayak et al. (2017) over `Multinest`, another sampler Lupu et al. (2016) considered that yielded consistent results. The albedo model is coded in Fortran; we convert it into a Python-callable library with the F2PY package. Each call to the forward model takes approximately 10 seconds of clock time on an 8-core processor. To visualize the MCMC results, we utilize the `corner` plotting package developed by Foreman-Mackey (2016).

Table 4.1 lists the priors for our parameters. We offer a generous range on the molecular abundances; we allow O_2 in particular to be able to dominate the atmospheric composition. Our choice of radius range (0.5–12 $R_\oplus$) reflects the range of of planetary sizes from Mars to Jupiter. Also, when performing retrievals, we impose two limiting conditions to maintain physical scenarios. First, we limit the mixing ratio of N_2, $f_{N_2} = 1 - \sum(\text{gas abundances})$, to be between 0 and 1. Second, for the cloud pressure terms, we reject any drawn value that does not satisfy $10^{p_t} + 10^{dp} < 10^{P_0}$ (i.e., that the cloud base cannot extend below the bottom of the atmosphere). Note that for the purposes of the retrieval, we consider pressures in log space.

We simulate noise in our observations following the expressions given in Robinson et al. (2016). For simplicity, we include only read noise and dark current, as (Robinson et al. 2016) showed that detector noise will be the dominant noise source in *WFIRST*-type spectral observations of exoplanets. Detector and instrument parameters for the HabEx and LUVOIR concepts are only loosely defined, and advances in detector technologies for these missions may move observations out of the detector-noise dominated regime. In the detector-noise dominated regime the signal-to-noise ratio is simply,

$$\mathrm{SNR} = \frac{c_p \times t_{\mathrm{int}}}{\sqrt{(c_d + c_r) \times t_{\mathrm{int}}}} \tag{4.13}$$

where t_{int} is the integration time, c_p is planet count rate, and c_d is the dark noise count rate, and c_r is the read noise count rate. More rigorously, it can be shown that, at constant spectral resolution, $\mathrm{SNR} \propto q\mathcal{T}A_g\Phi(\alpha)B_\lambda\lambda$, where q is the wavelength-dependent detector quantum efficiency, $\mathcal{T}$ is throughput, and B_λ is the host stellar specific intensity (taken here as a Planck function at the stellar effective temperature). We use a stellar temperature of 5780 K for the blackbody. When the SNR at one wavelength is specified, this scaling implies that the the calculation of the signal-to-noise ratio at other wavelengths is independent of the imaging raw contrast of the instrument. We can expect the noise at the redder end of our range to be large, as the detector quantum efficiency (taken to be appropriate for the *WFIRST*/CGI) rapidly decreases.

Since we treat only SNR rather than modeling exposure times, the exact mix of noise sources is not relevant (so that, e.g., dark current and readnoise are indistinguishable). The key relevant properties of the noise model is that it is uncorrelated between spectral channels and its magnitude only depends on wavelength via a dependence on point spread function area (Robinson et al. 2016, their Equation 26), which will be true for detector-limited cases but may not be true for large-aperture instruments limited by speckle noise.

For our study, we will consider multiple wavelength resolutions, R, and SNRs. Working in SNRs (instead of integration times) makes our investigations independent of telescope diameter, target distance, and other system-specific or observing parameters. Because the SNR is dependent on wavelength, we reference our values to be at V-band (550 nm) for all resolutions for HabEx/LUVOIR. Since the *WFIRST*/CGI spectrograph is currently planned to only extend to 600 nm at the blue end, we opt to reference our *WFIRST* SNRs to this wavelength. Unlike previous studies (Lupu et al. 2016; Nayak et al. 2017), our simulated *WFIRST* rendezvous data include two photometric points in the blue, which is consistent with current CGI designs. We set the SNR in the *WFIRST* filters to be equal to that at 600 nm.

Our simulation grid setup is shown in Table 4.2, where the spectral resolutions and SNRs assumed for different observing scenarios are indicated. Figure 4.5 demonstrates the *WFIRST* rendezvous scenario data along with $R = 70$ and $R = 140$ data points (for HabEx/LUVOIR) plotted over the forward model spectrum before noise is added. The scaling of SNR with wavelength for *WFIRST* rendezvous (normalized to

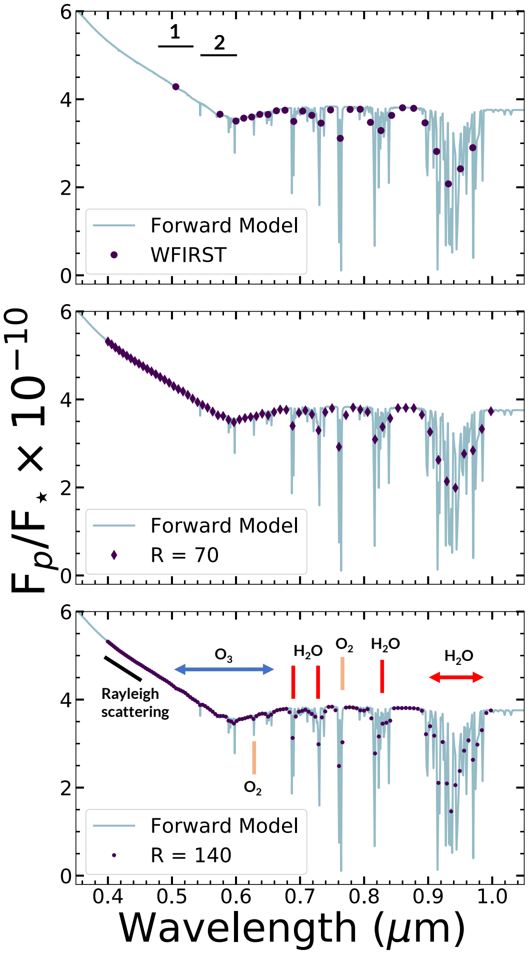


Figure 4.5: The high resolution (1000 wavelength points from $0.35 - 1.05$ μm) forward model spectrum, overplotted with simulated *WFIRST* rendezvous, $R = 70$, $R = 140$ data, from top panel to bottom. Key spectral features for atmospheric gases in our model are labeled. In the top panel, "1" and "2" mark the span of the *WFIRST* Design Cycle 7 filters (see Table 4.2).

unity at 600 nm) as well as our $R = 70$ and $R = 140$ cases (normalized to unity at 550 nm) is shown in Figure 4.6. The impact of the host stellar SED sets the overall shape of the SNR scaling, with additional influence from atmospheric absorption bands detector as well as quantum efficiency effects (that have strong impacts at red wavelengths). Thus, Figure 4.6 can be used to translate our stated SNR to the SNR at any other wavelength (e.g., a SNR= 10 simulation has a SNR in the continuum shortward of the 950 nm water vapor band of roughly $0.3 \times 10 = 3$).

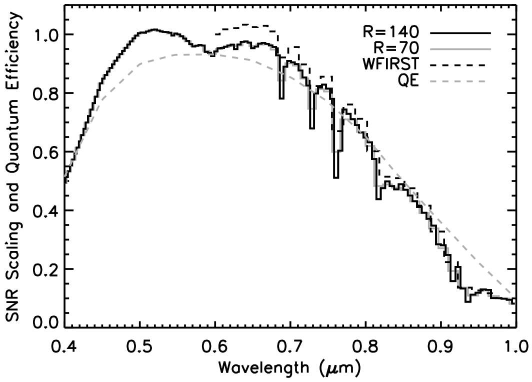


Figure 4.6: Scaling of SNR with wavelength for *WFIRST* rendezvous, $R = 70$, and $R = 140$ cases. The *WFIRST* curve is normalized to unity at 600 nm while the $R = 70$ and $R = 140$ curves are normalized to unity at 550 nm, following our definite of simulation SNR at these respective wavelengths. Also shown is the wavelength-dependent detector quantum efficiency (QE) that we adopt.

When generating simulated data with a noise model, there are several options for handling the placement/sampling of the mock observational data points. Previous studies (Lupu et al. 2016; Nayak et al. 2017) have generated a single, randomized dataset for a given SNR. The placement of a single spectral data point is determined by randomly sampling a Gaussian distribution whose width is determined by the wavelength-dependent SNR. While this treatment can accurately simulate a single observational instance, it also runs the risk (especially at lower spectral resolution and SNR) of biasing retrieval results, as the random placement of only a small handful of spectral data points can significantly impact the outcome. Given this, it is ideal to retrieve on a large number ($\gtrsim 10$) of simulated data sets at a given spectral resolution and SNR, where a comprehensive view of all the posteriors from the collection of instances will indicate expected telescope/instrument performance. Unfortunately, given the large number of R/SNR pairs in our study (10) and the long runtime of an individual retrieval (of order 1 week on a cluster), running ∼10 noise instances for each of our R/SNR pairs is computationally unfeasible (requiring ∼100 weeks of cluster time). Thus, we opt for an intermediate approach that maintains computational feasibility and avoids potential biases from individual noise instances. Here, we run only a single noise instance at a given R/SNR pair, but we do not randomize the placement of the individual spectral points. In other words, the individual simulated spectral points are placed on the "true" planet-to-star flux ratio point and are assigned error bars according to the SNR and noise model. While this approach prevents having a small handful of randomized data points from biasing retrieval results, it does lead to likely optimistic results, especially at modest SNR (i.e., SNR∼10), since data point randomization is, in effect, an additional "noise" source that we are omitting. This means that the posterior distributions will usually be centered on the true values in an unrealistic fashion. However, the width and shape of the posterior covariances will be representative of real observations, so the fidelity of retrievals can be assessed. We keep this optimism in mind when discussing results in later sections; in particular, we compare the performance of retrievals on multiple noise instances of a subset of the cases we consider to the non-randomized case in Section 4.5.2.

4.3 Retrieval Validation

Before using our framework on simulated data, we validate its accuracy and examine its performance. For this initial validation, we use non-randomized, wavelength-independent noise at a signal-to-noise ratio of 20 for a spectrum at a resolution of $R = 140$. Table 4.3 lists our four validation model variants, each increasing in complexity as we systematically explore how the addition of retrieved parameters influences the posterior distributions and correlations. In Model I, we fix all parameters except P_0 and A_s. In Model II, we add g and R_p; in Model III, we then add in gases as retrieved parameters (H_2O, O_3, O_2); and in Model IV, we add all cloud parameters. Incrementally increasing the number of free parameters (from 2 to 11) allows us to see the interconnections between them, and helps us understand how clouds can obscure our inferences.

In Figure 4.7, we present the posterior distributions for Model I. In the two-dimensional correlation histogram, a higher probability corresponds to a darker shade. With all else held constant, we see narrow posterior distributions and a slight correlation between P_0 and A_s. For lower values of surface pressure, which controls the turn off of the Rayleigh scattering slope, we need a brighter surface to maintain the measured brightness, especially in the red end of the spectrum, and vice versa. We mark the 16%, 50%, and 84% quantiles in the marginalized one-dimensional posterior distributions. The posterior distributions for Model II are shown in Figure 4.8, and are generally narrow (as only four parameters are being retrieved). There are two key correlations, one between g and P_0, and one between R_p and A_s. Both gravity and surface pressure influence the column mass, so that, when attempting to fit a spectrum, we can trade a larger gravity with a larger surface pressure and maintain a similar column mass (which controls, e.g., the Rayleigh scattering feature). Additionally, we can trade off a larger reflecting surface area (i.e., larger R_{p}) with a darker surface (lower A_{s}), which is a statement of the typical "radius/albedo degeneracy" problem. The posterior for surface albedo is now an upper limit instead of a constraint. As a result, the radius posterior distribution appears truncated at larger values given the tight correlation between these two parameters. The correlation seen originally in Model I, between P_0 and A_{s}, then acts as a chain between the other two, more prominent, correlations to induce correlations between parameters such as A_{s} and g or R_{p} and P_0.

Once we allow gases to be free parameters in Model III (Figure 4.9), the P_0 and A_{s} correlation becomes diminished as H_2O, due to its numerous bands across the spectral range, becomes a primary control of brightness. The significant impact of H_2O on the spectrum leads to a strong, positive correlation between H_2O abundance and planetary size, as additional water vapor absorption can be compensated by a larger planetary size to maintain fixed brightness. We now see gravity linked to the molecular abundances, which is expected as surface gravity directly influences the column abundance of a species. This key correlation also causes the individual gas abundances to be correlated with each other. The main correlations from Model II are still present. We note once more that we do not have constraint on the surface albedo, again leading to an asymmetric distribution for radius. Thus, from the strong correlation of H_2O with with R_p, and the fundamental correlation between R_{p} and A_{s}, we see correlations between planetary radius, surface albedo, and all gas abundances. Weak correlations between surface pressure and the gas abundances are due to column abundance effects.

Finally, as shown by Figure 4.10, we retrieve on the data with the full forward model, adding in the cloud parameters p_t, dp, τ, and f_c. This version of the model is what we apply when simulating direct-imaging data in the upcoming sections, and represents our most realistic (i.e., true to the actual Earth) scenario. The optical depth is shown to only have a lower limit constraint. Thus, the retrieval detects a cloud but cannot constrain the optical depth beyond showing that the cloud is optically thick. There is an expected correlation between τ and f_c; a higher cloudiness fraction can complement a less optically thick cloud, and vice versa. There is only an upper limit to dp, which is a result of the lack of vertical sensitivity given the constant-with-pressure abundance distributions. The posterior distribution for O_2 becomes a lower limit instead of a constraint as in Model III. Surface gravity is less precisely and less accurately constrained compared to the previous, less complex renditions of the model.

For optically thin clouds, we expect to better constrain surface albedo; however, we do not consider this scenario in our study. We examined instead the performance of a completely cloud-free model on data generated with our cloudy model. We find that while the model can fit the data and return accurate estimates of e.g., the mixing ratios, we get inaccurate estimates of the surface albedo and the surface pressure. These two parameters are biased, with lower surface pressure paired with higher surface albedo as the preferred configuration in the cloud-free case. As a result, we move forward with utilizing our cloudy forward model on our simulated data. However, we note that in realistic cases where we do not know the true state of a planet's atmosphere, we could obtain complementary information relating to the presence of clouds (e.g., variability) such that we may choose the most appropriate forward model.

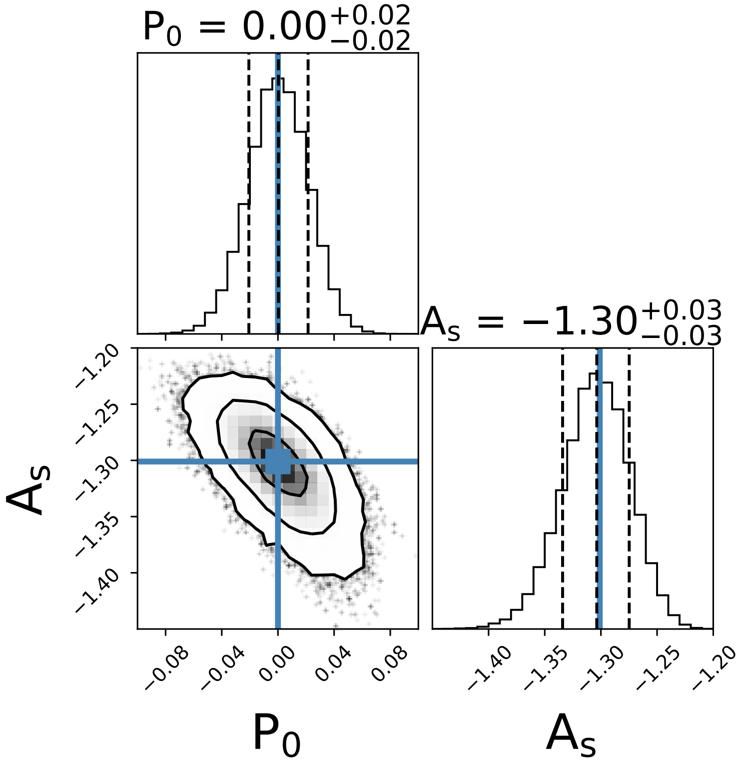


Figure 4.7: Posterior distributions of Model I from Table 4.3, where we fix all parameters but P_0 and $A_{\rm s}$. We retrieve on $R = 140$, SNR $= 20$ data with wavelength-independent noise. Overplotted in solid light-blue color are the fiducial parameter values. The 2D marginalized posterior distribution, used in interpreting correlations, is overplotted with the 1-, 2-, and 3-σ contours. Above the 1D marginalized posterior for each parameter, we list the median retrieved value with uncertainties that indicate the 68% confidence interval. Dashed lines (left to right) mark the 16%, 50%, and 84% quantiles.

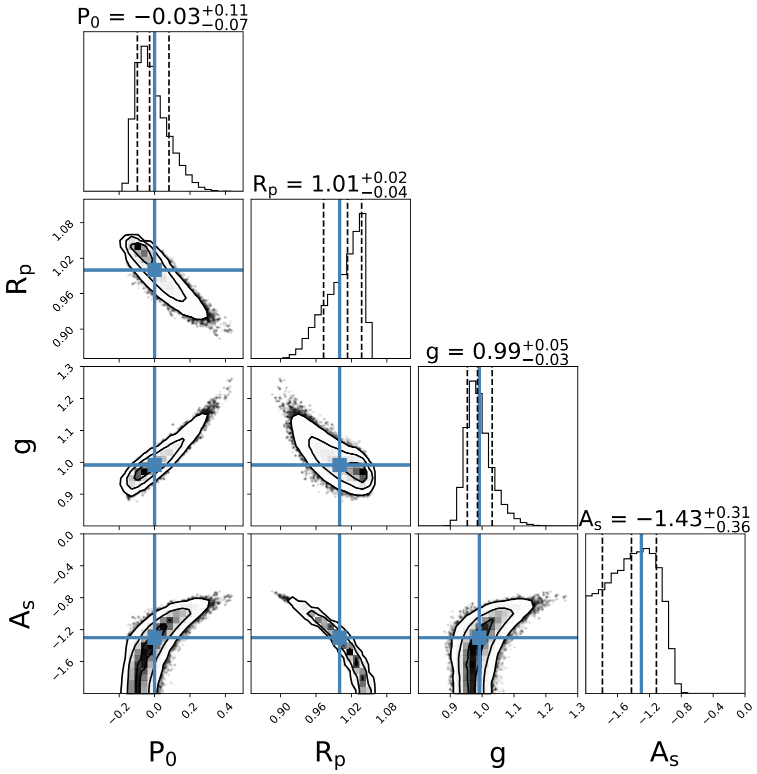


Figure 4.8: Posterior distributions of Model II from Table 4.3, where we fix all parameters except for P_0, A_s, g, and R_p. We retrieve on $R = 140$, SNR = 20 data with wavelength-independent noise. Overplotted in solid light-blue color are the fiducial parameter values. The 2D marginalized posterior distribution, used in interpreting correlations, is overplotted with the 1-, 2-, and 3-σ contours. Above the 1D marginalized posterior for each parameter, we list the median retrieved value with uncertainties that indicate the 68% confidence interval. Dashed lines (left to right) mark the 16%, 50%, and 84% quantiles.

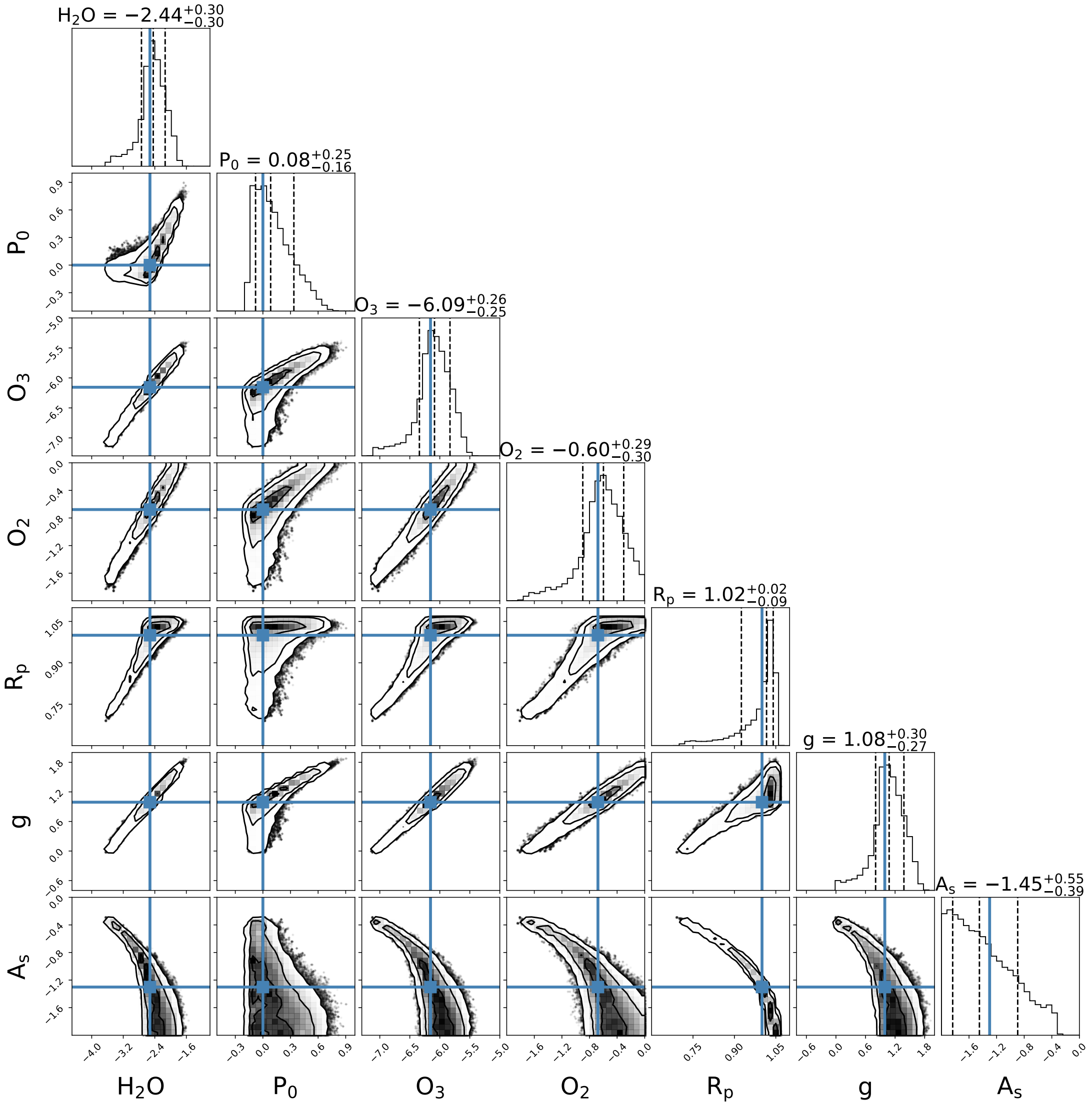


Figure 4.9: Posterior distributions of Model III from Table 4.3, where we retrieve P_0, A_s, g, R_p, $\mathrm{H_2O}$, $\mathrm{O_2}$, and $\mathrm{O_3}$. We retrieve on $R = 140$, SNR $= 20$ data with wavelength-independent noise. Overplotted in solid light-blue color are the fiducial parameter values. The 2D marginalized posterior distribution, used in interpreting correlations, is overplotted with the 1-, 2-, and 3-σ contours. Above the 1D marginalized posterior for each parameter, we list the median retrieved value with uncertainties that indicate the 68% confidence interval. Dashed lines (left to right) mark the 16%, 50%, and 84% quantiles.

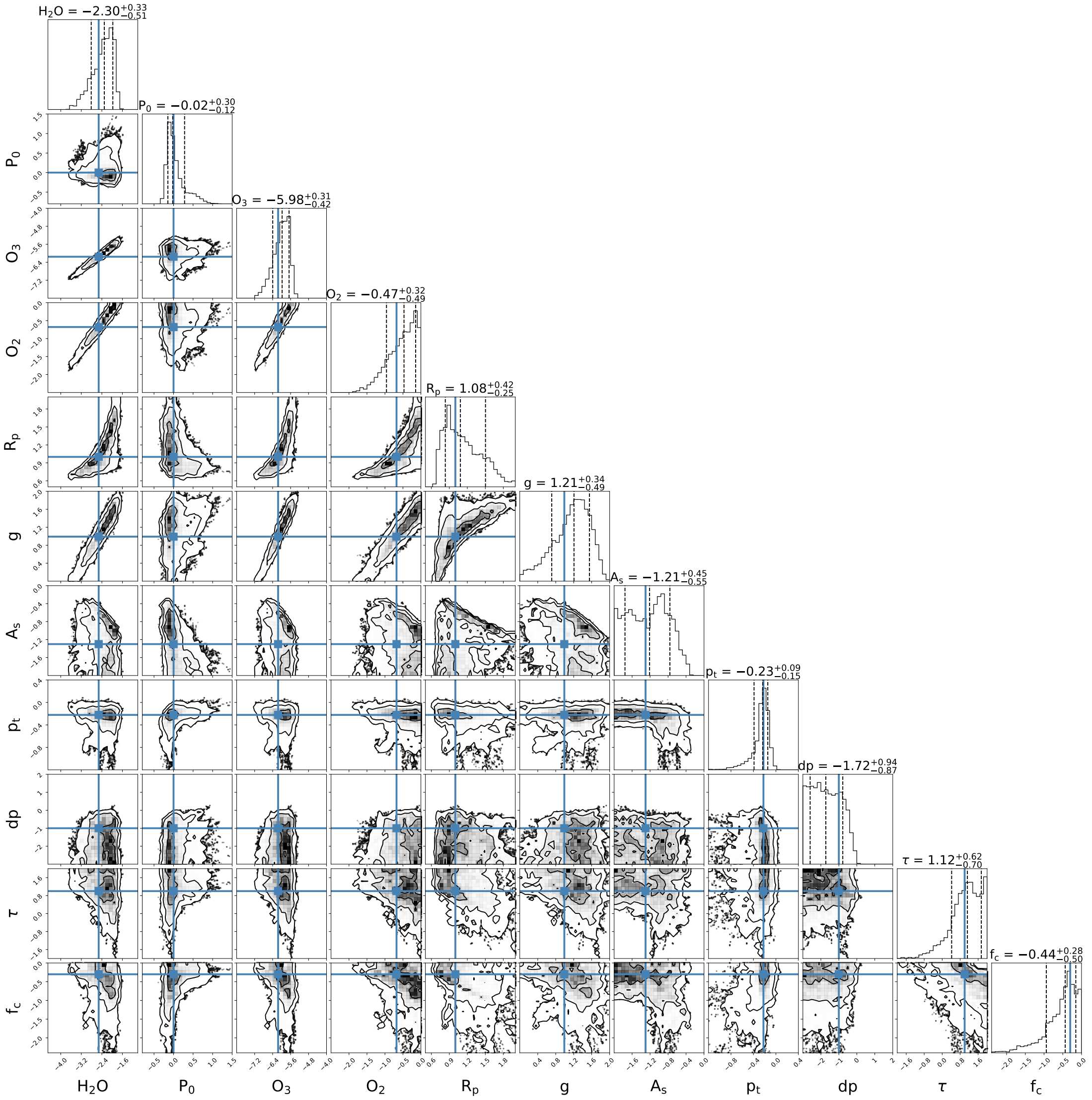


Figure 4.10: Posterior distributions of Model IV, or the complete model, from Table 4.3. We retrieve for 11 parameters: P_0, $A_{\rm s}$, g, $R_{\rm p}$, $\rm H_2O$, $\rm O_2$, $\rm O_3$, $p_{\rm t}$, dp, τ, and $f_{\rm c}$. We retrieve on $R = 140$, SNR = 20 data with wavelength-independent noise. Overplotted in solid light-blue color are the fiducial parameter values. The 2D marginalized posterior distribution, used in interpreting correlations, is overplotted with the 1-, 2-, and 3-σ contours. Above the 1D marginalized posterior for each parameter, we list the median retrieved value with uncertainties that indicate the 68% confidence interval. Dashed lines (left to right) mark the 16%, 50%, and 84% quantiles.

4.4 Results

We generate data sets for HabEx and LUVOIR-like missions ($0.4 - 1.0$ μm at $R = 70$, $R = 140$) at SNR$= 5,\ 10,\ 15,\ 20$, and for the *WFIRST* rendezvous scenario (two photometric points within $0.4 - 0.6$ μm plus a spectrum of $R = 50$ for $0.6 - 0.96$ μm) also at SNR$= 5,\ 10,\ 15,\ 20$. In all cases, we used the noise model to generate uncertainties expected for high-contrast imaging instead of the wavelength-independent noise for the validations in the previous section. As Section 4.2.4 described, the SNR refers to the value at 0.55 μm for $R = 70,\ 140$, and at 0.6 μm for *WFIRST*. We record the specific runs in Table 4.2. In place of showing the correlations for all parameters for all cases, we refer to Figure 4.10, which represents the ideal case correlations among the parameter posteriors. We only show the posterior probability distributions themselves to better highlight any trends with respect to SNR and/or R. We grouped the posteriors in terms of bulk atmospheric and planetary parameters (P_0, $R_{\rm p}$, g, $A_{\rm s}$), then cloud parameters ($p_{\rm t}$, dp, τ, $f_{\rm c}$), and finally gases (H_2O, O_3, O_2). For each case, `emcee` was run with 16 MCMC chains (walkers) per parameter for at least 12000 steps, the last 5000 of which are used to determine the posterior distributions. From those 5000 steps, we randomly selected 1000 sets of parameters to calculate their corresponding high resolution spectra. These spectra are plotted with the data to show the 1-σ, 2-σ, and median fits.

4.4.1 Results for $R = 70$, $R = 140$ simulated data

For both $R = 70$ and $R = 140$, we simulated data sets at SNR $= 5$, 10, 15, 20. Table 4.4 lists the median and 1-σ values of all retrieved parameters for each SNR at $R = 70$. Figure 4.11 shows the marginalized posterior distributions for the model parameters for all SNR cases for $R = 70$, plotted with the fiducial or "truth" values. Table 4.5 lists the median and 1-σ values of all retrieved parameters for each SNR at $R = 140$. Figure 4.12 shows the posterior distributions for $R = 140$ for the model parameters for all SNR cases compared against their input values. Figure 4.14 shows the corresponding spread in fits and the median fit to the data for each SNR for both resolutions.

4.4.2 Results for *WFIRST* rendezvous simulated data

For the *WFIRST* rendezvous scenario, we utilized the Design Cycle 7 instrument parameters to set the locations of the two photometric points and the range and resolution of the spectrometer ($R = 50$; see Table 4.2). Because of this particular set-up, we reference the SNRs in our grid (5, 10, 15, 20) at 600 nm, and assign the photometric points the same SNR as at 600 nm. Table 4.6 lists the median and 1-σ values of all retrieved parameters for each SNR variant. Figure 4.13 presents the posterior distributions for the four *WFIRST* rendezvous variants with respect to the input values. Figure 4.14 shows the spread in fits and median fit to the data for each variant.

Table 4.4: $R = 70$ retrieval results, with median value and 1-σ uncertainties of the parameters.

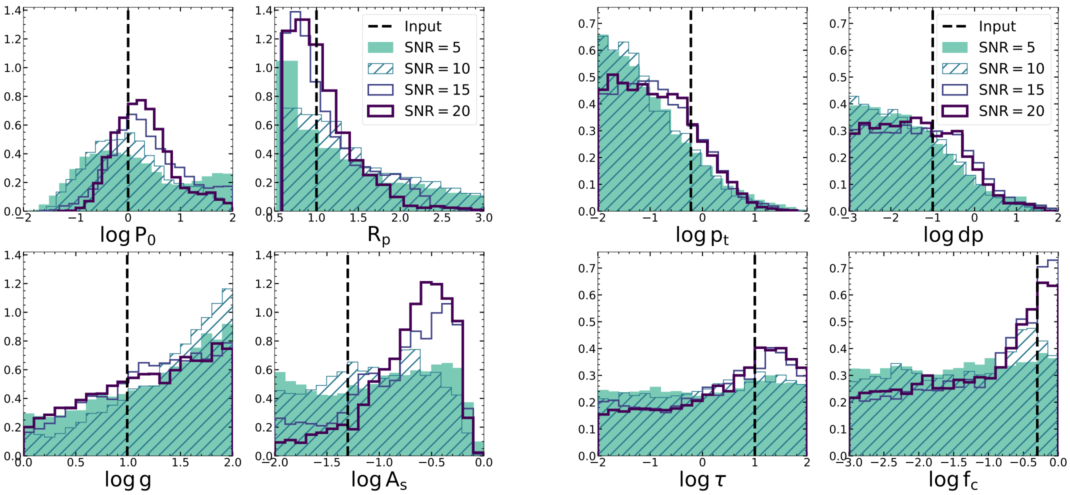
Parameter	Input	SNR= 5	SNR= 10	SNR= 15	SNR= 20
log H_2O	-2.52	$-5.07^{+2.34}_{-1.92}$	$-3.85^{+1.77}_{-2.60}$	$-3.12^{+0.97}_{-1.71}$	$-2.76^{+0.62}_{-0.88}$
log O_3	-6.15	$-7.55^{+1.49}_{-1.46}$	$-6.79^{+0.93}_{-1.81}$	$-6.37^{+0.55}_{-0.84}$	$-6.24^{+0.47}_{-0.60}$
log O_2	-0.68	$-5.12^{+3.25}_{-3.23}$	$-4.51^{+3.24}_{-3.61}$	$-1.86^{+1.29}_{-3.99}$	$-1.00^{+0.66}_{-1.01}$
log P_0	0.0	$0.02^{+1.35}_{-0.84}$	$-0.03^{+0.87}_{-0.70}$	$0.28^{+0.85}_{-0.56}$	$0.25^{+0.56}_{-0.49}$
R_p	1.0	$1.23^{+1.54}_{-0.58}$	$1.33^{+1.23}_{-0.52}$	$0.97^{+0.68}_{-0.27}$	$0.98^{+0.44}_{-0.25}$
log g	0.99	$1.33^{+0.48}_{-0.77}$	$1.48^{+0.38}_{-0.68}$	$1.28^{+0.51}_{-0.66}$	$1.24^{+0.55}_{-0.69}$
log A_s	-1.3	$-0.96^{+0.58}_{-0.74}$	$-1.05^{+0.55}_{-0.59}$	$-0.70^{+0.37}_{-0.62}$	$-0.63^{+0.29}_{-0.46}$
log p_t	-0.22	$-1.14^{+0.97}_{-0.61}$	$-1.19^{+0.93}_{-0.56}$	$-0.92^{+0.86}_{-0.71}$	$-0.94^{+0.84}_{-0.73}$
log dp	-1.0	$-1.67^{+1.24}_{-0.92}$	$-1.71^{+1.18}_{-0.91}$	$-1.35^{+1.17}_{-1.14}$	$-1.43^{+1.11}_{-1.06}$
log τ	1.0	$0.10^{+1.30}_{-1.43}$	$0.21^{+1.23}_{-1.48}$	$0.49^{+1.03}_{-1.66}$	$0.61^{+0.93}_{-1.66}$
log f_c	-0.3	$-1.43^{+0.99}_{-1.07}$	$-1.33^{+0.94}_{-1.12}$	$-0.93^{+0.71}_{-1.32}$	$-1.05^{+0.80}_{-1.27}$

Table 4.5: $R = 140$ retrieval results, with median value and 1-σ uncertainties of the parameters.

Parameter	Input	SNR= 5	SNR= 10	SNR= 15	SNR= 20
log H_2O	-2.52	$-4.56^{+2.14}_{-2.35}$	$-2.74^{+0.69}_{-1.07}$	$-2.61^{+0.47}_{-0.65}$	$-2.43^{+0.39}_{-0.56}$
log O_3	-6.15	$-7.36^{+1.26}_{-1.65}$	$-6.26^{+0.53}_{-0.68}$	$-6.18^{+0.42}_{-0.48}$	$-6.03^{+0.34}_{-0.48}$
log O_2	-0.68	$-4.45^{+3.08}_{-3.69}$	$-1.06^{+0.76}_{-1.43}$	$-0.76^{+0.51}_{-0.79}$	$-0.60^{+0.43}_{-0.59}$
log P_0	0.0	$0.07^{+1.01}_{-0.84}$	$0.20^{+0.72}_{-0.49}$	$0.12^{+0.49}_{-0.36}$	$0.07^{+0.39}_{-0.31}$
R_p	1.0	$1.25^{+1.16}_{-0.52}$	$1.01^{+0.60}_{-0.28}$	$0.99^{+0.42}_{-0.23}$	$1.05^{+0.42}_{-0.27}$
log g	0.99	$1.36^{+0.46}_{-0.74}$	$1.31^{+0.49}_{-0.77}$	$1.14^{+0.56}_{-0.65}$	$1.20^{+0.50}_{-0.64}$
log A_s	-1.3	$-0.98^{+0.54}_{-0.60}$	$-0.67^{+0.32}_{-0.50}$	$-0.68^{+0.29}_{-0.44}$	$-0.79^{+0.34}_{-0.69}$
log p_t	-0.22	$-1.23^{+1.03}_{-0.55}$	$-0.96^{+0.80}_{-0.71}$	$-0.79^{+0.70}_{-0.82}$	$-0.66^{+0.53}_{-0.85}$
log dp	-1.0	$-1.72^{+1.25}_{-0.91}$	$-1.43^{+1.13}_{-1.09}$	$-1.55^{+1.10}_{-1.00}$	$-1.49^{+1.00}_{-0.98}$
log τ	1.0	$0.18^{+1.31}_{-1.49}$	$0.50^{+1.09}_{-1.66}$	$0.61^{+0.98}_{-1.61}$	$0.79^{+0.87}_{-1.40}$
log f_c	-0.3	$-1.30^{+0.93}_{-1.13}$	$-1.31^{+0.94}_{-1.21}$	$-0.99^{+0.76}_{-1.27}$	$-0.76^{+0.54}_{-1.26}$

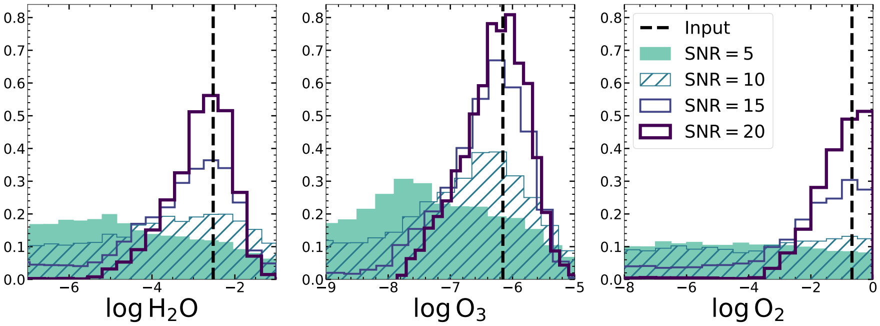
Table 4.6: *WFIRST* rendezvous retrieval results, with median value and 1-σ uncertainties of the parameters.

Parameter	Input	SNR= 5	SNR= 10	SNR= 15	SNR= 20
log H_2O	-2.52	$-4.94^{+2.35}_{-2.05}$	$-4.89^{+2.48}_{-2.11}$	$-4.03^{+1.87}_{-2.52}$	$-3.11^{+1.17}_{-1.71}$
log P_0	0.0	$-0.16^{+1.32}_{-0.80}$	$-0.19^{+1.03}_{-0.71}$	$0.03^{+1.16}_{-0.74}$	$0.45^{+1.01}_{-0.85}$
log O_3	-6.15	$-7.66^{+1.65}_{-1.59}$	$-7.53^{+1.54}_{-1.66}$	$-7.16^{+1.19}_{-1.66}$	$-6.80^{+0.94}_{-1.30}$
log O_2	-0.68	$-5.05^{+3.26}_{-3.39}$	$-4.89^{+3.43}_{-3.54}$	$-3.43^{+2.50}_{-4.41}$	$-2.26^{+1.71}_{-3.88}$
log P_0	0.0	$-0.16^{+1.32}_{-0.80}$	$-0.19^{+1.03}_{-0.71}$	$0.03^{+1.16}_{-0.74}$	$0.45^{+1.01}_{-0.85}$
R_p	1.0	$1.13^{+1.60}_{-0.50}$	$1.13^{+1.27}_{-0.48}$	$1.02^{+1.10}_{-0.38}$	$0.80^{+0.81}_{-0.19}$
log g	0.99	$1.42^{+0.42}_{-0.82}$	$1.45^{+0.41}_{-0.75}$	$1.41^{+0.43}_{-0.83}$	$1.26^{+0.52}_{-0.83}$
log A_s	-1.3	$-0.89^{+0.63}_{-0.74}$	$-0.84^{+0.56}_{-0.70}$	$-0.95^{+0.64}_{-0.68}$	$-0.76^{+0.53}_{-0.79}$
log p_t	-0.22	$-1.26^{+0.98}_{-0.52}$	$-1.24^{+0.83}_{-0.55}$	$-1.23^{+0.84}_{-0.57}$	$-0.84^{+0.89}_{-0.73}$
log dp	-1.0	$-1.75^{+1.16}_{-0.87}$	$-1.70^{+1.12}_{-0.87}$	$-1.46^{+1.14}_{-1.06}$	$-1.49^{+1.49}_{-1.03}$
log τ	1.0	$0.03^{+1.33}_{-1.39}$	$0.05^{+1.42}_{-1.40}$	$0.61^{+0.94}_{-1.55}$	$0.99^{+0.73}_{-1.44}$
log f_c	-0.3	$-1.41^{+0.96}_{-1.07}$	$-1.42^{+1.00}_{-1.07}$	$-0.82^{+0.60}_{-1.28}$	$-0.58^{+0.41}_{-0.97}$



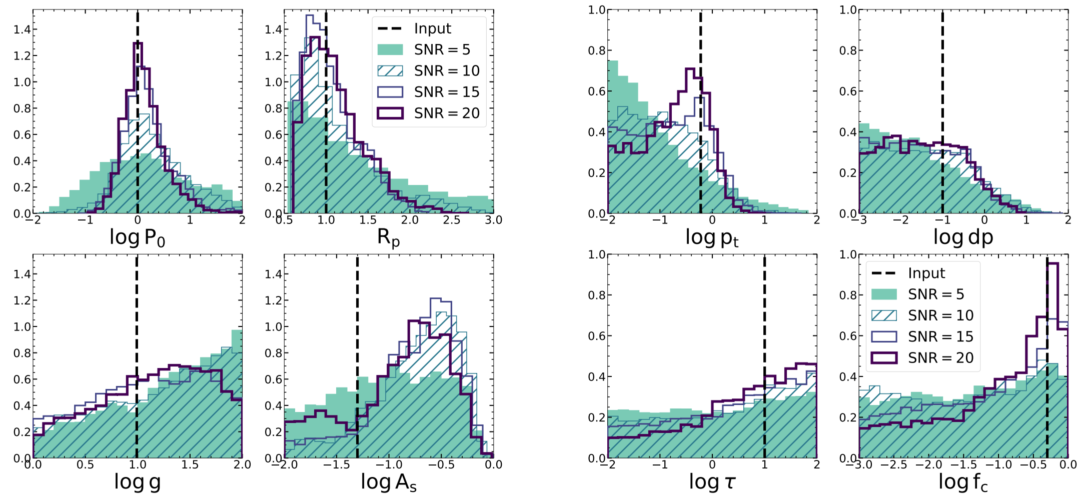
(a) $R = 70$ bulk parameters

(b) $R = 70$ cloud parameters



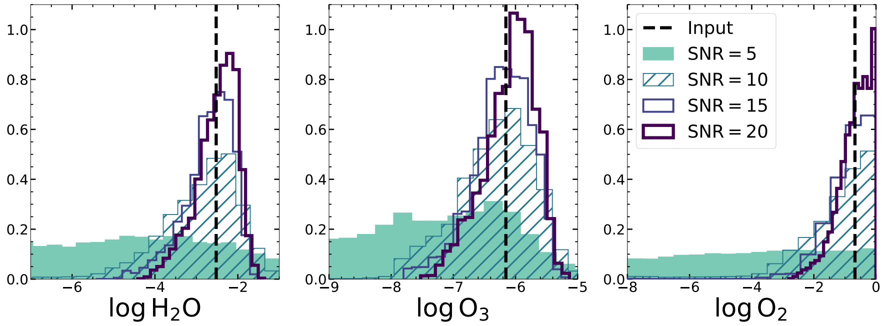
(c) $R = 70$ gas mixing ratios

Figure 4.11: Comparing 1D marginalized posterior distributions for all parameters for all SNR cases of $R = 70$. See Table 4.4 for corresponding median retrieved value with uncertainties that indicate the 68% confidence interval. Overplotted dashed line represents the fiducial values from Table 4.1.



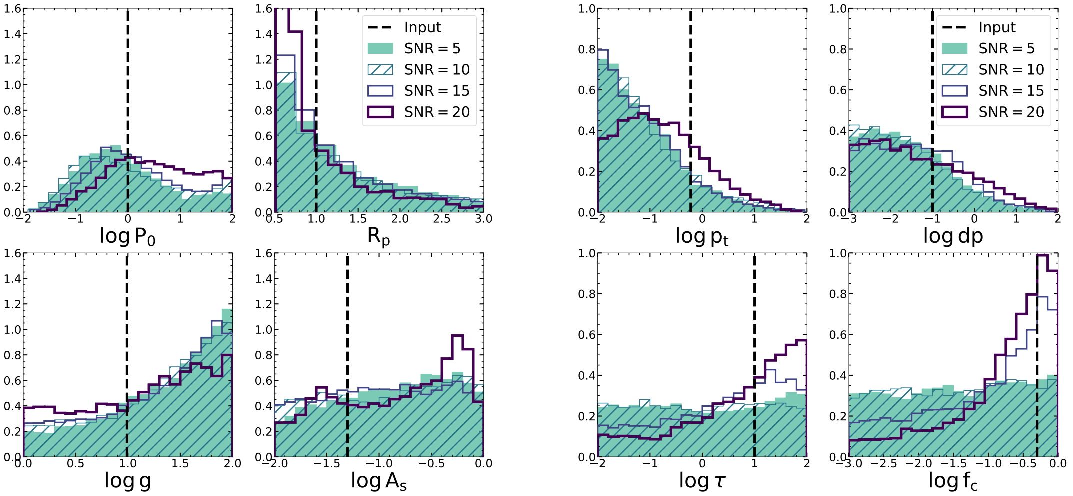
(a) $R = 140$ bulk parameters

(b) $R = 140$ cloud parameters

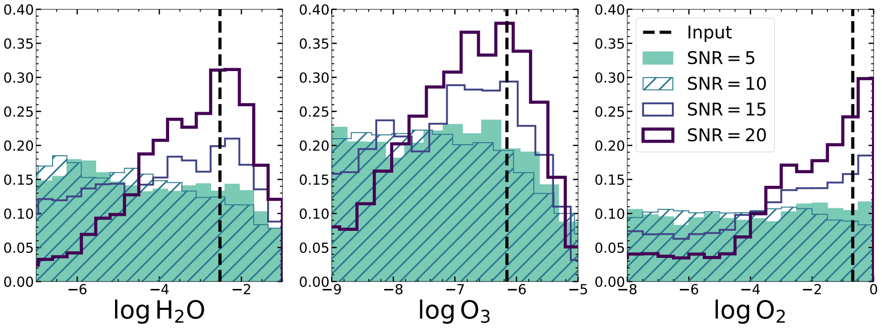


(c) $R = 140$ gas mixing ratios

Figure 4.12: Comparing 1D marginalized posterior distributions for all parameters for all SNR cases of $R = 140$. See Table 4.5 for corresponding median retrieved value with uncertainties that indicate the 68% confidence interval. Overplotted dashed line represents the fiducial values from Table 4.1.



(a) *WFIRST* bulk parameters (b) *WFIRST* cloud parameters



(c) *WFIRST* gas mixing ratios

Figure 4.13: Comparing 1D marginalized posterior distributions for all parameters for all SNR cases of a *WFIRST* rendezvous scenario. See Table 4.6 for corresponding median retrieved value with uncertainties that indicate the 68% confidence interval. Overplotted dashed line represents the fiducial values from Table 4.1.

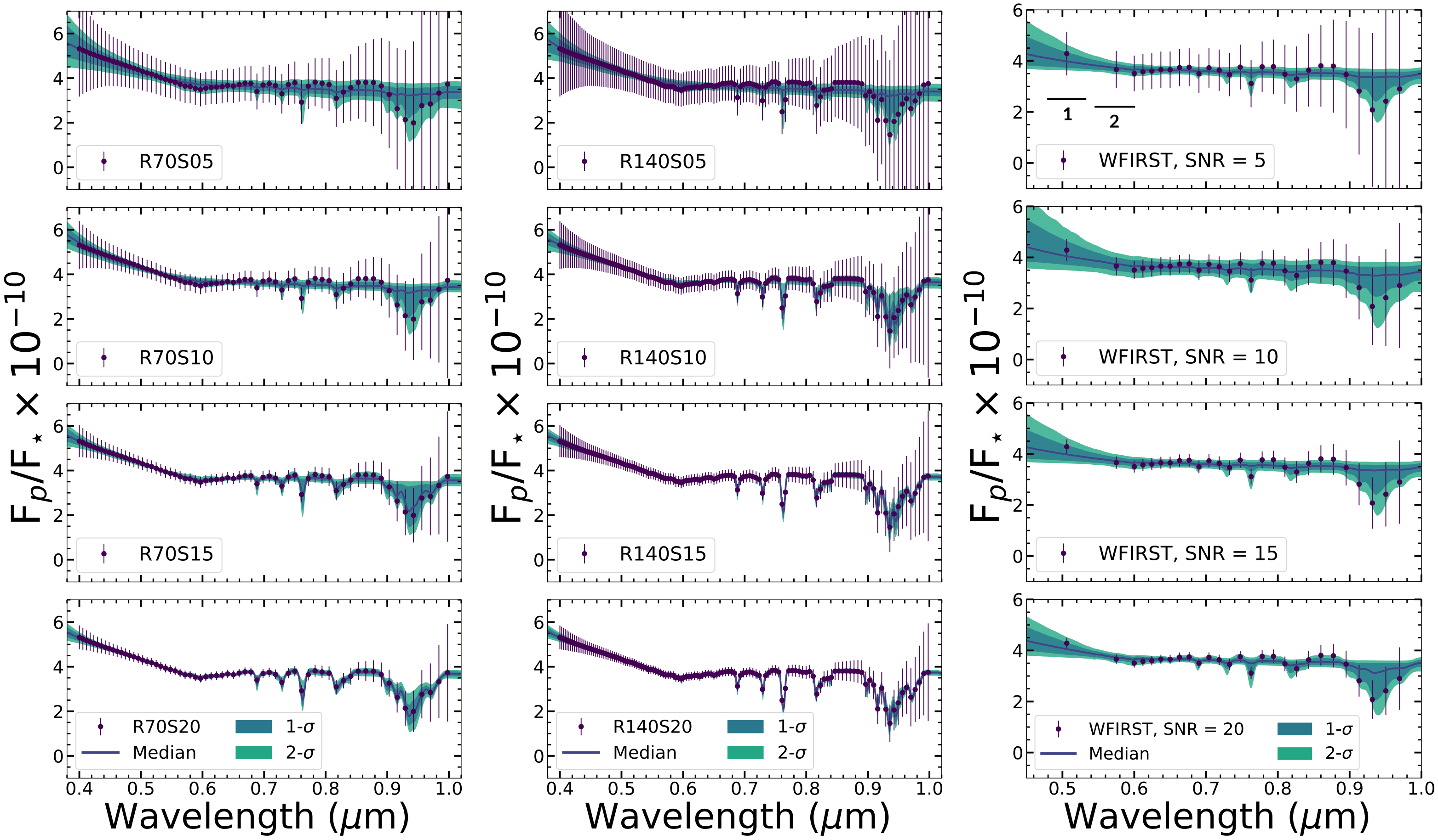


Figure 4.14: Spectra generated with 1000 randomly drawn sets of parameters sampled with the retrievals plotted with left: $R = 70$ data for SNR= 5, 10, 15, 20; middle: $R = 140$ data for SNR= 5, 10, 15, 20; and right: *WFIRST* rendezvous data at SNR = 5, 10, 15, 20. “1” and “2” mark the span of the *WFIRST* Design Cycle 7 filters (see Table 4.2). Lighter contour (light green) represents 2-σ fits while darker contour (blue-green) represents 1-σ fits. Solid line represents the median fit.

4.5 Discussion

The results from our retrieval analyses enable us to identify the SNR required, at a given spectral resolution, to constrain key planetary and atmospheric quantities. These findings have important implications for the development of future space-based direct imaging missions. We discuss these ideas below, and also touch on impacts of certain model assumptions and ideas for future research directions.

In what follows, we define a "weak detection" for a given parameter as having a posterior distribution that has a marked peak but which also has a substantial tail towards extreme values (indicating that, e.g., for a gas we could not definitively state that the gas is present in the atmosphere). A "detection" implies a peaked posterior without tails towards extreme values but whose 1-σ width is larger than an order of magnitude. We use the term "constraint" to indicate a detection whose posterior distribution has 1-σ width smaller than an order of magnitude. A non-detection would be indicated by a flat posterior distribution across the entire (or near-entire) prior range. For planetary radius, which is not retrieved in logarithmic space, we distinguish between a "detection" and a "constraint" when the 1-σ uncertainties are small enough to firmly place the planet in the Earth/super-Earth regime (i.e., with a radius below $1.5R_{\oplus}$, Rogers 2015; Chen & Kipping 2017). A visual summary of weak detections, detections, and constraints as a function of SNR for our different observing scenarios and for a selection of key parameters are given in Tables 4.7, 4.8, and 4.9.

4.5.1 Influence of SNR on Inferred Properties

For $R = 70$ at SNR $= 5$, Figure 4.11 shows there is only a weak detection of P_0 and a detection of R_p, which merely suggests the planet has an atmosphere and is not a giant planet. As SNR increases to 10, the O_3 posterior distribution has a weak peak near the fiducial value, and the gas is only weakly detected. Once the SNR is equal to 15, we weakly detect H_2O, O_3, and O_2. At a SNR of 20, it is possible to detect each of H_2O, O_3, and O_2. At this SNR, the oxygen mixing ratio is estimated to be above roughly 10^{-3}, indicating that we are unable to determine if O_2 is a major atmospheric constituent (i.e., present at the 1% level or more). Gravity (and, thus, planetary mass) remain undetected at all SNRs, similar to the findings of Lupu et al. (2016). The surface albedo is unconstrained (or worse) at all SNRs, but shows a weak bias toward a higher value of $A_s \approx 0.3$ ($\log A_\mathrm{s} \approx -0.5$) at the highest SNRs, which is likely due to the relatively large error bars at red wavelengths (driven primarily by low detector quantum efficiency) where we have the most sensitivity to the surface. We are able to get weak detections of τ and f_c, which are shown in Figure 4.10 to be correlated. Yet, with these posteriors, we cannot rule out scenarios without cloud cover. We note the drop-off in the posteriors of p_t and dp at higher pressure values likely result from the limiting conditions that the cloud base cannot extend below the surface pressure and the upper limit of the P_0 prior. The improved signal-to-noise ratio leads to a posterior more concentrated around the true value for p_t, dp, and R_p. For improved constraints on cloud properties, it may be beneficial to observe time variability with photometry (e.g., Ford et al. 2001) or use polarimetry (e.g., Rossi & Stam 2017).

At a higher spectral resolution ($R = 140$), the improvement in detections and constraints begin at a lower SNR, as illustrated by Figure 4.12. Gravity remains undetected for all SNRs. At a SNR equal to 5, P_0 and R_p have a weak detection and a detection, respectively. At SNR = 10, it is possible to detect H_2O, O_3, and O_2. As with the $R = 70$ case, surface albedo is unconstrained (or worse) at all SNRs, and, at the highest SNRs, the model is biased towards $A_\mathrm{s} \approx 0.3$ (as with $R = 70$). Moving to SNR = 15 adds a constraint to R_p, P_0, and O_3, as well as weak detections of cloud parameters. Increasing the SNR to 20 does not dramatically change the posterior distributions, although the posteriors for H_2O and O_2 become narrow enough to offer constraints. Here, the constraint on O_2 suggests it is a major constituent in the atmosphere. In spite of the generous SNR, though, the 1-σ uncertainties on the gas mixing ratios are not more precise than roughly an order of magnitude (see Table 4.5).

Considering both $R = 140$ and $R = 70$, we see that SNR = 5 data offer very little information about the planetary atmosphere. In the case of $R = 140$, SNR = 10 data offer detections but no constraints, and SNR = 20 data are required to constrain all included gas species. In other words, the conclusions we would draw about the planet (e.g., the amount of gases, the bulk and cloud properties) improve significantly between SNR = 10 and SNR = 20. With $R = 70$, the boost from SNR = 10 to SNR = 15 provides weak detections of key atmospheric and surface parameters, and SNR = 20 data offer detections but few constraints (i.e., except on planetary radius).

For the *WFIRST* rendezvous data sets, we are able to infer very little information at a SNR of 5 or 10 except for weak detection of surface pressure and a detection of the planetary radius. All gases remain undetected at these SNRs. The posterior distributions for most parameters do not vary much as SNR improves, although there are weak detections of cloud optical depth and fractional coverage at the highest SNRs. Like all previous cases, we do not detect the surface gravity. At SNR = 15, 20, the detection of f_c is unable to rule out scenarios with little cloud cover. To obtain weak detections of the atmospheric gases we require a SNR of 20, but, even here, the posteriors have tails that extend to near-zero mixing ratios.

To compare the performance of a *WFIRST* rendezvous scenario against HabEx or LUVOIR scenarios at $R = 70$ and $R = 140$, we plot together the posterior distributions of the parameters for the SNR = 10 results from *WFIRST* rendezvous, $R = 70$, and $R = 140$ in Figure 4.15. While this comparison sheds light on the corresponding trade-off in terms of parameter estimation for the same SNR, these cases do not represent equal integration times, which scales with resolution and SNR. If the dominant noise source does not depend on resolution (e.g., detector noise), the cases of $R = 140$ at SNR = 10, $R = 70$ at SNR = 20, and $R = 50$ at SNR = 28 would be roughly equal in integration times. However, if the dominant noise source does depend on resolution (e.g., exozodiacal dust), the cases of $R = 140$ at SNR = 10, $R = 70$ at SNR = 14, and $R = 50$ at SNR = 17 would roughly have equivalent integration times. Tables 4.7 through 4.9 allow approximate comparisons of these different scenarios, excluding a *WFIRST* rendezvous scenario at high SNR = 28 that we have not considered.

From Figure 4.15, we see that the performance of the *WFIRST* rendezvous retrieval is similar to that of $R = 70$ at SNR = 10. The noticeable difference is a weak

Table 4.7: $R = 70$: Strength of detection for a set of key parameters as a function of SNR.

Parameter	SNR= 5	SNR= 10	SNR= 15	SNR= 20
H_2O	–	–	W	D
O_3	–	W	W	D
O_2	–	–	W	D
P_0	W	W	W	D
R_p	D	D	D	C

Note. — Weak detection ("W") corresponds to a posterior distribution with a marked peak but also a substantial tail towards extreme values. Detection ("D") refers to a peaked posterior without tails towards extreme values but a 1-σ width larger than an order of magnitude. Constraint ("C") is defined as a peaked posterior distribution with a 1-σ width less than an order of magnitude. Non-detection, or flat posteriors across the entire (or near-entire) prior range, are marked with "–".

Table 4.8: $R = 140$: Strength of detection for a set of key parameters as a function of SNR.

Parameter	SNR= 5	SNR= 10	SNR= 15	SNR= 20
H_2O	–	D	D	C
O_3	–	D	C	C
O_2	–	D	D	C
P_0	W	D	C	C
R_p	D	D	C	C

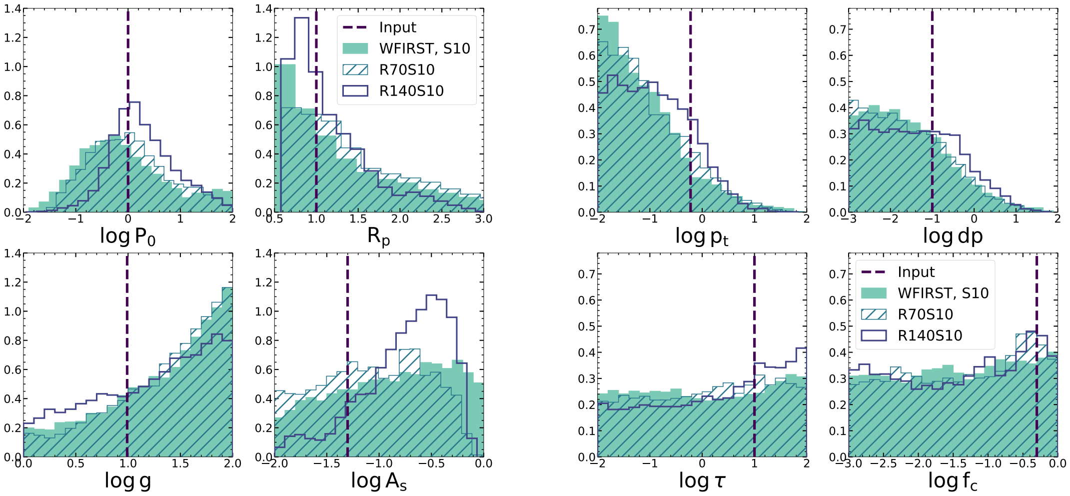
Note. — Weak detection ("W") corresponds to a posterior distribution with a marked peak but also a substantial tail towards extreme values. Detection ("D") refers to a peaked posterior without tails towards extreme values but a 1-σ width larger than an order of magnitude. Constraint ("C") is defined as a peaked posterior distribution with a 1-σ width less than an order of magnitude. Non-detection, or flat posteriors across the entire (or near-entire) prior range, are marked with "–".

Table 4.9: *WFIRST*: Strength of detection for a set of key parameters as a function of SNR.

Parameter	SNR= 5	SNR= 10	SNR= 15	SNR= 20
H_2O	–	–	–	W
O_3	–	–	–	W
O_2	–	–	W	W
P_0	W	W	W	W
R_p	D	D	D	D

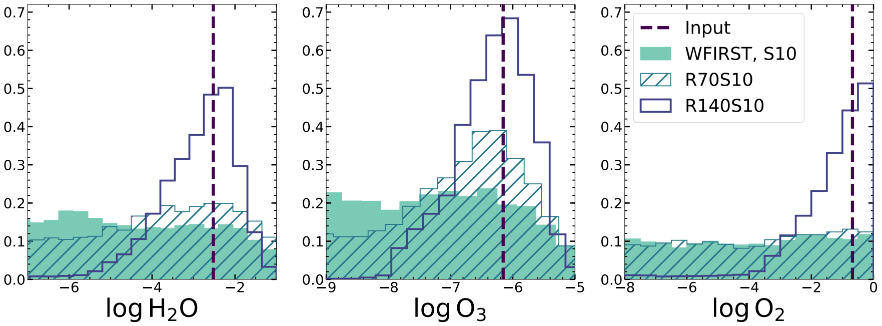
Note. — Weak detection (“W”) corresponds to a posterior distribution with a marked peak but also a substantial tail towards extreme values. Detection (“D”) refers to a peaked posterior without tails towards extreme values but a 1-σ width larger than an order of magnitude. Constraint (“C”) is defined as a peaked posterior distribution with a 1-σ width less than an order of magnitude. Non-detection, or flat posteriors across the entire (or near-entire) prior range, are marked with “–”.

detection of O_3 with $R = 70$. Because we adopt the photometric setup from *WFIRST* Design Cycle 7 through the shorter wavelengths, the data do not provide complete spectroscopic coverage across the significant O_3 feature from $0.5 - 0.7$ μm, as in the case of HabEx/LUVOIR simulated data. Figure 4.5 shows the sampling of the forward model spectrum for the three types of data sets we considered. We compare the spectral fits in Figure 4.14 and note the much wider spread in the possible fits for wavelengths shorter than 0.6 μm for *WFIRST* rendezvous versus $R = 70$ or $R = 140$, which have continuous coverage in the full range. The $R = 140$, SNR $= 10$ data set was able to offer *detections* of all atmospheric gases, setting it apart from the other two. We stress, however, that *constraints* were only found at SNR $= 20$ and $R = 140$.



(a) Mixed runs SNR= 10 bulk parameters

(b) Mixed runs SNR= 10 cloud parameters



(c) Mixed runs SNR= 10 gas mixing ratios

Figure 4.15: Comparing the posteriors for all parameters for SNR = 10 cases of *WFIRST* rendezvous, $R = 70$, and $R = 140$. Overplotted dashed line represents the fiducial values from Table 4.1.

4.5.2 Considering Multiple Noise Instances

Our parameter estimations are likely to be optimistic as a consequence of our adoption of non-randomized spectral data points in our faux observations. Thus, the requisite SNRs for detection detailed above should be seen as lower limits. Ultimately, our decision to use non-randomized data points stemmed from computational limitations (preventing us from running large numbers of randomized faux observations for each of our R/SNR pairs) and from a desire to avoid the biases that can occur from attempting to make inferences from retrievals performed on a single, randomized faux observation (Lupu et al. 2016).

However, we deemed it necessary to investigate the consistency of our findings with respect to different noise instances. To work within our computational restrictions, we realized that cases such as $R = 70$ with SNR $= 5$ yielded little detection information for any parameter even in the ideal scenario of non-randomized data. We then decided to focus on two "threshold" cases based on the results from the non-randomized data: $R = 140$ with SNR $= 10$ and $R = 70$ with SNR $= 15$. We ran 10 noise instances of these two cases where it is likely the optimistic non-randomized data makes the difference between detection and constraint for several parameters (see Tables 4.7 and 4.8).

Each noise instance is run for at least 10000 steps in `emcee`. Figure 4.16 shows all the individual posteriors for the gas mixing ratios from each noise instance for $R = 70$, SNR $= 15$. We highlight the posteriors from one "outlier" case where there is no oxygen detection. The corresponding set of data points are shown as well. This highlights the fact that single noise instances can mislead our interpretation and the benefit of having many noise instances run to obtain a more comprehensive understanding of the state of an atmosphere.

To summarize the noise instance results, we concatenate samples from the last 1000 steps in each noise instance and construct an averaged set of posteriors. We are able to do this because the noise instances are equally likely, having been drawn in the same manner from a Gaussian with set parameters (i.e., the same SNR as the standard deviation). In Figure 4.17, we plot up the combined posteriors of the 10 noise instances of $R = 70$, SNR = 15 and compared them to the posterior from the last 5000 steps of the non-randomized data case. We illustrate the same comparison for $R = 140$, SNR = 10 in Figure 4.18. We overplot the truth values as well as the 68% confidence interval and median value for each parameter from the combined noise-instances posterior and the non-randomized data posterior.

For all parameters in both the $R = 70$ and $R = 140$ cases, we find that the average posterior from the 10 noise instances agree with the posterior from the non-randomized data set qualitatively. Their medians and 68% confidence interval ranges are also similar with significant overlap. The overall conclusions we can draw from the average posteriors do not appear to differ much from those using the non-randomized data set posteriors.

4.5.3 Implications for Future Direct Imaging Missions

Future space-based direct imaging missions will have a diversity of goals for exoplanet studies, and will likely emphasize the detection and characterization of Earth-like exoplanets. For the detection of oxygen and ozone—which are key biosignature gases—in the atmospheres of Earth twins, our results indicate that spectra at a minimum characteristic SNR of 10 will suffice if at $R = 140$, while data at SNR of at least 15–20 would be needed at $R = 70$. For a *WFIRST* rendezvous-like observing setup, these gases would only be weakly detected even at a SNR of 20. Methane, which is another important biosignature gas, has no strong signatures in the visible wavelength range for the modern Earth, so we did not consider detection of this gas. Thus, we could not use our simulated data and retrievals to argue for detections of atmospheric chemical disequilibrium (Sagan et al. 1993; Krissansen-Totton et al. 2016).

Key habitability indicators include atmospheric water vapor and surface pressure. Detecting the former requires a SNR of 15–20 at $R = 70$, but only a SNR of 10 at $R = 140$. Surface pressure can be constrained to within an order of magnitude for SNR $\gtrsim 15$ at $R = 140$, although the overall lack of temperature information in these reflected-light spectra would make it impossible to use pressure/temperature data to argue for habitability (Robinson 2017). Surface temperature information may then need to come from climate modeling investigations that are constrained by retrieved gas mixing ratios.

For all of our observing setups, the data yield detections of, and in some cases constraints on, the planetary radius. Except at SNR of 20 for $R = 70$ or SNR > 15 for $R = 140$, the posterior distributions are not well-enough constrained to distinguish a Earth/super-Earth ($R_p < 1.5\,R_\oplus$) from a mini-Neptune based on size alone, although the data do rule out planetary sizes larger than Neptune. Additional atmospheric information (e.g., composition) could potentially be used to help distinguish between terrestrial planets and mini-Neptunes. These findings are consistent with the gas giant-focused work of Nayak et al. (2017), who note that observations at multiple phase angles can also help to better constrain planetary size. Our overall lack of surface gravity constraints, paired with the weak constraints on planet size, implies that we do not have a constraint on the planetary mass. Follow-up (or precursor) radial velocity observations (or, potentially, astrometric observations) could offer additional constraints on planet mass.

We can make rough comparisons of our R/SNR results to those of Brandt & Spiegel (2014), who used minimally parametric models to investigate detections of O_2 and H_2O for Earth twins. These comparisons are not direct, however, as Brandt & Spiegel (2014) were fitting for fewer parameters (8 versus our 11) and also only assumed that SNR was proportional to planetary reflectance (versus our more complicated scaling, as shown in Figure 4.6). For O_2, Brandt & Spiegel (2014) find $R = 150$ and SNR $= 6$ for a 90% detection probability, which is consistent with our $R = 140$ posteriors moving from a non-detection at SNR $= 5$ to a detection at SNR $= 10$. When investigating H_2O, Brandt & Spiegel (2014) find $R = 40$ and SNR $= 7.5$ or $R = 150$ and SNR $= 3.3$ for a 90% detection probability. Using Figure 4.6 to scale our SNRs to 890 nm (i.e., to the continuum just shortward of 950 nm water vapor band), at $R = 50$ we only find a weak detection of H_2O for $\mathrm{SNR}_{890\ \mathrm{nm}} = 10$, and at $R = 140$ we transition from a water vapor non-detection to detection between a $\mathrm{SNR}_{890\ \mathrm{nm}}$ of 2.5–5. Taken altogether, these comparisons indicate that we agree with Brandt & Spiegel (2014) at higher spectral resolution ($R = 140$–150), but that detection of H_2O at lower spectral resolution ($R = 50$) will likely require higher SNRs than originally indicated.

The discussion above emphasizes mere detections, not constraints (which, again, we define as having peaked posterior distributions with 1-σ widths less than an order of magnitude). While uncertain, we anticipate that characterization of climate, habitability, and life likely require constraints, not simple detections. Here, as is shown in Table 4.8, only $R = 140$ and SNR = 20 observations offer the appropriate constraints. Thus, future space-based high contrast imaging missions with goals of characterizing Earth-like planetary environments are likely to need to achieve $R = 140$ and SNR = 20 observations (or better). Of course, combining near-infrared capabilities, which would provide access to additional gas absorption bands, may help loosen these requirements.

4.5.4 Impacts of Model Assumptions

Several key assumptions adopted in this study warrant further comment. First, as noted earlier, we do not retrieve on planetary phase angle and planet-star distance, both of which influence the planet-to-star flux ratio. Thus, in effect, we are assuming that the planetary system has been revisited multiple times for photometric and astrometric measurements, such that the planetary orbit is reasonably well-constrained (i.e., that the orbital distance and phase angle are not the dominant sources of uncertainty when interpreting the observed planet-to-star flux ratio spectrum). If the orbit is not well-constrained, Nayak et al. (2017) showed that strong correlations can exist between the retrieved phase angle and the planet radius.

Second, we have assumed detector-dominated noise and a quantum efficiency appropriate for the *WFIRST*/CGI for all of our observational setups. While this is likely a fair assumption for our *WFIRST* rendezvous studies, it is likely that detector development will lead to major improvements in instrumentation for a HabEx/LUVOIR-like mission. Here, the rapid decrease into the red due to detector quantum efficiency may not be as dramatic, implying that spectra would have relatively more information content at red wavelengths as compared to the present study. Furthermore, a HabEx/LUVOIR-like mission may no longer be in the detector-dominated noise regime. In the limit of noise dominated by astrophysical sources (e.g., exo-zodiacal light or stellar leakage), the SNR only varies as $\sqrt{q\mathcal{T}B_\lambda}$.

Finally, we adopt a relatively simple parameterization of cloud three-dimensional structure. Specifically, we allow for only a single cloud deck in the atmosphere, and we then permit these clouds to have some fractional coverage over the entire planet. This parameterization of fractional cloudiness implies uniform latitudinal and longitudinal distribution of patchy clouds. In reality, clouds on Earth have a complex distribution in altitude, latitude, and longitude (Stubenrauch et al. 2013), and variations in time also have an observational impact (Cowan et al. 2009; Cowan & Fujii 2017). However, given the overall inability of our retrievals to constrain cloud parameters (at least at the SNRs investigated here; see also Lupu et al. 2016; Nayak et al. 2017), it seems challenging for future space-based exoplanet characterization missions to detect (or constrain) more complex cloud distributions with the types of observations studied here and data of similar quality.

4.5.5 Future Work

Our current forward model is able to include both CO_2 and CH_4, although we did not retrieve on these gases in the current study due to their overall lack of strong features in the visible wavelength range for modern Earth. However, these species do have stronger features in the near-infrared wavelength range. As both of the HabEx and LUVOIR concepts are considering near-infrared capabilities, it will be essential to extend our current studies to longer wavelengths and to investigate whether or not constraints on additional gases (i.e., beyond water, oxygen, and ozone) can be achieved at these wavelengths.

Additionally, given the likely huge diversity of exoplanets that will be discovered by future missions (and that have already been identified and studied by *Kepler*, *Hubble*, and *Spitzer*), it will be necessary to extend our parameter estimation studies to include a wider range of worlds. Both super-Earths and mini-Neptunes are more-favorable targets for a *WFIRST* rendezvous mission, and may also be easier targets for HabEx/LUVOIR-like missions. Our forward model is already capable of simulating these types of worlds, and we anticipate emphasizing a variety of exoplanet types in future studies. Such future studies may also include retrievals on planetary phase angle, which would be relevant to observing scenarios where the planetary orbit is poorly constrained.

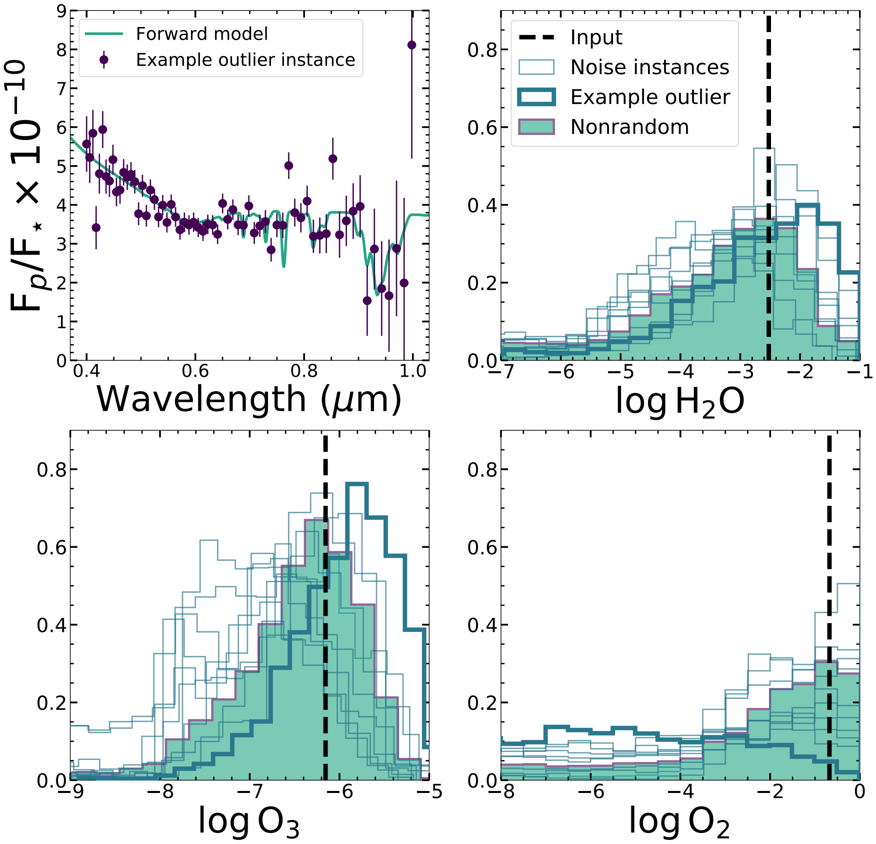
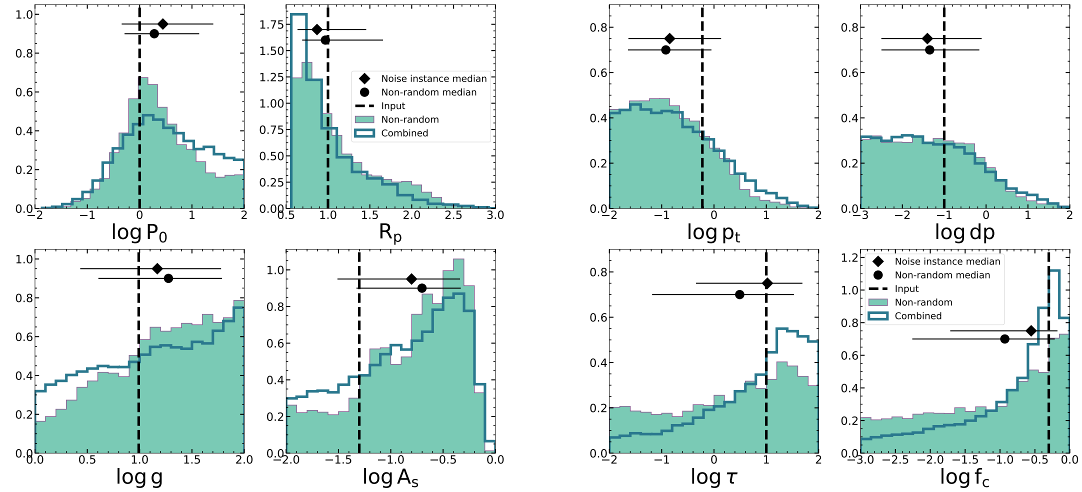
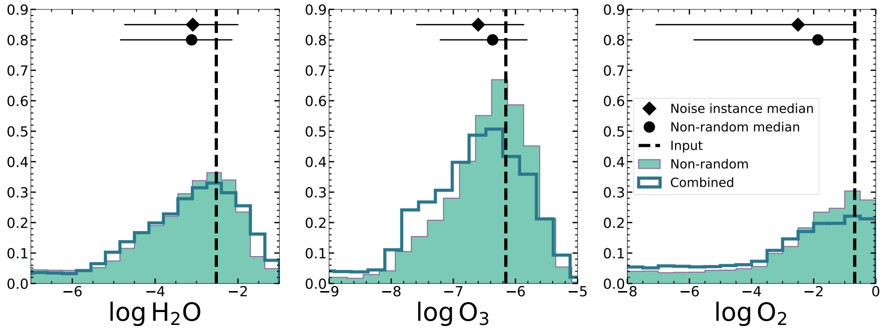


Figure 4.16: The top left panel shows one of the 10 noise instances we retrieved on for $R = 70$, SNR = 15 data, plotted along with the forward model spectrum at $R \sim 70$. The remaining three panels show the gas mixing ratio posteriors (H_2O, O_3, O_2) of all the 10 noise instances of $R = 70$, SNR = 15. In addition, we are showing the corresponding posterior distributions from the non-randomized data set (seen originally in Figure 4.11) for comparison. The set of posteriors that correspond to the noise instance in the top left panel is the set of bolded distributions. The vertical dashed lines represent the input values of the parameters.



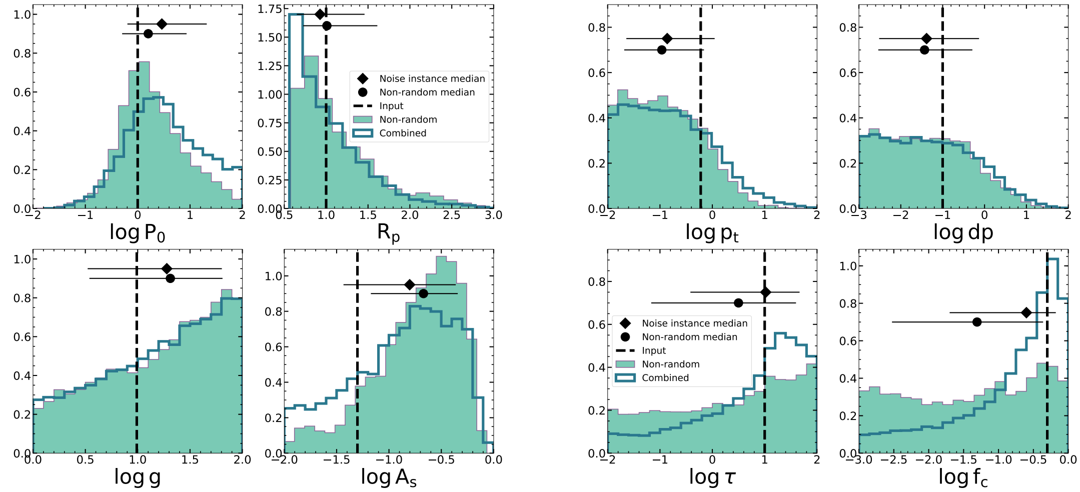
(a) $R = 70$, SNR= 15: Combined bulk parameter posteriors from 10 noise instances

(b) $R = 70$, SNR= 15: Combined cloud parameter posteriors from 10 noise instances



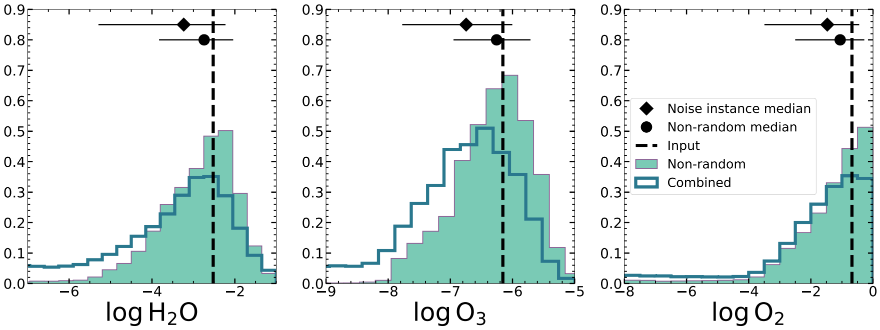
(c) $R = 70$, SNR= 15: Combined gas mixing ratio posteriors from 10 noise instances

Figure 4.17: The combined posteriors distributions from 10 noise instances of $R = 70$, SNR= 15 compared to the posteriors from the non-randomized data set (see also Figure 4.11). The diamond represents the median value of each combined posterior, while the circle is the median of the non-randomized data set posterior. Each median is plotted along with the 68% confidence interval from the same distribution. The vertical dashed lines represent the input values of the parameters.



(a) $R = 140$, SNR$= 10$: Combined bulk parameter posteriors from 10 noise instances

(b) $R = 140$, SNR$= 10$: Combined cloud parameter posteriors from 10 noise instances



(c) $R = 140$, SNR$= 10$: Combined gas mixing ratio posteriors from 10 noise instances

Figure 4.18: The combined posteriors distributions from 10 noise instances of $R = 140$, SNR$= 10$ compared to the posteriors from the non-randomized data set (see also Figure 4.12). The diamond represents the median value of each combined posterior, while the circle is the median of the non-randomized data set posterior. Each median is plotted along with the 68% confidence interval from the same distribution. The vertical dashed lines represent the input values of the parameters.

4.6 Summary

We have developed a retrieval framework for constraining atmospheric properties of an Earth-like exoplanet observed with reflected light spectroscopy spanning the visible range ($0.4 - 1.0\mu$m). We have upgraded an existing, well-tested albedo model to generate high-resolution geometric albedo spectra used to simulate data at resolutions and quality relevant to future telescopes, such as the HabEx and LUVOIR mission concepts. We combined our albedo model with Bayesian inference techniques and applied MCMC sampling to perform parameter estimation. The data we considered were for *WFIRST* paired with a starshade (i.e., the rendezvous scenario), $R = 70$, and $R = 140$ at SNR$= 5, 10, 15, 20$. We validated our forward model, and we demonstrated the successful application of our retrieval approach by gradually adding complexity to our inverse analyses.

Following work by Lupu et al. (2016) and Nayak et al. (2017), who have constructed a retrieval framework for gas giants in refleted light, we made several modifications to the albedo model featured in these previous studies. Our model has a reflective surface, absorption due to water vapor, oxygen, and ozone, Rayleigh scattering from nitrogen and other key gases, pressure-dependent opacities, an adaptive pressure grid, and a single-layer water vapor cloud layer with fractional cloudiness. We performed our retrievals with the goal of estimating our ability to detect and constrain the atmosphere of an Earth twin. We found that $R = 70$, SNR $= 15$ data allowed us to weakly detect surface pressure as well as water vapor, ozone, and oxygen. At $R = 140$, we found that SNR $= 10$ was needed to more firmly detect these parameters. At $R = 140$, a SNR of 20 was needed to constrain key planetary parameters, and $R = 70$ data at this SNR offered extremely few constraints. A *WFIRST* rendezvous scenario, with its photometric points and lower resolution spectrum ($R = 50$), is only able to offer limited diagnostic information. For example, at SNR = 10, we only weakly detect and detect surface pressure and planetary radius, respectively. To weakly detect the gases, *WFIRST* rendezvous data needed to be at least SNR = 20. Throughout our runs, we found that we are unable to accurately constrain surface albedo or place estimates on the surface gravity, although we can straightforwardly rule out planetary sizes above roughly the radius of Uranus or Neptune.

Our findings demonstrate that direct imaging of Earth-like exoplanets in reflected light offers a promising path forward for detecting and constraining atmospheric biosignature gases. Instrument spectral resolution for future missions strongly impacts requisite SNRs for detection and characterization, and this must be taken into consideration during mission design. Thus, the scientific yield of future space-based exoplanet direct imaging missions can only be maximized by simultaneously considering mission characterization goals, integration time constraints, and instrument spectral performance.

Chapter 5

Summary and Future Directions

For as long as we observe and ponder the universe, the challenge of interpreting remote data will remain a hurdle we must clear. Advancements in engineering will continue to bolster our access to distant objects and phenomena. Meanwhile, developments in theoretical models will map further destinations for our voyage of cosmic exploration.

This thesis, in particular, examined the intricate symbiosis between data and models in the context of exoplanet atmosphere characterization. Retrievals are a data-driven modeling tool that will only become more utilized as future missions improve the quality of data. I first studied the implications that arise from our assumption that spectra from unresolved planets can be adequately interpreted via 1D models in retrievals. After demonstrating one such significant consequence – biased abundance estimation – from hot Jupiters in emission, I investigated ways to use 2D models to understand spectroscopic phase curves of transiting planets. Finally, I showed that retrievals are not only valuable in inferring properties from data but also identifying the data we should obtain from future missions. To this end, I built a retrieval framework for reflected light spectra from rocky exoplanets. I outlined observational set ups capable of meeting science goal thresholds.

Astronomers pioneered the techniques for characterizing exoplanet atmospheres in the last decade, laying the groundwork for upcoming and proposed missions that will give us unprecedented insight to exoplanet diversity. The next few decades hold exciting discoveries for planets from hot Jupiters to terrestrial worlds, and retrievals will play a key role along the way.

5.1 Prospects for Transit Characterization of Irradiated Exoplanets

Transiting hot Jupiters have been excellent subjects of state-of-the-art theoretical models and observing facilities. However, less than a dozen have had spectroscopic data sets with high signal-to-noise ratio in emission (Madhusudhan 2018). Fewer still data sets exist for less massive planets. The much-aniticipated JWST will vastly expand the pool of exoplanets with well-characterized atmospheres. We also look forward to the ARIEL mission, expected to launch in the late 2020s. ARIEL will perform photometry and spectroscopy of around 1000 close-in planets around host stars of various spectral types.

With such exciting data on the horizon, now is the time to hone our modeling tools. In chapter 2, I was motivated to inspect whether simplified retrieval model assumptions can lead us astray, and how that may take shape given modern data, compared to more precise future data. I generated synthetic spectra of a planet with two dominant TP profiles, representative of a tidally-locked hot Jupiter with a day-night temperature contrast observed at first or third quarter phase with respect to the observer. I then retrieved on simulated observations from modern Hubble and Spitzer Space Telescopes and the future JWST.

I compared the performance of a 1D (1TP) model against a 2D (2TP) model. The modern data told a cautionary tale: CH_4 can appear artificially constrained with the 1D model, but the data do not strongly justify the inclusion of a second profile in the fit. Biased abundances can mislead the interpretation of the atmosphere, possibly causing speculation of disequilibrium chemistry where there is none. The future data, on the other hand, showed a more clearcut message: a 1D model is insufficient at fitting the data for a planet with thermal inhomogeneity, and all gas species suffered biased abundances.

While I also applied the 2D model to actual WASP-43b observations in chapter 2, it became clear that modeling the full phase curve can illuminate trends of inhomogeneity from retrievals. This need propelled the work in chapter 3. I used spherical trigonometry to appropriately capture limb darkening in a computationally efficient way for 2D retrievals of spectroscopic phase curves, where the observed day and night contributions can be asymmetrical. I then applied 1TP and 2TP retrievals to phase curves from simulated Hubble and Spitzer data, observed WASP-43b Hubble and Spitzer data, and simulated JWST data.

For modern data, almost half the orbit justified the inclusion of 2TP in the fits. H_2O constraints are consistent from phase to phase. CH_4 is biased to higher values for half the orbit. When considering actual WASP-43b observations, I also saw evidence of the offset hotspot (asymmetry in the phase curve) when phase geometry was allowed to be a free parameter. For JWST data, the only phase where 1TP was acceptable was secondary eclipse. Joint retrievals from simultaneously fitting for the full phase curve data set revealed increase in precision for molecular abundances and TP constraints. However, additional work is needed to further develop the simultaneous retrieval approach to tease apart artificial biases from true phenomenon when applying to real data sets.

More phase curves of wider wavelength coverage and increased precision are indeed expected when JWST and ARIEL launch. Finally, we will have exquisite access to multidimensional information from exoplanet atmospheres. We can anticipate mapping the distribution of clouds. This will enable detailed study of atmospheric dynamics and chemical processes. We will constrain more molecules than H_2O, bringing us closer to robust measurements of C/O ratios which link to planetary formation conditions. To prepare for these developments, the field will certainly shift more heavily toward the two following approaches:

- Using 3D General Circulation Models (GCMs) to benchmark retrievals. The work presented in this thesis relies on computationally efficient forward models that are simple by design to perform parameter estimation. GCMs, however, can include more complex self-consistent physics and chemistry. While GCMs are time intensive to run and thus unsuitable as the forward model in a Bayesian retrieval, they can be used to generate synthetic data sets with which to test retrieval performance (Blecic et al. 2017). The use of model comparison tools can identify when features from GCMs are robustly detected, from quenched abundances to the extent of a hot spot.

- Increasing vetting and ease of access to retrieval frameworks. The value of open source code is clear: it offers transparency, version control, ease of community use, and invites collaboration. As new retrieval codes continue to be developed, there needs to be benchmark cases to vet their results (Barstow et al. 2020). This collaborative effort will allow modelers to more clearly interpret results from their own retrievals and that of other teams.

5.2 Prospects for Direct Imaging Characterization of Rocky Exoplanets

Space-based direct imaging is one of the most technologically challenging ways to study exoplanets. Yet it will be our only access to the atmospheres of rocky planets orbiting in the Habitable Zone of Sun-like stars. Planned and proposed missions for the 2030s onward will perform reflected light spectroscopy of directly imaged targets.

In Chapter 4, I presented a retrieval framework to quantify our ability to constrain atmospheric properties using simulated data that such missions might collect. I focused on the visible wavelength range (0.4 - 1μm). For the Earth, this range captures Rayleigh scattering and absorption features of H_2O and key potential biosignatures, O_2 and O_3. I examined data sets of different R and SNR. Detection was achieved for the gases at SNR = 20, R = 70; SNR = 20, R = 140 data constrain them. These are configurations applied with HabEx/LUVOIR in mind. For scenarios relevant to the WFIRST starshade rendezvous mission, SNR = 20, the highest considered, only allowed weak detection of the gases.

I demonstrated the value in providing uncertainties for parameters to help quantify the tradeoff between instrument resolution and target SNR. These results informed set-ups presented in the HabEx/LUVOIR final reports for the 2020 astronomy decadal survey (Gaudi et al. 2020; The LUVOIR Team 2019). The final reports also consider the near-infrared (NIR) region of the spectrum. As described in Chapter 4, molecules like CH_4 and CO_2 have features at such wavelengths. A natural next step would be to investigate the value added by simulated NIR observations.

The inclusion of CH_4 and CO_2 will be significant because both are possible biosignatures. Although much of Chapter 4 is based on modern Earth and its characteristics, Earth's atmosphere has evolved much over the past 4.6 billion years. One example of major development was the buildup of O_2 only 2.5 billion years ago (e.g., review by Kasting & Catling 2003). We should anticipate anoxic biological activity on exoplanets; Krissansen-Totton et al. (2018) presented the case for using simultaneous CO_2 and CH_4 detection within an anoxic environment as evidence for atmospheric disequilibrium due to methanogenic life. It is expected that the exoplanets we will study are at different ages and thus stages in their own atmospheric history. Missions like HabEx and LUVOIR could capture biospheres that were once prevalent on earth, and the improvement of tools such as retrievals go hand in hand in making that a reality.

In addition to an extended wavelength range, two other enhancements to the retrieval framework will help meet the needs of pioneering science possible with future missions. The first is the implementation of a wavelength-dependent surface albedo.

The surface albedo presented in Chapter 4 utilizes a constant surface albedo. With the presence of oceans and continents, the Earth's albedo depends on wavelength. The vegetation "red edge," in particular, corresponds to a jump in reflectivity around 0.7μm (Seager et al. 2005). A parameterized step function can model a straightforward wavelength-dependence in albedo. However, given the radius-albedo degeneracy and the inclusion of additional free parameters, we need to perform a careful model comparison.

We need to establish how detectable a "red edge" or other biological pigment-induced albedo discontinuities (e.g., Hegde et al. 2015) can be, given the data. The retrieval framework from Chapter 4 is well-suited to tackle this investigation and evaluate the robustness of surface albedo as a biosignature. In addition, a variety of surface minerals that exist for rocky bodies in our Solar System. It behooves us to consider a slew of minerals or surface types (e.g., olivine, water ice) and their reflection features when considering rocky exoplanets in general (Hu et al. 2012). Retrievals can tell us if future data can constrain surface composition.

Another exciting prospect is monitoring for temporal variability of the surface. From season to season on Earth, vegetation coverage - and hence albedo - changes. Keeling et al. (1976) showed seasonal variations in CO_2 concentrations associated with photosynthetic activity and decomposition following increase or decrease of solar radiation. Long-term observations of planets in the Habitable Zone can reveal this behavior, if present. They can also be used to infer planetary rotation and reconstruct surface maps (Cowan et al. 2009; Lustig-Yaeger et al. 2018).

The second promising enhancement is retrieving background gas as a free parameter. While N_2 is the expected background gas for Earth, it will be important to study how well we can discern the properties of the background gas for a planet of interest. The *Kepler* mission determined that planets between the sizes of Earth and Neptune (~ 4 Earth-radii) are commonplace in our Galaxy (Petigura et al. 2018). Such planets are called super-Earths or mini-Neptunes, and their atmospheric composition is still under study. They may have H_2-rich atmospheres, and radius is not a well-constrained property based on Chapter 4 results. Consequently, we will need to rely on any constraining power the data may have for background gas to shed light on the nature of these planets. In the process of studying potentially habitable targets, HabEx and LUVOIR expect to capture plenty of data (following the long exposures) for sub-Neptune planets present in the same system. This data set would offer access to a variety of atmospheric compositions and evidence of different clouds and hazes (e.g., Morley et al. 2015). The comparative planetology enabled by future large space-based telescopes will undoubtedly transform our perspective on planet formation and evolution in systems around different stellar types.

Certainly there are a plethora of new parameters we can add to our retrieval framework. However, chapters 2 and 3 warn us of adopting more complexity than necessitated by the data. An essential extension of reflected light retrievals will be to study the conditions where a 1D model is appropriate. Recently, Batalha et al. (2019) developed PICASO (Planetary Intensity Code for Atmospheric Scattering Observations), an albedo model that has the advantage of being python-based and more easily compatible with existing Bayesian inference tools. PICASO can be incorporated into a retrieval framework to examine whether inhomogeneities are justified for a given data set. Yet, chapter 4 showed us that cloud parameters are difficult to constrain for the HabEx/LUVOIR-like data considered. While retrievals of reflected light spectra from HabEx/LUVOIR will likely remain 1D for the near future, these data will be thoroughly groundbreaking and drive the iterative improvement of models.

We are at the frontier of space missions that could discover evidence of life elsewhere in the universe. Both HabEx and LUVOIR outline a systematic approach to achieve this goal. The initial steps aim to confirm that a target is a planet in the Habitable Zone of its host. The next steps aim to detect the presence of atmospheric water. To eventually validate the biosignatures of promising candidates, we need to quantify the thresholds. The retrieval framework I developed in this thesis has already clarified the ease with which we can screen candidates for water and ultimately search for signs of life (Feng et al. 2018; Smith et al. 2020). With this tool and future advancements to be added, we may finally help answer the question: Are we alone?

Appendix A

Incorporating and validating phase geometry

Because our goal in chapter 3 is to perform computationally intensive retrievals on spectroscopic phase curve data, we need the forward model to remain simple enough while capturing the complexity of the geometry. Here we describe the construction of the 2TP-Crescent model geometry.

We consider the case of an orbit observed at N phases, where phase angle is defined as α. At a given phase, we leverage the fact that limb darkening is treated for intensities at the relevant viewing angles following Line et al. (2013). Radiative transfer and fluxes are calculated with integration by Gaussian quadrature. For an approximation of N_{Gauss} points, we have the same number of intensities, each at an angle of μ_i with a weight of w_i. The upwelling intensity is given by $I_\lambda = \sum B_\lambda e^{-\sum \Delta\tau} \Delta\tau$ (Line et al. 2013, Eq. 8). We integrate these intensities over 2π with Gaussian quadrature, leading to an annulus at each μ_i, similar to the formalism from Barman et al. (2005).

Each annulus can be a linear combination of arbitrary TP profiles: $2\pi \sum \mu_i w_i I$. In our study, $I = AI_{\text{hot}} + (1 - A)I_{\text{cold}}$, where A is the fractional area that is emitting with the hotter profile in a 2TP setup. I_{hot} represents the intensities calculated using a profile based on the TP parameters including β_{day}, while I_{cold} uses the same parameters except β_{night}.

We take advantage of a hot Jupiter's (assumed) tidally locked configuration, which leads to a large temperature contrast, as well as our knowledge of the phase geometry, i.e., the amount of a planet that is seen to be illuminated at angle α. The equation for illuminated fraction (k) as a function of phase angle (α) is

$$k = \frac{1 - \cos\alpha}{2}. \tag{A.1}$$

The next step is determining the fraction within each annulus at a given phase that corresponds to the hotter emission on the visible hemisphere. As seen in Figure A.2(a) , for the Gaussian quadrature integration of N points, there is a set of weights w_i and corresponding θ_i that indicates the direction of the radiation beam for each point i. The arc that spans between two θ_i is the angular width of an annulus. We use the width to determine the area of each annulus, summing up to 2π, the surface area of half a sphere (the visible hemisphere). The covered area of a spherical lune of angle α (at phase α) is 2α (Fig A.2(b)). We can calculate the fractional area in each annulus that then sum up to 2α for one profile.

An important choice in balancing forward modeling speed and model accuracy

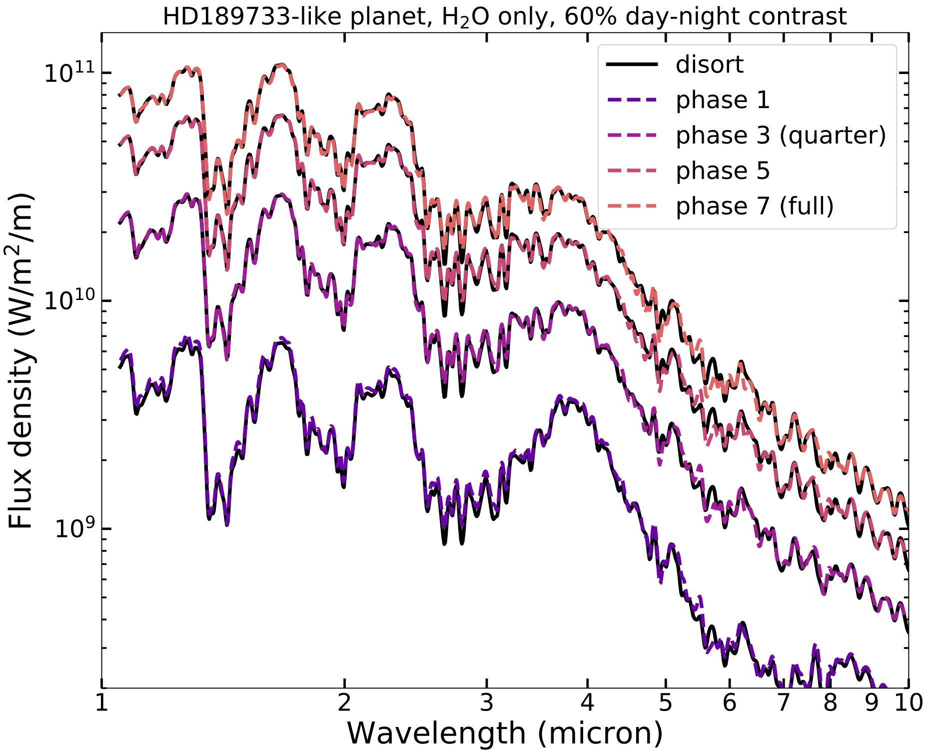
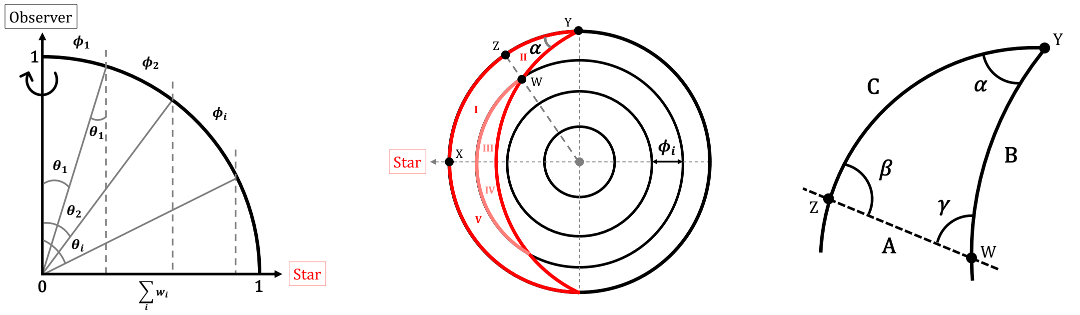


Figure A.1: Comparing Morley et al. (2015) `disort` 3D model's spectra (dashed) to this study's annulus model's spectra (solid) at four phases from after transit to secondary eclipse. The simpler annulus model is able to match the 3D model's output well throughout the orbit.



(a) Sideview, defining ϕ_i.

(b) Front view, annuli.

(c) AAS spherical triangle solution.

Figure A.2: Detailed schematics of annulus geometry. (a) 1D sideview of a section of the atmosphere in the Gaussian quadrature setup with w_i for N points. We use a unit circle as an example. For each $\sum_i w_i$, there is a corresponding angle θ_i. The span of each arc between θ_i is ϕ_i. Within each arc is a beam of radiation. This quadrant is integrated azimuthally over 2π to determine the total outgoing radiation of the hemisphere facing the observer. (b) View of hemisphere visible to observer at phase α. The emitting region (in red) intersects the annuli at different points. By determining the areas of these segments, we can calculate how much of each annulus is emitting as described in the text. (c) A zoom-in of spherical triangle ZYW for phase α. The known variables are α, β, and A. This sets up the solution for AAS (angle-angle-side) spherical triangles. See Equations A.2 to A.4.

is setting N_{Gauss}. We first verify that our annulus model produces the emitting fractions as determined by Equation A.1 for the set of phase angles in Table 3.2. A 3D model with higher spatial resolution and individually calculated fluxes would better simulate a realistic planet atmosphere at partially illuminated phases. However, 3D models are time-consuming to run even once. If one attempts to fold that into a retrieval framework, where numerous calls to the forward model are necessary, then it will not be an effective way to estimate properties of the atmosphere. However, we can make use of 3D models to validate our annulus approach. If the spectrum from the annulus method matches that of a 3D output, then we would be confident in the retrieval inferences (similar to Blecic et al. 2017).

We compare the spectra from a 3D model to spectra generated with the annulus model using $N_{\text{Gauss}} = 4$. Our 3D model combines the 1D radiative transfer code `disort` as adapted by Morley et al. (2015) with a 3D longitude-latitude grid. We assign atmospheric properties and a TP profile to each point in the grid and integrate for the emerging flux from the planet. For the visible hemisphere, the grid has 16 longitudes and 32 latitudes. As orbital phase increases from transit to secondary eclipse, points along additional longitudes adopt the day-side temperature profile.

Our test case is an atmosphere with only water vapor as an opacity source with 60% day-night temperature contrast ($\beta_{\text{night}} = 0.4$) for an HD 189733b-like planet. Past phase angle 22.5° (0.0625 if one phase goes from 0 to 1; see also Table 3.2), we find good agreement between the 3D spectra and our annulus model, as seen in Figure A.1. This phase is in fact the smallest we consider in our paper, as it is the first phase in the

WASP-43b data we use (Stevenson et al. 2017). We attribute the small mismatches to differences in opacity libraries. Furthermore, our simulated data will be generated and retrieved using the same forward model, providing us with self-consistency in evaluating the results. Thus, $N_{\text{Gauss}} = 4$ offers accurate spectra for a given phase geometry and is computationally efficient within our retrieval framework. We use $N_{\text{Gauss}} = 4$ for 2TP-Crescent and for all models in this work.

Figure A.2(b) shows a view of $N_{\text{Gauss}} = 4$ where we divide the visible hemisphere into annuli, and the planet is viewed at phase α. Because the top and bottom halves are symmetric, we only need to determine the illuminated areas of the top and multiply by two. To calculate the area of Region II as marked in Figure A.2(b), we utilize the AAS (angle-angle-side) solution to solving spherical triangles. Spherical AAS is where one Side, one adjacent Angle, and one opposite Angle are known for a triangle on a sphere, as illustrated in Figure A.2(c). Because angle β is marked by the intersection between a great circle arc from the center of the sphere and the equator (viewed top down), its value is $\beta = \frac{\pi}{2}$. We know α as the phase angle. Side A is the width of the annulus, ϕ_i. To calculate side B, side C, and angle γ, we follow Equations A.2 through A.4. Once side C, or arc ZY in Figure A.2(b), is obtained, we get arc $ZX = XY - ZY = \frac{\pi}{2} - C$. The area of a spherical triangle is the sum of the angles; for Region II, the area is thus $\alpha + \beta + \gamma$. The area of Region I is calculated as the area of a longitude-latitude patch, using $I = ZX \cdot |\sin \text{lat1} - \sin \text{lat2}|$, where $\text{lat1} = \frac{\pi}{2}$, $\text{lat2} = (\frac{\pi}{2} - \theta_1)$ for the outermost annulus in Figure A.2(a). The area of Region III is then $\alpha - (I + II)$. Based on these methods, we calculate the set of areas for hot regions in the annuli as a function of phase.

$$B = \frac{\sin A \sin B}{\sin \alpha} \tag{A.2}$$

$$C = 2\arctan\Big[\tan\big(\frac{1}{2}(A-B)\big)\frac{\sin\big(\frac{1}{2}(\alpha+\beta)\big)}{\sin\big(\frac{1}{2}(\alpha-\beta)\big)}\Big] \tag{A.3}$$

$$\gamma = 2\,\mathrm{arccot}\Big[\tan\big(\frac{1}{2}(\alpha-\beta)\big)\frac{\sin\big(\frac{1}{2}(A+B)\big)}{\sin\big(\frac{1}{2}(A-B)\big)}\Big] \tag{A.4}$$

Bibliography

Agúndez, M., Parmentier, V., Venot, O., Hersant, F., & Selsis, F. 2014, A&A, 564, A73

Agúndez, M., Venot, O., Iro, N., et al. 2012, A&A, 548, A73

Allen, C., & Cox, A. 2000, Allen's Astrophysical Quantities, Allen's Astrophysical Quantities

Arcangeli, J., Désert, J.-M., Parmentier, V., et al. 2019, A&A, 625, A136

Artigau, É., Bouchard, S., Doyon, R., & Lafrenière, D. 2009, ApJ, 701, 1534

Balasubramanian, K., White, V., Yee, K., et al. 2016, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011005

Barman, T. S., Hauschildt, P. H., & Allard, F. 2005, ApJ, 632, 1132

Barman, T. S., Konopacky, Q. M., Macintosh, B., & Marois, C. 2015, ApJ, 804, 61

Barman, T. S., Macintosh, B., Konopacky, Q. M., & Marois, C. 2011, ApJ, 733, 65

Barstow, J. K., Changeat, Q., Garland, R., et al. 2020, MNRAS, 493, 4884

Batalha, N. E., Marley, M. S., Lewis, N. K., & Fortney, J. J. 2019, ApJ, 878, 70

Bean, J. L., Stevenson, K. B., Batalha, N. M., et al. 2018, PASP, 130, 114402

Benneke, B. 2015, arXiv e-prints, arXiv:1504.07655

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

—. 2013, ApJ, 778, 153

Bézard, B., Lellouch, E., Strobel, D., Maillard, J.-P., & Drossart, P. 2002, Icarus, 159, 95

Blecic, J., Dobbs-Dixon, I., & Greene, T. 2017, ApJ, 848, 127

Borucki, W. J., Koch, D. G., Lissauer, J. J., et al. 2003, in Proc. SPIE, Vol. 4854, Future EUV/UV and Visible Space Astrophysics Missions and Instrumentation., ed. J. C. Blades & O. H. W. Siegmund, 129–140

Borucki, W. J., Koch, D. G., Basri, G., et al. 2011, ApJ, 728, 117

Brandt, T. D., & Spiegel, D. S. 2014, Proceedings of the National Academy of Science, 111, 13278

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

Buenzli, E., Apai, D., Radigan, J., Reid, I. N., & Flateau, D. 2014, ApJ, 782, 77

Buenzli, E., Saumon, D., Marley, M. S., et al. 2015, ApJ, 798, 127

Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, The Astrophysical Journal, 809, 8

Burrows, A. 2014, ArXiv e-prints, arXiv:1412.6097

Burrows, A., Hubeny, I., Budaj, J., Knutson, H. A., & Charbonneau, D. 2007, ApJ, 668, L171

Cady, E., Prada, C. M., An, X., et al. 2016, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011004

Cahoy, K. L., Marley, M. S., & Fortney, J. J. 2010, ApJ, 724, 189

Caldas, A., Leconte, J., Selsis, F., et al. 2019, A&A, 623, A161

Cash, W. 2006, Nature, 442, 51

Changeat, Q., Edwards, B., Waldmann, I. P., & Tinetti, G. 2019, ApJ, 886, 39

Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L. 2002, ApJ, 568, 377

Chen, J., & Kipping, D. 2017, ApJ, 834, 17

Cooper, C. S., & Showman, A. P. 2006, ApJ, 649, 1048

Cornish, N. J., & Littenberg, T. B. 2007, Phys. Rev. D, 76, 083006

Cowan, N. B., & Fujii, Y. 2017, ArXiv e-prints, arXiv:1704.07832

Cowan, N. B., Agol, E., Meadows, V. S., et al. 2009, ApJ, 700, 915

Crill, B., & Siegler, N. 2017, ArXiv e-prints, arXiv:1709.06660

de Wit, J., Wakeford, H. R., Gillon, M., et al. 2016, Nature, advance online publication,

Dobbs-Dixon, I., & Agol, E. 2013, MNRAS, 435, 3159

Drummond, B., Mayne, N. J., Manners, J., et al. 2018, ApJ, 869, 28

Drummond, B., Hebrard, E., Mayne, N. J., et al. 2020, arXiv e-prints, arXiv:2001.11444

Eccleston, P., Tinetti, G., Beaulieu, J.-P., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9904, Proc. SPIE, 990433

Enoch, M. L., Brown, M. E., & Burgasser, A. J. 2003, AJ, 126, 1006

Espinoza, N., Fortney, J. J., Miguel, Y., Thorngren, D., & Murray-Clay, R. 2017, ApJ, 838, L9

Evans, T. M., Sing, D. K., Kataria, T., et al. 2017, Nature, 548, 58

Feng, Y. K., Line, M. R., Fortney, J. J., et al. 2016, ApJ, 829, 52

Feng, Y. K., Robinson, T. D., Fortney, J. J., et al. 2018, AJ, 155, 200

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

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

Ford, E. B., Seager, S., & Turner, E. L. 2001, Nature, 412, 885

Foreman-Mackey, D. 2016, The Journal of Open Source Software, 24

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

Fortney, J. J., Cooper, C. S., Showman, A. P., Marley, M. S., & Freedman, R. S. 2006, ApJ, 652, 746

Fortney, J. J., Lodders, K., Marley, M. S., & Freedman, R. S. 2008, ApJ, 678, 1419

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

Gaudi, B. S., Seager, S., Mennesson, B., et al. 2020, arXiv e-prints, arXiv:2001.06683

Goodman, J., & Weare, J. 2010, Commun. Appl. Math. Comput. Sci., 5, 65

Gordon, C., & Trotta, R. 2007, MNRAS, 382, 1859

Greco, J. P., & Burrows, A. 2015, ApJ, 808, 172

Greene, T. P., Line, M. R., Montero, C., et al. 2016, ApJ, 817, 17

Gregory, P. C. 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support

Groff, T., Gong, Q., Mandell, A. M., et al. 2018, in American Astronomical Society Meeting Abstracts, Vol. 231, American Astronomical Society Meeting Abstracts, #355.47

Guillot, T. 2010, A&A, 520, A27

Hansen, J. E., & Travis, L. D. 1974, Space Sci. Rev., 16, 527

Hegde, S., Paulino-Lima, I. G., Kent, R., Kaltenegger, L., & Rothschild, L. 2015, Proceedings of the National Academy of Science, 112, 3886

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

Horak, H. G. 1950, ApJ, 112, 445

Horak, H. G., & Little, S. J. 1965, ApJS, 11, 373

Hu, R., Ehlmann, B. L., & Seager, S. 2012, ApJ, 752, 7

Irwin, P. G. J., Parmentier, V., Taylor, J., et al. 2020, MNRAS, 228

Irwin, P. G. J., Teanby, N. A., de Kok, R., et al. 2008, J. Quant. Spec. Radiat. Transf., 109, 1136

Iyer, A. R., & Line, M. R. 2020, ApJ, 889, 78

Kasdin, N. J., Lisman, D., Shaklan, S., et al. 2012, in Proc. SPIE, Vol. 8442, Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave, 84420A

Kasting, J. F., & Catling, D. 2003, ARA&A, 41, 429

Kasting, J. F., Whitmire, D. P., & Reynolds, R. T. 1993, Icarus, 101, 108

Kataria, T., Showman, A. P., Fortney, J. J., et al. 2015, ApJ, 801, 86

Kattawar, G. W. 1975, J. Quant. Spec. Radiat. Transf., 15, 839

Keeling, C. D., Bacastow, R. B., Bain-Bridge, A. E., et al. 1976, Tellus, 28, 538

Knutson, H. A., Charbonneau, D., Allen, L. E., Burrows, A., & Megeath, S. T. 2008, ApJ, 673, 526

Komacek, T. D., Showman, A. P., & Tan, X. 2017, ApJ, 835, 198

Konopacky, Q. M., Barman, T. S., Macintosh, B. A., & Marois, C. 2013, Science, 339, 1398

Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, ApJ, 765, 131

Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014a, ApJ, 793, L27

—. 2014b, Nature, 505, 69

Kreidberg, L., Line, M. R., Parmentier, V., et al. 2018, AJ, 156, 17

Krissansen-Totton, J., Bergsman, D. S., & Catling, D. C. 2016, Astrobiology, 16, 39

Krissansen-Totton, J., Garland, R., Irwin, P., & Catling, D. C. 2018, AJ, 156, 114

Krist, J., Nemati, B., & Mennesson, B. 2016, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011003

Lee, J.-M., Fletcher, L. N., & Irwin, P. G. J. 2012, MNRAS, 420, 170

Lin, D. N. C., Bodenheimer, P., & Richardson, D. C. 1996, Nature, 380, 606

Line, M. R., Knutson, H., Wolf, A. S., & Yung, Y. L. 2014, ApJ, 783, 70

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

Line, M. R., Vasisht, G., Chen, P., Angerhausen, D., & Yung, Y. L. 2011, ApJ, 738, 32

Line, M. R., Zhang, X., Vasisht, G., et al. 2012, ApJ, 749, 93

Line, M. R., Wolf, A. S., Zhang, X., et al. 2013, ApJ, 775, 137

Line, M. R., Stevenson, K. B., Bean, J., et al. 2016, AJ, 152, 203

Lupu, R. E., Marley, M. S., Lewis, N., et al. 2016, AJ, 152, 217

Lupu, R. E., Zahnle, K., Marley, M. S., et al. 2014, ApJ, 784, 27

Lustig-Yaeger, J., Meadows, V. S., Tovar Mendoza, G., et al. 2018, AJ, 156, 301

Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, Proceedings of the National Academy of Science, 111, 12661

Macintosh, B., Graham, J. R., Barman, T., et al. 2015, Science, 350, 64

Madhusudhan, N. 2018, Atmospheric Retrieval of Exoplanets, 104

—. 2019, ARA&A, 57, 617

Madhusudhan, N., Knutson, H., Fortney, J. J., & Barman, T. 2014, in Protostars and Planets VI, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 739

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

—. 2010, ApJ, 725, 261

Madhusudhan, N., Harrington, J., Stevenson, K. B., et al. 2011, Nature, 469, 64

Mai, C., & Line, M. R. 2019, ApJ, 883, 144

Marley, M., Lupu, R., Lewis, N., et al. 2014, ArXiv e-prints, arXiv:1412.8440

Marley, M. S., Gelino, C., Stephens, D., Lunine, J. I., & Freedman, R. 1999, ApJ, 513, 879

Marley, M. S., & Robinson, T. D. 2015, ARA&A, 53, 279

Marley, M. S., Saumon, D., Cushing, M., et al. 2012, ApJ, 754, 135

Mawet, D., Ruane, G., Xuan, W., et al. 2017, ApJ, 838, 92

Maxted, P. F. L., Anderson, D. R., Doyle, A. P., et al. 2013, MNRAS, 428, 2645

Mayor, M., & Queloz, D. 1995, Nature, 378, 355

McClatchey, R. A., Fenn, R. W., Selby, J. E. A., Volz, F. E., & Garing, J. S. 1972,

McKay, C. P., Pollack, J. B., & Courtin, R. 1989, Icarus, 80, 23

Meador, W. E., & Weaver, W. R. 1980, Journal of Atmospheric Sciences, 37, 630

Meadows, V. S., & Crisp, D. 1996, J. Geophys. Res., 101, 4595

Mendonça, J. M., Malik, M., Demory, B.-O., & Heng, K. 2018a, AJ, 155, 150

Mendonça, J. M., Tsai, S.-m., Malik, M., Grimm, S. L., & Heng, K. 2018b, ApJ, 869, 107

Mennesson, B., Gaudi, S., Seager, S., et al. 2016, in Proc. SPIE, Vol. 9904, Space

Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave, 99040L

Mollière, P., van Boekel, R., Dullemond, C., Henning, T., & Mordasini, C. 2015, ApJ, 813, 47

Mollière, P., Wardenier, J. P., van Boekel, R., et al. 2019, A&A, 627, A67

Morello, G., Danielski, C., Dickens, D., Tremblin, P., & Lagage, P. O. 2019, AJ, 157, 205

Morley, C. V., Fortney, J. J., Marley, M. S., et al. 2015, ApJ, 815, 110

Morley, C. V., Marley, M. S., Fortney, J. J., & Lupu, R. 2014, ApJ, 789, L14

Moses, J. I., Visscher, C., Fortney, J. J., et al. 2011, ApJ, 737, 15

Mukherjee, P., Parkinson, D., & Liddle, A. R. 2006, ApJ, 638, L51

Nayak, M., Lupu, R., Marley, M. S., et al. 2017, PASP, 129, 034401

Noecker, M. C., Zhao, F., Demers, R., et al. 2016, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011001

Öberg, K. I., Murray-Clay, R., & Bergin, E. A. 2011, ApJ, 743, L16

Ohno, K., & Zhang, X. 2019, ApJ, 874, 2

Parmentier, V., & Crossfield, I. J. M. 2018, Exoplanet Phase Curves: Observations and Theory, 116

Parmentier, V., & Guillot, T. 2014, A&A, 562, A133

Peterson, B. M., Fischer, D., & LUVOIR Science and Technology Definition Team. 2017, in American Astronomical Society Meeting Abstracts, Vol. 229, American Astronomical Society Meeting Abstracts, 405.04

Petigura, E. A., Marcy, G. W., Winn, J. N., et al. 2018, AJ, 155, 89

Qi, C., Öberg, K. I., Wilner, D. J., et al. 2013, Science, 341, 630

Radigan, J., Lafrenière, D., Jayawardhana, R., & Artigau, E. 2014, ApJ, 793, 75

Rajan, A., Rameau, J., De Rosa, R. J., et al. 2017, AJ, 154, 10

Robinson, T. D. 2017, ArXiv e-prints, arXiv:1701.05205

Robinson, T. D., Ennico, K., Meadows, V. S., et al. 2014, ApJ, 787, 171

Robinson, T. D., & Marley, M. S. 2014, ApJ, 785, 158

Robinson, T. D., Stapelfeldt, K. R., & Marley, M. S. 2016, PASP, 128, 025003

Robinson, T. D., Meadows, V. S., Crisp, D., et al. 2011, Astrobiology, 11, 393

Rodgers, C. D. 1976, Reviews of Geophysics and Space Physics, 14, 609

Rogers, L. A. 2015, ApJ, 801, 41

Rossi, L., & Stam, D. M. 2017, A&A, 607, A57

Rothman, L. S., Gordon, I. E., Babikov, Y., et al. 2013, J. Quant. Spec. Radiat. Transf., 130, 4

Sagan, C., Thompson, W. R., Carlson, R., Gurnett, D., & Hord, C. 1993, Nature, 365, 715

Savransky, D., & Garrett, D. 2016, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011006

Seager, S., Turner, E. L., Schafer, J., & Ford, E. B. 2005, Astrobiology, 5, 372

Seager, S., Turnbull, M., Sparks, W., et al. 2015, in Proc. SPIE, Vol. 9605, Techniques and Instrumentation for Detection of Exoplanets VII, 96050W

Seo, B.-J., Gordon, B., Kern, B., et al. 2016, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011019

Shaw, J. R., Bridges, M., & Hobson, M. P. 2007, MNRAS, 378, 1365

Showman, A. P., Fortney, J. J., Lian, Y., et al. 2009, ApJ, 699, 564

Showman, A. P., & Guillot, T. 2002, A&A, 385, 166

Showman, A. P., & Polvani, L. M. 2011, ApJ, 738, 71

Sing, D. K., Fortney, J. J., Nikolov, N., et al. 2016, Nature, 529, 59

Skemer, A. J., Marley, M. S., Hinz, P. M., et al. 2014, ApJ, 792, 17

Skilling, J. 2004, in American Institute of Physics Conference Series, Vol. 735, American Institute of Physics Conference Series, ed. R. Fischer, R. Preuss, & U. V. Toussaint, 395–405

Smith, A. J. R. W., Feng, Y. K., Fortney, J. J., et al. 2020, AJ, 159, 36

Spergel, D., Gehrels, N., Breckinridge, J., et al. 2013, ArXiv e-prints, arXiv:1305.5422

Spergel, D., Gehrels, N., Baltay, C., et al. 2015, ArXiv e-prints, arXiv:1503.03757

Steinrueck, M. E., Parmentier, V., Showman, A. P., Lothringer, J. D., & Lupu, R. E. 2019, ApJ, 880, 14

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

Stevenson, K. B., Line, M. R., Bean, J. L., et al. 2017, AJ, 153, 68

Stubenrauch, C. J., Rossow, W. B., Kinne, S., et al. 2013, Bulletin of the American Meteorological Society, 94, 1031

Taylor, J., Parmentier, V., Irwin, P. G. J., et al. 2020, arXiv e-prints, arXiv:2002.00773

The LUVOIR Team. 2019, arXiv e-prints, arXiv:1912.06219

Thorngren, D., Gao, P., & Fortney, J. J. 2019, ApJ, 884, L6

Tinetti, G., Drossart, P., Eccleston, P., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9904, Proc. SPIE, 99041X

Tinetti, G., Drossart, P., Eccleston, P., et al. 2018, Experimental Astronomy, 46, 135

Toon, O. B., McKay, C. P., Ackerman, T. P., & Santhanam, K. 1989, Journal of Geophysical Research, 94, 16287

Traub, W. A., Breckinridge, J., Greene, T. P., et al. 2016, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011020

Trauger, J., Moody, D., Krist, J., & Gordon, B. 2016, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011013

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

Venot, O., Parmentier, V., Blecic, J., et al. 2020, arXiv e-prints, arXiv:2001.04759

von Paris, P., Hedelt, P., Selsis, F., Schreier, F., & Trautmann, T. 2013, A&A, 551, A120

Waldmann, I. P., Rocchetto, M., Tinetti, G., et al. 2015a, ApJ, 813, 13

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

Wang, J., Mawet, D., Ruane, G., et al. 2017a, Proc.SPIE, 10400, 10400

Wang, J., Mawet, D., Ruane, G., Hu, R., & Benneke, B. 2017b, AJ, 153, 183

Wong, I., Knutson, H. A., Kataria, T., et al. 2016, ApJ, 823, 122

Zhang, X., & Showman, A. P. 2014, ApJ, 788, L6

Zhou, Y., Apai, D., Schneider, G. H., Marley, M. S., & Showman, A. P. 2016, ApJ, 818, 176