

ABSTRACT

Title of Dissertation: **CONNECTING THEORY AND OBSERVATIONS
OF EXOPLANET ATMOSPHERES AND SURFACES
AT THE INDIVIDUAL AND POPULATION LEVEL
WITH JWST**

Jegug Ih
Doctor of Philosophy, 2024

Dissertation Directed by: **Professor Eliza Miller-Ricci Kempton**
Department of Astronomy

Observing an exoplanet's atmosphere via photometry and spectroscopy has provided the main window to understanding its properties and processes, as the atmospheric spectra encompass information about the chemistry, thermal structure, surfaces, as well as formation history and even biology. To this end, one key science goal of the James Webb Space Telescope (JWST) is to establish whether rocky planets around M dwarfs can host atmospheres or not. JWST offers unprecedented signal-to-noise and unlocks new parameter space regimes of planets available for characterizing not only the atmosphere but also the surface. This advancement in observing capability simultaneously poses novel challenges to atmospheric characterization. My dissertation addresses some of the new challenges to atmospheric retrievals in the era of JWST and the characterization of rocky planets.

Firstly, I quantified the effects of wavelength-correlated systematics on atmospheric retrievals. Wavelength-correlated noise can occur due to instrumental systematics or stellar effects

and the merging of discrete data sets. I investigated the effect of correlated noise and constrained the potential biases incurred in the retrieved posteriors by performing retrievals on multiple noise instances of synthetic data. The study found that correlated noise allows for overfitting the spectrum, thereby yielding a better goodness of fit on average but degrading the overall accuracy of retrievals by roughly 1σ . In particular, correlated noise can manifest as an apparent non-Rayleigh slope in the optical range, leading to an incorrect estimate of cloud/haze parameters. Finally, I show that while correlated noise cannot be reliably distinguished with Hubble Space Telescope observations, inferring its presence and strength may be possible with JWST.

Secondly, I studied the how the choice in parameterization of the atmospheric composition can influence the posterior when performing retrieval analyses on terrestrial planet atmospheres, for which the mean molecular weight is not known a priori. By performing self-retrievals and varying the parameterization, I found that the centered log-ratio transform, commonly used for this application, tends to overestimate the abundances of spectroscopically active gases when inactive ones are present. Over multiple noise instances, I found that no one parameterization method always outperforms others. Comparing the Bayesian evidences from retrievals on multiple noise instances, I found that for a given spectrum, the choice in parameterization can affect the Bayes factor of whether a molecule should be included. Alongside astrophysical effects, this remains a fundamental challenge to atmospheric retrievals for small planet and can be addressed by more observations.

Finally, I constrained the atmospheric thickness and characterized the surface from the first JWST measurement of thermal emission from a rocky exoplanet, TRAPPIST-1 b. I compared TRAPPIST-1 b's measured secondary eclipse depth to predictions from a suite of self-consistent radiative-convective equilibrium models. I found that plausible atmospheres (i.e., those that con-

tain at least 100 ppm CO₂) with surface pressures greater than 0.3 bar are ruled out at 3 σ , regardless of the choice of background atmosphere, and a Mars-like thin atmosphere with surface pressure 6.5 mbar composed entirely of CO₂ is also ruled out at 3 σ . I modelled the emission spectra for bare-rock planets of various compositions and found that a basaltic surface best matches the measured eclipse depth to within 1 σ .

Connecting Theory and Observations of Exoplanet Atmospheres
and Surfaces at the Individual and Population Level with JWST

by

Jegug Ih

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

Advisory Committee:

Professor Eliza M.-R. Kempton, Chair/Advisor

Professor M. Coleman Miller

Professor Benedikt Diemer

Professor Ricardo David Arevalo

Dr. Josh Lothringer

© Copyright by
Jegug Ih
2024

Preface

The work presented in Chapters 2 and 4 of this dissertation has been previously published as first author papers. The work in Chapter 3 is an early draft of a paper currently in-prep that will be submitted shortly as a first author paper. Chapter 5 contains a figure from an in-prep co-first author paper.

The work in Chapter 2 was published in the *Astronomical Journal* (AJ) as "Understanding the Effects of Systematics in Exoplanetary Atmospheric Retrievals" [1] and is presented here with minimal modification. I conducted the simulations and produced all graphics and tables. My co-author (advisor) provided the initial topic and contributed feedback on methods and text.

The work in Chapter 3 is a draft of a paper that will be submitted shortly as a first author paper. The work was presented in a number of virtual conferences and meetings in 2021. I conducted the simulations and produced all graphics and tables. My co-author (advisor) contributed feedback on methods and text.

Chapter 4 was published in the *Astrophysical Journal Letters* (ApJL) as "Constraining the Thickness of TRAPPIST-1 b's Atmosphere from Its JWST Secondary Eclipse Observation at 15 μm " and is presented here with minimal modification. The observation that the work characterizes was published in Greene et al. [2], and my co-author Emily Whittaker assisted with setting up and troubleshooting the radiative transfer code HELIOS.

Chapter 5 contains a figure from an in-prep co-first author paper. Co-author Brandon Park Coy collated the dataset.

Dedication

박혜선 선생님께.

Acknowledgments

The work in this dissertation took place over 6 years of graduate study at the University of Maryland which I could not have possibly made through without the support of so many individuals. First of all, I would like to thank my graduate advisor, Eliza Kempton. Thank you for the academic, professional, and funding support throughout my graduate career and patiently never losing confidence in me, especially when I struggled to find it myself. I also thank the my committee members: Cole Miller for answering all of my silly questions about statistics, Drake Deming for so much insight into how observations work. Thank you all.

There are countless people in the exoplanet community at UMD and in the collaborator network that I always appreciated working with and whose expertise and camaraderie that I am grateful for – especially Matej, Sandra, Tad, Matt, Anjali, Peter, Jacob, and Megan, thank you for helping me navigate this strange world of academia.

The past six years have been a period defined not only by intellectual and profession growth but also a personal one. There are many friends and family to whom I owe an immense gratitude...

...to Vicente, Arjun, Ramsey, Liz, Guangwei, Milena, Julian, and all the fellow graduate student friends that I made at the UMD Astronomy– I deeply appreciate having had the opportunity to meet you and am grateful for all the ways in which you enriched the experience.

...to AJ, Alex, Ari, Ben, Marty, Melissa, Ryan– I can only aspire to be as creative and fearless off stage as I am when I am with you. Especially to Cass, Ryan, and Sadie, thank you for

being my friends, neighbors, and adopted family.

...그리고 승찬, 창훈, 진혁, 택민, 보리, 서현, 형석, 주형, 병제, 승찬, 동구, 민수, 동륜, 기홍, 서호, 재서에게- 꽤나 힘들었던 순간, 너희들 덕분에 이겨낼 수 있었어.

...to Aya, Callingham, Ed, Els, Fred, Isaac, Lucrezia, Maude, Priyasha, and all of my Clare friends- thank you for reminding me that I will always have a third place I can call home in you.

To Tony, Sebastian, Giovanni, and Cathie, thank you for shaping me as a scientist before graduate school. To memories of Katherine- when I was a wee little intern, you have taught me that the most fundamental part of being a good scientist, truly, is to be a good person. I still think of you when I struggle to abide by that. May you rest in peace.

To my parents, thank you for your unending support and instilling curiosity in me from an early age. To my brother Dodam, thank you for being someone I can always learn from. To Kate, thank you for bringing new warmth to my family; you are officially my favorite sister. To Midnight and Dawn, thank you for always being there.

To you, dear reader, for taking your time.

Thank you all.

Table of Contents

Preface	ii
Dedication	iii
Acknowledgements	iv
Table of Contents	vi
List of Tables	viii
List of Figures	ix
Chapter 1: Introduction	1
1.1 Why Observe Exoplanets and Their Atmospheres?	1
1.2 Observations of Transiting Exoplanet Atmospheres and Surfaces	3
1.2.1 Observing geometries and modes	3
1.3 Atmospheric Modelling	7
1.3.1 Physical structure of the atmosphere	8
1.3.2 Radiative-Convective Equilibrium	9
1.3.3 Bare rock planets	10
1.3.4 Accounting for Redistribution in a 1D Model	12
1.4 Atmospheric Retrieval and Bayesian Methods	13
1.4.1 Non-Bayesian Methods	13
1.4.2 Bayesian Methods	14
1.5 Formation and Evolution of Terrestrial Planet Atmospheres	15
Chapter 2: Understanding the Effects of Systematics in Exoplanetary Atmospheric Retrievals	17
2.1 Introduction	17
2.2 Methods	22
2.2.1 Framework for Statistics of Retrievals	22
2.2.2 Forward Model Parameters	23
2.2.3 Noise Model	28
2.2.4 Retrieval Setup	31
2.3 Results	32
2.3.1 Overfitting Due to Correlated Noise	32
2.3.2 Which Parameters are Affected?	36

2.3.3	Extensions to Other Planet Parameters	38
2.4	Can We Tell if Systematics are Present?	45
2.5	Discussion	52
2.5.1	Model Limitations	52
2.5.2	Instrumental Systematics	53
2.5.3	Data Outliers and Free Retrieval	56
2.6	Summary and Future Work	59
Chapter 3: On the Composition Parameterization Problem for High MMW Atmospheric Retrievals		60
3.1	Introduction	60
3.2	Methods	64
3.2.1	Parameterization methods	65
3.2.2	Model atmospheres	71
3.3	Results	74
3.3.1	Comparison of parameterization methods at low S/N	74
3.3.2	Comparison at higher precision	78
3.3.3	Atmospheres with aerosols	81
3.3.4	Comparison of Bayesian Evidences	82
3.4	What is an Uninformative Prior?	82
3.5	Summary and Future Work	90
Chapter 4: Constraining the Thickness of TRAPPIST-1 b’s Atmosphere from its JWST Secondary Eclipse Observation at 15 micron		92
4.1	Introduction	92
4.2	Methods	94
4.3	Results	98
4.3.1	Atmospheric Thickness and Surface Composition	98
4.3.2	Prospects for Future Observations	103
4.4	Discussion and Summary	105
Chapter 5: Conclusions and Future Work		108
5.1	Conclusions	108
5.2	Future Work	109
Appendix A: Regarding the Study of Correlated Noise		113
A.1	Retrieval Histograms for Various Planet Realizations	113
Appendix B: Facilities & Software used in this Thesis		118
B.1	Facilities	118
B.2	Software	118
Bibliography		120

List of Tables

2.1	Summary of Equilibrium Chemistry Retrievals and the Input Parameters and Priors Used for the Hot Jupiter cases	27
2.2	Summary of Equilibrium Chemistry Retrievals and the Input Parameters and Priors Used for the Warm Neptune Case	28
2.3	K-S Statistic of χ_r^2 and PIT Values. The K-S statistic, D , measures the maximum vertical discrepancy between the cumulative distributions (see Figure 7), of the goodness of fit, and the retrieval accuracy for each planet scenario. The first column shows the two-sample D between the distributions of fit χ_r^2 for the Gaussian and correlated noise. The next two columns contain the D between the expected uniform distribution and the PIT value distributions from retrievals with Gaussian and correlated noise, respectively. The final column shows the D between the two distributions. In all cases, the discrepancy is due to overfitting and worsening of retrievals.	41
3.1	Table summarizing the four cases of compositions and the forward model parameters used to generate the unpolluted spectrum. The mean molecular weight is calculated self-consistently from the abundances. We run two sets of each models, with and without clouds.	72

List of Figures

1.1	Confirmed planets to date plotted in the orbital period-radius space. The solar system planets are also indicated. This figure was created using the NASA Exoplanet Archive.	2
1.2	Observations in transit and eclipse geometry. Figure taken from [3].	4
1.3	The first <i>JWST</i> transmission spectrum of rocky planet, LHS 475, showing a “featureless” spectrum. Figure taken from Lustig-Yaeger et al. [4].	5
1.4	In the solar system, bodies with and without atmospheres are separated by the so-called “Cosmic Shoreline”, or a line in the escape speed-insolation plane [5]. This figure shows the Cosmic Shoreline but instead against cumulative XUV history, thought to be the driving force behind atmospheric evaporation. The triangles correspond to planned (colored) and potential (blank) targets observable with <i>JWST</i> and are scaled to the Transmission and Emission Spectroscopy Metrics [6] The further a target is to the bottom right, the less likely it is to have an atmosphere.	16
2.1	The observed WFC3 data of HD 97658b [7] and the best-fit PLATON spectrum in full-resolution and binned (solid line). The residuals from the fit are also shown in the bottom panel. The best-fit χ^2 is 56, and reduced $\chi_r^2 = 2.5$. An unusual upward trend in the residuals is present in the long wavelength limit, and the residuals appear to be correlated with wavelength.	19
2.2	Schematic diagram of our method. We generate multiple (~ 500) observational instances of a given planetary scenario and noise model, and perform atmospheric retrieval on each spectrum using PLATON.	22
2.3	Simulated unpolluted spectra of the hot Jupiter cases. Both full-resolution and binned depths are plotted. The offset case is identical to the clear case in the WFC3 band. The main effect of clouds is the truncation of the bottom-most depths compared to the clear case.	26
2.4	Same as Figure 2.3, but for the warm Neptune case.	26

2.5	Comparison of Gaussian (left column) and correlated (right column) noise models for the clear hot Jupiter case. In the top row of panels, 5 randomly selected spectrum realizations are plotted in color and the unpolluted spectrum is plotted in black. The disconnection in lines indicate discrete instruments. The <i>Spitzer</i> data are shown separately in the inset plot. The assumed 75-ppm error bar is shown for scale. In the bottom row of panels, the residuals relative to the unpolluted spectrum are shown. The effect of wavelength-correlation is discernible in the slightly redder noise (i.e. the residuals appear subtly sparser due to less zero-crossings as neighboring residuals are more likely to have the same sign) in the bottom right panel, but would not be discernible in the top right panel without prior knowledge of the ground-truth spectrum.	30
2.6	Two-dimensional distribution comparing retrievals with Gaussian noise (black) and correlated noise (blue). The three parameters shown are: the chi-squared between the unpolluted spectrum and the observation instance (True χ^2); the reduced chi-squared between the best-fit spectrum and the observation instance (Fit χ_r^2); and the PIT values for the retrieved posterior. The correlated noise allows for overfitting the spectrum, while simultaneously degrading the accuracy of the retrieval. We remind the reader that this corner plot is not showing a single retrieved posterior result, but a composite of multiple posteriors.	33
2.7	Same distribution as in the right panel of Figure 2.6, but shown as a cumulative distribution. The horizontal axis corresponds to the probability integral transform (PIT) values for each retrieval; the vertical axis corresponds to the cumulative fraction of all retrievals. The K-S metric measures the maximum vertical discrepancy between two cumulative distributions, such as above, and values of this metric are reported in Table 2.3. The expected uniform distribution (red), and our independent noise (black) and correlated noise (blue) simulation cases are shown.	35
2.8	The distribution of retrieved means per each parameter, for independent noise (black) and for correlated noise (blue), for the baseline hot Jupiter case. The input values used to generate the spectrum are shown in red.	38
2.9	The distribution of retrieved standard errors per each parameter, for independent noise (black) and for correlated noise (blue), for the baseline hot Jupiter case.	39
2.10	Summary of the goodness-of-fit and retrieval accuracy for the five planetary cases; black is for Gaussian noise, and blue is for correlated noise. The filled symbols mark the peak in both distributions, and the error bars denote the 1- σ spread. Values of χ_r^2 closer to unity imply a better fit. Similarly, lower PIT values correspond to a more accurate retrieval result. This plot shows the information in bottom middle panel of Figure 2.6 for all five cases. For the clear hot Jupiter with offset, the retrieval accuracy is measured as the error in an 11-parameter Gaussian distribution; the rest are quantified in 9 parameters. The correlated retrievals clearly have distinct distributions of goodness-of-fit and accuracy of retrieval, but have enough overlap such that one cannot discern whether a single retrieval instance has correlated noise.	40

2.11	Distributions of each parameter, normalized by the retrieved uncertainty, for the hot Jupiter retrievals. A value of 0 indicates that the retrieved mean coincides with the input value. The dashed lines show the high-precision (10 ppm) case, and the solid line shows the baseline (75 ppm) case. Only the Gaussian noise retrievals are shown here, but the correlated noise case displays similar behavior.	44
2.12	Same as Figure 2.7, but for the high-precision case. The smaller data error results in a better agreement between the ideal uniform distribution and the retrieval accuracy in the Gaussian noise retrievals.	45
2.13	Histograms showing marginalized posterior distribution for the correlation strength parameter for a hot Jupiter on a WFC3-like wavelength grid. The relevant input values for error multiple and correlation strength for each run is shown in legend. Two posteriors are shown for each: one where only correlation strength was retrieved (blue) and one where both error multiple and correlation strength were retrieved (orange). The corresponding marginalized posteriors for the error multiple are not shown. We note that our original model in Section 2.2 roughly corresponds to $\epsilon \sim 0.37$ (see text).	48
2.14	Same as Fig 2.13, but for NIRSPEC-like wavelength grid. The horizontal scale has been kept the same as Fig 2.13 and reflects the full width of the uniform prior used.	49
2.15	Bayes factors B_{10} of retrieval models that exclude error multiple (top) or correlation strength (bottom) for WFC3-like data. The model that includes both hyperparameters is used as the baseline. A positive value indicates that the model better fits the data while spanning a smaller prior volume and hence supports the removal of the parameter. The interpretation of the strength of evidence is shown in the colorbar. The Bayes factors are calculated as the difference in estimated Bayesian log-evidence; the values in the table have a resulting uncertainty of ~ 0.3	50
2.16	Same as Fig 2.15, but for NIRSPEC-like spectrum.	50
2.17	Two observation instances (blue and orange) of the ground-truth spectrum (black) for our baseline hot Jupiter case detailed in Table 2.1. The best-fit spectrum for each observation are also shown.	58
2.18	The retrieved posterior from the two input datasets from Figure 2.17. We only show the relevant parameters. The colors corresponding to each dataset are the same as in Figure 2.17. Black crosshairs indicate the ground-truth input values.	58
3.1	The marginalized priors probability distributions for different trace-filler parameterizations, shown in log scale (left) and in linear scale (right). Solid lines and dashed lines correspond to log and linear species, respectively; the filled histogram correspond to the filler species. The DF method “corrects” the prior of the filler such that it more resembles a uniform distribution, with a slight taper around unity due to the sum of log trace species.	67
3.2	The marginalized priors probability distributions for single Trace-Filler and the Centered Log-Ratio parameterization, shown in log scale (left) and in linear scale (right). Solid lines and dashed lines correspond to log and linear species, respectively; the filled histogram correspond to the filler species.	69

3.3	The implicit prior on mean molecular weight set by the different methods of parameterizing the composition. The peaks corresponds to specific gases, and the different methods vary on what gases are allowed to be major components. For the independent MMW retrieval, we set a uniform prior.	70
3.4	The forward model spectra corresponding to the 4 cases used. Spectra of clear and cloudy atmospheres are shown with solid and dashed lines, respectively. The simulated data error from 2 transit observations with NIRISS and NIRSpec is shown around a median featureless spectrum. The NIRSpec uncertainty is typically 40 ppm at $R = 100$; in other words, all 8 spectra are appreciably “flat lines”.	73
3.5	Comparison of the posteriors for the 4 cases using 4 parameterizations, for 2 transits each of NIRISS and NIRSPEC of GJ 1132 b. The vertical bars represent the ground truth values. The posterior for the MMW is inferred if not explicitly retrieved.	75
3.6	Same as Figure 3.5, but now showing linear abundance (from 0 to 1) to show the retrieved major component of the atmosphere.	76
3.7	Same as Figure 3.5, but now includes posteriors for 5 and 10 transits each of NIRSPEC and NIRISS. The posteriors are now shown in a box plot style as the median with their first and third quartiles, with the full range omitted. We note that the posteriors may not resemble Gaussian distributions and show quartiles rather than the 1-sigma uncertainties.	79
3.8	Same as Figure 3.7, but now in linear scale (from 0 to 1). The posteriors are now shown in a box plot style as the median with their first and third quartiles, with the full range omitted. We note that the posteriors may not resemble Gaussian distributions and show quartiles rather than the 1-sigma uncertainties.	80
3.9	Same as Figure 3.5, but now shows posteriors from atmospheres with opaque clouddecks. The posteriors are now shown in a box plot style as the median with their quartiles, with the full range omitted. We note that the posteriors may not resemble Gaussian distributions.	81
3.10	Violin plot comparing the Bayesian evidence ($\log \mathcal{Z}$) for retrievals for 20 noise instances. The connected lines indicate each noise instance. A higher $\log \mathcal{Z}$ indicates that the prior set by the parameterization weighs the peak likelihood higher. The values shown here have typical uncertainties of 0.3. No one parameterization method outperforms others across the board, as most of the evidence is contained in the likelihood peak.	83
3.11	Violin plot comparing the changes in Bayesian evidence ($\Delta \log \mathcal{Z}$) with and without CO_2 for retrievals for 20 noise instances. The connected lines indicate each noise instance. A smaller $\Delta \log \mathcal{Z}$ indicates that the inclusion of CO_2 is favored. The values shown here have typical uncertainties of 0.4.	84
3.12	The density distributions of spectra sampled from each parameterization. Each panel shows 3000 random instances of compositions drawn from the prior for a GJ 1132 b-like planet. In the distribution for CLR (bottom right), the three visually apparent modes correspond to H_2O -, CO_2 -, and CH_4 -dominated atmospheres.	86

3.13	CO ₂ feature strength as a function of the CO ₂ mixing ratio for a different mixtures of background gas. The feature strength is obtained by the ratio of the transit depth at 4.38 μm (in band) and 1.39 μm (out of band). To show the effects of changing opacity and mean molecular weight (MMW), the solid lines correspond to atmospheres in which MMW is calculated self-consistently from the mixing ratios, while the dashed lines correspond to atmospheres for which MMW was fixed to that calculated from the background gas, even at higher mixing ratios of CO ₂ . The small black vertical bar represents the uncertainty on the features strength obtained from the uncertainties on the transit depth for the simulated observations as described in 3.2.2.	88
4.1	The eclipse spectra of various models run in this study. We show: a suite of atmospheric models that are 1-σ consistent with the observation (<i>top left</i>); bare surface models, which are all consistent with the observation (<i>top right</i>); 100 % CO ₂ atmosphere models at various surface pressures (<i>bottom left</i>); and models with surface pressures of 0.1 bar, varying the compositions (<i>bottom right</i>). The compositions denote that the first species is the dominant species, with the second species in indicated trace amounts. The binned depths at F1500W and F1280W are shown as markers, as well as each bandpass function weighted by the stellar spectrum. We also show, in dashed lines, the eclipse depths resulting from blackbodies at 508 K (blue) and 400 K (red), corresponding to no redistribution ($f = 2/3$) and full redistribution ($f = 1/4$), respectively. On the upper right panel, dashed lines indicate grey albedo surface models. The features in the blackbody eclipse spectrum arise due to spectral features in the <i>stellar</i> spectrum.	97
4.2	The binned eclipse depths and their brightness temperature in the F1500W band for all of the atmospheric models run, varying the pressure of the atmosphere at the surface. Models atmospheres that do and do not include CO ₂ are shown in the left and the right panel, respectively. The measured eclipse depth from Greene et al. [2] is shown as the solid black line, and its 1-σ (grey) and 3-σ (red) uncertainties are also shown, as well as the corresponding brightness temperatures. The compositions denote that the first species is the dominant species, with the second species in indicated trace amounts. Atmospheres with ≥100 ppm CO ₂ are consistent with the measurement at 1σ only if the atmospheric pressure is less than 0.1 bar.	98

4.3	The temperature-pressure (T-P) profiles of the model atmospheres in radiative-convective equilibrium. Models atmospheres that do and do not include CO ₂ are shown in the left and the right panel, respectively, similarly to Figure 4.2. The optically thick region of the T-P profiles below the photosphere ($\tau = 2/3$) at $\lambda = 14.79 \mu\text{m}$ are shown with thick lines. The markers indicate the surface pressure of each model atmosphere. The compositions denote that the first species is the dominant species, with the second species in indicated trace amounts. The N ₂ and O ₂ -dominated atmospheres completely overlap in the left panel. It can be seen that while near-infrared absorbers such as H ₂ O can cause thermal inversions, they occur at regions where the atmosphere is optically thin and hence will not result in emission features in the spectra. For most of the models that do not contain CO ₂ , the atmosphere is optically thin in the F1500W bandpass down to the surface.	99
4.4	A color-color-like diagram of predicted binned eclipse depths in the F1280W band (horizontal axis) and the binned F1500W eclipse depths for all of the model atmospheres, along with their brightness temperatures (T_b) in each band. Models atmospheres that do and do not include CO ₂ are shown in the left and the right panel, respectively. The measured eclipse depth from Greene et al. [2] is shown as the solid black line, and its 1- σ (grey) and 3- σ (red) uncertainties are also shown. The vertical axis is identical to Figure 4.2, but is zoomed to focus on models consistent with the F1500W observation, alongside the expected F1280W uncertainty (~ 100 ppm) shown as an errorbar. The binned eclipse depths for a blackbody over a range of temperatures is shown as a multi-colored line. The temperature of the blackbody can be read off from the T_b in either axes, by definition. The corresponding T_b and bond albedo (A) at each confidence interval is also shown. All models that include CO ₂ (in the left panel) lie on the right side of the blackbody line, indicating a higher T_b in the F1280W than in F1500W due to the CO ₂ absorption at 15 micron. The compositions denote that the first species is the dominant species, with the second species in indicated trace amounts. As one follows each composition line, atmospheric pressure starts at 10^{-4} bar close to the observed F1500W measurement and increases in 1-dex intervals as in Figure 4.2 with generally decreasing 15- μm eclipse depths. We do not show the bare surface depths in this figure, but they lie close to the blackbody line and deviate less than 25 ppm in either bandpass.	102
5.1	Observed trend in normalized brightness temperatures of rocky planets observed so far, as a function of their irradiation temperature.	110
A.1	Same as Figure 2.8, but for the hot Jupiter case including offsets. This figure now includes panels for both instrumental offsets in addition to the original set of retrieved parameters.	114
A.2	Same as Figure 2.9, but for the hot Jupiter case including offsets. This figure now includes panels for both instrumental offsets in addition to the original set of retrieved parameters.	114
A.3	Same as Figure 2.8, but for the cloudy hot Jupiter case.	115
A.4	Same as Figure 2.9, but for the cloudy hot Jupiter case.	115

A.5	Same as Figure 2.8, but for the high-precision hot Jupiter case.	116
A.6	Same as Figure 2.9, but for the high-precision hot Jupiter case.	116
A.7	Same as Figure 2.8, but for the warm Neptune case.	117
A.8	Same as Figure 2.9, but for the warm Neptune case.	117

Chapter 1: Introduction

1.1 Why Observe Exoplanets and Their Atmospheres?

Hypothesizing alien worlds on planets orbiting stars other than our own Sun, known as extrasolar planets or exoplanets, has deeper roots than what may be apparent from how relatively young the field of exoplanetary science is. In one of the earliest examples, Democritus (c. 460 - c. 370 BCE) posited, “In some worlds there is no Sun and Moon, in others they are larger than in our world, and in others more numerous. In some parts there are more worlds, in others fewer (...); in some parts they are arising, in others failing. There are some worlds devoid of living creatures or plants or any moisture.” The actual *study* of exoplanets did not take off until observations much later, of course, when the first exoplanet around a pulsar was found in 1992 [8]; two years later, the first exoplanet around a main sequence star was discovered [9]. Since then, 5,632 exoplanets have been confirmed as of this writing according to the NASA Exoplanet Archive [10], spanning various sizes, mass, and orbital periods, as shown in Figure 1.1.

Of course, studying exoplanets extends beyond the mere cataloging of distant worlds. We also wish to *characterize* them; that is, identify the ongoing processes, composition, and structure of the atmosphere and the interior. In the bigger picture, there are a few related scientific goals here. First and foremost, we aim to understand where *we* come from. How do planets form, evolve, and possibly lead to the presence life? The diverse outcomes of planet formation

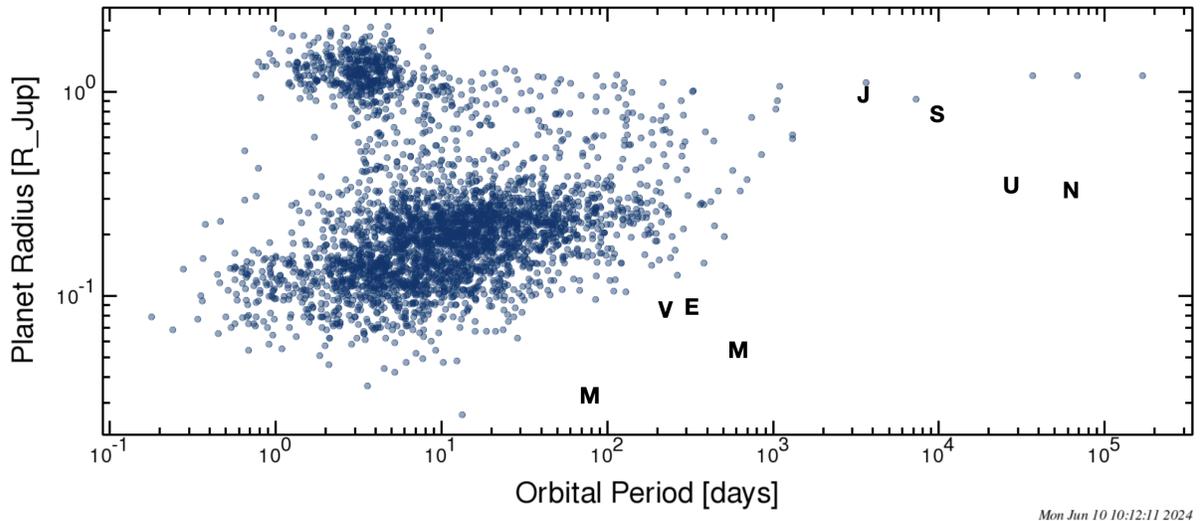


Figure 1.1: Confirmed planets to date plotted in the orbital period-radius space. The solar system planets are also indicated. This figure was created using the NASA Exoplanet Archive.

seen in exoplanets offer a broader sample size than the solar system (by a few order of magnitudes) for answering basic questions about planet formation (Figure 1.1). Secondly, we test our knowledge of planetary and atmospheric physics, utilizing the various extreme environments that exoplanets provide as scientific labs, especially those conditions that we cannot easily recreate experimentally. This includes, for instance, fluid dynamics [11], radiative transfer [12], molecular and atomic line lists [13, 14], and high pressure equations of states [15, 16]. Finally, we wish to study the diversity of worlds and habitability. By understanding what planetary conditions are possible, we can establish a context in which we can understand our own.

For this purpose, the atmospheres of exoplanets have been the primary target of observation beyond their bulk properties such as mass and radius. An atmosphere is the layer of gas that shroud a planet, and, being the outermost part of the planetary environment, is the most amenable to being observed via remote sensing [17, 18]. Atmospheric observations provide measurable information on the planet’s climate and chemistry, which can in turn inform the characterization

of e.g. the formation history [19], interior structure [20, 21], surfaces [22] and oceans [23], and biological processes [24]

1.2 Observations of Transiting Exoplanet Atmospheres and Surfaces

1.2.1 Observing geometries and modes

Transiting exoplanets are exoplanets whose orbital inclination is aligned with our line of sight such that they pass in front of and behind the host star. For these targets, the geometry allows for a number of differential measurement techniques to separate the planetary atmosphere from the stellar emission. One can observe the change in the observed flux at each wavelength as the planet passes in front of or behind the star, called transmission and emission observations, respectively.

1.2.1.1 Transmission spectroscopy

Transmission spectroscopy involves observing the starlight that passes through the planet's atmosphere during a transit. As the planet passes in front of its host star, it blocks the stellar flux equal to the projected area of the planet relative to the star, producing a drop in flux, called transit depth $\delta_\lambda = (R_{p,\lambda}/R_s)^2$. In practice, the stellar disk is not uniformly bright, and the transit depth is obtained via simultaneous modelling of the limb darkening and the transit in the light curve.

Information about the atmosphere is encoded in how this transit depth varies with wavelength; the starlight blocked by the planet filters through the planet's atmospheric layers, leaving imprints in the form of absorption lines. These spectral lines correspond to specific wavelengths where atmospheric gases absorb light; for instance, as water is opaque at 1.4 μm , the atmosphere

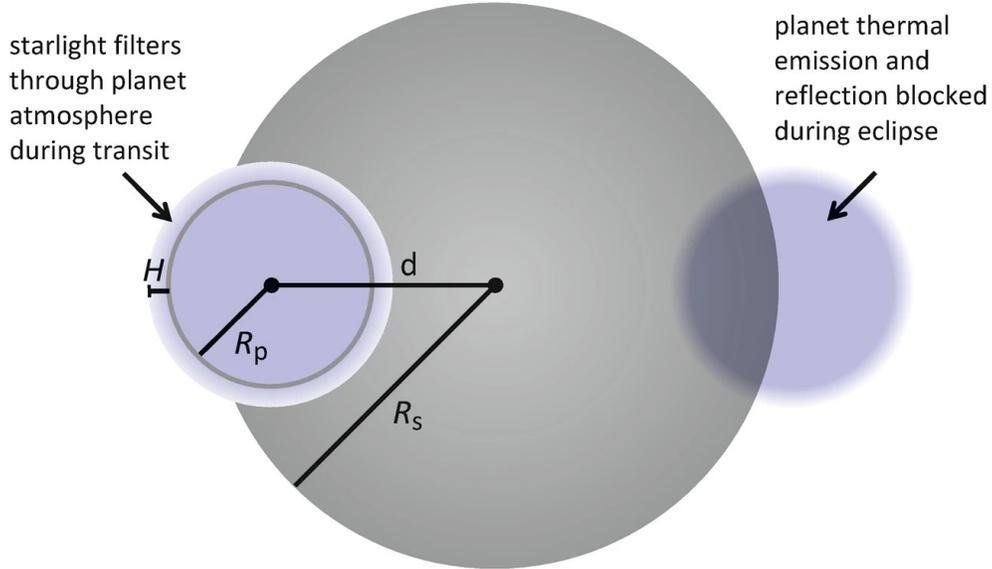


Figure 1.2: Observations in transit and eclipse geometry. Figure taken from [3].

will appear larger than compared to, say $1.2 \mu\text{m}$, outside the absorption band of water. If the height at which the atmosphere is opaque can be described as n times the atmospheric scale height $H = k_B T / \mu g$, or the height over which atmospheric pressure changes by an e-fold, the spectral feature size can be written as:

$$\delta_\lambda = \frac{(R_p + nH)^2}{R_s^2} - \frac{R_p^2}{R_s^2} = 2nR_p H / R_s^2. \quad (1.1)$$

Typically n is of order unity to ten.

It can be seen here that the spectral feature sizes will be larger for planets with: (a) larger planetary radius; (b) hotter temperature; (c) low mean molecular weight; (d) low mass (or gravity); and (e) small host star radius. Historically, this has naturally pushed characterization via transmission observations towards hot Jupiters, with nominal values (for say, HD 209458 b) being $\delta \sim 1\%$, and $\delta_\lambda \sim 0.1\%$ [25, 26]. For Earth-like planets, with smaller radius, cooler temperature, and higher mean molecular weight, both of these values could easily be 3 orders of magnitude

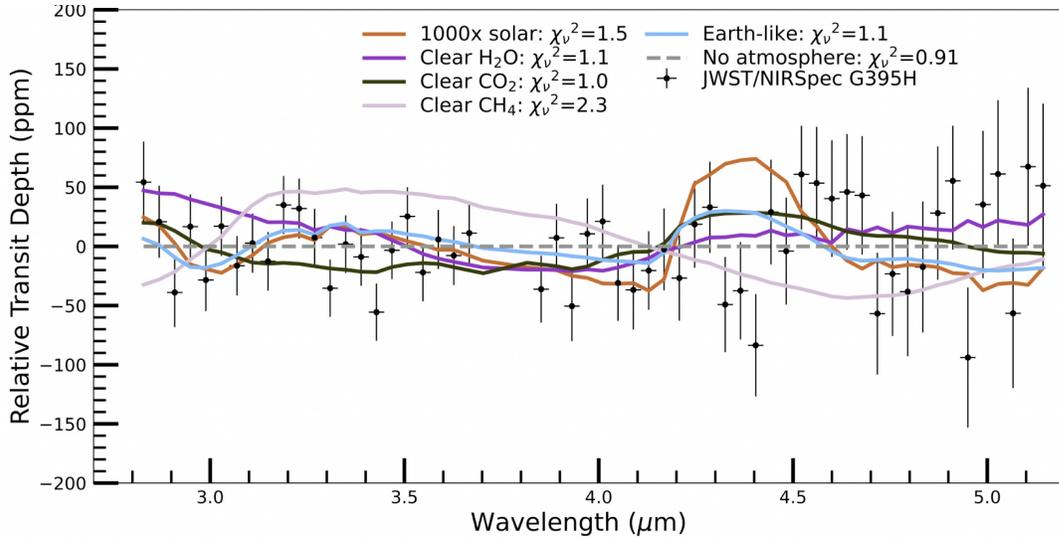


Figure 1.3: The first *JWST* transmission spectrum of rocky planet, LHS 475, showing a “featureless” spectrum. Figure taken from Lustig-Yaeger et al. [4].

smaller. This also pushes observations towards planets around smaller stars, as smaller R_s increases the spectral feature size. As such, despite the superior sensitivity of *JWST*, transmission spectra of small, rocky planets have frequently been observed to be “featureless” spectra, where the data is consistent with both a bare rock or a cloudy atmosphere (a flat line) or an atmospheric spectrum with small scale height (Figure 1.3).

1.2.1.2 Emission spectroscopy

Emission observations, on the other hand, disentangle the light emitted or reflected by the planet itself during secondary eclipses when the planet passes behind its star [25, 27]. Just before the secondary eclipse, the observed light from the system comprises the flux from both the host star and the emitted and reflected dayside of the planet. During the eclipse, the planet’s emitted and reflected light is blocked, thus allowing for a differential measurement. At longer wavelengths, where reflection contributes negligible flux, the thermal emission from the planet

produces eclipse depth F_p/F_s :

$$\frac{F_p}{F_s} = \frac{R_p^2 B_\lambda(T_{p,\lambda})}{R_s^2 B_\lambda(T_s)}, \quad (1.2)$$

where it is assumed that the planet and star emit as perfect blackbodies with the wavelength-dependent flux $B_\lambda(T)$.

Here, the information about the atmosphere is encoded in how the eclipse depth varies with wavelength. Due to the changing opacity with wavelength, the height of the photosphere of the atmosphere, where optical depth is of order unity, varies with wavelength. As such, a vertical temperature gradient in the atmosphere gives rise to spectral features. For a decreasing temperature with altitude, one sees a cooler part of the atmosphere at wavelengths where a gas absorbs, producing shallower eclipse depths at those wavelengths. For an increasing temperature with altitude, called thermal inversion, one sees a hotter part of the atmosphere, producing deeper eclipse depths [28, 29].

For transmission spectroscopy, the *absolute* depth is generally of less interest as planetary radius can be found without the use of a spectral observation. On the other hand, for eclipse observations, the absolute eclipse depth can be of interest as it can constrain the dayside heat budget of the planet. This allows for the technique of measuring the thickness of the atmosphere for rocky planets, as a thick atmosphere will cool the dayside via circulation, which I describe in detail in Chapter 4 [2, 30, 31, 32, 33].

1.3 Atmospheric Modelling

Exoplanetary atmospheres are always observed in an integrated sense, as opposed to solar system planets where remote sensing can resolve the disk of the planet or in situ measurements allow for stratified measurement in the local environment. As such, we must invoke fundamental physics to carefully model the planetary atmosphere to disentangle the various information in the observation. Additionally, a theoretical model spectrum is necessary to evaluate the feasibility of an observation and assess what information can be gained from which set of observations [34]. Towards both of these purposes, broadly speaking, modelling codes come in two flavors: *forward* models and *retrieval* models. I will illustrate the difference between the two by describing the two codes used in this dissertation.

In this dissertation, I use two planetary modelling codes, HELIOS [35, 36, 37] and PLATON [38, 39]. HELIOS is a self-consistent 1-D forward modelling code that iteratively solves for the thermal structure of a planetary atmosphere in radiative-convective equilibrium with its chemical composition; the thermal structure can then be processed to produce the emergent planetary emission spectra and hence an eclipse spectrum. The self-consistency is particularly important for modeling emission spectra, as the vertical temperature gradients produce the emergent spectral features.

On the other hand, PLATON is a retrieval code for transmission and eclipse observations, and, as such, the forward model therein does not enforce self-consistency and makes simplifying assumptions, such as a vertically fixed chemical abundances in the atmosphere. The simplifying assumptions are necessary to run the model a large number (10^{5-7}) of times, embedded in a parameter estimation scheme. As the transmission spectra are less sensitive to the vertical tem-

perature gradients, I use PLATON for generating transmission spectrum. A full description of the two codes can be found in the respective sources; I will outline the physics the two codes solve for in this section and unpack the jargon in this section.

1.3.1 Physical structure of the atmosphere

Both models are 1-D models and treat the atmosphere as vertically stratified slabs. The atmospheric modelling simultaneously solves for four variables that depend on each other: pressure, temperature, chemical composition, and heat transport. A reasonable assumption is to assume hydrostatic equilibrium, in which gravity is balanced by the pressure gradient:

$$\frac{dP}{dz} = -\rho(z)\frac{GM_p}{z^2}, \quad (1.3)$$

where the gravity varies with altitude while the enclosed mass is constant. This implicitly assumes the atmosphere contributes negligible mass.

Additionally, the models assume ideal gas as the equation-of-state describing the pressure:

$$P(z) = n(z)k_B T(z) = \frac{\rho(z)k_B T(z)}{\mu(z)m_H}, \quad (1.4)$$

where $\mu(z)$ describes the mean molecular weight (MMW).

Both HELIOS and PLATON allow for thermochemical equilibrium (or just chemical equilibrium in common parlance), in which the composition of the atmosphere, given an elemental composition, is determined entirely by the local pressure and temperature by minimizing the Gibbs free energy [40, 41]. The so-called non-equilibrium effects, such as vertical mixing or photochemistry, are not natively considered [42, 43]. However, most of the models in this disser-

tation concern small planets, for which equilibrium chemistry is not expected. For these planets, a common simplifying assumption is to assume a vertically fixed composition [44].

1.3.2 Radiative-Convective Equilibrium

As stated earlier, a key difference between the two codes is how $T(z)$ is treated. On one hand, PLATON, being a retrieval code, treats the thermal structure parametrically using analytical functions (or assumes a simple isotherm) [45, 46] and therefore does not enforce any energy balance, global or local. On the other hand, HELIOS iteratively solves for the thermal profile such that local energy balance is satisfied at each layer. That is, the divergence of the net bolometric flux is zero at every vertical layer, absent of any additional heating sources such as latent heat or horizontal transport:

$$\frac{\partial F_-}{\partial z} = \frac{\partial(F_{\uparrow} - F_{\downarrow})}{\partial z} = \rho(z)c_p \frac{dT(z)}{dt} = 0, \quad (1.5)$$

at each vertical layer, where the arrows indicate upwelling and downwelling stream fluxes and c_p is the specific heat capacity. Equilibrium is reached by timestepping through changes in temperature at each layer.

Once the radiative equilibrium has been reached, HELIOS then checks if each layer is stable against convection, using the Schwarzschild criterion [36, 47]:

$$\frac{dT}{dz} < \frac{g}{c_p}, \quad (1.6)$$

where the right hand term corresponds to the dry adiabatic lapse rate. Currently, the public version HELIOS does not account for moist adiabats [48, 49] or convective overshoot [50]. HELIOS makes

convective adjustment to satisfy convective equilibrium in the layers with superadiabatic lapse rate.

1.3.3 Bare rock planets

Here we define a number of temperatures, equilibrium temperature, surface temperature, and brightness temperature as they apply to observations of bare rock planets. For a bare rock planet, spectral flux at the surface, $F_{p,\lambda}$, is given by:

$$F_{p,\lambda} = F_{p,\lambda_{\text{SW}}} + F_{p,\lambda_{\text{LW}}} = A_{\lambda}F_{\downarrow,\lambda} + (1 - A_{\lambda})\pi B_{\lambda}(T_{\text{surf}}), \quad (1.7)$$

where the terms $F_{p,\lambda_{\text{SW}}}$ and $F_{p,\lambda_{\text{LW}}}$ are the shortwave reflection of the incident stellar flux $F_{\downarrow,\lambda}$ and the longwave outgoing planetary flux, respectively. Notice the emissivity term $(1 - A_{\lambda})$.

Surface temperature

The maximally hot dayside temperature is found by setting $A_{\lambda} = 0$:

$$T_{\text{max}} = \left(\frac{2}{3}\right)^{\frac{1}{4}} \left(\frac{R_{*}}{a}\right)^{\frac{1}{2}} T_{*\text{eff}}, \quad (1.8)$$

where the numerical factor of $\frac{2}{3}$ accounts for the spherical geometry and arises due to integration of cosines, and since $\mathcal{F}_{\text{p}} = \mathcal{F}_{\downarrow}$, the surface temperature becomes:

$$T_{\text{surf}} = \left(\frac{1 - A_{\text{SW}}}{1 - A_{\text{LW}}}\right)^{\frac{1}{4}} T_{\text{max}} \quad (1.9)$$

and $A_{\text{SW}} > A_{\text{LW}}$ cools the surface temperature.

Equilibrium temperature

The blackbody flux at the equilibrium temperature matches only the longwave planetary emission and not the bolometric (otherwise the albedo would just cancel out). This also implicitly assumes that the reflected and emitted spectrum don't overlap. T_{eq} is found by:

$$T_{\text{eq}} = (1 - A_{\text{sw}})^{\frac{1}{4}} \left(\frac{1}{4}\right)^{\frac{1}{4}} \left(\frac{R_*}{a}\right)^{\frac{1}{2}} T_{*\text{eff}}. \quad (1.10)$$

The equilibrium temperature only depends on the Bond albedo A_{sw} because by definition T_{eq} is the surface temperature when $A_{\text{LW}} = 0$. For a realistic surface with $A_{\text{LW}} \sim 0$, the surface temperature is similar to the equilibrium temperature, but see Appendix B of [51] for how non-unity emissivity affects estimating the Bond albedo A_{sw} .

Brightness temperature

The brightness temperature, $T_{\text{b},\lambda}$, is found by matching the blackbody at a given wavelength:

$$B_{\lambda}(T_{\text{b},\lambda}) = (1 - A_{\lambda})B_{\lambda}(T_{\text{surf}}), \quad (1.11)$$

for longwave emission. In general, even for a grey albedo $A_{\lambda} = A$, the brightness temperature varies with wavelength and is not equal to the equilibrium temperature, unless of course $A = 0$. In practice, this is a pretty minor effect [52].

If we go out to longer wavelengths, $T_{\text{b},\lambda} \rightarrow (1 - A)T_{\text{max}}$ for bare rock planets as we approach the Rayleigh-Jeans limit, but for MIRI observations of cold rocky planets the limit may

not always hold.

Additionally, filters are not infinitely narrow, so one has to actually integrate $P_\lambda B_\lambda d\lambda$ over the filter bandpass (weighted by the bandpass function P_λ) for a grid of temperatures to find the brightness temperature.

1.3.4 Accounting for Redistribution in a 1D Model

As HELIOS is a vertical 1-D model, it does not directly model the effects of horizontal atmospheric transport between the dayside and nightside. Instead, it employs a scaling developed by Koll [53] to calculate the dayside energy budget, where the redistribution factor f for a tidally locked (or slowly rotating) planet depends on the surface pressure P_{surf} , longwave optical depth τ_{LW} , and equilibrium temperature T_{eq} :

$$f = \frac{2}{3} - \frac{5}{12} \times \frac{\tau_{\text{LW}} \left(\frac{P_{\text{surf}}}{1 \text{ bar}}\right)^{2/3} \left(\frac{T_{\text{eq}}}{600 \text{ K}}\right)^{-4/3}}{k + \tau_{\text{LW}} \left(\frac{P_{\text{surf}}}{1 \text{ bar}}\right)^{2/3} \left(\frac{T_{\text{eq}}}{600 \text{ K}}\right)^{-4/3}}, \quad (1.12)$$

where k is a term that captures all planetary parameters other than the three variables and can be assumed to be ≈ 2 . The scaling follows the expectations that, in the uniform redistribution limit for thick atmospheres, as $P_{\text{surf}}, \tau_{\text{LW}} \rightarrow \infty$, $f \rightarrow \frac{1}{4}$, and, in the no redistribution limit for thin atmospheres, as $P_{\text{surf}}, \tau_{\text{LW}} \rightarrow 0$, $f \rightarrow \frac{2}{3}$.

In the current implementation of HELIOS, this scaling is employed by subtracting the approximated transported heat budget from the incident stellar flux to calculate the dayside energy budget, and therefore does not consider the vertical dependence of the day-to-night heat flow. Regardless, this scaling has been validated against global circulation models that consider full three-dimensional flow to be able to successfully reproduce observations [32, 52].

1.4 Atmospheric Retrieval and Bayesian Methods

Characterizing the atmosphere of an exoplanet from its observations constitutes an *inverse problem*; instead of solving for the atmosphere and its observation from physical properties, one aims to derive statistical constraints on the unknown physical properties from the given measurements via a forward model tied to robust parameter estimation methods. This is done via running a large number of forward models to simulate spectra in a multidimensional parameter search and comparing them to the observed measurements, thereby finding the set of parameters and their uncertainties [34, 46].

1.4.1 Non-Bayesian Methods

I will instead briefly describe *non*-Bayesian methods, and why one might use them as I do in Chapter 4. When there is insufficient information in the data to constrain the model parameters (either due to resolution or signal-to-noise), however, one typically has to resort to a simpler approach such as grid-fitting, where one runs a suite of self-consistent models on a grid of parameters of interest to find solutions that best fit the data. In this scheme, compared to a retrieval, one is making more physical assumptions about the atmosphere by imposing self-consistency and, generally, the uncertainties on physical parameters is typically not retrieved. This approach was popular, for instance, in characterizing exoplanet atmospheres from Spitzer observations and was used to infer e.g. thermal inversions in hot Jupiters [28, 54, 55]. The observation of smaller targets with *JWST* necessitates this approach once again.

1.4.2 Bayesian Methods

Bayesian approaches to inference make use of Bayes' theorem to estimate a set of parameters $\underline{\theta}$ and to select for models. Given a model M and a set of observations \underline{y} , Bayes' theorem states that:

$$\Pr(\underline{\theta}|\underline{y}, M) = \frac{\Pr(\underline{y}|\underline{\theta}, M)\Pr(\underline{\theta}|M)}{\Pr(\underline{y}|M)}, \quad (1.13)$$

from which we define the terms posterior probability $\mathcal{P}(\underline{\theta})$, likelihood $\mathcal{L}(\underline{\theta})$, prior probability $\pi(\underline{\theta})$, and evidence \mathcal{Z} , respectively, as:

$$\mathcal{P}(\underline{\theta}) = \frac{\mathcal{L}(\underline{\theta})\pi(\underline{\theta})}{\mathcal{Z}}. \quad (1.14)$$

Here, while the likelihood corresponds to the probability of the data given the model and the parameter, the posterior probability corresponds to the probability distribution of the parameters given the data and the model. The prior probability corresponds to the probability distribution of the parameters given the model, before the data has been observed. The prior could be construed as supplying additional information to shape the posterior that is based on the data.

The evidence is calculated via integrating the likelihood function over the prior space:

$$\mathcal{Z} = \int \mathcal{L}(\underline{\theta})\pi(\underline{\theta})d\underline{\theta}. \quad (1.15)$$

For posterior estimation, the evidence is only a normalizing term, ensuring the posterior sums to unity over the parameter space. The evidence can also be used for model selection prob-

lems, where the relative probability of two different models given the data, or the odds ratio, can be obtained via the evidence ratio. This is justified by:

$$\frac{\Pr(M_1|\underline{y})}{\Pr(M_0|\underline{y})} = \frac{\Pr(\underline{y}|M_1)\Pr(M_1)}{\Pr(\underline{y}|M_0)\Pr(M_0)} = \frac{\mathcal{Z}_1}{\mathcal{Z}_0}, \quad (1.16)$$

which makes the assumption that the prior relative probability of the models is unity.

Bayesian methods have become the standard in atmospheric retrievals, used to infer the required model complexity, e.g. whether a gas species should be considered or not [46, 56, 57].

1.5 Formation and Evolution of Terrestrial Planet Atmospheres

While gas giant planets retain the primordial, H-dominated atmospheres from formation, terrestrial planets are prone to losing their primordial atmospheres and forming secondary atmospheres, the composition of which can reflect the various pathways to forming these atmospheres.

Volatiles can be trapped during early formation or be delivered by planetary embryos and planetesimals during the late stages of planet formation [58]. These volatiles can then be outgassed from the interiors, forming secondary atmospheres [21, 59].

X-ray and ultraviolet (XUV) radiation from stars can drive atmospheric escape processes that significantly influence the volatile content of terrestrial planets, potentially leading to complete atmospheric loss in some cases [60, 61]. Nonetheless, substantial atmospheric escape can be counterbalanced by vigorous outgassing from the interior or from magnetic shielding [62], rendering it difficult to predict the atmospheric extent of terrestrial exoplanets.

As such, whether terrestrial exoplanets have atmospheres at all is still an open question, especially those around M stars, whose cumulative XUV irradiation is greater than around Sun-

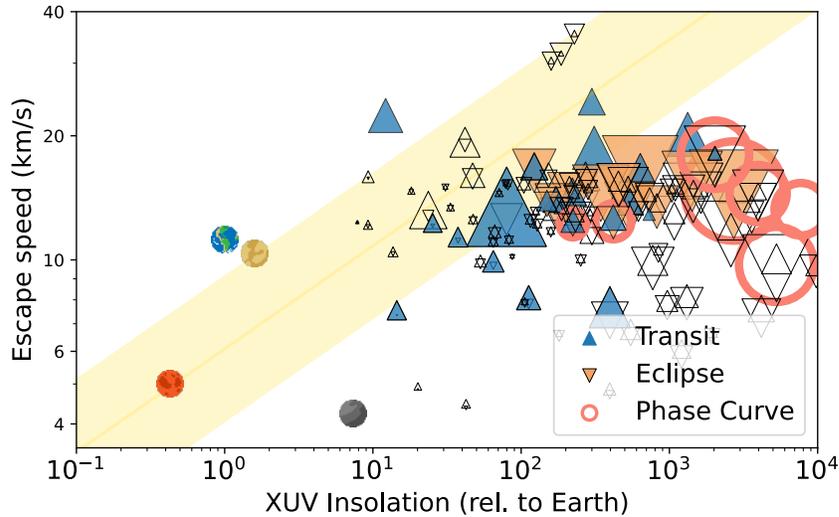


Figure 1.4: In the solar system, bodies with and without atmospheres are separated by the so-called “Cosmic Shoreline”, or a line in the escape speed-insolation plane [5]. This figure shows the Cosmic Shoreline but instead against cumulative XUV history, thought to be the driving force behind atmospheric evaporation. The triangles correspond to planned (colored) and potential (blank) targets observable with *JWST* and are scaled to the Transmission and Emission Spectroscopy Metrics [6]. The further a target is to the bottom right, the less likely it is to have an atmosphere.

like stars, due to more of their bolometric flux coming from XUV and being longer-lived [63]. For instance, Figure 1.4 shows targets that can be observed with *JWST* in the escape speed-XUV insolation plane. In the solar system, bodies with and without atmospheres are separated by the so-called “Cosmic Shoreline”, or a line in the escape speed-insolation plane [5]. All of the *JWST* targets are around M stars due to the favorable signal size, and it can be seen that, scaled relative to the solar system, they are all on the side of the Cosmic Shoreline that is unlikely to have an atmosphere. The lack of a clear picture underscores the need for ongoing and expanded observational studies.

Chapter 2: Understanding the Effects of Systematics in Exoplanetary Atmospheric Retrievals

2.1 Introduction

Inverse modelling the transmission spectra of exoplanets allows for extracting information about various properties and processes in the atmosphere. This is commonly done by pairing a forward model, which generates a spectrum from atmospheric parameters, and a parameter estimation scheme, which samples the parameter space to compute the probability distribution of the set of parameters. This method of analyzing observed spectra, called atmospheric retrieval, originally developed for remote sensing in Solar System bodies, [e.g. 64] has become a standard tool in characterizing exoplanetary atmospheres and has allowed for measuring abundances of various species and identifying atmospheric phenomena such as the presence of clouds and thermal inversions [e.g. 65, 66, 67, 68, 69, 70, 71].

Recently, there has been a growing body of work that addresses the potential to be misled by incomplete physics or simplifying assumptions used in retrievals, often invoked to speed up the computation and make the retrieval feasible [e.g. 72, 73, 74, 75, 76, 77]. These studies are especially germane in preparation for interpreting *James Webb Space Telescope (JWST)* observations, the precision of which will now render the finer details of the model consequential.

Such details include 3-dimensional atmospheric structure, host star effects, aerosol models, and disequilibrium chemistry. The general method used in the aforementioned works is to retrieve on a synthetic spectrum generated by a more complex and complete forward model and then inspect the retrieval result to quantify what bias may be incurred, with the aim of making judicious choices as to how and which model complexity and compromises should be introduced to the retrieval's forward model.

Adopting a similar approach, we focus on a separate but related issue in this work. A universal assumption made in atmospheric retrieval is that the reported errors in the observed spectrum are Gaussian and independent. This assumption is encoded into the cost function one tries to minimize during the parameter estimation, which is invariably a chi-squared statistic that does not take the covariance between the residuals into account [65, 69, 78]. While this assumption is a reasonable starting point for analyzing observations, as data quality reaches unprecedented precision and as retrievals incorporate increasingly sophisticated forward models and more rigorous statistical methods, it is necessary to understand the significance of the assumption of independent errors.

As further motivation for this work, there have been observations that hint at the extent to which errors may be correlated with wavelength. To pull one such example from the literature, we identify the observed spectrum of HD 97658b with the WFC3 instrument on *Hubble* [7]. The retrieval on this dataset strongly favors either high-metallicity or cloudy atmospheres, corresponding with a nearly featureless transmission spectrum (a flat line). However, none of the forward models considered by Guo et al. [7] provide an excellent fit to the data. For example, we show the best-fit PLATON [69] spectrum in Figure 2.1, which yields a reduced chi-squared of 2.5 (with 21 degrees-of-freedom) and is ruled out by the data at $4.9\text{-}\sigma$. (We note that Guo et al. [7]

can find models that provide a high-quality fit to their data by scaling their formal error bars by a factor > 1 — a practice that we weigh in on later in this chapter and that we do not endorse.) As can be seen in Figure 2.1, the best-fit spectrum produces residuals that are possibly correlated. One rudimentary method of quantifying correlation is to count the number of zero-crossings, which should follow a symmetric binomial distribution if the noise were independent. Additionally, an unusual upward slope is seen in the residual in the redmost edge. A similar behavior was seen in the observed spectrum of KELT-11b [see 79, Figure 20]. This could be attributed to either complicated physics unaccounted for in the forward model or the presence of correlated noise in the data.

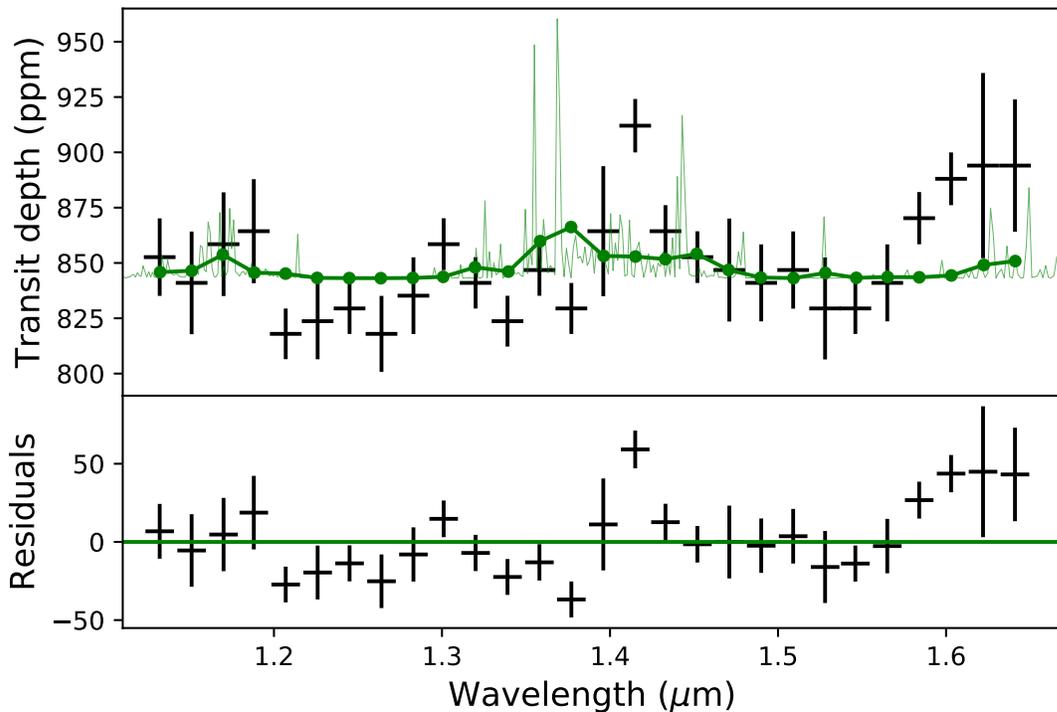


Figure 2.1: The observed WFC3 data of HD 97658b [7] and the best-fit PLATON spectrum in full-resolution and binned (solid line). The residuals from the fit are also shown in the bottom panel. The best-fit χ^2 is 56, and reduced $\chi_r^2 = 2.5$. An unusual upward trend in the residuals is present in the long wavelength limit, and the residuals appear to be correlated with wavelength.

Potential sources of correlated noise are numerous and include faulty data reduction due to

incorrect orbital parameters or incomplete subtraction of stellar contributions [80, 81]. Choices made during the data reduction, such as the choice of a model to remove visit-long trends, can potentially produce a wavelength-dependent correlated effect in the spectrum [see 7, Figure 7]. More fundamentally, the removal of instrumental systematics such as ramping or horizontal shifts in *Hubble* data has intrinsic uncertainties and may potentially manifest themselves in wavelength-dependent manners [82].

For space-based observation facilities, there are some reasons from observer experience to suspect that correlated noise is more likely in the case of a very bright host star, for which the instrumental systematics either behave differently or become more apparent due to lower photon noise. In Stevenson and Fowler [83, Figure 9], it can be seen that no observations with bright host stars of J-band magnitude ≤ 8.5 meet the ideal precision per orbit, instead maxing out at ~ 35 ppm. (Interestingly, our previously-discussed example case of HD 97658b fits into this category, with a J-magnitude of 6.2.) This effect has been attributed to unaccounted for wavelength-dependent systematics, that have no guaranteed way of being completely modelled out. In the case of ground-based observations, the highly time-dependent telluric contamination may also be a potential source of correlated systematics.

A separate, but related cause of wavelength-dependent correlation in data errors arises when combining data from various instruments to gain a wider wavelength coverage. Each instrument has its own instrumental systematics and data reduction pipeline, leading to distinct noise statistics among datasets. More fundamentally, these observations are not simultaneous and are hence subject to differing conditions with respect to stellar and planet variability. An insufficient number of observed transits may admit such variability in the data, even if the desired signal-to-noise ratio is formally achieved. Some retrieval analyses have included *ad hoc* “offset” parameters,

which vertically shift all measured transit depths from one dataset by a variable amount, to fit for the discrepancies between datasets, but doing so can induce bias in the estimation of other parameters [84].

Another issue arises in how outliers in the observed data are interpreted in a retrieval. After fitting a spectrum to data, the presence of anomalous outliers in the residuals is certainly within expectation of what can happen, and statistical methods such as bootstrapping, though rarely used in retrieval studies, do offer objective criteria to exclude these outliers. However, the fact that we rely on the best-fit spectrum to determine outliers raises the question if a statistically equivalent datum could have been accommodated for as a detected feature if it happened to have occurred at where we expect one. This problem is especially pertinent in the context of retrievals with non-equilibrium models. It particularly affects resolution-limited observations and species with single, narrow features, e.g. atomic lines such as Na or K for which only one or two data points dictate the retrieved abundance.

Given the number of potential issues raised above, in the present work we aim to address the question: how reliable are our atmospheric retrievals and what are best practices in the face of these idiosyncratic data systematics? We perform atmospheric retrievals on simulated spectra, while varying the noise properties, and conduct a detailed analysis of the retrieved parameters. Such an analysis provides a statistical context in which one can assess the credibility of a retrieval beyond a raw retrieved posterior. In what follows, in §2.2 we describe our planet parameters and noise models used; in §2.3, we present our findings in how retrievals are affected; in §2.4, we test whether correlation can be retrieved during retrieval; in §2.5, we discuss how we might be able to better understand the sources of these systematics and implications for future telescopes; in §2.6 we summarize and conclude.

2.2 Methods

2.2.1 Framework for Statistics of Retrievals

To study how non-Gaussian error can bias the retrieval results, we perform retrievals on simulated data generated with and without correlation in the noise. Here, by using the same forward model to create the synthetic spectrum and to retrieve on it, we remove model dependencies as much as possible and are able to examine the bias due to noise in isolation. To obtain statistics of retrievals, we use the following procedure (also shown schematically in Figure 2.2):

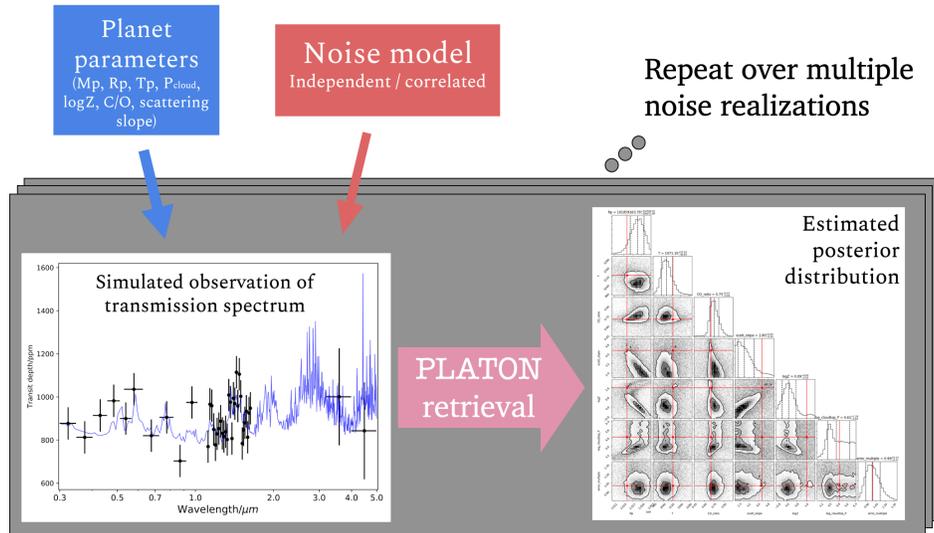


Figure 2.2: Schematic diagram of our method. We generate multiple (~ 500) observational instances of a given planetary scenario and noise model, and perform atmospheric retrieval on each spectrum using PLATON.

1. Choose the input parameters of a planet, such as radius, temperature, metallicity.
2. Run the forward PLATON [69] model to produce the unpolluted spectrum of the atmosphere.
3. Bin the full-resolution unpolluted spectrum to the chosen wavelength bins.

4. Choose a noise model (independent or correlated) and noise parameters that simulate the noise of a real observation.
5. Sample a noise realization of the binned spectrum using the noise model.
6. Perform retrieval analysis on the simulated data.
7. Repeat steps 5 and 6 a sufficient number of times such that the retrieved parameters can be combined to generate reliable statistics.

We use PLATON [version 3.0; 69], an open-source retrieval tool, as the forward and retrieval model for transmission spectra. PLATON has the advantage of being extremely fast for a retrieval code (under 30 minutes per run), which is suitable for our application as we perform hundreds of retrievals with randomly sampled noise realizations. We perform this process only on transmission spectroscopy, as the geometry allows for assuming an isothermal atmosphere, greatly reducing the number of free parameters in our model as well as the computation time per run.

We repeat the above procedure for five cases of observation: a clear hot Jupiter, a clear hot Jupiter with offsets between datasets, a cloudy hot Jupiter, a clear hot Jupiter at higher (*JWST*-like) precision, and a warm Neptune. In what follows, we describe the forward model parameters, the noise model, and the retrieval setup. A summary of the input parameters and the assumed priors for each set of retrievals is provided in Table 2.1.

2.2.2 Forward Model Parameters

To best imitate retrievals on actual observations, we choose input parameters and spectrum bins similar to *Hubble* and *Spitzer* observations of the canonical hot Jupiter HD 209458b and

exo-Neptune GJ436b which have reliable data and have been studied in the context of retrievals [85, 86, 87, 88, 89, 90]. We adopt the measured mass, radius, and temperature of each planet, and set the log-metallicity to 0.3 and carbon-to-oxygen ratio to the solar value of 0.53 [91].

We also choose to include clouds and hazes in our model. Such aerosols have been found to be ubiquitous in exoplanetary atmospheres [e.g. 92, 93] and produce a spectral signature that can be degenerate with other parameters [94, 95]. PLATON accounts for clouds and hazes using a parametric model. The user can specify a grey cloud-top pressure, the atmosphere absorption below which is truncated, and an amplitude and slope in the optical end of the spectrum to account for a non-Rayleigh slope caused by Mie scattering. In summary, the aerosol opacity κ_{aer} is given as:

$$P > P_{\text{cloud}} : \kappa_{aer} = \infty$$

$$P < P_{\text{cloud}} : \kappa_{aer} \propto a\lambda^{-\gamma},$$

where $a = 1$ and $\gamma = 4$ corresponds to Rayleigh scattering from the gaseous atmosphere.

For our cloud-free simulations, we choose a low-altitude cloud-top pressure of 0.5 bar. For our cloudy simulations, we choose 0.1 mbar such that clouds obscure roughly half of the spectrum while preserving some molecular features. Similarly, we adopt a nearly-Rayleigh slope of 4.3 and amplitude of 1, indicating no excess Rayleigh scattering from aerosol particles. We stress that since we use the same forward model in the retrieval to isolate the effects of noise from model systematics, the specific choices in parameters are not of great importance as long as they can be correctly retrieved and as long as we select a set of forward models that span a representative set of exoplanet atmospheres.

We choose a wavelength range that spans observations from the *Hubble Space Telescope* (*HST*) spectrographs most commonly used for exoplanet atmospheric investigations — specifically the Space Telescope Imaging Spectrograph (STIS) and Wide-Field Camera 3 (WFC3) — as well as photometric observation from the *Spitzer Space Telescope*. In the STIS wavelength range we follow the bins of Knutson et al. [87]; in the WFC3 wavelength we choose 33 equal sized bins between 1.01 μm and 1.64 μm [96]; for Spitzer we include observations in the photometric bands of the IRAC instrument at 3.6 μm and 4.5 μm . The resulting wavelength bins are comparable to a complete panchromatic dataset from current space-based observations. The resulting simulated spectra are shown in Figure 2.3 and Figure 2.4.

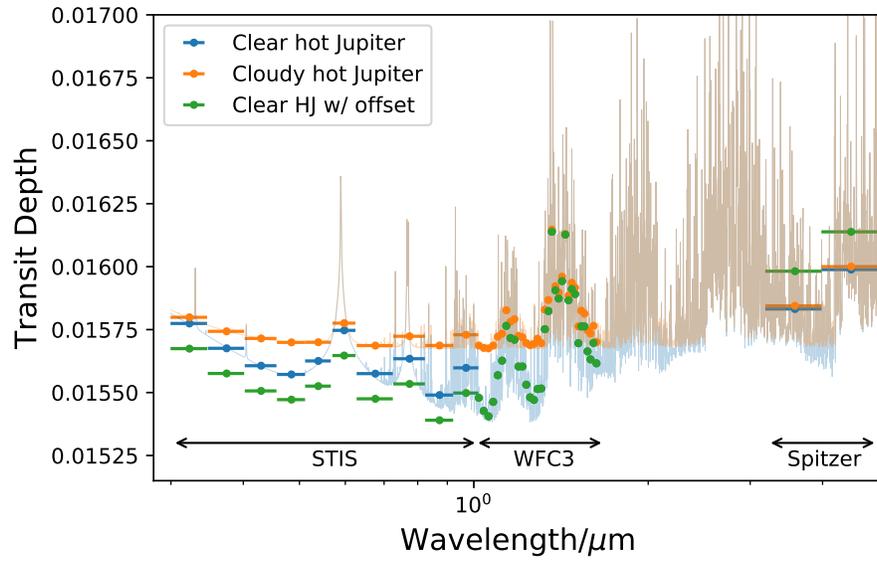


Figure 2.3: Simulated unpolluted spectra of the hot Jupiter cases. Both full-resolution and binned depths are plotted. The offset case is identical to the clear case in the WFC3 band. The main effect of clouds is the truncation of the bottom-most depths compared to the clear case.

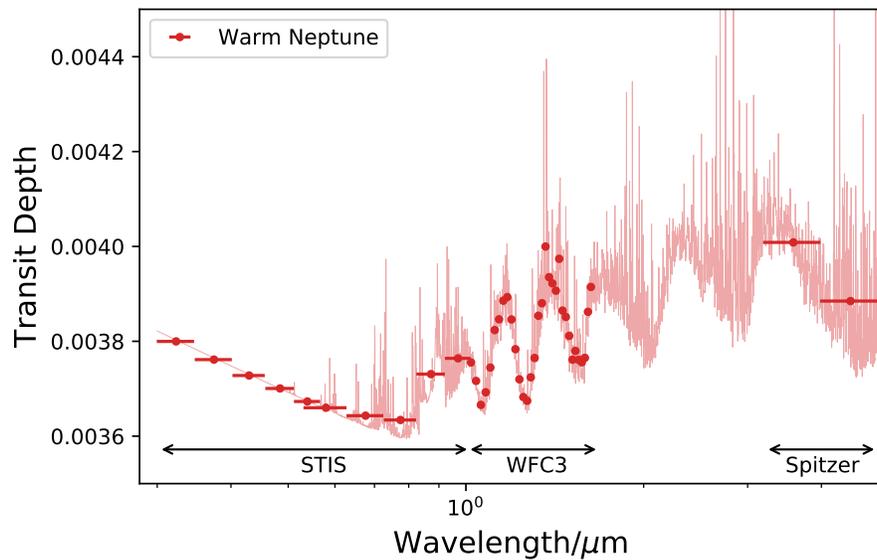


Figure 2.4: Same as Figure 2.3, but for the warm Neptune case.

Table 2.1: Summary of Equilibrium Chemistry Retrievals and the Input Parameters and Priors Used for the Hot Jupiter cases

Parameter		Clear HJ		Clear HJ with Offset	
Name		Truth Value	Prior	Truth Value	Prior
M_p	Planetary mass (M_J)	0.73	15%	0.73	15%
R_p	Planetary radius (R_J)	1.42	15%	1.42	15%
T	Temperature (K)	1130	[300, 1500]	1130	[300, 1500]
C/O	Carbon-to-oxygen ratio	0.53	[0.2, 2.0]	0.53	[0.2, 2.0]
log a	Log-scattering factor	0	[-0.3, 0.3]	0	[-0.3, 0.3]
γ	Scattering slope	4.3	[2.0, 5.5]	4.3	[2.0, 5.5]
log Z	Log-metallicity	0.3	[-1, 3]	0.3	[-1, 3]
log P_{cloud}	Cloud-top pressure (log(Pa))	4.7	[2.5, 6.5]	4.7	[2.5, 6.5]
Offset 1	STIS offset (ppm)	-100	[-300, 300]
Offset 2	Spitzer offset (ppm)	150	[-300, 300]
	Error multiple	...	Unity [0.5, 2.0]	Unity	[0.5, 2.0]
Other parameters					
R_s	Stellar radius (R_\odot)	1.19		1.19	
T_s	Stellar temperature (K)	6090		6090	
	Data error (ppm)	75		75	
	# of runs	440		660	
Parameter		Cloudy HJ		Clear HJ, High-precision	
Name		Truth Value	Prior	Truth Value	Prior
M_p	Planetary mass (M_J)	0.73	15%	0.73	15%
R_p	Planetary radius (R_J)	1.42	15%	1.42	15%
T	Temperature (K)	1130	[300, 1500]	1130	[300, 1500]
C/O	Carbon-to-oxygen ratio	0.53	[0.2, 2.0]	0.53	[0.2, 2.0]
log a	Log-scattering factor	0	[-3.0, 3.0]	0	[-0.3, 0.3]
γ	Scattering slope	4.3	[2.0, 5.5]	4.3	[2.0, 5.5]
log Z	Log-metallicity	0.3	[-1, 3]	0.3	[-1, 3]
log P_{cloud}	Cloud-top pressure (log(Pa))	3.0	[0.0, 6.5]	4.7	[2.5, 6.5]
Offset 1
Offset 2
	Error multiple	Unity	[0.5, 2.0]	Unity	[0.5, 2.0]
Other parameters					
R_s	Stellar radius (R_\odot)	1.19		1.19	
T_s	Stellar temperature (K)	6090		6090	
	Data error (ppm)	75		10	
	# of runs	660		400	

Table 2.2: Summary of Equilibrium Chemistry Retrievals and the Input Parameters and Priors Used for the Warm Neptune Case

Parameter		Warm Neptune	
Name		Truth Value	Prior
M_p	Planetary mass (M_J)	0.0736	15%
R_p	Planetary radius (R_J)	0.395	15%
T	Temperature (K)	700	[300, 1500]
C/O	Carbon-to-oxygen ratio	0.53	[0.2, 2.0]
log a	Log-scattering factor	0	[-2, 2]
γ	Scattering slope	4.3	[2.0, 5.5]
log Z	Log-metallicity	0.3	[-1, 3]
log P_{cloud}	Cloud-top pressure (log(Pa))	5.0	[2.5, 6.5]
Offset 1
Offset 2
	Error multiple	Unity	[0.5, 2.0]
Other parameters			
R_s	Stellar radius (R_\odot)	0.683	
T_s	Stellar temperature (K)	4780	
	Data error (ppm)	75	
	# of runs	400	

2.2.3 Noise Model

To simulate observed data, we sample multiple noise instances centered around the unpolluted spectrum, treating the simulated unpolluted spectra as a multivariate Gaussian distribution with the unpolluted depths as the mean and the reported error at each bin as the width. In addition, we adopt a non-diagonal covariance matrix with an exponential kernel to simulate correlated noise, such that the matrix element that correlates wavelength bin at λ_i and λ_j is given by:

$$K_{ij} = \epsilon_{ij} \sigma_i \sigma_j \exp\left(-\frac{|\lambda_i - \lambda_j|}{l}\right), \quad (2.1)$$

where σ_i is the reported error at wavelength λ_i , and ϵ_{ij} is 1 for pair of points from the same dataset and 0 otherwise. We choose the scaling factor l to be the distance to the neighboring bin. We select this covariance matrix in particular because it allows for the best-fit spectrum to

the WFC3 observations of HD 97658b in Guo et al. [7] to yield a reduced chi-squared of ~ 1 , as opposed to 2.5 when the errors are construed to be independent and Gaussian. We choose a random noise error of 75 ppm for all instruments, which represents a moderate quality data for STIS and WFC3 but is better than average for typical *Spitzer* observation [93, 97]. We also assume a uniform transmission across the wavelength range of each *Spitzer* filter in both the forward and retrieval models. In practice, we find that the two broadband *Spitzer* points provide little constraint on the retrieved parameters, and using the same error for all instruments does not give undue importance to the *Spitzer* points. For the high-precision hot Jupiter case, we use 10 ppm errors to match the best of current data quality [79].

We show a few randomly selected noise realizations in Figure 2.5 for the Gaussian and correlated noise models. It is discernible from the bottom row of panels that the correlated noise has slightly redder residuals compared to the Gaussian noise; that is, there are less zero crossings as neighboring points are correlated. We also stress that overall this is a rather subtle effect; without knowing the unpolluted depths *a priori* to produce the residuals, from the spectrum alone it is hardly obvious that there is a distinction between the two.

Additionally, to examine the effects of including the offset between datasets as a retrieved parameter, we create spectra with and without a fixed offset between datasets. Namely, we add a vertical offset of -100 ppm to the STIS points and $+150$ ppm to *Spitzer* points, holding the WFC3 points constant. The specific amount of offset is a somewhat arbitrary choice. The relevant heuristic is that the offset should be exactly retrieved in the absence of degeneracy.

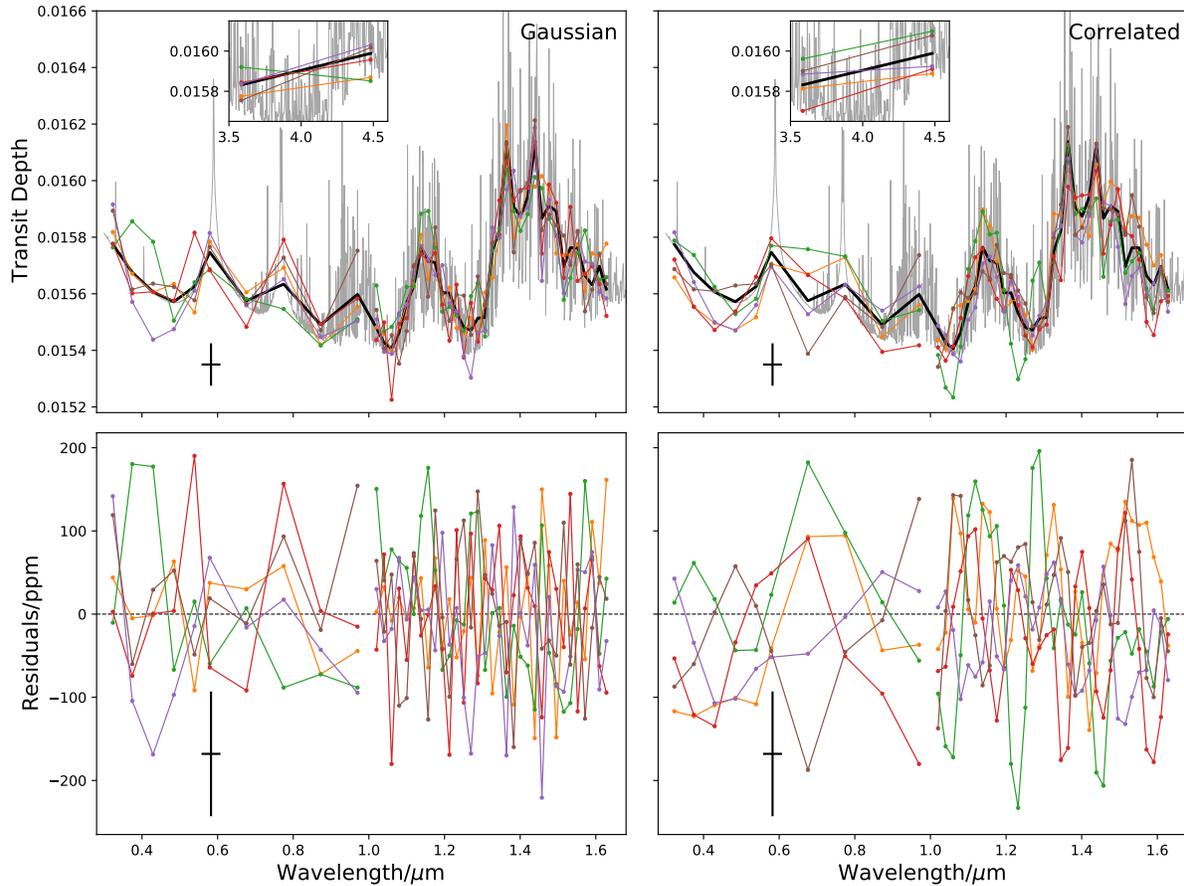


Figure 2.5: Comparison of Gaussian (left column) and correlated (right column) noise models for the clear hot Jupiter case. In the top row of panels, 5 randomly selected spectrum realizations are plotted in color and the unpolluted spectrum is plotted in black. The disconnection in lines indicate discrete instruments. The *Spitzer* data are shown separately in the inset plot. The assumed 75-ppm error bar is shown for scale. In the bottom row of panels, the residuals relative to the unpolluted spectrum are shown. The effect of wavelength-correlation is discernible in the slightly redder noise (i.e. the residuals appear subtly sparser due to less zero-crossings as neighboring residuals are more likely to have the same sign) in the bottom right panel, but would not be discernible in the top right panel without prior knowledge of the ground-truth spectrum.

2.2.4 Retrieval Setup

Using PLATON, setting up the retrieval involves choosing the priors and the statistical sampling method to be used. We choose the priors to be comparable to a real retrieval analysis. PLATON accepts either a Gaussian prior or a uniform prior in a user-specified interval. We set Gaussian priors for planet radius and mass, as these are often constrained via other methods such as radial velocity and transit measurements prior to observing the transmission spectrum. The Gaussian prior is centered around the input value and with a standard error of 15%. This precision is comparable to or slightly overestimates the typical uncertainty in measurement of mass via the RV method and provides sufficient tolerance for the retrieved value to deviate from the input value, if necessary. For the instrumental offsets, to ensure that the prior is broad enough, we choose a uniform prior offset to be 2- and 3-folds of the offsets. For all other parameters we opt for uniform priors that are as wide and uninformative as possible and adopt the full parameter range supported.

For now, we only choose to do retrievals with equilibrium chemistry, where the composition of the atmosphere at a given temperature and pressure is dictated by the the global elemental abundances set by metallicity and carbon-to-oxygen ratio. While disequilibrium chemistry is indeed expected for planets below $T_{\text{eq}} \leq 1200\text{K}$, most of its effects take place below the altitude that typically probed by transmission spectroscopy and have no easily discernible effect on the spectrum at the data precision of current instruments. The actual discrepancy in the relevant pressure range (~ 1 mbar) is smaller than the uncertainties we can current obtain [e.g. 98, 99, 100].

PLATON supports either Markov chain Monte Carlo (MCMC) [101] or nested sampling

methods for the posterior estimation [102]. We note that, while both are statistically robust and widely used, we observe a minor discrepancy in the resulting posteriors between the two methods, in which the posteriors estimated by MCMC tend to be slightly broader than those by nested sampling. This does not pose a major issue for this work inasmuch as we are concerned with biases due to data idiosyncrasies, and we consistently use one algorithm across our analyses. Nevertheless, we draw attention to this point as it requires extra scrutiny when comparing retrieval results. We choose nested sampling as it is known to perform better in estimating multi-modal or oddly-shaped distributions.

2.3 Results

In this section we first present the overall effects of correlated systematics on our retrievals, using the clear hot Jupiter case as a baseline. We then examine which parameters in particular are affected. We finally show how the results for the baseline case also extend to the other retrieval cases, and point out additional effects that arise.

2.3.1 Overfitting Due to Correlated Noise

Here we present the effects of correlation in data on the retrieval overall. To do this, we must reduce a retrieval result into a simpler metric. In retrievals on actual data this is commonly done by using the reduced chi-squared between the observed data and the best-fit or median spectrum. Here, as we begin from a known simulated ground truth, we can also compare the retrieved posterior directly to the input values to measure the accuracy of retrieval by using a probability integral transform (PIT), which is the cumulative distribution function evaluated at the

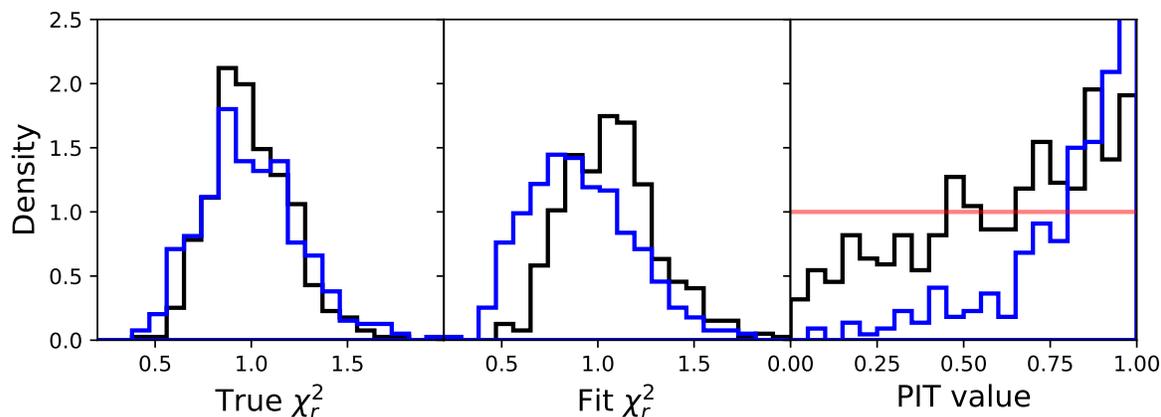


Figure 2.6: Two-dimensional distribution comparing retrievals with Gaussian noise (black) and correlated noise (blue). The three parameters shown are: the chi-squared between the unpolluted spectrum and the observation instance (True χ^2); the reduced chi-squared between the best-fit spectrum and the observation instance (Fit χ_r^2); and the PIT values for the retrieved posterior. The correlated noise allows for overfitting the spectrum, while simultaneously degrading the accuracy of the retrieval. We remind the reader that this corner plot is not showing a single retrieved posterior result, but a composite of multiple posteriors.

input value. To do this, in each retrieved posterior distribution, we draw an iso-likelihood contour of the input parameters and sum the relative weights contained within, producing a confidence interval between 0 (the input was the most likely sample) and 1 (the least likely). The distribution of PIT values should follow a uniform $\mathcal{U}(0, 1)$ were the retrievals accurate.

Our main finding is that, on average, correlation in the data allows for overfitting the spectrum, thereby weakening the overall accuracy of the retrieval. In other words, the best-fit spectrum is more likely to achieve a reduced chi-squared lower than unity in the presence of correlated noise, whilst simultaneously the retrieved posterior distribution rules out the input with higher significance.

In Figure 2.6, we present the comparison between Gaussian and correlated noise for the clear hot Jupiter case. For both retrievals with Gaussian (black) and correlated noise (blue), we show the distributions of: χ^2 between the unpolluted spectrum and the data instances; the reduced

chi-squared between the best-fit spectrum and the data instances; and the PIT values showing the accuracy of retrieval, as described above. A few observations can be made:

- The fit χ_r^2 , or the **goodness-of-fit** of the retrieval, on synthetic data is on average skewed to better than unity in the presence of correlated noise. In other words, it is more likely that the retrieval will overfit the data with forward models.
- The accuracy of the retrieval, shown as the PIT value, on the other hand, is *worse* in the presence of correlated noise. We also show the cumulative distribution in Figure 2.7 to demonstrate the worsening of the accuracy. The K-S statistic between the two cases is 0.19.
- For the normal noise case, even in the absence of the correlated noise, the retrieval accuracy is close to but slightly worse than the expected uniform distribution. This minor discrepancy is likely due to the fact that the retrieved posterior distribution is already non-Gaussian for a few parameters (discussed in Subsection 2.3.2), as well as to degeneracy between certain parameters.
- While the goodness-of-fit is on average better in the presence of correlated noise, it is not the case that the distributions of fit χ_r^2 are so discrepant that one can deduce the presence of correlated noise from the goodness-of-fit alone. That is, a given overfit spectrum can plausibly be construed as either a consequence of correlated noise or as an unlucky instance of Gaussian noise that happens to lie at the tail of the χ_r^2 distribution. As such, we stress that the effect of correlated noise is manifest statistically, and no individual value of χ_r^2 , good or bad, is uniquely diagnostic of correlated noise in a single retrieval instance.

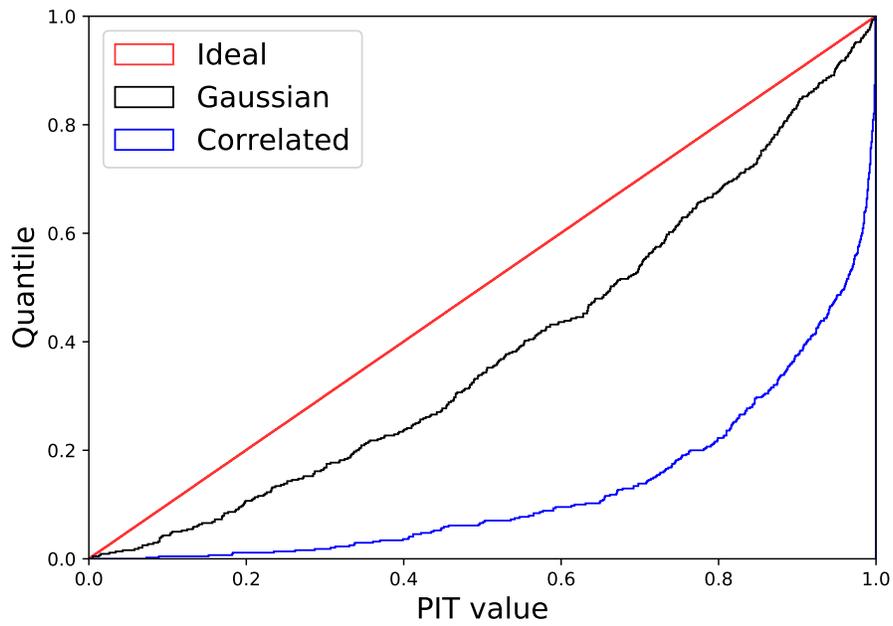


Figure 2.7: Same distribution as in the right panel of Figure 2.6, but shown as a cumulative distribution. The horizontal axis corresponds to the probability integral transform (PIT) values for each retrieval; the vertical axis corresponds to the cumulative fraction of all retrievals. The K-S metric measures the maximum vertical discrepancy between two cumulative distributions, such as above, and values of this metric are reported in Table 2.3. The expected uniform distribution (red), and our independent noise (black) and correlated noise (blue) simulation cases are shown.

2.3.2 Which Parameters are Affected?

Given that the correlated noise degrades the overall accuracy of retrievals, it is necessary to then look at which parameters are affected. We do this by marginalizing the posterior distribution over each parameter, as is typically done in retrieval analyses. The effect on each marginalized distribution can be twofold — the mean can be biased, the estimated error can be affected, or both. Either a shift in mean away from the input parameter or an underestimation of the error can worsen the accuracy of retrieval. As such, we examine the retrieved mean and the retrieved error separately for each parameter.

The distribution of retrieved means is shown in Figure 2.8. For the case in which noise is independent (black), the retrieved means form clean normal distributions around the input values (red) for most parameters, as expected from the central limit theorem. The two exceptions are *C/O* ratio and the cloudtop pressure. This is most likely due to the fact that the retrieved distributions for these parameters are not Gaussian in the first place. For cloudtop pressure, the retrieved distribution is at best a flat distribution with a lower bound, ruling out a cloudy atmosphere as per the clear atmosphere in the input used. For *C/O* ratio, we suspect that the distribution is skewed due to the increasing influence the parameter has over its range. That is, as one sweeps through *C/O* ratio, the spectrum changes more rapidly over the range above the solar value of 0.53, and thus the means are naturally skewed to values lower than the solar value where there is a greater density of near-consistent solutions.

The presence of correlated noise has a few interesting effects on the retrieved means. First off, the error multiple parameter is biased to less than unity. This intuitively follows from the global result that correlated noise allows for overfitting the spectrum, tricking the error multi-

ple parameter to believe that the error bars are overestimated. This means that the error multiple parameter is more likely to behave pathologically in a situation where one may expect it to be useful, such as if the reported error bars truly were underestimated due to unknown and unaccounted systematics. The retrieval instead selects a less-than-unity value of the error multiple, incorrectly implying that the data precision is better than initially reported. This is possible if the domain of input parameters and the forward model can still reproduce the spectrum polluted with systematics.

In the correlated noise case, the retrieved means generally show a wider distribution to varying degrees for each parameter. Specifically, the radius, mass, and temperature are the most affected, while the effect is the least pronounced for metallicity and cloudtop pressure. This result may be explained by considering the wavelength-scale the former three parameters have on the spectrum. Mass and temperature affect the scale height of the atmosphere, which affects the overall vertical extent of the transmission spectrum. Radius affects the baseline transit depth as well as the scale height. These are “global” parameters in the sense the transit depths in all bins are affected together. As such, a wavelength-dependent correlation can bias these parameters. On the other hand, metallicity, while it also affects the scale height (via the mean molecular weight), directly controls the individual transit depths. This has a more local effect in that it changes the actual shape of the spectrum.

The distribution of retrieved errors is shown in Figure 2.9. The effect of correlated noise is clearly visible for all parameters in that the retrieved error bars show a tendency to be underestimated. For instance, the retrieved error on log-metallicity is on average underestimated by ~ 0.2 dex. While this disparity is smaller still than typical constraints, it is worth bearing in mind as this is a statistical effect; the spread over the retrieved error is by itself broad enough that the

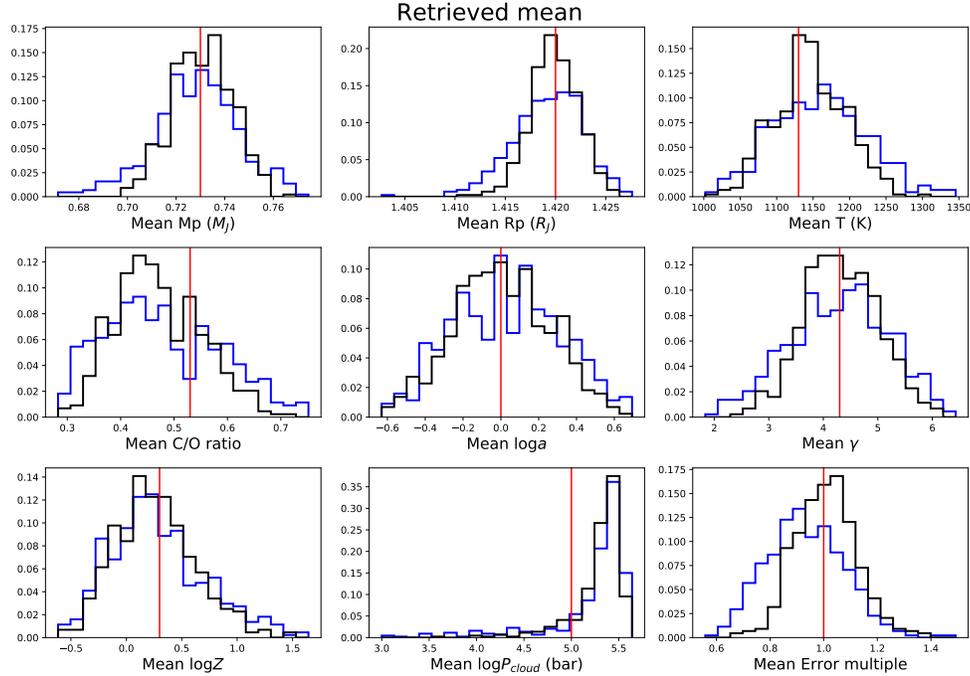


Figure 2.8: The distribution of retrieved means per each parameter, for independent noise (black) and for correlated noise (blue), for the baseline hot Jupiter case. The input values used to generate the spectrum are shown in red.

actual effect of a given instance can be much larger than this value. Additionally, when *JWST* allows for precise measurement of metallicity, this level of uncertainty may not be negligible when one considers analyzing archival data simultaneously. The same consideration applies to other parameters. As such, in this context we suggest that the retrieved constraints for parameters, in the face of the potential for correlated noise, are best understood as lower limits.

2.3.3 Extensions to Other Planet Parameters

In this section, we present our results for planet scenarios other than the baseline clear hot Jupiter case to understand the sensitivities of our results to various system and dataset parameters. We show histograms of the retrieved mean and retrieved standard error for each parameter in Appendix for the remaining planetary scenarios (the hot Jupiter with offsets, cloudy hot Jupiter,

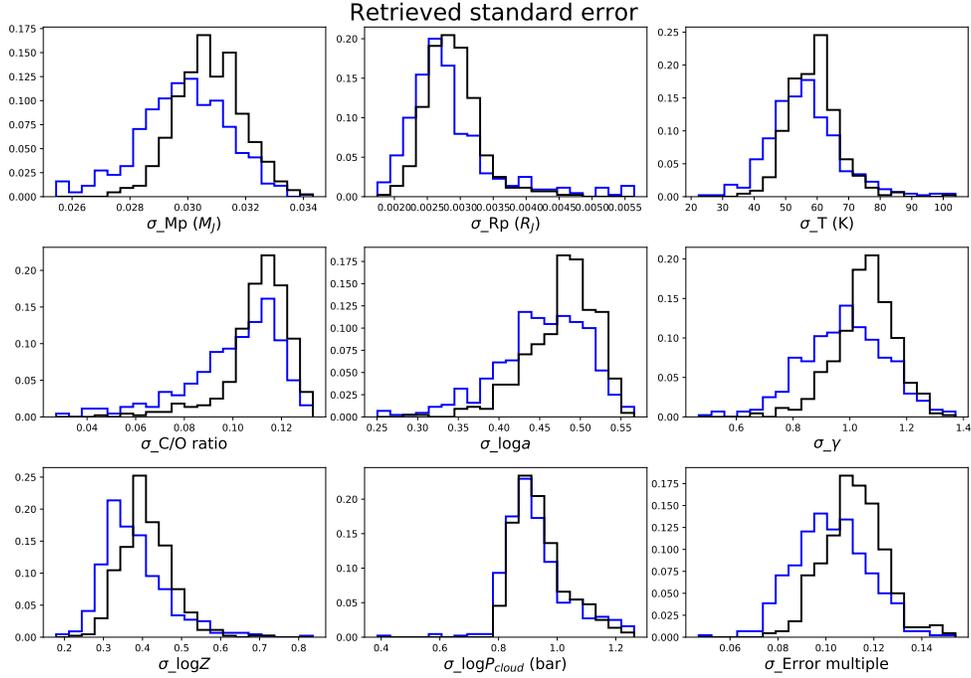


Figure 2.9: The distribution of retrieved standard errors per each parameter, for independent noise (black) and for correlated noise (blue), for the baseline hot Jupiter case.

high precision hot Jupiter, and warm Neptune). We generally find that the main results stated so far hold true for all cases: correlated noise causes both overfitting in χ_r^2 and worsening of the accuracy-of-retrieval (i.e. larger PIT values). This point is summarized in Figure 2.10, in which we show the medians of the χ_r^2 and PIT value distributions for each planet realization, i.e. distilling down the results of Figure 2.6 and the like to values quantifying the peak and the spread.

To further quantify this point, we perform a Kolmogorov-Smirnov (K-S) test for the goodness-of-fit and the accuracy-of-retrieval metrics to measure the discrepancy between the results of Gaussian and correlated noise. In Table 2.3, we show the K-S statistics, D , for the fit χ_r^2 and PIT value representing the accuracy-of-retrieval. We find that the clear hot Jupiter happens to be the best-case for the smallest discrepancy of accuracy-of-retrieval between Gaussian and correlated noise, and that other cases generally result in further discrepancy between results with Gaussian

and correlated noise.

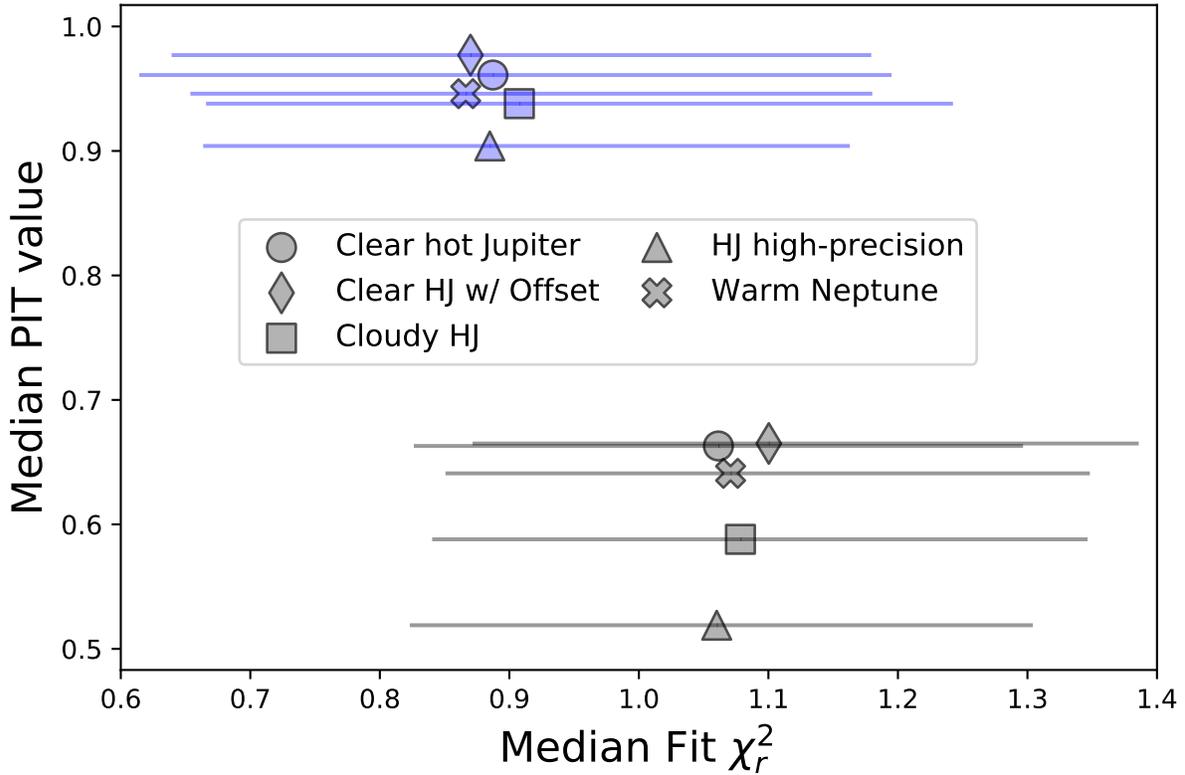


Figure 2.10: Summary of the goodness-of-fit and retrieval accuracy for the five planetary cases; black is for Gaussian noise, and blue is for correlated noise. The filled symbols mark the peak in both distributions, and the error bars denote the 1- σ spread. Values of χ_r^2 closer to unity imply a better fit. Similarly, lower PIT values correspond to a more accurate retrieval result. This plot shows the information in bottom middle panel of Figure 2.6 for all five cases. For the clear hot Jupiter with offset, the retrieval accuracy is measured as the error in an 11-parameter Gaussian distribution; the rest are quantified in 9 parameters. The correlated retrievals clearly have distinct distributions of goodness-of-fit and accuracy of retrieval, but have enough overlap such that one cannot discern whether a single retrieval instance has correlated noise.

2.3.3.1 Bias in Non-Rayleigh Scattering Slope

In the retrievals of the hot Jupiter with instrumental offsets and the warm Neptune, we find that the retrieved haze properties also show the potential to be biased. Correlated noise can bias the scattering slope, γ , away from the Rayleigh value of 4, misleading the retrieval to infer

Case	χ_r^2	PIT value		
	G-C	U-G	U-C	G-C
Clear hot Jupiter	0.28	0.19	0.60	0.47
Clear hot Jupiter w/offset	0.36	0.19	0.67	0.58
Cloudy hot Jupiter	0.27	0.12	0.54	0.48
Clear hot Jupiter, high precision	0.28	0.04	0.48	0.47
Warm Neptune	0.35	0.16	0.58	0.52

Table 2.3: K-S Statistic of χ_r^2 and PIT Values. The K-S statistic, D , measures the maximum vertical discrepancy between the cumulative distributions (see Figure 7), of the goodness of fit, and the retrieval accuracy for each planet scenario. The first column shows the two-sample D between the distributions of fit χ_r^2 for the Gaussian and correlated noise. The next two columns contain the D between the expected uniform distribution and the PIT value distributions from retrievals with Gaussian and correlated noise, respectively. The final column shows the D between the two distributions. In all cases, the discrepancy is due to overfitting and worsening of retrievals.

the presence of aerosols. This bias makes intuitive sense as, if a handful of points in the optical wavelengths align due to correlated noise, those points can mimic the behavior of non-Rayleigh slope [103]. As such, we caution that a spurious detection of haze can be possible in interpreting data in which the presence of correlated noise is either expected or suspected.

We suspect that this bias happens more readily for the warm Neptune case compared to the hot Jupiter retrievals because the overall signal is smaller while the data error used to scramble the spectrum was held constant at 75 ppm, resulting in a larger relative error. The warm Neptune spectra consequently have greater potential for large (apparent) optical slopes to manifest.

2.3.3.2 Retrieving Offsets

We ran a set of retrievals that includes non-zero “offset” parameters between datasets from different instruments. We find that, while the presence of correlated noise does cause underestimation of the uncertainty in the offset in an identical manner to other parameters, it does not

worsen the retrieval of the means. The offsets are accurately retrieved in both Gaussian and correlated noise retrievals and do not pose any obvious degeneracies.

This is somewhat surprising as, in our formulation of correlated noise, offsets can be regarded as correlated noise with high correlation and long wavelength-order. For instance, in Figure 2.5, the data instances with correlated noise in the *Spitzer* band mimics the presence of an offset.

It should be obvious that the influence of offset data points will strongly depend on the specific wavelength those points occupy, as well as the offset magnitude and sign. As such, we present here only one possible manifestation of how real data could behave. For instance, we have only considered offsets between datasets disjointed in wavelength, but, say, merging data from ground-based and space-based observations can produce offset data with overlapping wavelength coverage. Hou Yip et al. [84] found that if there is overlapping data with non-zero offsets and if free retrievals are used, such offsets can be degenerate with the estimated abundances if equilibrium chemistry is not assumed.

2.3.3.3 Effects of Clouds

The main effect of adding gray clouds to the model, from the point of view of the retrieval, is washing out information contained in the spectrum that originates from the high-pressure portion of the atmosphere. In Figure 2.3, roughly half of the *Hubble* points are covered by clouds, no longer constraining, say, a baseline radius or metallicity. We find that the broad effect of underestimating uncertainty and biasing means due to correlated noise still holds for cloudy hot Jupiter retrievals.

For the retrieved cloud-top pressure parameter, the main effect is a bias in the retrieved mean. Specifically, the presence of correlated noise disrupts the distribution of retrieved means of cloud-top pressure by extending the tail in the high-pressure direction. In other words, the spectrum is more likely to be understood as having a clear atmosphere. Upon examining the spectra for the retrievals that populate this tail, we find that the correlated noise happens to manifest as a number of data points dipping under the opaque cloud-top where the atmosphere is normally optically thick, thereby mimicking the behavior of a clear atmosphere.

2.3.3.4 Effects of Higher Precision

We find that in the hot Jupiter retrieval with high-precision (10 ppm) data, the broad conclusions again still hold. Correlated noise leads to an underestimation of retrieved uncertainty for all parameters. Compared to other cases however, correlated noise does not shift the retrieved means as much, which is to be expected since every noise instance only has a minor deviation from the unpolluted spectrum (specifically a factor of $7.5\times$ smaller than in our baseline case), even with correlation.

Comparing the high-precision case to the baseline case with 75 ppm errors, we find that, naturally, both the estimated means are retrieved closer to the input values and the retrieved parameter uncertainties are concurrently smaller. Interestingly, the uncertainties shrink more than the means approach the input values; consequently, in the high-precision case, the retrieval more readily rules out the input. This is shown in Figure 2.11, in which the retrieved means are normalized by their retrieved uncertainty to show the number of standard errors the input value is retrieved within for each parameter. The high-precision case (dashed line) actually has more re-

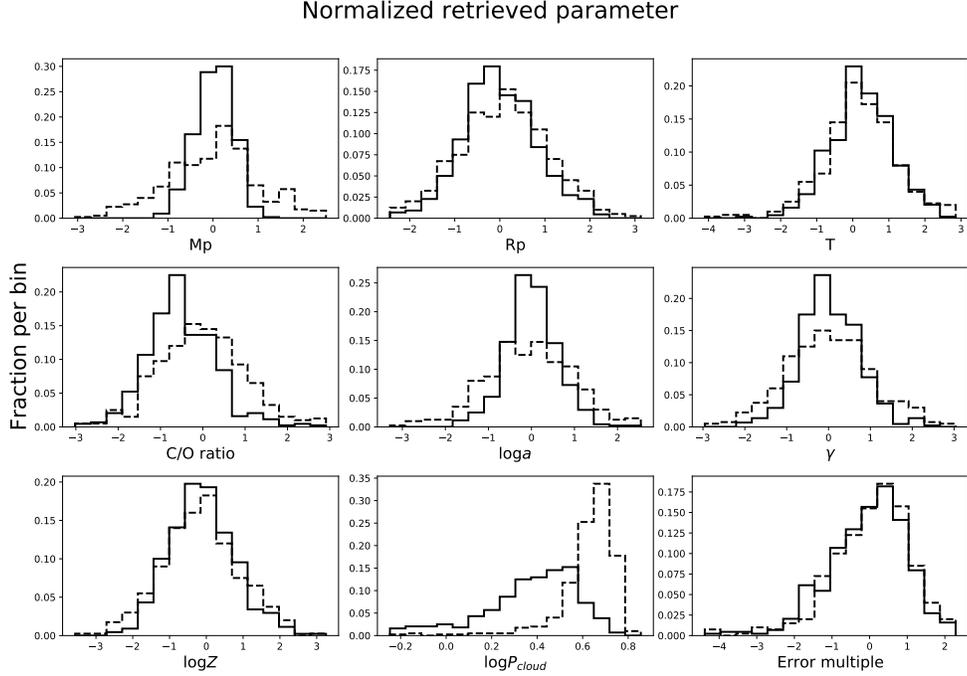


Figure 2.11: Distributions of each parameter, normalized by the retrieved uncertainty, for the hot Jupiter retrievals. A value of 0 indicates that the retrieved mean coincides with the input value. The dashed lines show the high-precision (10 ppm) case, and the solid line shows the baseline (75 ppm) case. Only the Gaussian noise retrievals are shown here, but the correlated noise case displays similar behavior.

retrievals *further* from the input value when normalized. We find an identical trend for the retrievals with correlated noise.

Additionally, we find the PIT value distribution for retrievals with Gaussian noise much more successfully follows the ideally expected uniform distribution, with a K-S statistic of $D = 0.04$ (see Table 2.3 and Figure 2.12). Given that the accuracy-of-retrieval for all other cases (with 75 ppm error bars instead) are at least somewhat discrepant from the expected distribution even for the Gaussian noise retrievals, this result suggests that, even for hot Jupiters, 75 ppm error bars are too large to assume *a priori* that the retrieved posterior will follow a multivariate Gaussian. This has implications for parameter estimation methods that need this assumption of Gaussian posterior, such as optimal estimation or some of recent machine learning-based retrievals [65,

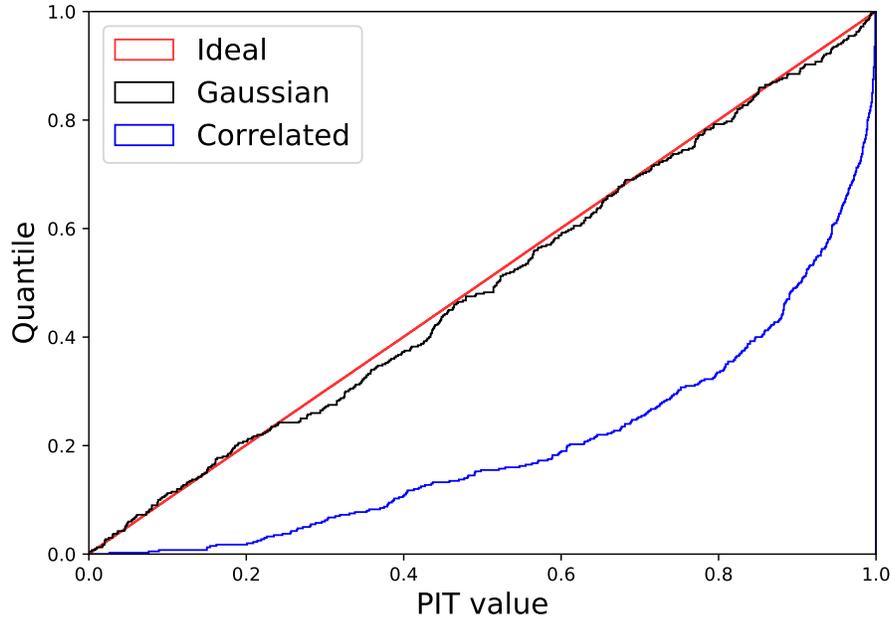


Figure 2.12: Same as Figure 2.7, but for the high-precision case. The smaller data error results in a better agreement between the ideal uniform distribution and the retrieval accuracy in the Gaussian noise retrievals.

104]. These methods require that the data uncertainty is small enough such that the forward model behaves linearly over the parameter uncertainties. Retrievals using these methods must be trusted only when the data has exceptional SNR.

2.4 Can We Tell if Systematics are Present?

In the motivating example spectrum of HD 97658b in Section 2.1, the presence of correlated noise was suspected based on the fact that no forward model can produce a satisfactory fit under the assumption of randomly scattered residuals. If we are to presume that the retrieval is indeed correct, and the residuals evidence correlated noise contaminating a genuine featureless spectrum, then we should also consider how prevalent unnoticed correlated noise can be in the observed data of other planets. The natural question then is to ask whether there is a more

robust and comprehensive way of distinguishing correlated noise within the framework of a retrieval. Especially, given that correlated noise can give rise to overfitting, it is of special interest whether correlated noise can be distinguished from merely overestimated errorbars. A natural way to achieve this is to modify the likelihood function such that it can reward or penalize when residuals are correlated.

To test this, we implement a parameterized covariance matrix and let the retrieval estimate the hyperparameter that measures the correlation strength. We use a nearest-neighbor correlated noise model as in Sivia and Skilling [105], where the correlation strength is parameterized by ϵ , such that the covariance matrix element between points at i and j is given by:

$$K_{ij} = \sigma_i \sigma_j \epsilon^{|i-j|} \quad (2.2)$$

This differs slightly from the correlated noise model used in Section 2.2 in that this model does not depend on the wavelength difference between two points but depends instead on the difference in indices. The two implementations would be identical if the wavelength grid was regularly spaced, with the exponential base giving correlation strength $\epsilon = e^{-1} \sim 0.37$. While extension to accommodate wavelength-dependent correlation is certainly possible, as a first test this simplification provides a reasonable starting point for exploring whether correlated noise can be retrieved.

This simplification allows for writing the likelihood function as:

$$\ln \mathcal{L} = -\frac{1}{2} \left[(N-1) \ln(1-\epsilon^2) + \sum_{i=1}^N \ln 2\pi\sigma_i^2 + \frac{Q}{1-\epsilon^2} \right], \quad (2.3)$$

in which Q is the modified chi-squared-like term related to the error-scaled residuals R_i by:

$$Q = (1 + \epsilon^2) \sum_{i=1}^N R_i^2 - \epsilon^2 (R_1^2 + R_N^2) - 2\epsilon \sum_{i=1}^{N-1} R_i R_{i+1}. \quad (2.4)$$

We perform a small grid of retrievals to study when the correlated noise can be distinguished from overestimated errorbars and correctly retrieved. The input parameters for the planet remain the same as Case 1 in Section 2.2, while we vary the noise properties. Our grid consists of three values of error multiple – $\eta = 0.8, 1, 1.25$, three values of correlation strength – $\epsilon = 0, 0.25, 0.5$, and two wavelength grids modeling HST WFC3 and JWST NIRSPEC prism grids. Our NIRSPEC grid consists of 133 linearly spaced points between 0.6 and 5 microns. From a few initial tests we find that the varying the absolute size of the error bar does not change the results, and hence we keep them fixed at 75 ppm. For each combination of noise parameters, we run three retrievals: one in which both error multiple and correlation strength were included as retrieved parameters, and two in which either was removed. We set a uniform prior between 0 and 1 for the correlation strength, as we do not expect an anti-correlation between neighboring points.

We show the marginalized posterior distribution for correlation strength in Figures 2.13 and 2.14 for the WFC3 and NIRSPEC grids, respectively. From the posterior distributions it is clear that the correlation strength can be adequately retrieved, and the posterior width is dependent on the number of points in the data, as expected. It can also be seen that underestimated error bars when unaccounted for can be mistaken for the presence of correlated noise, and vice versa, as demonstrated in the previous section. The main difference between the two instruments from the point of view of the retrieval is simply the number of points.

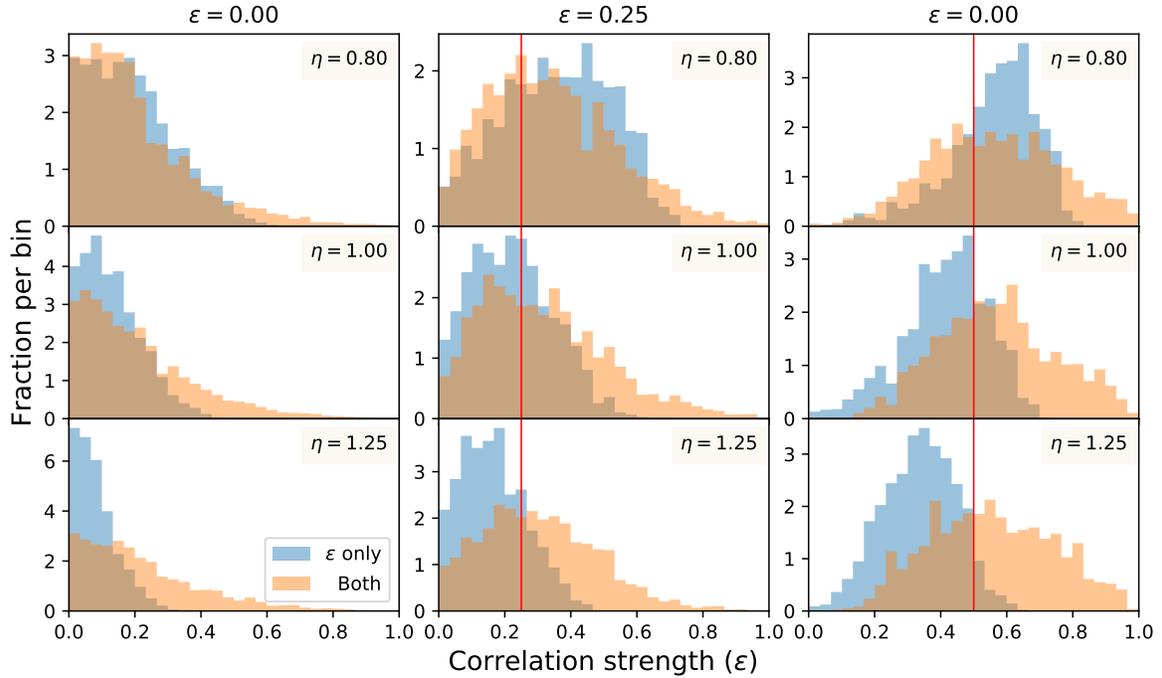


Figure 2.13: Histograms showing marginalized posterior distribution for the correlation strength parameter for a hot Jupiter on a WFC3-like wavelength grid. The relevant input values for error multiple and correlation strength for each run is shown in legend. Two posteriors are shown for each: one where only correlation strength was retrieved (blue) and one where both error multiple and correlation strength were retrieved (orange). The corresponding marginalized posteriors for the error multiple are not shown. We note that our original model in Section 2.2 roughly corresponds to $\epsilon \sim 0.37$ (see text).

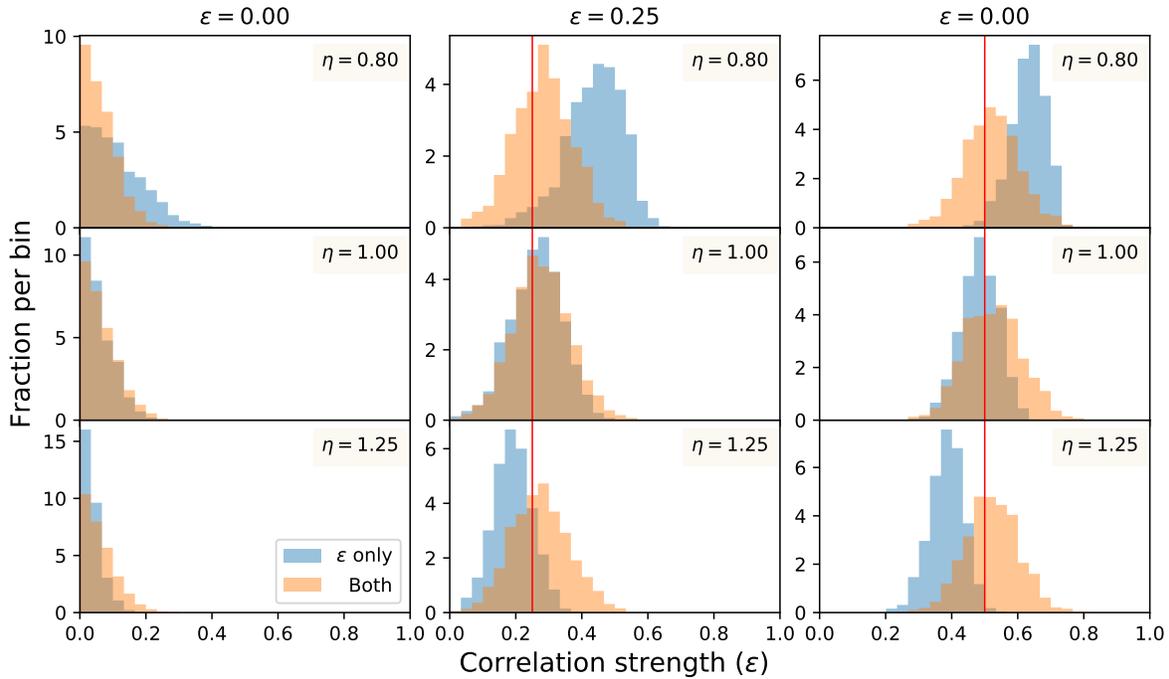


Figure 2.14: Same as Fig 2.13, but for NIRSPEC-like wavelength grid. The horizontal scale has been kept the same as Fig 2.13 and reflects the full width of the uniform prior used.

Additionally, we calculate the Bayes factor among retrievals to determine whether the inclusion of each parameter is warranted. The results for WFC3- and JWST-like data is shown in Figure 2.15 and 2.16. The case in which both parameters were included is used as the baseline, and the log-ratio (in base 10) of Bayesian evidence is shown for each case in the grid. A positive value indicates that the model better fits the data while spanning a smaller prior volume, indicating that the parameter should be removed.

For WFC3-like data (Figure 2.15), the inclusion of correlation strength is supported with strong evidence only in the strong correlation ($\epsilon = 0.5$) cases and is otherwise not easily ruled out either way. Interestingly, the inclusion of error multiple is disfavored with substantial or stronger evidence not only when the errorbars are correct, but also even in the case of underestimated errorbars. This shows that, for WFC3-like data, the error multiple parameter is not warranted in general.

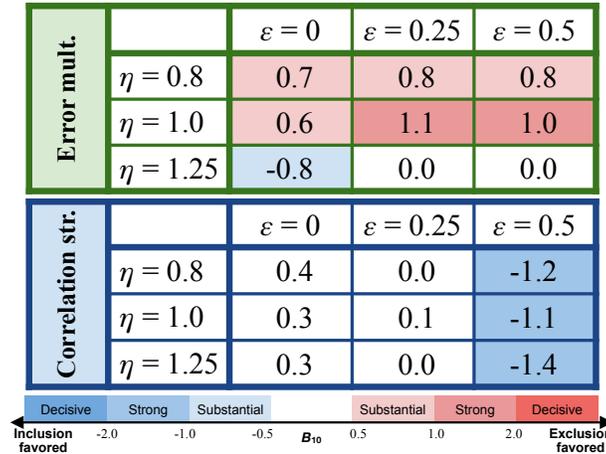


Figure 2.15: Bayes factors B_{10} of retrieval models that exclude error multiple (top) or correlation strength (bottom) for WFC3-like data. The model that includes both hyperparameters is used as the baseline. A positive value indicates that the model better fits the data while spanning a smaller prior volume and hence supports the removal of the parameter. The interpretation of the strength of evidence is shown in the colorbar. The Bayes factors are calculated as the difference in estimated Bayesian log-evidence; the values in the table have a resulting uncertainty of ~ 0.3 .

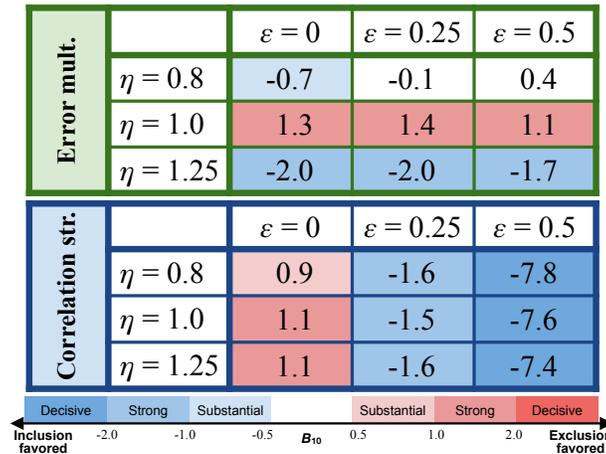


Figure 2.16: Same as Fig 2.15, but for NIRSPEC-like spectrum.

For NIRSPEC-like data (Figure 2.16), the inclusion of the correlation strength can be more robustly judged. Its inclusion is supported with strong to decisive evidence when correlation is present. Conversely, its removal is supported with substantial to strong evidence when it is not present. This indicates that, for JWST-like data, there is the possibility that we can characterize the correlation during the retrieval. The inclusion of the error multiple can be more robustly judged as well. Its removal is supported when the errorbars are correct, and its inclusion is favored when the errorbars are overestimated. However, when the errorbars are underestimated, its inclusion is supported with substantial evidence only when there is no correlation. This behavior matches with the results from the previous section that correlation can be mistaken for underestimated error bars. This shows that the error multiple is generally not effective at indicating accommodating underestimated errorbars.

The above results show that while it is difficult to conclusively infer the presence of correlated noise with *HST* data, it is certainly within possibility that its presence and strength can be measured with *JWST* data. A few caveats must be made. Our preliminary test presented here treats the data over the entire NIRSPEC wavelength range as sharing correlation, hence providing an abundant number of points for the correlation strength to be measured; in reality, if its discrete grisms are used, any correlation of instrumental origin will be per each wavelength range. Additionally, as we will discuss in Section 2.5.1, missing physics in the model acts as a source of systematics, which we do not consider here. Furthermore, we used the same noise model to generate the observation instance as well as to retrieve its parameter. While doing so is obviously a gross simplification, especially considering that numerous sources of correlated noise can operate simultaneously, this provides a reasonable starting point towards using a more complex likelihood function to fit for correlated noise. Additionally, adding hyperparameters to a retrieval further di-

lutes the noise budget, broadening the retrieved uncertainties of other parameters. Ascertaining what degeneracy this incurs on the estimation of other parameters is left for future work.

2.5 Discussion

2.5.1 Model Limitations

A major compounding issue is that, when retrieving on real data, model assumptions and unknowns contribute to and act as systematic errors in addition to the data systematics themselves. In short, bad data are degenerate with bad models. In our study we generated the synthetic observations using the identical forward model as that used in the retrieval in order to minimize any model-dependent effects and to isolate the effects of data systematics. In interpreting real data, the fact that our forward models are a simplified incomplete representation of complex atmospheric phenomena will act as a source of systematic error that will remain pervasive, even if the observed data were perfect and free from their own systematics. We therefore remain open to the possibility that the observed examples of potential systematic noise in the data are in fact due to unaccounted for obscure physics.

For the same reasons discussed above, this will adversely affect high SNR observations in particular, in which the fine (and the not-so-fine) details of the model become discernible. There has recently been a growing body of work that studies the biases incurred by model assumptions and parameterization. To list a few examples for demonstration, MacDonald et al. [106] performed 1-D retrievals on 3-D synthetic spectra to show that the retrieval biases the limb temperature to few hundred Kelvins cooler than the actual day-night mean temperature. Lacy and Burrows [76] extended this study to cloudy atmospheres, finding that the presence of aerosols

exacerbate the biases induced by 3-D effects when not accounted for. Changeat et al. [74] found that using a vertically constant chemical abundance profiles may no longer be sufficient to fully capture signatures of disequilibrium processes in the spectrum. Perhaps most inconspicuously, Barstow et al. [75] compared and performed cross-retrieval between retrieval codes developed by three groups and found that *JWST*-quality data is now sensitive to rather rudimentary model unknowns such as the line lists used to generate the opacities and the precision of fundamental constants used.

In our framing of describing biases, these considerations generally result in shifts in the estimated means of retrieved parameters. Incorporating our conclusion that correlated noise generally leads to underestimated error of retrieved parameters means that biases due to model limitation now strike with a stronger statistical significance. Furthermore, the wavelength-dependent effect of both missing physics and systematics now leave possibility for degenerate interpretations.

2.5.2 Instrumental Systematics

To better understand the significance of our results, it would be useful to consider the different sources of how systematics can arise and evaluate their prevalence especially in the context of future space missions such as *JWST*. While other sources of systematics are possible, such as starspots [80, 81], inaccurate orbital parameters, or time-dependent telluric contamination in ground-based observations, the key source of systematics that we will discuss here is instrumental. However, we remind the reader that our formulation of correlated noise can be generalized to any effect that results in wavelength-dependent correlation.

While instrumental systematics are expected to be ubiquitous to some extent, the exact

magnitude of their effect in generating wavelength-correlated noise has not been fully understood. Yet a handful of observations exist that hint at the existence of such systematics. Colón et al. [79] argued that, with current facilities, these systematics are visible at the highest level of precision (~ 15 ppm), inferred from an unusual behavior of residuals in the H_2O band. In the spectrum of HD 97658b in Guo et al. [7], our motivating example in §2.1, we inferred from the inability to fit the data as well as no obvious physics being missing that there must be some systematics present, even at a lower precision (~ 25 ppm). These examples indicate that some wavelength-correlated noise must be present, unless there is unaccounted for physics in the retrieval model.

There is some reason to surmise that these systematics are more prevalent (or, at least, more noticeable) in the case of bright host stars. A brighter host star allows for a better SNR and higher precision and thereby naturally makes the presence of these systematics more conspicuous compared to a lower SNR data. Additionally, even at the same data quality, a brighter host star requires fewer stacking of observations; for a dimmer host star, by contrast, the number of stacking required to achieve the same SNR naturally averages out any *non-repeatable* correlated noise. Finally, as alluded to in §2.1, given that instrumental systematics can also behave differently with bright sources, it is not out of question that there is a separate effect at play here beyond SNR which may persist through multiple observations in a repeatable fashion.

Comparing the WFC3 spectra of GJ 1214b [92] and of HD 97658b [7] illustrates this point. While both planets are comparable sub-Neptunes with featureless spectra and have a similar level of precision, the spectrum of the latter displays an unusual upward trend in transit depth in the redder end and other wavelength-correlated residuals throughout the WFC3 bandpass. The relevant difference here may be the host star brightness (9.8 versus 6.2 in J-band magnitude, respectively). The spectrum of GJ 1214b is the combination of stacking 15 transits, whereas that

of HD 97658b has 4. Further, the spectrum of HD 97685b presented in Knutson et al. [107], which only had the first 2 visits, shows the most obvious possible example of correlated noise due to systematics.

These types of systematics may be even more pernicious for future high SNR observations from *JWST* for a few reasons. First, as the noise floor is lower, correlated noise will be relatively more prominent even if it actually manifests at weaker levels. Then one can no longer reliably assume that the observed noise is strictly photon-dominated. This requires an additional step of modeling out now wavelength-dependent systematics during the data reduction, which is necessarily (although perhaps not practically) incomplete. Secondly, the high SNR per transit means that stacking will be unnecessary for most targets. As such, non-repeating systematics do not get averaged out. Thirdly, we have demonstrated that higher precision leads to biases of stronger significance. This is true even in the absence of systematics in the sense that a retrieval will be more sensitive to the observational instance. Fourthly, we can predict that our understanding of the characteristics of *JWST* instruments and their appropriate data reduction tools will be only partially correct, at least during the initial few cycles of *JWST* before practical experience accumulates. Finally, as planets around bright host stars allow for achieving high SNR, they will make attractive targets for *JWST* observation. However, if the above intuition that bright host stars can exacerbate instrumental systematics is true, it adds another dimension to consider when selecting targets for observation, in addition to the SNR.

Given that this is the case, it would be worthwhile to put the above heuristic that bright stars bring about correlated noise to a more formal test. This can be accomplished with *JWST* if the correlation strength can actually be measured, and by marginalizing over the magnitude of the host star to obtain a trend. While this would not comprise a main scientific objective of any

program, correlation strength is a parameter we can try to measure for all observations, so this is a test we can perform at no extra cost in observation time.

2.5.3 Data Outliers and Free Retrieval

PLATON originally supports equilibrium chemistry retrievals only. Using this method the molecular abundance of each species at a given temperature and pressure is set by the metallicity and carbon-to-oxygen ratio. A popular alternative method to constrain chemistry is to use “free” retrievals, in which the abundances of each species are allowed to vary independently. This accounts for any non-equilibrium chemistry effects in the atmosphere, brought on by vertical mixing or photochemical interactions. To establish how wavelength-dependent systematics or outliers can bias certain species, we implement free retrievals in PLATON and perform some basic tests in addition to the suite of retrievals already presented that used equilibrium chemistry models.

PLATON already natively supports inputting custom chemical profiles to its forward models, but only accepts equilibrium chemistry parameters during retrievals. We extend its capability by allowing it to accept custom chemical abundances during retrievals as well. As such, all other details regarding how the spectrum is calculated during the retrieval remain exactly the same as the original implementation in PLATON. We assume that each species has a vertically fixed mixing ratio. This is a reasonable approximation over the pressure range probed by transit spectroscopy at present data quality, and most current 1-D free retrieval codes parameterize the composition using this assumption [73]. *JWST*-quality data may merit a more complex prescription, such as a 2-part vertical abundance profile [74], but for now we do not consider these possibilities. Additionally,

for these tests we remove the simulated *Spitzer* data points from our synthetic spectra to limit the wavelength range, thereby reducing the choice of necessary species to be included. We assume a H/He-dominated background atmosphere. We include H₂O, Na, K, TiO, and VO as they are the primary detectable species in the remaining wavelength and temperature range; we are mostly interested in how differently species with broad absorption features (e.g. H₂O) versus species with narrow ones (e.g. Na) can be biased.

We show a free retrieval on two observational instances of the same baseline hot Jupiter as described in Section 2.2, where we did not add correlated noise or any instrumental offsets. Figure 2.17 shows the two random input data realizations and the best fit spectra, and Figure 2.18 shows the retrieved posteriors.

The point of interest here is how the retrieved abundances compare for species with broad spectral features (H₂O, 1 – 2 μm) and with a narrow feature (Na, 0.58 μm). As one may expect, for both of the two random spectra, the retrieved posterior distribution for water shows a tighter constraint than for Na, which is dictated almost entirely by one data point. Consequently, between the two posteriors as well, the retrieved distributions for Na show little overlap, ruling out each other by $\geq 2\sigma$. These are simply two randomly drawn samples, but it demonstrates the point that measurements of water abundance or metallicity are more robust compared to that of say Na or K abundances, because of the impact the former parameters have across a broader wavelength range.

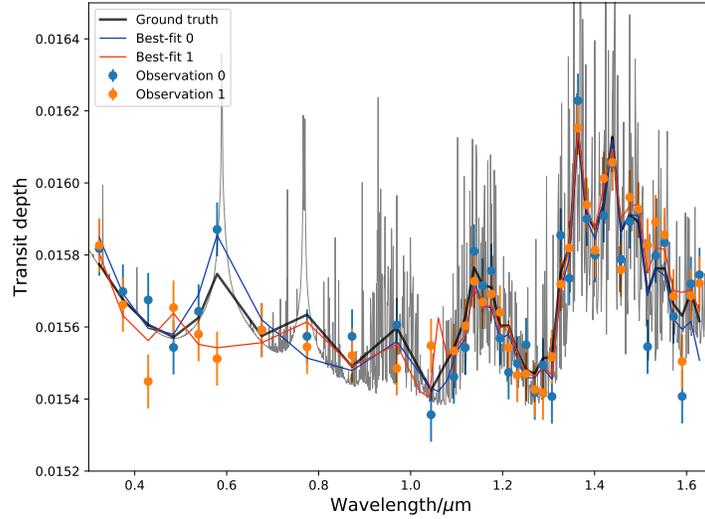


Figure 2.17: Two observation instances (blue and orange) of the ground-truth spectrum (black) for our baseline hot Jupiter case detailed in Table 2.1. The best-fit spectrum for each observation are also shown.

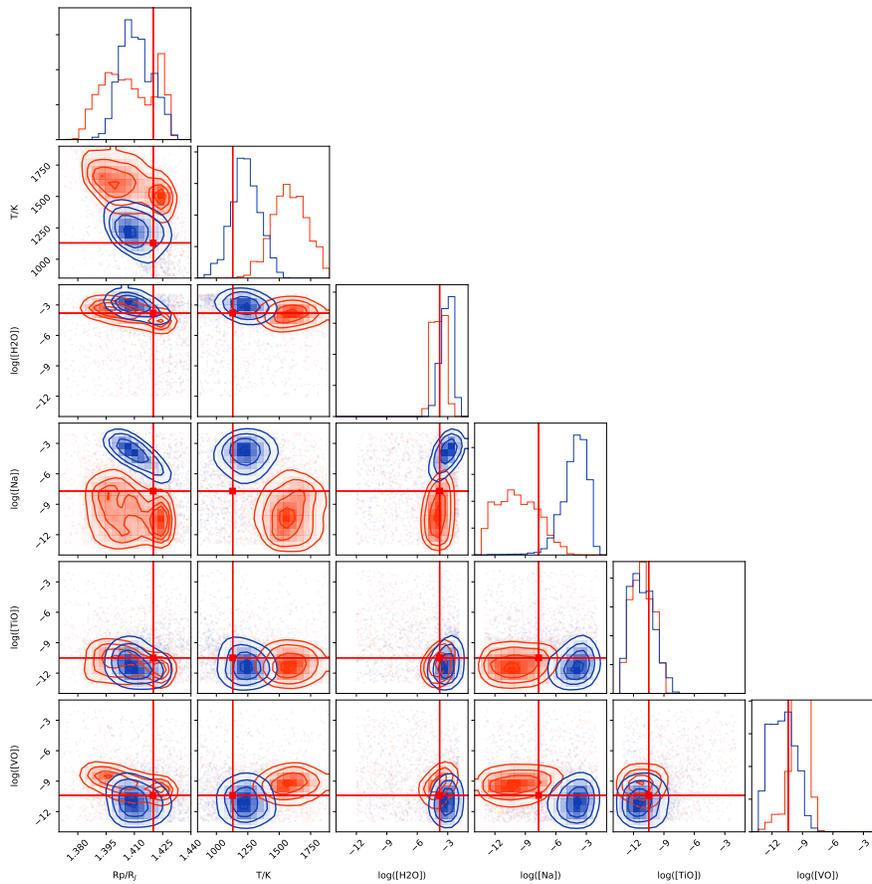


Figure 2.18: The retrieved posterior from the two input datasets from Figure 2.17. We only show the relevant parameters. The colors corresponding to each dataset are the same as in Figure 2.17. Black crosshairs indicate the ground-truth input values.

2.6 Summary and Future Work

Atmospheric retrieval provides a robust framework to interface theory and observations and is a key tool to furthering our understanding of exoplanets. One major outstanding issue is disentangling the effects of systematic biases that may be in operation, and in response there is a growing body of work in the literature that investigates the consequences of biases that arise from forward model assumptions.

This chapter instead presents an assessment of biases that arise from systematic noise in data, while remaining agnostic as to the source of such systematics. We stress that, although our implementation of correlated noise (using Gaussian process) is just one mathematical option, the general results remain robust. We find that the presence of correlated noise can mislead us in various ways. We are more likely on average to obtain better goodness-of-fit, but obtain worse retrieval accuracy overall. This is due to both the parameter mean being biased and the retrieved error being underestimated. Specifically, we observe that correlated noise can bias the retrieved aerosol properties, mimicking non-Rayleigh slopes or misrepresenting the location of a cloud deck. Additionally, we find that offsets between datasets can be correctly retrieved and are not degenerate with retrieved chemistry when equilibrium chemistry is assumed, so long as the forward model is an accurate depiction of the atmosphere. We also find that while correlated noise cannot be characterized during retrieval for *HST* data, there is potential (and perhaps necessity) for *JWST* data, even though our tests reflect optimistic conditions. Additionally, we validate the intuition that retrievals are sensitive to individual noise instances, and, especially in the context of free retrievals, that statistical outliers can have significant effects on the retrieved chemistry, especially in the phase space the retrieved uncertainties do not scale linearly with the data error.

Chapter 3: On the Composition Parameterization Problem for High MMW Atmospheric Retrievals

3.1 Introduction

Atmospheric spectra have been the primary target of observations in the effort to characterize exoplanets from ground- and space-based observatories. The spectrum of an exoplanet atmosphere encodes significant information about its properties, such as composition and thermal structure, which in turn reveal the processes within, e.g., photochemistry, dynamics, and surfaces. An observed spectrum is then decoded by using inverse modelling (called atmospheric retrievals), which finds a distribution of solutions that match the observed data by repeatedly running forward models that simulate physical processes in the atmosphere. Atmospheric retrievals and Bayesian statistics have become a standard tool in characterizing exoplanet atmospheres, and there is an ever growing body of work dedicated to understanding their limitations and potential pitfalls [e.g. [1](#), [45](#), [46](#), [56](#), [108](#), [109](#), [110](#)].

As the endeavor to characterize exoplanets continues in the era of James Webb Space Telescope (JWST), the parameter space of observable planets will expand to smaller, cooler, and rockier worlds. From a theory point of view, the atmospheres of these planets are expected to be more diverse in composition relative to the giant planets observed so far, reflecting various

formation histories that are possible [111]. Most importantly, the primordial atmospheres of hot Jupiters can be safely assumed *a priori* to be dominated with H and He while the rest of the constituents are all assumed to exist in trace amounts, and therefore all spectroscopically active species, such as H₂O and CO₂, can also be assumed to be in trace amounts. In terrestrial planet atmospheres, however, the bulk constituent gas is not known *a priori* and multiple major components (say, $\gtrsim 10\%$) can exist simultaneously, regardless of whether they are spectroscopically active or not, such that there may not be one gas to designate as the “dominant” gas at all.

The *a priori* H-rich assumption for hot Jupiters is encoded into the parameterization of Bayesian priors used by most atmospheric retrieval codes [1, 46, 108, 110]. The assumption is captured by setting the abundances of the spectroscopically active species as the free parameters against a background atmosphere. The abundance of the “filler” gas—usually a mixture of H, H₂, and He—is not a free parameter and is inferred from the constraint that the composition must sum to unity. The abundances of the active species are typically parameterized in ranges of trace amounts and over multiple orders of magnitudes, warranting the appropriate use of a log scale for abundance constraints. The prior is set to be uniform over a range of log-abundances, with the minimum log-abundance typically set to reflect detectability of a particular gas species from its spectral features given the data uncertainty. The upper limit is set such that the gas cannot be a major component of the atmosphere [110]; otherwise there is a joint prior on the free parameters set by the implicit constraint that the sum of the mixing ratios cannot exceed unity. We will refer to this parameterization method throughout current work as the “trace and filler” (TF) method, and will describe in more detail in §3.2.1.1.

The expected dissimilarity of high-mean molecular weight (MMW) atmospheres to those of hot Jupiters pose a challenge to this parameterization being effective if the *a priori* H-rich as-

sumption is no longer valid. Benneke and Seager [56] (BS12 hereafter in this work) discussed this problem in detail. BS12 recognized that using the trace-filler method to set the priors for abundances of gases lead to an asymmetric prior between the trace and filler gases and argued against its usage when the dominant species is not known *a priori*. As an alternative, BS12 suggested a reparameterization based on the centered log-ratio (CLR) transformation of the abundance space such that the prior is symmetric between all gases. They suggested that via this reparameterization, the retrieval no longer makes a strong assumption about the atmosphere and is therefore applicable to retrievals of both H-rich and non H-rich atmospheres. The authors argue that this is a true uninformed or “ignorance” prior. By testing the retrieval on simulated spectra of super-Earth GJ 1214 b with H₂O-rich, N₂-dominated and H-dominated atmospheres, they showed that the using the reparameterization correctly retrieves the compositions under ideal observation scenarios. In a follow-up work, Benneke and Seager [112] then demonstrated that Bayesian models using this new parameterization allow for correctly inferring the chemical makeup of the atmosphere and hence distinguishing between cloudy H-rich atmospheres and high-MMW atmospheres.

Barstow et al. [108] compared three retrieval codes developed by different groups and found that it is for cloudy or high-MMW super-Earth atmospheres where the retrieval codes begin to disagree, whereas retrievals for hot Jupiters or clear H-rich super-Earths resulted in only minor discrepancies. As the simulated spectra for these atmospheres showed good agreement among the codes when used as forward models, the parameterization of composition during retrieval was identified as the main source of disagreements in the high-MMW case, which in turn also affects retrieved distributions of other parameters such as temperature via degeneracy. The latter effect was especially apparent when simulated data had a high signal-to-noise (30ppm at R=100 for an Ariel-like wavelength range), as the estimated uncertainties shrink around different estimated

means.

Welbanks and Madhusudhan [110] brought attention back to the framework presented in [BS12](#) and performed comparisons between retrievals using the trace-filler parameterization and the [BS12](#) reparameterization, for the transmission spectra of hot Jupiter HD 209458 b, mini-Neptune K2-18b, and TRAPPIST-1 e. They found that given sufficient signal-to-noise, the CLR method can be used to infer the abundance of O₂.

Importantly, the choice of parameterization may affect any tests involving Bayesian evidences, or the likelihood function averaged over the prior space. The implicit prior set by the parameterization choice weighs the different parts of the parameter space differently, and hence if—only by coincidence—the prior already favors where the likelihood function is the greatest and disfavors where it is low, the retrieval will estimate a higher Bayesian evidence. As it is very typical to employ Bayesian evidence testing to, e.g., identify which species are justified in being included or to estimate the evidence of a rocky planet hosting an atmosphere, a precise understanding of how the parameterization choice affects evidence will be useful.

In this work, we extend the comparison between parameterization choices to a more diverse set of compositions of terrestrial planets, as previous works in the literature have focused on the end-member cases of composition. We do so by performing self-retrievals on simulated transmission spectra of atmospheres of plausible compositions observed by *JWST*. Retrieval on simulated data allows for comparing the retrieved results directly to the ground truth and quantifying their accuracy. Furthermore, performing the retrievals on simulated data bypasses biases due to model assumptions (e.g. lack of 3D effects) that can arise when retrieving on real data [108, 113, 114, 115, 116]. By doing so we aim to ascertain what is the appropriate parameterization of chemistry when retrieving on atmospheres whose mean molecular weight is not known *a*

priori, and investigate what constraints are realistically possible with greater number of transits. In the current work, we limit our analysis to retrievals on transmission (rather than emission) spectra to avoid degeneracies with retrieval of thermal profiles and focus on chemistry only, and also because *JWST* Cycle 1 programs included a number of observed small planets in transmission [e.g. 117, 118, 119, 120]

This chapter is structured as follows. We describe the various parameterization choices considered and the forward model atmospheric composition in §3.2. We present the comparison results for each case in §3.3. We discuss what an ignorance prior is in the context of retrievals in §3.4. We summarize our conclusions in §3.5 and discuss related problems in retrievals.

3.2 Methods

We perform self-retrievals on simulated spectra, as used to test retrieval schemes in e.g. [109, 110]. To both generate the spectrum and to retrieve from it, we use a modified version of the publicly available code PLATON, adapted to perform non-equilibrium chemistry (“free”) retrievals using various parameterizations that we describe below. Natively, PLATON already supports forward models with custom abundance profiles but is not configured to do so during retrievals. We allow the vertically fixed abundances of each species to be set during a retrieval according to the retrieved parameter [1]. Additionally, we allow for setting and retrieving on the mean molecular weight profile separately from the chemistry in the forward and retrieval models, and comment on the reasoning for this choice, below.

In this section, we describe the retrieval methods and forward model parameter choices. First, we describe the different parameterization methods. Then, we present the cases of terrestrial

atmospheres we perform comparisons on.

3.2.1 Parameterization methods

The variety of parameterization methods arise due to the simplicial nature of atmospheric composition. That is, for N species, there are $N - 1$ degrees-of-freedom, as the sum of composition must be unity. The parameterization methods therefore differ in how $N - 1$ numbers sampled from a standard uniform distribution are mapped to N composition values that sum to unity. We use four different parameterization methods:

1. Using N_2 as a single filler gas (SF)
2. Using H_2 and N_2 as two filler gases (DF)
3. Centered log-ratio (CLR)
4. Mixing ratios with no filler gas but μ as an independent parameter

In this subsection, we describe each parameterization method and what choices need to be made when using each method. We look at the prior set on the composition and on mean molecular weight.

3.2.1.1 Trace and filler methods

The most commonly used parameterization for composition in retrieval codes is the trace and filler (TF) method. In the TF method, the $N - 1$ parameters correspond directly to the (log-) mixing ratios of $N - 1$ “trace” species. Here and throughout the work, we use the term “trace species” without necessarily implying that they are in negligible amount (which we will refer to

as “trace amount”), but rather that they are tracked directly by the parameterization rather than inferred implicitly. To satisfy the unity sum requirement, the TF method assumes the identity of a filler gas and backfills the remainder from unity. This assumption also sets an implicit common prior that the sum of the trace species cannot exceed unity, shown as a minor downturn in the prior close to unity when viewed in log space (Figure 3.1). We implement this common prior by rejecting samples when the sum of trace species exceed unity. We note that DYNESTY automatically excludes these samples when calculating evidence. We choose the background gas to be N_2 . In principle, any gas can be the filler gas, but it is typical and prudent to use a spectroscopically inactive species, as abundances of a spectroscopically active species can be measured directly.

The trace species are typically sampled in log-abundance, such that the parameter spans multiple orders of magnitude of mixing ratios. We compare two variations for the TF method, one in which all trace species are sampled from log-uniform distributions, and another in which a linear uniform distribution is used for H_2 . We call the latter variation a two-filler method, as the H_2 and N_2 now have the same priors. As the change in H_2 affects the spectrum primarily via changing the MMW, it should ideally be in linear scale. Additionally, setting one trace species in linear abundance changes the implicit prior for the filler (N_2 in this case) such that it now also resembles a uniform distribution between 0 and 1, with a slight tapering off near unity due to the sum of the log trace species.

A choice needs to be made regarding the minimum and maximum abundance to be retrieved, i.e. the cutoff for the uniform prior. The minimum value is often chosen empirically as the value above which a change in the mixing ratio has an effect on the spectrum above the noise floor, typically $\sim 10^{-12}$ [45, 46, 108]. For hot Jupiter atmospheres, the upper limit has been set lower than unity in some retrieval analyses [46, 110] to ensure that the atmosphere is dominated by the filler

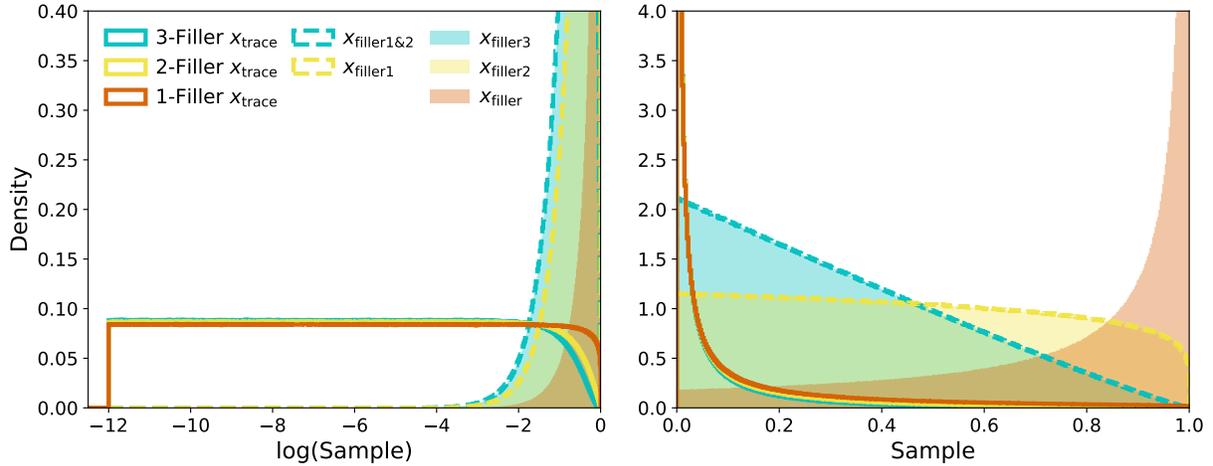


Figure 3.1: The marginalized priors probability distributions for different trace-filler parameterizations, shown in log scale (left) and in linear scale (right). Solid lines and dashed lines correspond to log and linear species, respectively; the filled histogram correspond to the filler species. The DF method “corrects” the prior of the filler such that it more resembles a uniform distribution, with a slight taper around unity due to the sum of log trace species.

(H/He), and this is a justified choice as trace species are indeed expected to be in trace amounts. For terrestrial planets, however, an upper limit set lower than unity may rule out plausible bulk composition of the atmosphere (such as 100% CO₂), and hence we use unity for all upper limits.

Additionally, the rejection sampling used to satisfy the common prior incurs an efficiency of 98% for a single-filler gas, 86% for two-filler gas, and 42% for three-fillers when $N = 6$. We find that this adds to the computation time only during the early steps of the retrieval before it has converged somewhat, after which the overall efficiency of the sampling algorithm is generally lower. We discuss the effects of rejection sampling on the Bayesian evidence in §3.5.

3.2.1.2 Centered log-ratio transform (CLR) method

The centered log-ratio transform (CLR) method, widely used for compositional data [121], was first introduced to Bayesian atmospheric retrievals in BS12 and has been implemented in recent codes [109, 110]. The CLR method parameterizes the abundances as the log-ratio to the

geometric mean.

$$\xi_i \sim \mathcal{U}(l, h)$$

$$x_i = \frac{e^{\xi_i}}{\sum_i e^{\xi_i}},$$

where ξ_i is the log-ratio of the abundances to the geometric mean and is sampled from a uniform distribution between the interval $[l, h]$, where the limits are set to a minimum abundance. The sampled values of ξ_i is then reparameterized to the actual mixing ratios x_i via the centered log-ratio transform.

By employing the reparameterization, the CLR method bypasses some foibles of the TF method. Firstly, it ensures that the prior is symmetric between trace species and background species, as no species is designated (or assumed) to be a filler. This may not necessarily be the appropriately uninformative choice, given that the ideally uninformative prior of a species depends on how the change in its abundance modulates the spectrum, as discussed in §3.4. Secondly, it removes the choice between log and linear parameterization. Instead, the resulting prior is one that, viewed in log space, maintains a gentle slope over the range $\lesssim 10^{-2}$ and ramps up closer to unity (see Figure 3.2, in blue). The prior of the log-sample of the CLR method thus resembles the trace prior of the TF method over the range $[\frac{l}{2}, -2]$ and the filler prior over the range $[-2, 0]$. Viewed in linear space, the prior is bimodal and strongly favors values close to either zero or unity.

The CLR method requires a choice in the cutoff range for the uniform prior from which the reparameterized values ξ_i are sampled. Choosing to ensure that the resulting minimum abundance

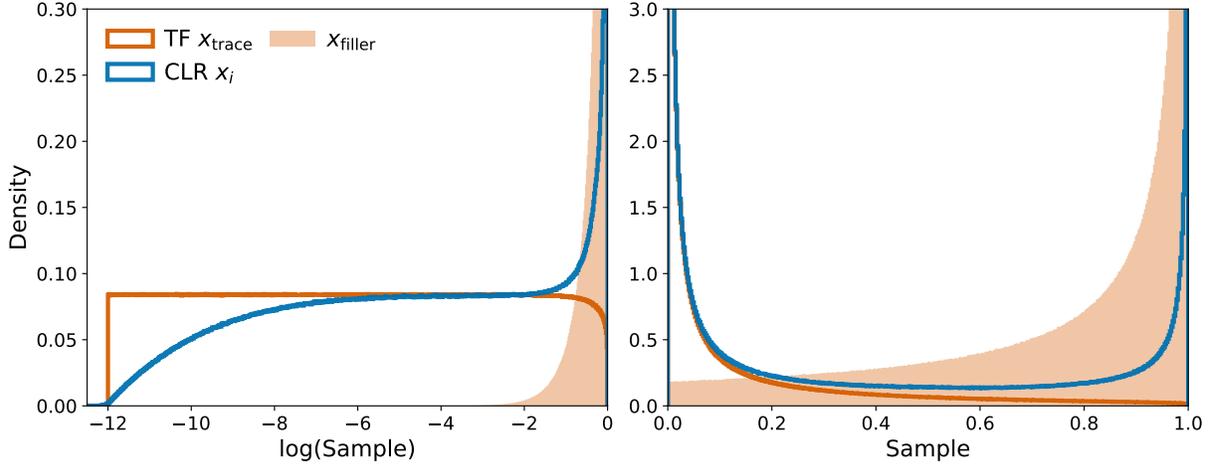


Figure 3.2: The marginalized priors probability distributions for single Trace-Filler and the Centered Log-Ratio parameterization, shown in log scale (left) and in linear scale (right). Solid lines and dashed lines correspond to log and linear species, respectively; the filled histogram correspond to the filler species.

in the marginalized prior corresponds to a specific value ($\sim 10^{-12}$ in both [BS12](#) and Welbanks and Madhusudhan [110]) sets both cutoff values for the uniform prior.

3.2.1.3 Inclusion of mean molecular weight (MMW) as an independent parameter

We also include a method in which no gas is used to backfill the composition, but we additionally retrieve on MMW as an independent parameter. In our model we assume that the MMW is fixed vertically. By separating MMW from the abundances, the change in abundance of a spectroscopically active species has a monotonic effect on the strength of its features, as it now affects the spectrum via opacity only. We choose the prior on the MMW to be uniform over $[2, 44]$, where the limits correspond to MMW of H_2 - and CO_2 -dominated atmospheres, respectively.

3.2.1.4 Comparison of the priors on MMW

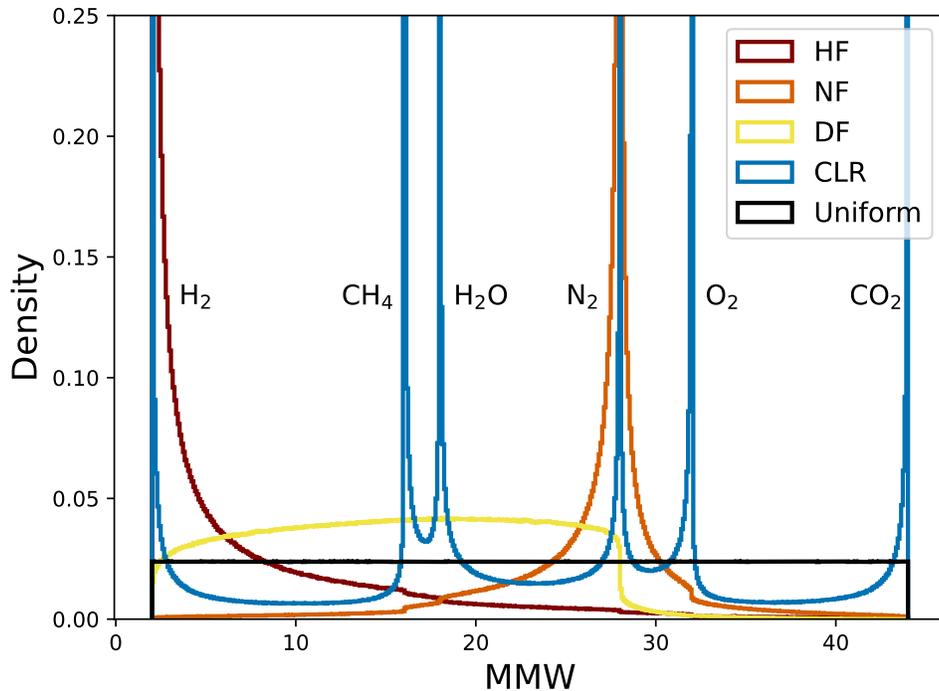


Figure 3.3: The implicit prior on mean molecular weight set by the different methods of parameterizing the composition. The peaks corresponds to specific gases, and the different methods vary on what gases are allowed to be major components. For the independent MMW retrieval, we set a uniform prior.

The prior on the mean molecular weight propagated from the prior on the composition is shown in Figure 3.3. The figure demonstrates the awkward tension between priors in composition and prior in mean molecular weight. The various peaks in the prior on MMW in turn set a peaked prior on the scale height. If one were to think analogously to the prior on temperature, the primary effect of which is also controlling the scale height, only the uniform prior on MMW may be considered an uninformative prior.

3.2.2 Model atmospheres

To compare the different parameterization methods, we run retrievals on simulated spectra of GJ 1132 b. From analyses of *Hubble* observations, GJ 1132 b has shown to be lacking a clear, primordial atmosphere, indicating the potential for a thick, secondary atmosphere of high mean molecular weight, although a cloudy atmosphere or an airless body are also plausible explanations for a spectrum that is featureless over the observed range. Controversially, other analysis have indicated that the planet may have a H/He-rich atmosphere [122]. Recently, *JWST* NIR-SPEC observations have conclusively shown either a high mean molecular weight or a featureless spectrum [123].

We simulate four plausible compositions of the atmosphere of GJ 1132 b corresponding to: an (1) atmosphere evenly split between CO₂-O₂ dominated atmosphere, in which one spectroscopically active species is a major component; an (2) atmosphere evenly split between H₂O-CO₂-O₂, in which two spectroscopically active species are major components; an (3) an N₂-dominated atmosphere, in which all spectroscopically active species exist only in trace amounts; an (4) Earth-like atmosphere, in which two spectroscopically inactive species are major components. These compositions are all plausible for the atmospheres of rocky planets [21]. We note that in all of our parameterizations, we treat the O₂ as a spectroscopically active species throughout, as O₂ has a spectral feature at 0.76 micron, albeit weak and narrow; the expectation is that it will act like a spectroscopically inactive species at lower S/N, and an active one at higher S/N. We use the stellar parameters from Berta-Thompson et al. [124]. We show the forward model parameters and the composition used in Table 3.1.

We also repeat all of our models with clouds, which can mask the bottom of the atmosphere

Case	H ₂ O	CO ₂	CH ₄	O ₂	H ₂	N ₂	MMW (amu)
CO ₂ -O ₂	0	50%	0	50%	0	0	38.00
H ₂ -CO ₂ -O ₂	33.3%	33.3%	0	33.3%	0%	0	31.34
N ₂ -rich	100 ppm	100 ppm	100 ppm	0%	0%	99.97%	28.01
Earth-like	1%	100ppm	0	20.99%	0	78%	28.75
	M _p	R _p	T _p	Clouds	R*	T*	
	1.659 M _E	1.160 R _E	580 K	10 mbar	0.207 R _☉	3270 K	

Table 3.1: Table summarizing the four cases of compositions and the forward model parameters used to generate the unpolluted spectrum. The mean molecular weight is calculated self-consistently from the abundances. We run two sets of each models, with and without clouds.

(and the spectrum), thereby prohibiting the estimation of the scale height and hence the MMW of the atmosphere. Clouds are implemented in PLATON is such that they are wavelength-independent and have a sharp cutoff in opacity at the assigned cloudtop pressure. Notionally a grey planetary surface would also have the same effect on a transmission spectrum, in which case the cloudtop pressure parameter can be understood as a surface pressure [56]. We choose the cloudtop pressure for these models to be at 10 mbar, resulting in obscuring roughly half of the spectral features for each case.

We use the forward model of PLATON to generate the unpolluted true spectrum, and then use PANDEXO [125] to generate the simulated noise and bin to the desired resolution. We retrieve on both the unpolluted spectrum and 20 instances of polluted spectra with a Gaussian scatter around the simulated data. The polluted spectra are necessary to ascertain more realistic ranges of Bayesian evidences. We simulate observations for NIRISS and NIRSpec. To compare the parameterization methods at different signal-to-noise levels, we simulate 2, 5, and 10 transits for both instruments. We run the retrievals on the data binned to R=100 using a model resolution of R=10,000, using a nested sampler of DYNesty with 300 live points. The retrieval includes plan-

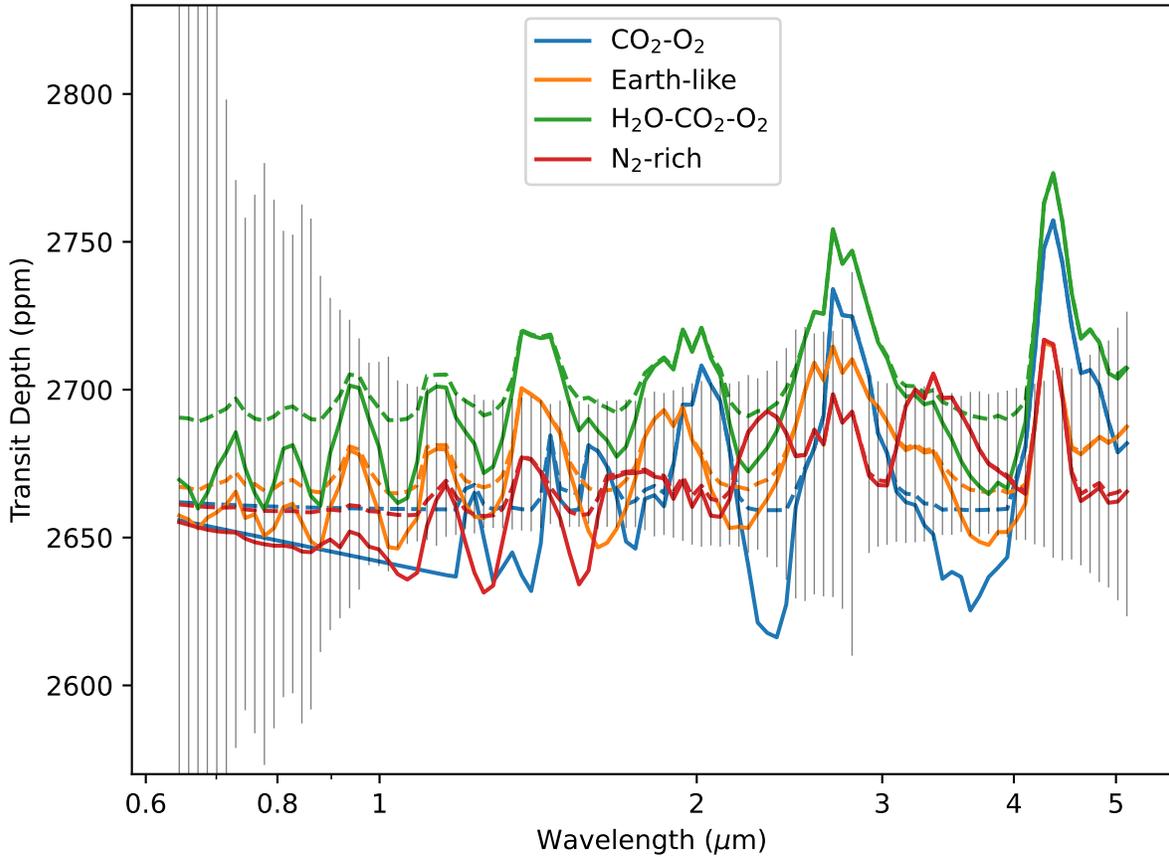


Figure 3.4: The forward model spectra corresponding to the 4 cases used. Spectra of clear and cloudy atmospheres are shown with solid and dashed lines, respectively. The simulated data error from 2 transit observations with NIRISS and NIRSpec is shown around a median featureless spectrum. The NIRSpec uncertainty is typically 40 ppm at $R = 100$; in other words, all 8 spectra are appreciably “flat lines”.

etary radius, an (isothermal) temperature, and cloudtop pressure as free parameters in addition to the parameters for composition. To ensure that we are using sufficient live points, we ran multiple identical retrievals with this setup and found that both the posterior distribution and estimated evidences are repeatable.

3.3 Results

Here we present and compare the retrieval results for each parameterization – SF, DF, and CLR, with and without μ as a retrieved parameter. We choose $N_{transits} = 2$ and comment on the degree to which the prior implementation impacts the retrieved parameters. In the following subsection, we investigate how our results change in the higher signal-to-noise regime by simulating data resulting from $N_{transits} = 5$ and 10 stacked transits. In general, at low signal-to-noise, there is the risk that a data set will fail to constrain certain parameters. The resulting posterior distribution will then mimic the shape of the prior. If the prior itself does not have a broadly flat distribution, then it may appear that the parameter has been constrained, when in fact it has not. As such, we will examine the marginalized posteriors by comparing them to the marginalized priors in Figures 3.1 and 3.2.

3.3.1 Comparison of parameterization methods at low S/N

Here we look at the benchmark case of comparing the different parameterizations, and examine the higher precision and cloudy cases in the following subsections. We first present the retrievals on the unpolluted spectrum in Figure 3.5.

It is also of interest to test if one could obtain constraints on not merely log-abundances but on the abundance in linear scale (e.g., “is the atmosphere 10% or 50% CO₂?”). For this reason, we show the retrieved posteriors for the spectroscopically active species also in linear abundances in Figure 3.6.

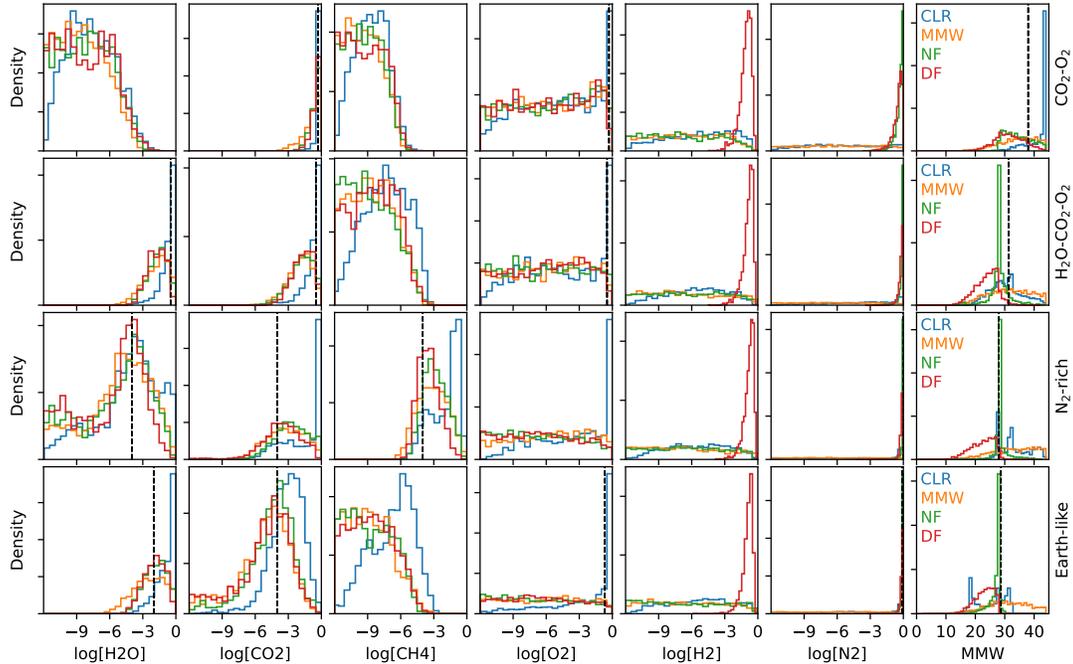


Figure 3.5: Comparison of the posteriors for the 4 cases using 4 parameterizations, for 2 transits each of NIRISS and NIRSPEC of GJ 1132 b. The vertical bars represent the ground truth values. The posterior for the MMW is inferred if not explicitly retrieved.

N₂-rich and Earth-like atmosphere We find notable disagreement between the parameterization methods for N₂-rich atmosphere, where all spectroscopically active species are in trace amounts. Here, the CLR method, rather than inferring that there is N₂ in the atmosphere, incorrectly retrieves a CO₂-dominated atmosphere with a major amount of CH₄ abundance to constitute the same MMW (third row, second and third columns). Because N₂ lacks a spectral signature, the CLR method favors constituting the the high MMW with a mixture of CO₂ and CH₄, the presence of which can be inferred from its features. As the spectral features saturate past mixing ratios of $\sim 1\%$, it is free to increase the abundance of CO₂ such that the atmosphere is dominated by CO₂. Similarly, for the Earth-like atmosphere the CLR method finds an atmosphere that is either dominated by H₂O and O₂.

On the other hand, both the N₂-filler and two-filler methods retrieve posteriors that peak

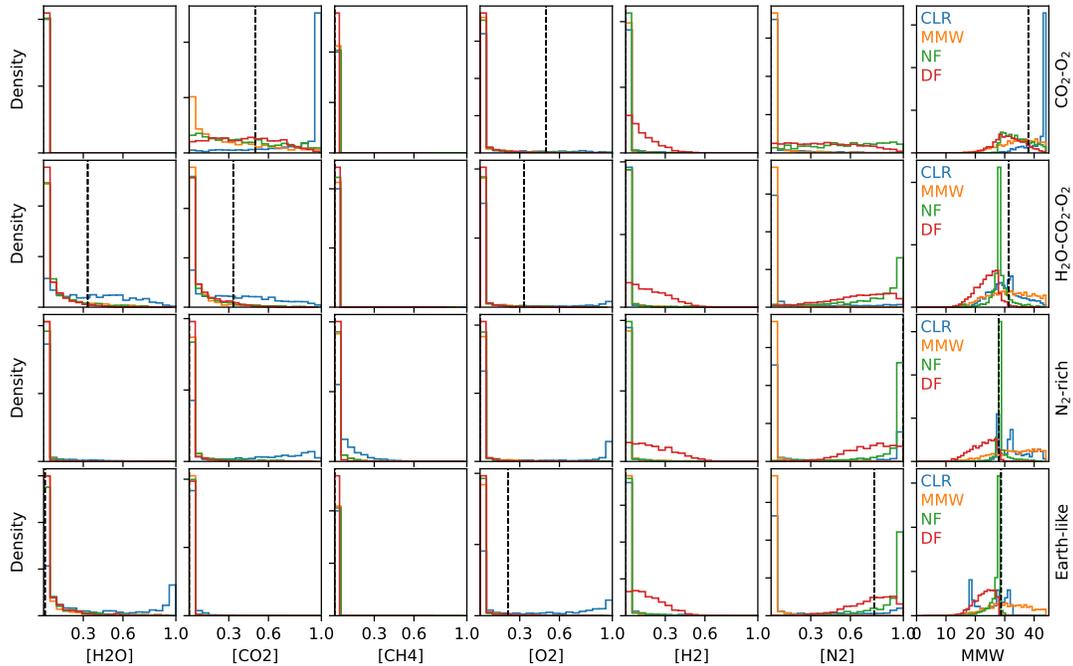


Figure 3.6: Same as Figure 3.5, but now showing linear abundance (from 0 to 1) to show the retrieved major component of the atmosphere.

at the true value for the spectroscopically active species. One difference between the two is that, while both methods correctly infer that the atmosphere is N_2 -dominated, the N_2 -filler method correctly retrieves N_2 abundance merely due to the formulation of the prior, whereas the two-filler method correctly retrieves the N_2 abundance based on actually distinguishing between a H_2 - and N_2 -rich atmosphere. This is best seen in linear scale abundances, shown in Figure 3.6, where the major component gas can be identified. Here H_2 abundance, retrieved as a free parameter in linear scale, acts as a proxy for MMW. The posterior for H_2 and N_2 are broader for the two-filler method.

Our retrieval result for the N_2 -dominated is in disagreement with the retrieval analysis done in BS12, which employed the CLR method on the simulated spectrum of super-Earth GJ 1214 b with N_2 -dominated atmosphere and found good constraints on all species, with 10 transits. We

identify two sources of disagreement, both relating to S/N. Firstly, the overall S/N is much lower in our test, including for the infrared absorbers. The simulated spectrum of GJ 1214 b in [BS12](#) has a much larger transit depth with a median depth of ~ 13800 ppm and with a spectrum that spans roughly ~ 500 ppm, whereas the simulated spectrum of GJ 1132 b in the current work has a median depth of ~ 2700 ppm and a span of roughly 100 ppm. Secondly, the S/N is specifically far worse in the short wavelength range in our simulated NIRISS observation, compared to the uncertainties used in [BS12](#). The worse S/N in the optical wavelength range obscures the Rayleigh slope that provides the most information about the MMW of the planet.

CO₂-O₂ and H₂-CO₂-O₂ atmospheres. We find that in the unpolluted case, even with 2 transits, each method correctly identifies the presence of spectroscopically active gases, CO₂ and H₂O (top two rows of [Figure 3.5](#)), but not the spectroscopically inactive gas, O₂. Here, the main difference is in the retrieved uncertainty and the retrieved mean molecular weight due to whether O₂ has been rejected or not. For the H₂-CO₂-O₂ case in the second row of [Figure 3.5](#), the CLR method finds a narrower uncertainty on H₂O and CO₂ abundance. It can be seen, in the retrieved MMW in the top rightmost panel, that the CLR method finds an atmosphere dominated by CO₂. This likely arises from the fact that O₂ is not detectable at this signal-to-noise and the CLR method instead substitutes the O₂ for CO₂ without changing the spectrum. This tendency for the CLR method to prefer spectroscopically active gases has also been seen in e.g. [Piette et al. \[126\]](#).

Both of these atmospheres have MMW higher than N₂, but we see discrepant behaviors in the N₂-filler method between the CO₂-O₂ case and the H₂-CO₂-O₂ case. N₂-filler method (green line) pushes the retrieved log-abundance of CO₂ to 0 and rejects a N₂-dominated atmosphere such that the MMW is no longer peaked at the N₂ value, as seen in the retrieved N₂ abundance in

Figure 3.6. For the $\text{H}_2\text{-CO}_2\text{-O}_2$ case, on the other hand, the MMW is closer to that of N_2 , and the N_2 -filler method instead finds an N_2 -dominated atmosphere.

In both of these cases, the free-MMW and the H_2/N_2 -filler parameterization methods find weak or incorrect constraints on the MMW. This is due to the fact that for both of these methods, the prior in CO_2 and H_2O abundances are biased towards trace amounts; hence the parameterizations favor low MMW (i.e. large scale height) solutions, as can be seen in Figure 3.6.

For these cases, we also ran a quick set of models that used broadened priors, resulting in prior log-abundance ranges of $[-24, 0]$ for both CLR and the filler-based methods. We find that the resulting posteriors do not differ in an appreciable way and reflect the prior when the likelihood function does not have a strong peak.

3.3.2 Comparison at higher precision

We repeat the above analysis with 5 and 10 transits instead of 2. The general expectation is that higher precision will lessen the importance of the prior, as the likelihood function is more sharply peaked, and hence will lead to better agreement between parameterizations. The resulting posteriors are shown in Figure 3.7 and Figure 3.8.

At 5 transits, the spectroscopically active species in trace amounts are well constrained. For the N_2 -rich atmosphere, the CO_2 abundance is tightly constrained enough such that the N_2 abundance is retrieved by all methods. In the $\text{CO}_2\text{-O}_2$ atmosphere, the aforementioned bias in which the CLR overestimates the abundance of CO_2 is still present.

This effect still persists at 10 transits, showing the difficulty of spectrally constraining the abundance of O_2 . We also find that, even at this high S/N, the log abundance of O_2 is not re-

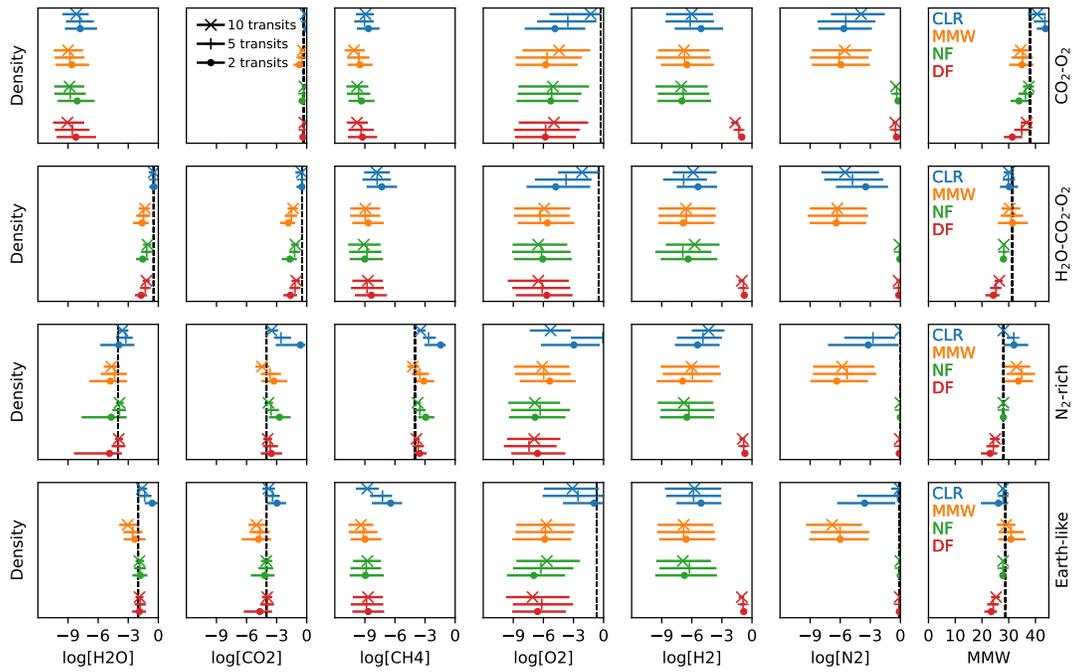


Figure 3.7: Same as Figure 3.5, but now includes posteriors for 5 and 10 transits each of NIRSPEC and NIRISS. The posteriors are now shown in a box plot style as the median with their first and third quartiles, with the full range omitted. We note that the posteriors may not resemble Gaussian distributions and show quartiles rather than the 1-sigma uncertainties.

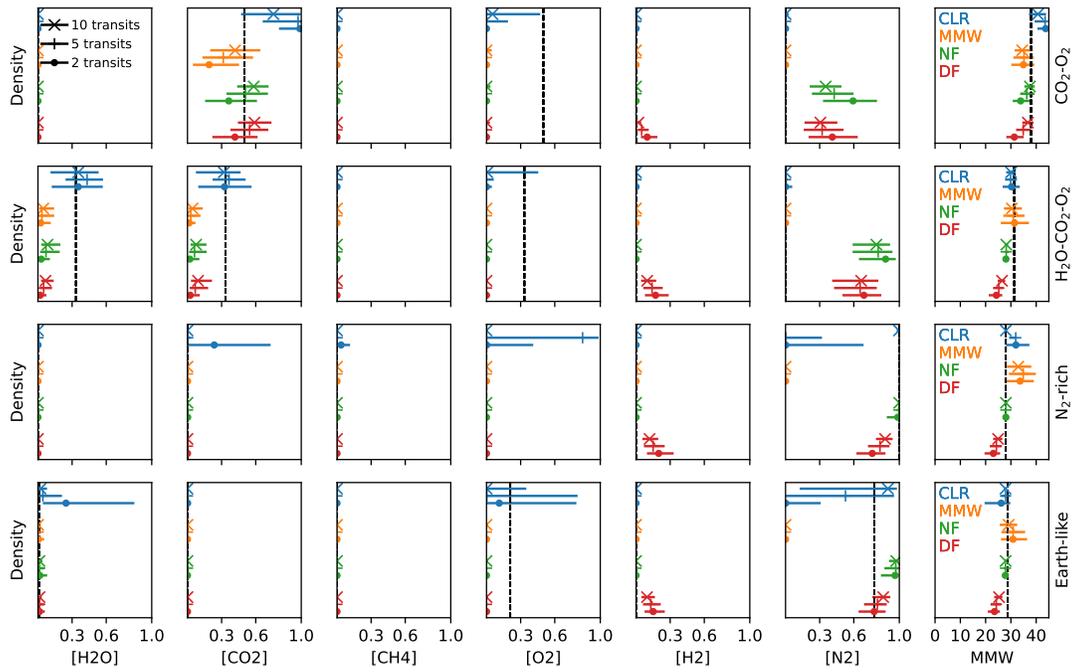


Figure 3.8: Same as Figure 3.7, but now in linear scale (from 0 to 1). The posteriors are now shown in a box plot style as the median with their first and third quartiles, with the full range omitted. We note that the posteriors may not resemble Gaussian distributions and show quartiles rather than the 1-sigma uncertainties.

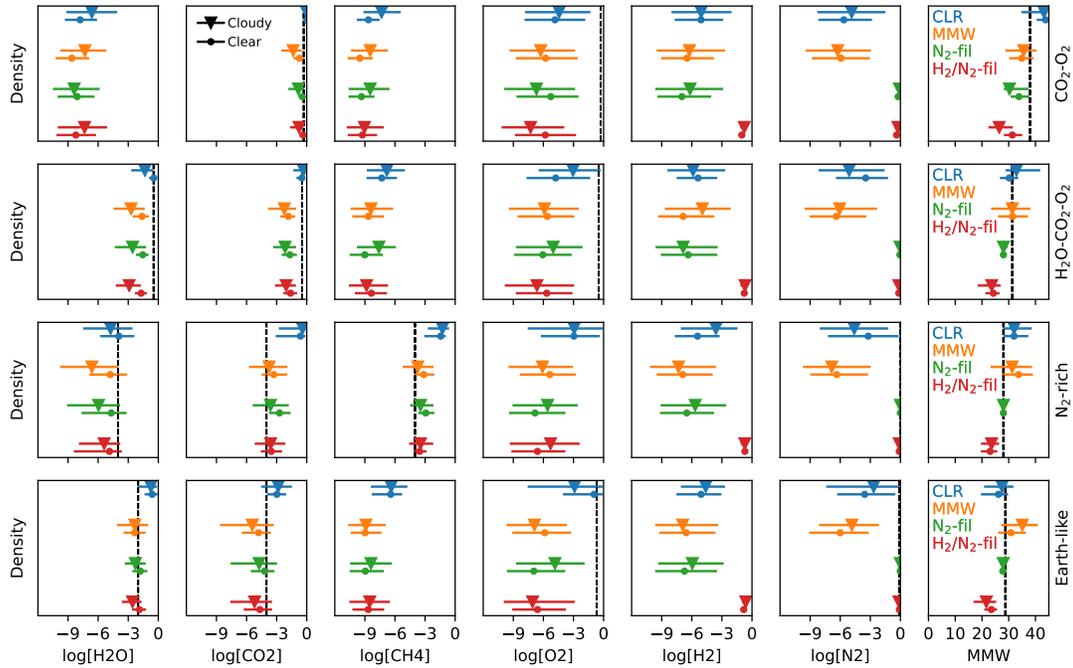


Figure 3.9: Same as Figure 3.5, but now shows posteriors from atmospheres with opaque cloud-decks. The posteriors are now shown in a box plot style as the median with their quantiles, with the full range omitted. We note that the posteriors may not resemble Gaussian distributions.

trieved by any of the methods. This may either be due to insufficient S/N or too much binning in wavelength resolution.

3.3.3 Atmospheres with aerosols

We repeat the above analysis with grey cloud deck present in the forward model atmosphere, placed such that it obscures roughly half of the spectral features. The resulting posteriors are shown in Figure 3.9.

In our retrieval simulations with grey clouds, we find that the general effect is, as expected, to increase the retrieved uncertainties of the abundances of spectroscopically active species. Nevertheless, the retrieved uncertainty on the MMW is not strongly affected, even when included as an independent parameter. The fact that the scale height—and hence the MMW—can be mea-

sured is reassuring. We do not deem necessary an extensive test with varying cloudtop pressure, but nevertheless we expect that the more the spectrum is obscured will naturally lead to broader uncertainties on the MMW, and ultimately to a posterior that simply resembles the prior.

3.3.4 Comparison of Bayesian Evidences

Here we compare the Bayesian evidences of each parameterization method for each case. As the Bayesian evidence for an unpolluted spectrum likely represents an unrealistic best case scenario, we iterate over 20 noise instances and show the Bayesian evidences in Figure 3.10. We find that, in general, there is no parameterization method is systematically favored. This suggests that most of the evidence is contained in the likelihood peak, and the prior volume away from the peak does not contribute significantly to the evidence.

Of more interest is the change in Bayesian evidence depending on whether a gas is included or not. We perform retrievals without CO₂ for each of the noise instances, to quantify the difference in Bayesian evidences in Figure 3.11. While we find no clear pattern, we find that the choice in the parameterization method basically adds random scatter to the $\Delta \log \mathcal{Z}$ to an amount comparable to the scatter due to the injected noise. The maximum discrepancy between parameterization methods given a noise instance is 8.8, 9.4 and 21.7 for 2, 5, and 10 visits, respectively.

3.4 What is an Uninformative Prior?

It is worth contemplating what, in general, our priors are when performing retrievals so that it informs our decisions from an *a priori* standpoint. The ultimate aim here is to seek out the truly uninformed—if theoretical—prior because we typically do not have prior knowledge of the

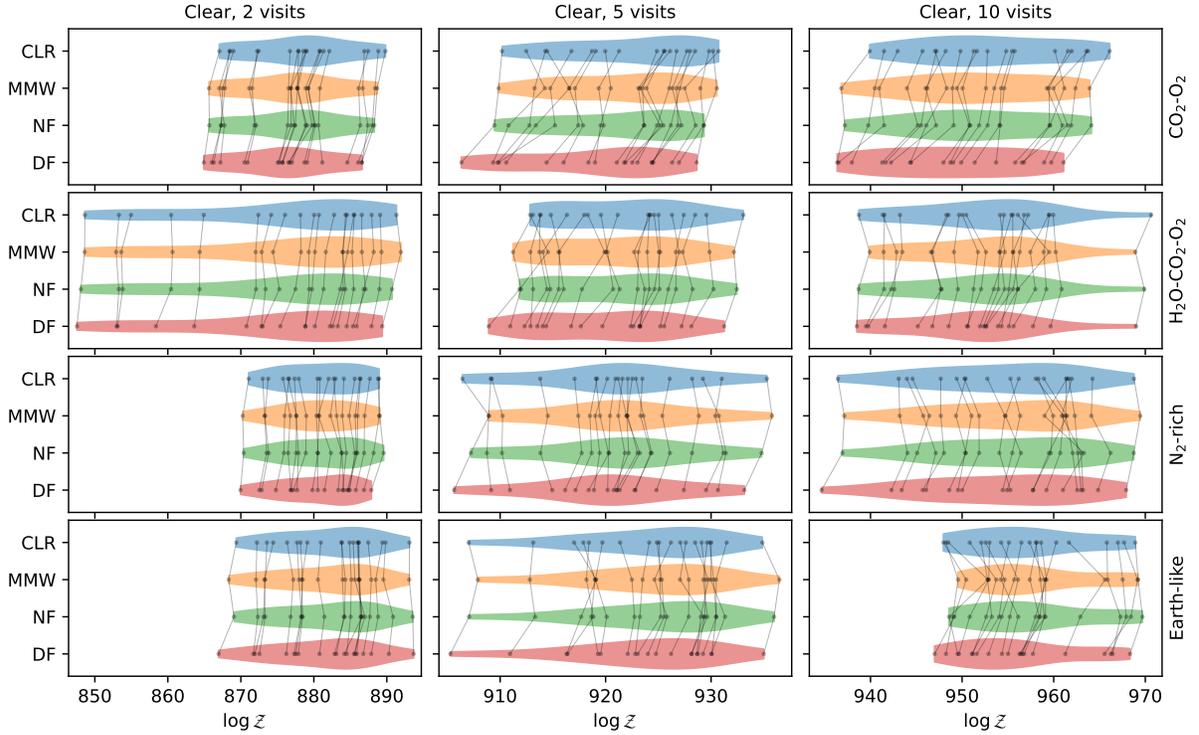


Figure 3.10: Violin plot comparing the Bayesian evidence ($\log \mathcal{Z}$) for retrievals for 20 noise instances. The connected lines indicate each noise instance. A higher $\log \mathcal{Z}$ indicates that the prior set by the parameterization weighs the peak likelihood higher. The values shown here have typical uncertainties of 0.3. No one parameterization method outperforms others across the board, as most of the evidence is contained in the likelihood peak.

atmospheric composition of a terrestrial exoplanet. This discussion extends to other parameters, such as choosing how the thermal profile is parameterized or how the 3D-ness of the atmosphere is parameterized, but we will focus on atmospheric composition in this work [113, 127, 128], as atmosphere composition is typically the first-order scientific question to be answered, especially for rocky planets [e.g. 4, 129].

In the absence of a previous observation, we want to adopt a prior that is either as uninformative as possible or reflects some broad intuition about the planet that is motivated by the current understanding of similar planets. An example of the former (ignorance prior) is using a uniform prior that is as wide as possible (wherein the limits are often set by physics or the limita-

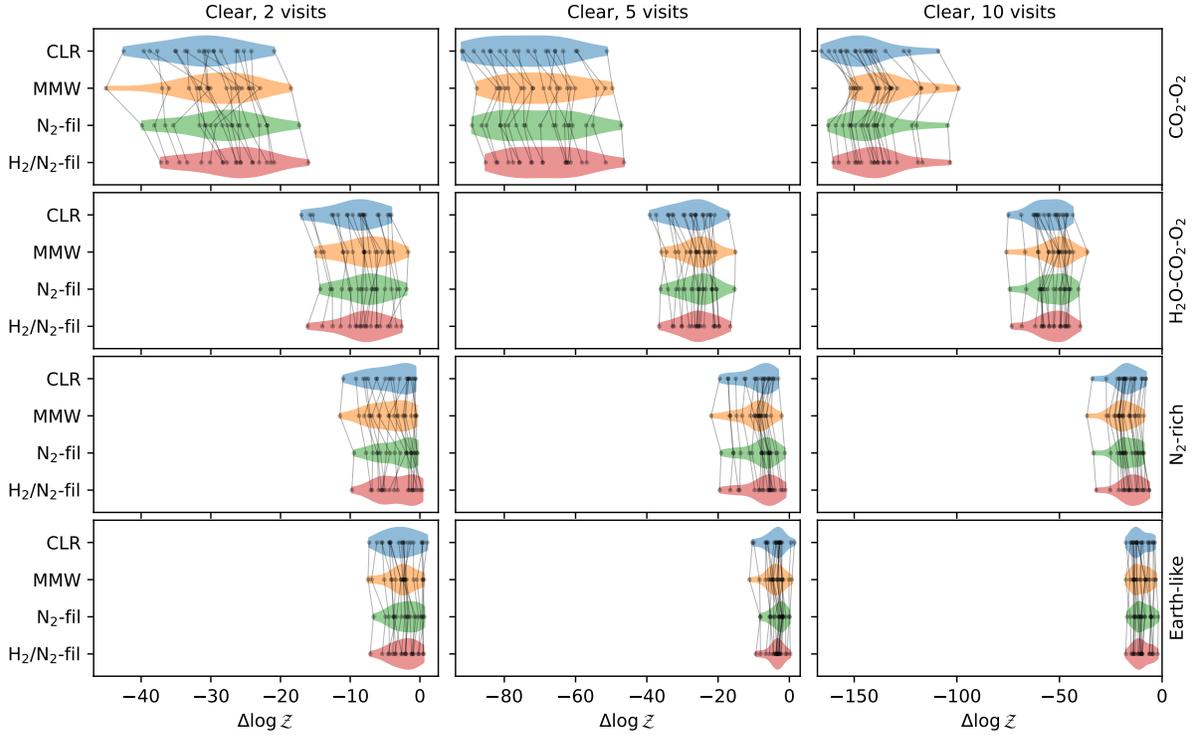


Figure 3.11: Violin plot comparing the changes in Bayesian evidence ($\Delta \log \mathcal{Z}$) with and without CO_2 for retrievals for 20 noise instances. The connected lines indicate each noise instance. A smaller $\Delta \log \mathcal{Z}$ indicates that the inclusion of CO_2 is favored. The values shown here have typical uncertainties of 0.4.

tions of the retrieval code). Here, the parameterization sets the prior; for example, when retrieving on a parameter such as the temperature, the choice remains whether to use a uniform prior, log-uniform, uniform for inverse temperature, or yet another reparameterization. An example of the latter (intuition prior) is the choice of background gas when parameterizing the chemistry of hot Jupiters, atmospheres of which one can reasonably expect are dominated by H and He. When there is a well-founded previous measurement for a parameter, such as for radius or mass of the planet, Gaussian priors are also a popular choice.

The way to achieve a generically objective, uninformative prior is the Jeffreys prior, which is defined to the prior that is invariant under a change of parameterization. This is achieved by

constructing the prior from the likelihood function. For a parameter vector $\vec{\theta}$, the prior density $p(\vec{\theta})$ is given by:

$$p(\vec{\theta}) \propto \sqrt{\det \mathcal{I}(\vec{\theta})}, \quad (3.1)$$

where $\mathcal{I}(\vec{\theta})$ is the Fisher information matrix, or the expected curvature of the likelihood function evaluated for a parameter vector. Generally, deriving the Jeffreys prior is not an analytically tractable problem, and extension to multiple parameters is not straightforward. Regardless, we invoke the Jeffreys prior to build intuition into what we deem to be an objectively uninformative prior.

The underlying intuition here is that, if we wish to not bias the sampler as it estimates the posterior by supplying an informative prior, we must set the prior such that the prior density is higher for the parameter space volume that is easier to find evidence against and lower for the volume that is harder to find evidence against. Otherwise the posterior will generically favor the parameter space volume that is harder to find evidence against. As such, the Jeffreys prior is proportional to the expected curvature of the likelihood function, which measures the sensitivity of the spectrum to the change in a parameter.

Applied to the exoplanetary atmospheric retrievals, we can treat the forward model as the change of coordinates from parameter space to spectrum space and the retrieval as an estimation of the mean of a multivariate normal distribution with a fixed standard deviation from which the observation data is sampled, as assumed in the likelihood function. The Jeffreys prior for estimating the mean of a multivariate normal distribution is a uniform distribution across the real space. This implies that, in the context of atmospheric retrievals, the maximally uninformative

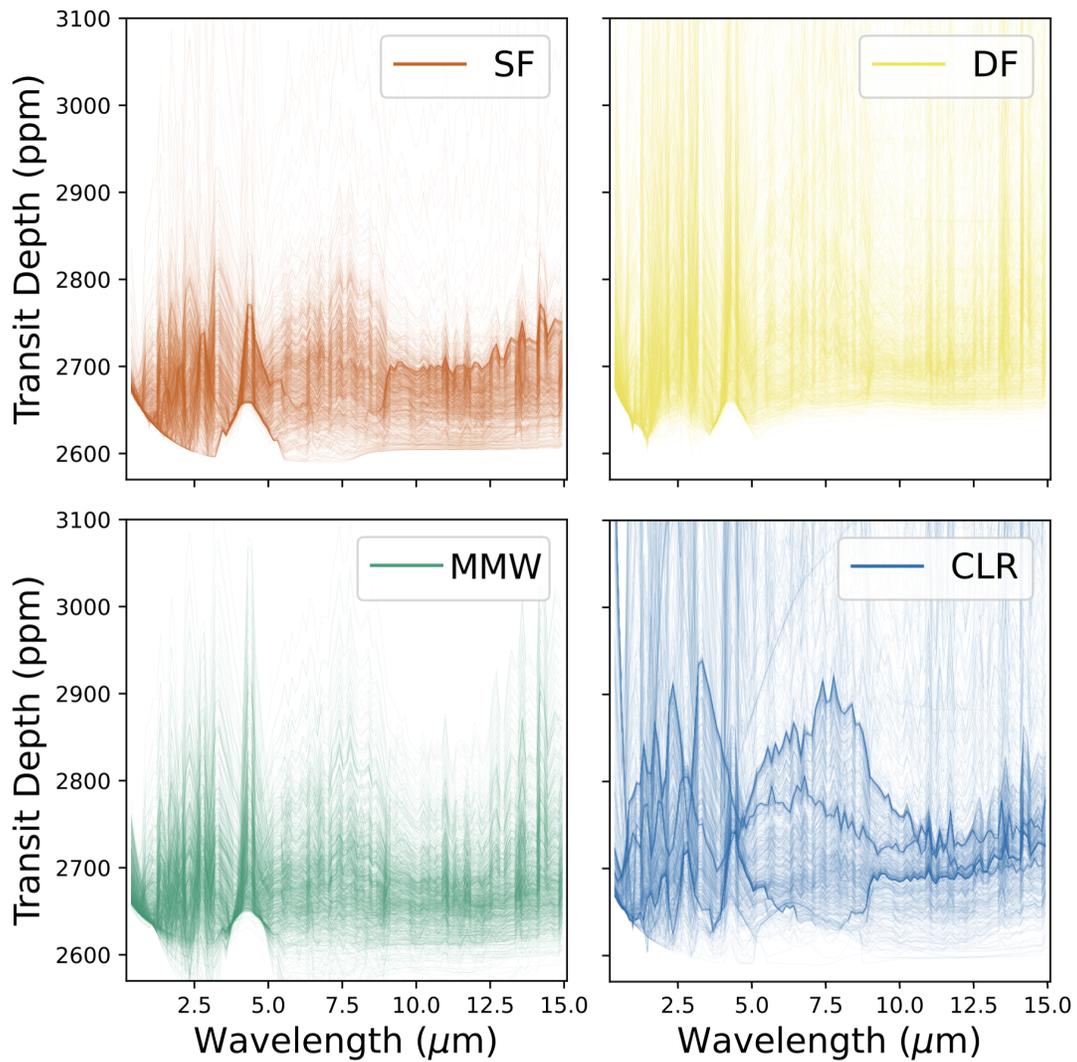


Figure 3.12: The density distributions of spectra sampled from each parameterization. Each panel shows 3000 random instances of compositions drawn from the prior for a GJ 1132 b-like planet. In the distribution for CLR (bottom right), the three visually apparent modes correspond to H₂O-, CO₂-, and CH₄-dominated atmospheres.

prior for a set of parameters is one that produces a uniform distribution of spectra in the spectrum space, at the very least once marginalized in each dimension. Notionally, this also corresponds to the ignorance prior on a spectrum before it is observed – the spectrum can be anything, and thus a uniform distribution over the entire spectrum space is appropriate. Of course, one could argue the last point is somewhat subjective; we know, for instance, that a transmission spectrum will not look like an emission spectrum. However, we argue that this still qualifies as an informed prior, even if it may be completely justified. Seen this way, an uninformative prior is ultimately reduced to what is physically or semantically allowed. The sum of mixing ratios cannot but be unity; a transit depth cannot be negative.

In Figure 3.12, we show distributions of transmission spectra of a planet the atmospheric abundances of which were sampled using each parameterization, with the same bulk properties. A few observations can be made. Firstly, it is visible that the CLR method is more likely to sample H₂O-, CH₄-, and CO₂-dominated atmospheres, as represented by the 3 thicker lines. The upturn near unity linear abundance space seen in Figure 3.2 constitutes the mode where the atmosphere is dominated by one species, which may be spectroscopically active. This mode is not existent in the trace-based methods (Figure 3.1), where we designate the filler gas(es) to be spectroscopically inactive. The resulting distribution of spectra therefore does not have the modes dominated by a spectroscopically active gas.

Secondly, the overall extent of the distribution, in terms of absolute transit depth, is set by the atmospheres with low-MMW, high-scale height, and thus large-features. As such, the two-filler and CLR methods have a greater extent than SF, as H₂ can be a major species in two-filler and CLR but not single-filler. Moreover, comparing the the two-filler method to the CLR method, the former has a higher density of spectrum of low-MMW atmospheres, as H₂ is continuously

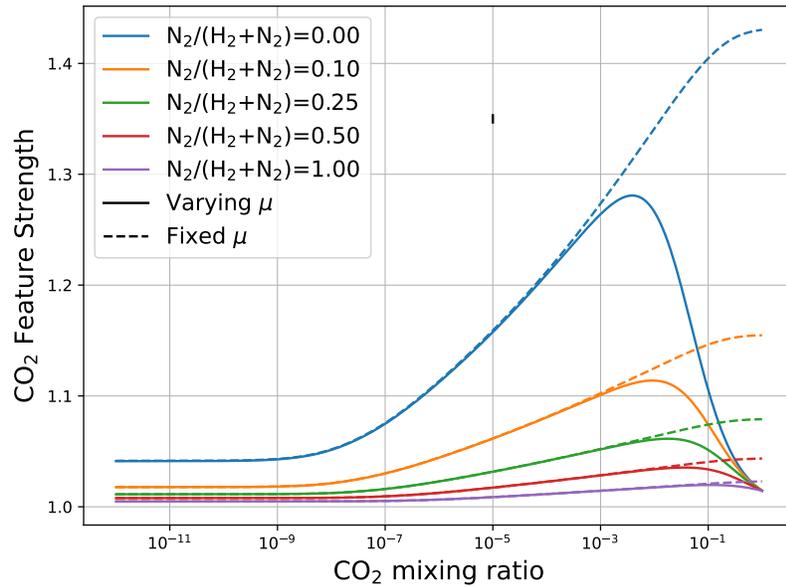


Figure 3.13: CO₂ feature strength as a function of the CO₂ mixing ratio for a different mixtures of background gas. The feature strength is obtained by the ratio of the transit depth at 4.38 μm (in band) and 1.39 μm (out of band). To show the effects of changing opacity and mean molecular weight (MMW), the solid lines correspond to atmospheres in which MMW is calculated self-consistently from the mixing ratios, while the dashed lines correspond to atmospheres for which MMW was fixed to that calculated from the background gas, even at higher mixing ratios of CO₂. The small black vertical bar represents the uncertainty on the features strength obtained from the uncertainties on the transit depth for the simulated observations as described in 3.2.2.

varied from 0 to 1 in the former, whereas H_2 has 1 in 6 chance to be the dominant gas in the latter.

Thirdly, the main effect of including MMW is that it vertically smears out the distribution of spectra, as the scale height is now freely varied. This effect is already present in the two-filler, whose scale height varies somewhat linearly according to the sampled H_2 abundance.

In practice, the Jeffreys prior is impossible to implement exactly due to complexity of the forward model, such as degeneracies or, for composition, simplicial structure. It does imply, however, that the uninformative prior for a parameter depends on its effect on the spectrum. Specifically, it can inform the critical choice such as whether the prior for the abundance of a species should resemble a log-uniform or uniform prior. In Figure 3.13, we show the feature strength of a spectrum (the difference between the depths in/out of a CO_2 feature) as a function of log-abundance of CO_2 . This feature strength is used as proxy for how the spectrum is changing overall. A mixture of varying proportion of N_2 and H_2 is used as the filler gas. The feature strength increases under $\log[\text{CO}_2] \lesssim -2$ as the opacity increases, and decreases over $\log[\text{H}_2\text{O}] \gtrsim -2$ as the change in CO_2 abundance influences the MMW.

Given that the Jeffreys prior is proportional to the rate of change of a parameter, a log-uniform prior should be used for species that dominantly affect the spectrum via opacity, whereas a linearly uniform prior should be used for species that dominantly affect the spectrum via mean molecular weight. Additionally, when using a log-uniform prior, the minimum abundance should be set where there is little change in the spectrum with the change in parameter compared to the data uncertainty [46].

3.5 Summary and Future Work

In this section, we summarize the conclusions from the current study.

We have demonstrated, when retrieving on the spectra of atmospheres of high mean molecular weight, the parameterization of the abundances can have consequential effects, as the bulk of the atmosphere is not known. This is most apparent when viewed in both log and linear abundance. The most prominent effect is that the symmetric method prefers to inflate the abundance of a spectroscopically active gas for which there is evidence in the spectrum to constitute the bulk of the atmosphere, rather than invoking the appropriate abundance of the spectroscopically inactive species. This is made possible because the atmospheric spectrum is not sensitive to the change in the abundance of the spectroscopically active gas near unity.

Unsurprisingly, the choice in parameterization method is the most relevant when retrieving low S/N data. At high S/N with 10 transits for GJ 1132 b, each parameterization correctly estimates the log-abundances of spectroscopically active species in all cases.

In the current work, we have treated O₂ as a spectroscopically active species, as it has a small feature near 0.76 micron. We find that, again unsurprisingly, this feature cannot be used to constrain the O₂ abundance.

We find that when using Bayesian evidence to establish whether a species should be included, the choice in the parameterization adds, effectively, a random scatter to the $\Delta \log \mathcal{Z}$. We stress that Bayesian evidence cannot be used to *infer* a correct parameterization.

We have also shown that, in the context of atmospheric retrievals, the Jeffreys prior corresponds to one that, when sampled from, produces the most uniform distribution of spectra. Amongst the priors compared in this work, the **two-filler** method most closely corresponds to the

Jeffreys prior. However, it is also not a given that it will produce the most correct posteriors in all cases, and we argue that there is no straightforward abundance parameterization that perfectly corresponds to the ideal Jeffreys prior for transmission spectroscopy.

Therefore, we argue that when applying atmospheric retrievals to terrestrial planets, the issue of priors remains a fundamental problem alongside complications of astrophysical origin. As such, we recommend that when performing retrievals, one test various parameterization methods to test the robustness of the inferences. We leave for future work to test whether techniques such as importance sampling [130, 131] can be applied here to swap out priors, post-inference.

Chapter 4: Constraining the Thickness of TRAPPIST-1 b's Atmosphere from its JWST Secondary Eclipse Observation at 15 micron

4.1 Introduction

We have now entered the era of JWST, and with it comes the potential to perform the first meaningful characterization of terrestrial (i.e., rocky) exoplanets. Among the possible rocky planet targets for JWST, those in the TRAPPIST-1 system are some of the most promising for atmospheric characterization due to their very favorable planet-to-star size ratios [132]. The system is also of extreme interest because it hosts multiple terrestrial planets, including several that reside in or near the habitable zone [133]. Recently, Greene et al. [2] measured the thermal emission from the innermost planet, TRAPPIST-1 b, and found that its 15- μm brightness temperature is consistent with the planet being a bare rock, devoid of any atmosphere at all.

Thermal emission measurements of presumed tidally-locked planets, such as those produced by Greene et al. [2] for TRAPPIST-1 b, are a productive avenue for confirming whether rocky exoplanets possess atmospheres [31, 32]. By measuring the planet's dayside temperature via secondary eclipse observations, one can constrain the presence and thickness of the atmosphere in the following sense: atmospheres serve to lower the dayside emission temperature below what would be expected for a bare (and dark) rocky surface. Even moderately thick atmospheres

transport considerable heat away from a tidally-locked planet’s dayside [53]. Reflective aerosols, another signpost of a planet possessing an atmosphere, also serve to lower the dayside temperature by reflecting incoming stellar radiation back to space [31]. The maximal dayside effective temperature, corresponding to no atmosphere and a zero-albedo surface is:

$$T_{max} = T_* \sqrt{\frac{R_*}{d}} \left(\frac{2}{3}\right)^{1/4} \quad (4.1)$$

where T_* and R_* are the stellar effective temperature and radius, and d is the planet-star separation. For TRAPPIST-1 b, $T_{max} = 508 \pm 6$ K, whereas the $15 \mu\text{m}$ brightness temperature reported by Greene et al. is 503^{+26}_{-27} K, fully consistent with the no-atmosphere scenario.

From a theoretical standpoint, it is unclear whether terrestrial planets orbiting M-dwarfs should be expected to possess atmospheres. There are studies that go both ways. Atmospheric loss processes should be efficient for planets orbiting active M-dwarf host stars, but some planets may be able to retain their atmospheres or renew them via outgassing following a decline in stellar activity with age [e.g. 5, 134, 135].

Observationally, to-date there are no studies that definitively confirm the presence of an atmosphere on a rocky exoplanet. Flat transmission spectra are the norm [e.g. 136, 137, 138, 139, 140, 141], and the few studies that have claimed detections of atmospheric spectral features for terrestrial exoplanets have been called into question or have ambiguous interpretation [e.g. 122, 142, 143]. Thermal emission measurements of the planets LHS 3844b [144] and GJ 1252b [145] have found dayside temperatures that are consistent with the no-atmosphere limit, the former by way of a full-orbit phase curve. It stands to reason that less irradiated planets should be less susceptible to atmospheric loss, but TRAPPIST-1 b is the coldest planet yet to be subjected to the

thermal emission test for possessing an atmosphere, yielding the same result of no apparent sign of a gaseous envelope.

In this Letter we quantify the range of atmospheres and surfaces that are consistent with the Greene et al. [2] measurement of TRAPPIST-1 b’s secondary eclipse depth at $15\ \mu\text{m}$. We show in what follows that thick atmospheres can be definitively ruled out by this single data point. Given the range of scenarios that we still find to be consistent with the data, we also predict the degree to which further observations, including planned measurements at $12.8\ \mu\text{m}$, will be able to distinguish among the remaining plausible atmospheres and surfaces.

4.2 Methods

In this section, we describe our model and parameter choices. To calculate the eclipse spectrum of different surfaces and atmospheres, we use HELIOS, an open-source 1D radiative transfer code that computes the thermal profile of a planetary atmosphere in radiative-convective equilibrium [35, 36, 37]. Most of our approach closely follows Whittaker et al. [51], which performed a similar analysis for the *Spitzer* observation of LHS 3844 b, and we refer the readers to that work for more details of the modelling.

One key detail worth mentioning here is that we calculate the heat redistribution factor (f) self-consistently with the radiative transfer using the analytical approximation in Koll [53, equation 10]. In the approximation, f depends on the equilibrium temperature, the surface pressure, and the longwave optical depth at the surface; HELIOS has the ability to iterate to a value of f that satisfies global energy balance. We note a caveat that this method subtracts the approximated transported heat from the incident stellar flux to calculate the dayside energy budget, but does not

consider the vertical dependence of the day-to-night heat flow; hence the redistribution could be construed to happen either uniformly or at the top of the atmosphere in our models.

We model a range of surface pressures that is broad enough span full redistribution ($f = 1/4$) to no redistribution ($f = 2/3$), resulting in a surface pressure grid of 10^{-4} bars to 10^2 bars, spaced at 1 dex. For the composition of the atmospheres, in addition to a 100% CO_2 atmosphere, we choose to vary the abundance of trace CO_2 , at 1 ppm, 100 ppm, and 1%, against background gases of N_2 , O_2 , and H_2O . Moreover, we also consider atmospheres containing a range of other trace gases plausible in secondary atmospheres [51, 134, 146], which may not necessarily absorb at $15 \mu\text{m}$ but may be detected via observations at other wavelengths. For this purpose, we adopt the same trace abundance grids (i.e. 1 ppm, 100 ppm, 1%) for CO , CH_4 , H_2O , and SO_2 , against a background gas of N_2 for the former two and O_2 for the latter. SO_2 is unique in that it has broad infrared absorption features just outside the $15\text{-}\mu\text{m}$ bandpass, which produce interesting implications for observations at $15 \mu\text{m}$; we discuss this further in Section 4.3. For all models, we assume an intrinsic temperature of $T_{\text{int}} = 0\text{K}$.

For all of the atmosphere models, we adopt a surface albedo of 0 (i.e. a true blackbody), to produce the maximum limit on the atmospheric pressure consistent with the observation; any value of non-zero albedo will dilute the energy budget and decrease the eclipse depth, thereby making a model at a given atmospheric pressure even less consistent with the observation.

Given that TRAPPIST-1 b's dayside temperature is consistent with the no-atmosphere limit, we also explore a number of bare surface models that have no atmospheres at all. Here the eclipse spectrum instead arises due to the wavelength-dependent albedo spectrum of the surfaces. We consider six surfaces that are plausible, given the level of irradiation received by TRAPPIST-1 b: basaltic, ultramafic, feldspathic, metal-rich, Fe-oxidized, and granitoid [31, 147]. We also run a

number of grey albedo surfaces at $A = 0.2, 0.4, 0.6, 0.8, 0.95$.

We adopt the stellar and planetary parameters as obtained in Agol et al. [148]. We use the SPHINX stellar model spectrum grid [149] interpolated to TRAPPIST-1 parameters assuming solar composition to calculate the thermal profile and the eclipse depth of the planet. SPHINX models are expected to better model the stellar spectra at such low temperature ranges than the typical PHOENIX models, using updated line lists [149]. Indeed, we find that the SPHINX model reproduces the observed stellar flux at $15 \mu\text{m}$ better than the PHOENIX model [to within 7% versus 13%; see Methods of 2]

After obtaining the eclipse spectra, we calculate the binned depth at the photometric band of F1500W; we integrate the planetary flux weighted by the bandpass function, then integrate the stellar flux weighted by the same function, and then obtain the ratio of the two. We perform the same calculation for F1280W to make predictions for upcoming observations. The F1280W bandpass lies outside the CO_2 absorption feature, and the difference between the two bandpasses serves as a metric to constrain either atmospheric pressure, CO_2 abundance, or both [30].

We calculate the brightness temperature (T_b) in the F1500W filter by determining the temperature of the blackbody whose eclipse depth (obtained via identical weighting and integrating as for the planetary flux) matches the observed eclipse depth. We note that this calculation differs slightly from the procedure followed by Greene et al. [2], who found the temperature of the blackbody whose per-frequency flux evaluated at the “effective” filter wavelength matched the observed per-frequency planetary flux. Our calculation leads to a best-fit brightness temperature of $T_b = 505 \pm 27\text{K}$, rather than the $T_b = 503_{-27}^{+26}\text{K}$ reported in Greene et al. [2]. Given the uncertainty, this minor discrepancy will not impact our analysis.

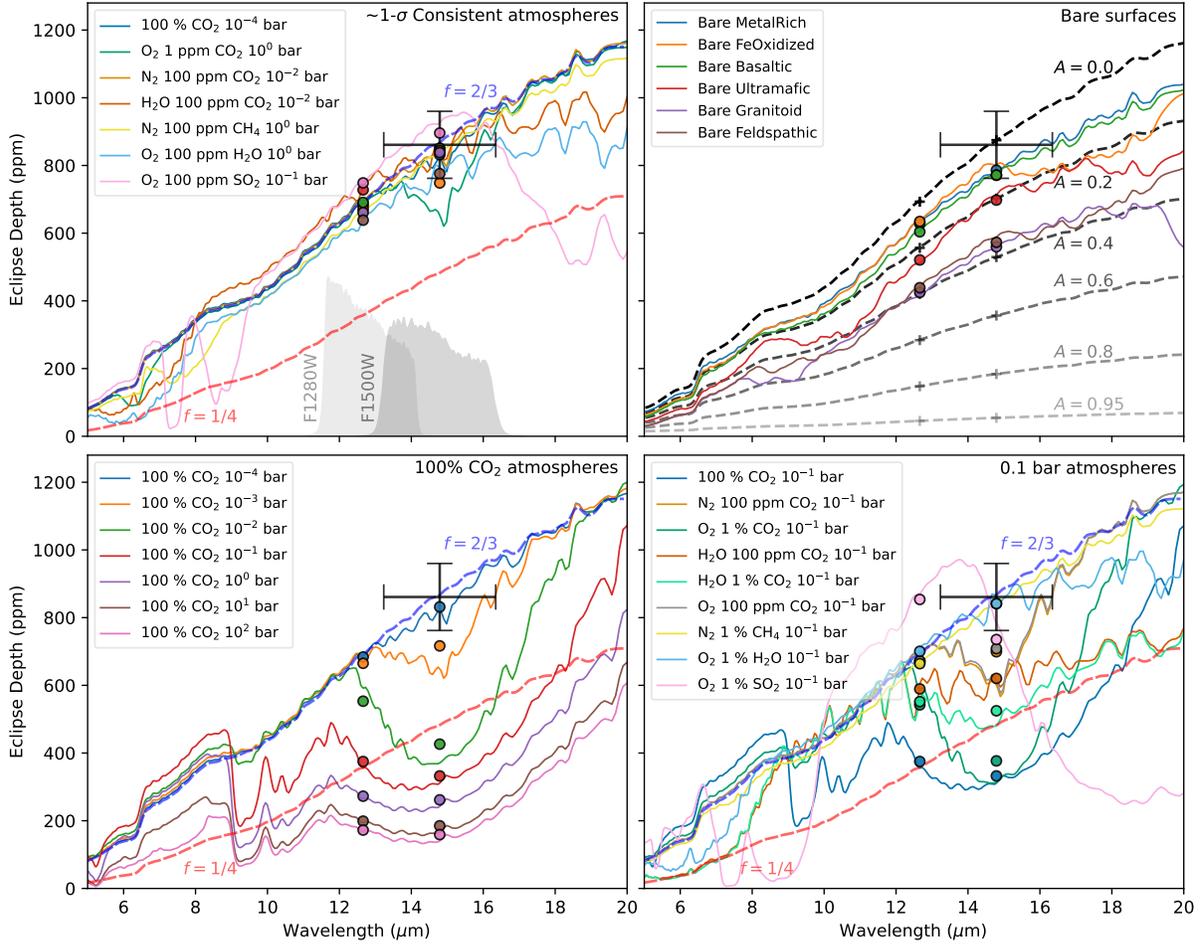


Figure 4.1: The eclipse spectra of various models run in this study. We show: a suite of atmospheric models that are $1\text{-}\sigma$ consistent with the observation (*top left*); bare surface models, which are all consistent with the observation (*top right*); 100 % CO_2 atmosphere models at various surface pressures (*bottom left*); and models with surface pressures of 0.1 bar, varying the compositions (*bottom right*). The compositions denote that the first species is the dominant species, with the second species in indicated trace amounts. The binned depths at F1500W and F1280W are shown as markers, as well as each bandpass function weighted by the stellar spectrum. We also show, in dashed lines, the eclipse depths resulting from blackbodies at 508 K (blue) and 400 K (red), corresponding to no redistribution ($f = 2/3$) and full redistribution ($f = 1/4$), respectively. On the upper right panel, dashed lines indicate grey albedo surface models. The features in the blackbody eclipse spectrum arise due to spectral features in the *stellar* spectrum.

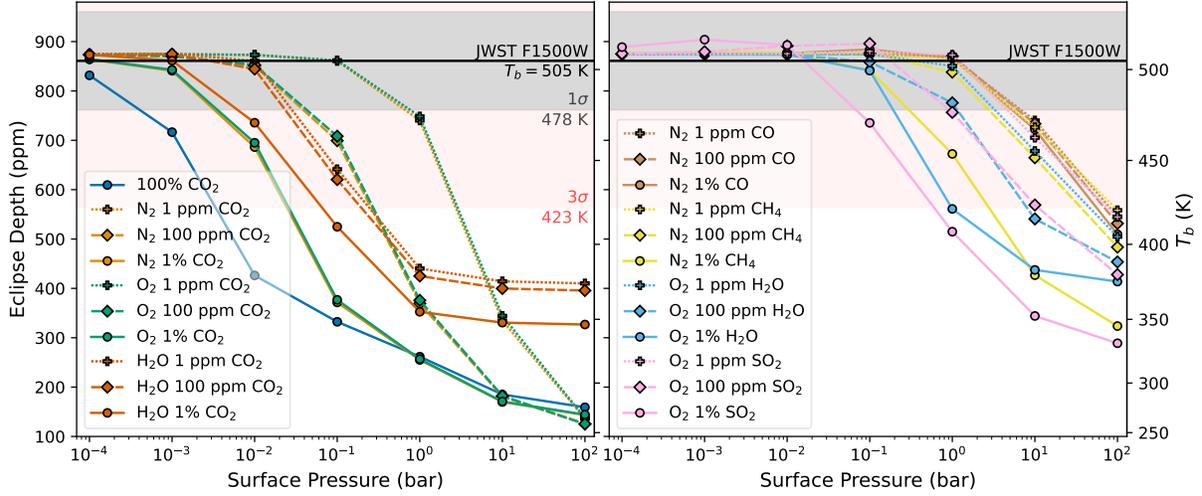


Figure 4.2: The binned eclipse depths and their brightness temperature in the F1500W band for all of the atmospheric models run, varying the pressure of the atmosphere at the surface. Models atmospheres that do and do not include CO₂ are shown in the left and the right panel, respectively. The measured eclipse depth from Greene et al. [2] is shown as the solid black line, and its 1- σ (grey) and 3- σ (red) uncertainties are also shown, as well as the corresponding brightness temperatures. The compositions denote that the first species is the dominant species, with the second species in indicated trace amounts. Atmospheres with ≥ 100 ppm CO₂ are consistent with the measurement at 1 σ only if the atmospheric pressure is less than 0.1 bar.

4.3 Results

4.3.1 Atmospheric Thickness and Surface Composition

Our results support the general conclusion from Greene et al. [2] that TRAPPIST-1 b does not possess a thick atmosphere. We will present the maximum atmospheric thickness consistent with the observed eclipse depth of 861 ± 99 ppm for each set of model composition and also highlight interesting behaviors from a theoretical perspective. We show the eclipse spectra for selected atmospheric and surface models in Figure 4.1 and the binned eclipse depths for all of the atmospheric models in Figure 4.2, varying the composition and the surface pressure. The accompanying temperature-pressure (T-P) profiles for each of the atmosphere models are shown

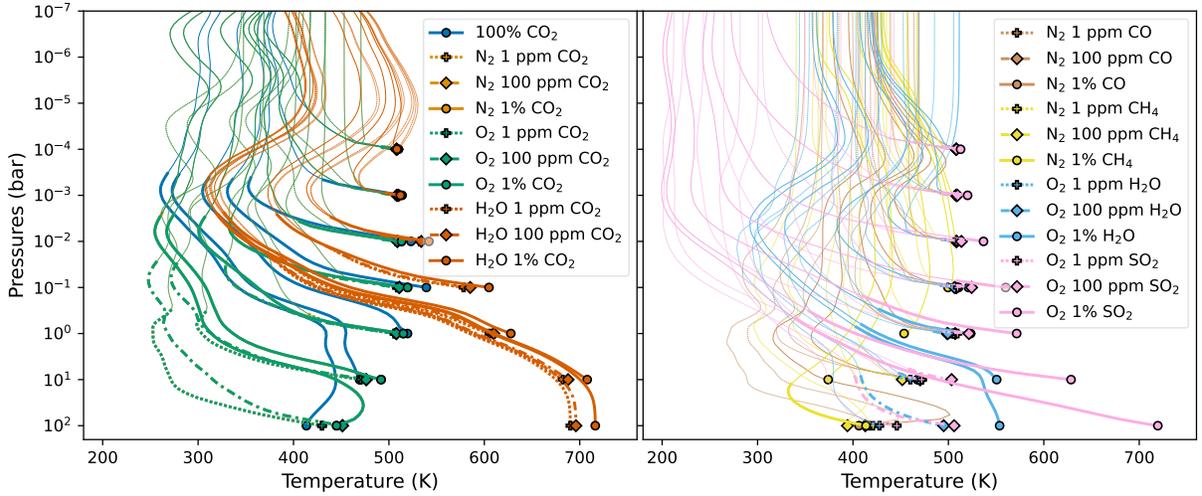


Figure 4.3: The temperature-pressure (T-P) profiles of the model atmospheres in radiative-convective equilibrium. Models atmospheres that do and do not include CO_2 are shown in the left and the right panel, respectively, similarly to Figure 4.2. The optically thick region of the T-P profiles below the photosphere ($\tau = 2/3$) at $\lambda = 14.79 \mu\text{m}$ are shown with thick lines. The markers indicate the surface pressure of each model atmosphere. The compositions denote that the first species is the dominant species, with the second species in indicated trace amounts. The N_2 and O_2 -dominated atmospheres completely overlap in the left panel. It can be seen that while near-infrared absorbers such as H_2O can cause thermal inversions, they occur at regions where the atmosphere is optically thin and hence will not result in emission features in the spectra. For most of the models that do not contain CO_2 , the atmosphere is optically thin in the F1500W bandpass down to the surface.

in Figure 4.3.

4.3.1.1 Atmospheres with CO_2

We posit that TRAPPIST-1 b should realistically have at least moderate amounts of CO_2 if it does possess an atmosphere. This statement is in line with theoretical studies of the atmosphere of TRAPPIST-1 b and in general of rocky exoplanets receiving a comparable degree of irradiation [134, 150, 151]. CO_2 is robustly expected to be present in non-hydrogen-dominated atmospheres [e.g., as indicated for TRAPPIST-1 b from its transmission spectrum; 152], and the gas is robust against various escape processes, although photodissociation can deplete its abundance.

Pure CO₂ atmospheres are 1- σ consistent with the eclipse measurement for surface pressures up to 0.4 mbar and 3- σ consistent up to 3 mbar (Figure 4.2), indicating that even a Mars-like thin atmosphere ($P_{\text{surf}} = 6.5$ mbar) composed entirely of CO₂ is unambiguously ruled out. To first order, the secondary eclipse depth depends on the *partial pressure* of CO₂, so the atmosphere may be thicker if the CO₂ abundance (i.e. its mixing ratio) is smaller. N₂ or O₂-dominated atmospheres with ≥ 100 ppm of CO₂ are 1- σ consistent at 0.04 bar at most, and 1 bar atmospheres are ruled out by more than 3 σ .

The presence of H₂O has a non-trivial effect on the eclipse spectrum as it both increases the absorption and changes the thermal structure. For instance, at a surface pressure of 0.1 bar, H₂O-dominated atmospheres with 1 ppm or 100 ppm CO₂ have deeper eclipse depths than the corresponding O₂ or N₂-dominated atmospheres, while the one with 1% CO₂ has a shallower depth than atmospheres with the other background gases. Additionally, the lower atmosphere becomes much hotter for the thicker H₂O-dominated atmospheres due to greenhouse heating being more effective than the cooling of day-night redistribution.

H₂O is also interesting in that it can generate thermal inversions in planets orbiting M stars [36]. Thermal inversions are interesting in the context of the Greene et al. [2] secondary eclipse measurement because they have the potential to reverse absorption features into emission, opening a possibility that the high observed 15- μm brightness temperature could be due to a CO₂ *emission* feature originating from from a thick(er) atmosphere. For TRAPPIST-1 b, we indeed find that H₂O causes thermal inversions (Figure 4.3), but they occur in the upper atmosphere well above the IR photosphere and thus do not significantly impact the shape of the 15- μm CO₂ feature, which uniformly appears in absorption in all of the models we have produced. We have also experimented with different mixtures of O₂, H₂O, and CO₂ (not shown), but find that no

combination leads to emission features. In fact, in Figure 4.4, one can see that the brightness temperature at 15 micron is lower than that at 12.8 micron for every model, indicating CO₂ absorption, rather than emission, is being observed.

4.3.1.2 Atmospheres with no CO₂

While less plausible chemically, atmospheres that do not contain any CO₂ at all remain consistent with the secondary eclipse measurement to higher surface pressures. Atmospheres that have CO or CH₄ as the trace gas are 1- σ consistent to 1 bar for all trace abundances, except the 1% CH₄ model which has a shallower depth that is 2- σ consistent. In Figure 4.3, it can be seen in the right panel that all of these atmospheres except the 1% CH₄ 10² bar model remain optically thin in the 15 μ m bandpass down to the surface, and the change in eclipse depth with surface pressure is due to the cooling effect of redistribution. Atmospheres with trace H₂O behave similarly except that the 1% H₂O atmospheres becomes optically thick at atmospheric pressures around 0.1 bar, and the eclipse depth is already > 3- σ inconsistent for a surface pressure of 1 bar.

Atmospheres with trace SO₂ behave somewhat differently since SO₂ has a broad absorption feature at wavelengths just redward of the 15- μ m bandpass. For moderate SO₂ abundances (e.g. the pink line for the 100 ppm 0.1 bar atmosphere in the top left panel of Figure 4.1), the strong absorption at \sim 18–20 μ m pushes more flux into the 15- μ m bandpass, leading to *increased* planetary emission over the wavelength range of the Greene et al. [2] secondary eclipse observation. The emission from a transparent spectral window is therefore a plausible mechanism for increasing the secondary eclipse depth in a single bandpass, but it comes at the cost of sharply reduced fluxes at other wavelengths; this effect can therefore be diagnosed with additional spectroscopic

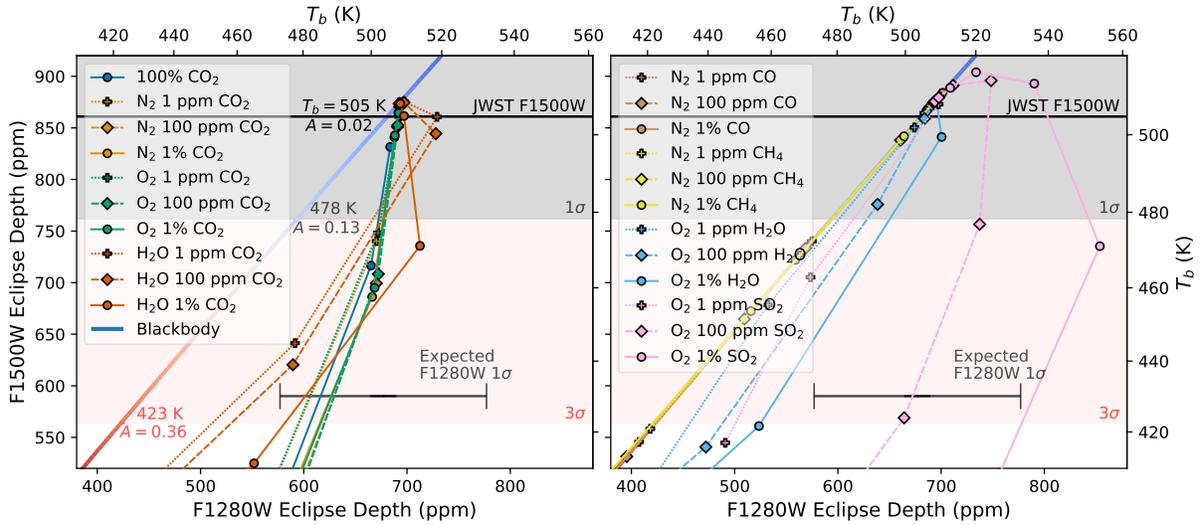


Figure 4.4: A color-color-like diagram of predicted binned eclipse depths in the F1280W band (horizontal axis) and the binned F1500W eclipse depths for all of the model atmospheres, along with their brightness temperatures (T_b) in each band. Models atmospheres that do and do not include CO_2 are shown in the left and the right panel, respectively. The measured eclipse depth from Greene et al. [2] is shown as the solid black line, and its $1\text{-}\sigma$ (grey) and $3\text{-}\sigma$ (red) uncertainties are also shown. The vertical axis is identical to Figure 4.2, but is zoomed to focus on models consistent with the F1500W observation, alongside the expected F1280W uncertainty (~ 100 ppm) shown as an errorbar. The binned eclipse depths for a blackbody over a range of temperatures is shown as a multi-colored line. The temperature of the blackbody can be read off from the T_b in either axes, by definition. The corresponding T_b and bond albedo (A) at each confidence interval is also shown. All models that include CO_2 (in the left panel) lie on the right side of the blackbody line, indicating a higher T_b in the F1280W than in F1500W due to the CO_2 absorption at 15 micron. The compositions denote that the first species is the dominant species, with the second species in indicated trace amounts. As one follows each composition line, atmospheric pressure starts at 10^{-4} bar close to the observed F1500W measurement and increases in 1-dex intervals as in Figure 4.2 with generally decreasing $15\text{-}\mu\text{m}$ eclipse depths. We do not show the bare surface depths in this figure, but they lie close to the blackbody line and deviate less than 25 ppm in either bandpass.

observations. For higher SO_2 abundances however, the absorption feature is strong enough to affect the $15\text{-}\mu\text{m}$ bandpass, and it therefore has the opposite effect of reducing the eclipse depth in the F1500W filter (Figure 4.1, pink line in bottom right panel). This indicates that the nature of the absorber needs to be very finely tuned to match the Greene et al. [2] measurement.

4.3.1.3 Bare surfaces

If TRAPPIST-1 b truly has no atmosphere whatsoever, we find that the F1500W measurement is consistent with a bare rock planet with a basaltic, Fe-oxidized, or metal-rich surface to within 1σ , while granitoid and feldspathic surfaces are ruled out at more than 3σ (Figure 4.1, top right panel). The latter two materials have high albedos around $1\ \mu\text{m}$ where the luminosity of the TRAPPIST-1 host star is greatest [31, 147], thus reducing the energy received by the planet and lowering the temperature at which it radiates. The fact that we can rule out some surface compositions demonstrates the utility of secondary eclipse spectroscopy for constraining the surface properties of rocky exoplanets. However, Mansfield et al. [31] point out that granitoid and feldspathic surfaces (the ones that we rule out here) are also among those that are implausible for hot rocky planets like TRAPPIST-1 b, as they either require liquid water to form or they are unlikely to be able to form on larger planets [153]. Among grey surfaces, we find that the best-fit surface albedo is 0.02 ± 0.11 .

4.3.2 Prospects for Future Observations

Given the various atmospheres and surfaces that remain consistent with the Greene et al. [2] $15\ \mu\text{m}$ secondary eclipse measurement, we investigate here the possibility that additional observations could help to further constrain the properties of TRAPPIST-1 b. In particular, five secondary eclipses are slated to be observed with MIRI F1280W filter centered on $12.8\ \mu\text{m}$ to provide a second spectroscopic data point for TRAPPIST-1 b's thermal emission. In Figure 4.4 we show the eclipse depths from our models binned to the F1280W bandpass against the binned eclipse depth in the F1500W bandpass.

The F1280W is intended to observe the eclipse depth out of the CO₂ band such that the difference between the two provides a constraint on the atmospheric pressure and possibly composition, but the very high eclipse depth of F1500W alone already provides a firm constraint on the brightness temperature and hence the atmospheric pressure. Assuming an observation uncertainty comparable to that of F1500W (99 ppm), the F1280W secondary eclipse is unlikely to help further distinguish between, for example, a very thin 10⁻⁴ bar 100% CO₂ atmosphere, a 1 bar O₂-dominated 1 ppm CO₂ atmosphere, a 1 bar N₂-dominated atmosphere with 100 ppm CH₄ as they all fall roughly within a span of 100 ppm. Therefore, we conclude that the F1280W observation will be most useful for validating the high brightness temperature of TRAPPIST-1 b as observed by F1500W.

Indeed, in Figure 4.1, most 1- σ consistent spectra follow the $f = 2/3$ blackbody spectrum (blue dashed line) closely down to 10 μm , and only at shorter wavelengths do spectroscopic absorption features appear. However, due to the small eclipse depth at these wavelengths, spectroscopy using MIRI LRS with nominal uncertainty of (say) 30 ppm at a spectral resolution of $R = 10$ will be able to distinguish only between end-member cases at best rather than tightly constraining the composition and the surface pressure. Namely, if the planet has H₂O, CH₄, or SO₂, absorption features between 5–10 μm , MIRI LRS could be used to distinguish between an airless blackbody and a thin atmosphere.

As for distinguishing among bare rock surfaces, the additional F1280W observation is unlikely to be helpful for this purpose as the binned eclipse depths of consistent surfaces are very similar (Figure 4.1). The surfaces are generally difficult to distinguish across all wavelengths that MIRI can observe in.

4.4 Discussion and Summary

We have shown that, based on the Greene et al. [2] secondary eclipse observation at $15\ \mu\text{m}$, TRAPPIST-1 b does not appear to host a thick atmosphere. Formally, our models rule out atmospheres with at least 100 ppm CO_2 thicker than 0.3 bars at 3σ . For a 100% CO_2 atmosphere (i.e., a Mars or Venus-like composition), the atmosphere must be less than 3 mbar thick at 3σ confidence to be consistent with the measured eclipse depth at $15\ \mu\text{m}$. We argue that TRAPPIST-1 b is unlikely to host an atmosphere devoid of CO_2 , and therefore atmospheres thicker than ~ 0.1 bar are ruled out. Various types of geophysically plausible rocky surfaces are all consistent with the Greene et al. [2] measurement, and the eclipse observation rules out less plausible granitoid and feldspathic surfaces. The best-fit grey surface albedo is 0.02 ± 0.11 .

The $1\text{-}\sigma$ consistent atmospheres and surfaces that we identify in this Letter will be difficult to distinguish with upcoming JWST observations except perhaps the very end-member scenarios. The predicted eclipse depths for the F1280W filter are close enough to each other to be within the uncertainty of the observation. MIRI LRS may be able to distinguish between a bare rock and a 0.1 bar H_2O -dominated atmosphere by measuring the eclipse spectrum from $5\text{-}10\ \mu\text{m}$, but there are many degenerate scenarios in between. Finally, the planned NIRISS SOSS observation of TRAPPIST-1 b via complementary measurements in *transmission* [154, Cycle 1 GO 2589] also aims to distinguish between a bare rock and a thin atmosphere. In the case of a clear atmosphere, transmission spectroscopy can generally provide a signal that is easier to interpret than that of thermal emission, since H_2O and CO_2 features should be detectable. Transmission spectroscopy is also more agnostic to the thermal structure of the atmosphere and could therefore provide a less ambiguous constraint on the composition. On the other hand, transmission spectroscopy of

small, rocky planets is challenging as the high mean molecular weight of secondary atmospheres and aerosols (if present) render the transmission spectrum closer to a flat spectrum, which is indistinguishable from a bare rock planet [155, 156, 157]. Additionally, host stellar effects also leave an imprint on the transmission spectrum, leading to spectral contamination that can be difficult to disentangle from *bona fide* atmospheric features [143, 158, 159].

We have neglected the radiative effects of clouds in our work. The clear atmosphere T-P profiles in Figure 4.3 do cross condensation curves such that water or sulfur clouds can form [150, 160]. However, clouds of appreciable column density will have higher albedos than rocky surfaces [31, Fig. 6] and are inconsistent with the observation, given such a low inferred albedo (even with the uncertainties taken into account). Additionally, climate modelling suggests that aerosols are unlikely to form in TRAPPIST-1 b [150]. As such, we find the scenario that the planet hosts an atmosphere with a reflecting cloud to be inconsistent with the Greene et al. [2] secondary eclipse measurement.

The F1500W observations of TRAPPIST-1 b demonstrate the utility of secondary eclipse observations for determining whether rocky planets possess atmospheres and for constraining their surface composition. Secondary eclipse observations will soon also be applied to other rocky planets around M dwarfs, with observation planned for more targets such as TRAPPIST-1 c [161, Cycle 1 GO 2304], Gl 486 b [162, Cycle 1 GO 1743], GJ 1132 b [163, Cycle 1 GTO], and LHS 3844 b [164, Cycle 1 GO 1846]. The latter three use MIRI LRS rather than F1500W; an identical analysis to the current work can be performed by binning the entire 8–12 μm LRS spectrum to create a single broad photometric bandpass [see e.g. §3 of 32], and the additional *spectral* information can be used to further constrain the composition of the atmosphere or the surface [51]. A larger sample of rocky planet targets observed in secondary eclipse will also help

to answer population-level questions of whether rocky planets around M dwarfs can really host atmospheres and identify the ideal parameter space for establishing regimes in which they can.

Chapter 5: Conclusions and Future Work

5.1 Conclusions

In this thesis, I demonstrate challenges and pathways in characterizing terrestrial planets with *JWST*. Characterizing terrestrial planets from their transmission spectra will be challenging, and using Bayesian approaches are inhibited. Instead, by focusing on studying eclipse observations of terrestrial planets at the population level, we may gain insight into their physical processes and formation histories.

As shown in Chapter 2, correlated systematics in transmission spectra, independent of whether their sources are instrumental or astrophysical in nature, can inhibit accurate characterization of the planet's atmospheres by mimicking aerosols or masking bona fide atmospheric features. For high signal-to-noise targets, this poses less of a problem, but for small planets, instrumental systematics or stellar contamination can be consequential. Notably, while Chapter 2 focused primarily on how correlated noise can impact retrievals for *HST* and *Spitzer* observations, recent *JWST* transmission observations of sub-Neptunes and terrestrial planets in e.g. May et al. [123], Moran et al. [143], Lim et al. [165], Wallack et al. [166] clearly show either stellar contamination or instrumental systematics.

As shown in Chapter 3, the problem of Bayesian priors and parameterization due to unknown mean molecular weight remains a challenge to applying robust retrieval techniques on

the atmospheric composition and inferring detection of molecules. This problem posits a fundamental challenge that can only be overcome by broader wavelength coverage and stacking more transits, making transmission spectroscopy an expensive approach to characterizing terrestrial planet atmospheres, especially if it is unknown a priori whether they have atmospheres at all.

Instead, using eclipse observations to constrain the dayside temperature can be used as an effective and efficient way to probe whether a terrestrial planet has an atmosphere or not, as shown in Chapter 4. Due to the narrow wavelength coverage and low signal-to-noise, it remains difficult to perform detailed characterization of their atmosphere or surfaces. This calls for a more population-level approach.

5.2 Future Work

Solving the supposed problem of rocky planets around M stars is an achievable goal within the lifetime of *JWST*, but it will require both a judicious planning of future observations, as well as better modelling of formation and evolution and physical processes to help understand the observations.

We need a framework for a judicious planning of targets, that will maximize the population-level information and will also take into account what observations are already available. Figure 5.1 shows the trend in the normalized brightness temperatures of all rocky planets observed in emission so far, as a function of their irradiation temperatures. The analytical models are fit to the data to show that there is a statistically significant drop in the brightness temperature at irradiation temperatures cooler than 800 K.

The population-level trend here is stitched together from a decade of observations, span-

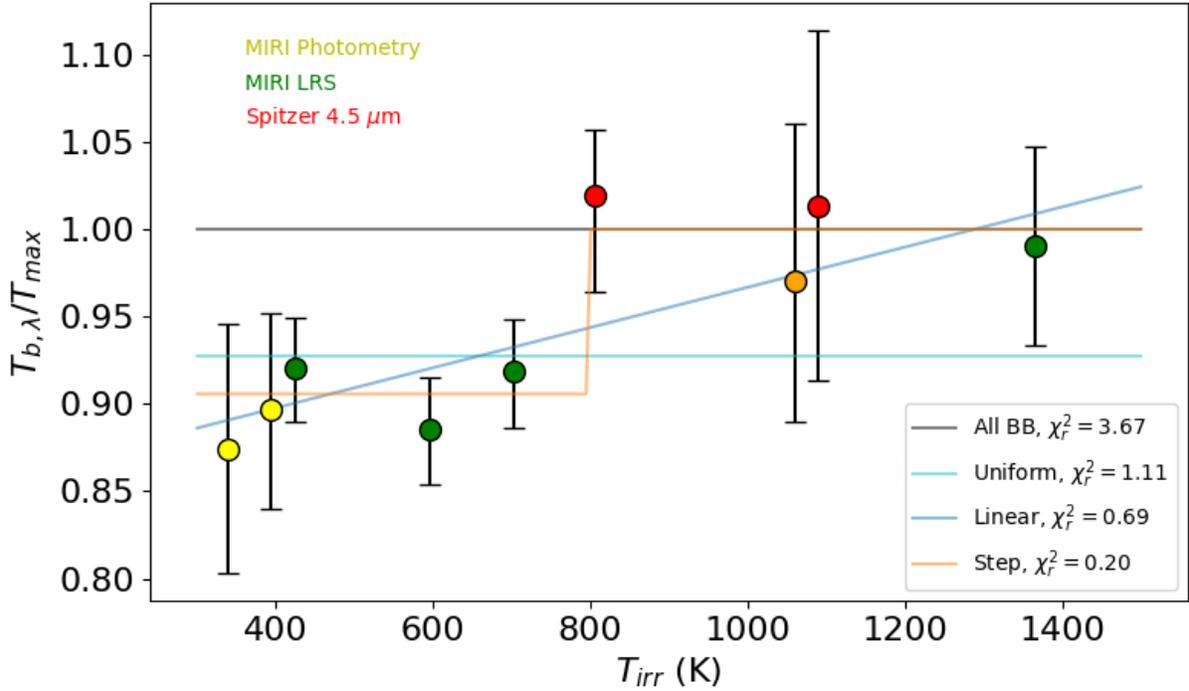


Figure 5.1: Observed trend in normalized brightness temperatures of rocky planets observed so far, as a function of their irradiation temperature.

ning different instruments and wavelength ranges, with the targets chosen primarily based on their signal size. Recent studies such as Batalha et al. [167] show that the set of targets chosen exclusively for the highest signal-to-noise may not be the best set of targets given a science goal. For instance, choosing targets based on some observation metric tends to bias the samples selection—this is somewhat obvious in our current context, as this will choose for the targets with the hottest equilibrium temperature and coolest stellar temperatures that produce the largest eclipse depths, when spanning both of these parameters is very much the science of interest.

There is also a seemingly conflicting consideration here: we also need to ensure that our results are not merely at the population level. As a rather crude illustration, observing 10 rocky planets with null results that each indicate 50 % chance of having an atmosphere is equivalent to 1 in 1000 chance that they are all bare rocks, but this is an unsatisfying conclusion as to whether

we have found an atmosphere around a rocky planet or not. Therefore, to ensure that we have a number of individually robust conclusions, we need to follow up on the most promising targets. Considerations for both breadth (to obtain the best population-level result) and depth (to ensure some strong individual results) calls for a dynamic approach in observation planning.

Relatedly, population-level simulations are typically done assuming an absence of previous observations. In practice, we are always adding to a set of observations that already exist. As such, population-level simulations should be flexible enough to incorporate the current observations and choose future observations conditioned on what we have. Bayesian approaches would be particularly useful here.

Moreover, targets in multi-planet systems should be prioritized. Whether a planet has retained its atmosphere depends the most strongly on its XUV irradiation history, which in turn depends on the pre-main sequence evolution of the host star and its subsequent spin-down history [168, 169]. This information can at best only be loosely inferred from the current rotation rate and high-energy emission. This critical unknown incurs a scatter in the population-level trend. However, since both planets in one system share the same birth environment, current instellation is a direct indicator of its XUV irradiation history; constraining how much atmosphere they have allows for isolating the effect of instellation and provides an up-close look at the Cosmic Shoreline.

On the modelling front as well, there are a few planned improvements to be made to HELIOS, aimed towards identifying potential degeneracies. Most relevant to rocky planet observations, nearly all forward models in the field use the same set of surface albedo spectra generated in Hu et al. [147], with some recent work incorporating space weathering in the context of exoplanets [170]. The single scatter albedo spectra are generated using the Hapke model [171] and

assumes a specific mineral composition and grain size for each crust. Incorporating a more diverse set of surfaces, as well as considering effects of darkening and brightening space weathering processes are necessary for better understanding observations of potential bare rock planets.

Appendix A: Regarding the Study of Correlated Noise

A.1 Retrieval Histograms for Various Planet Realizations

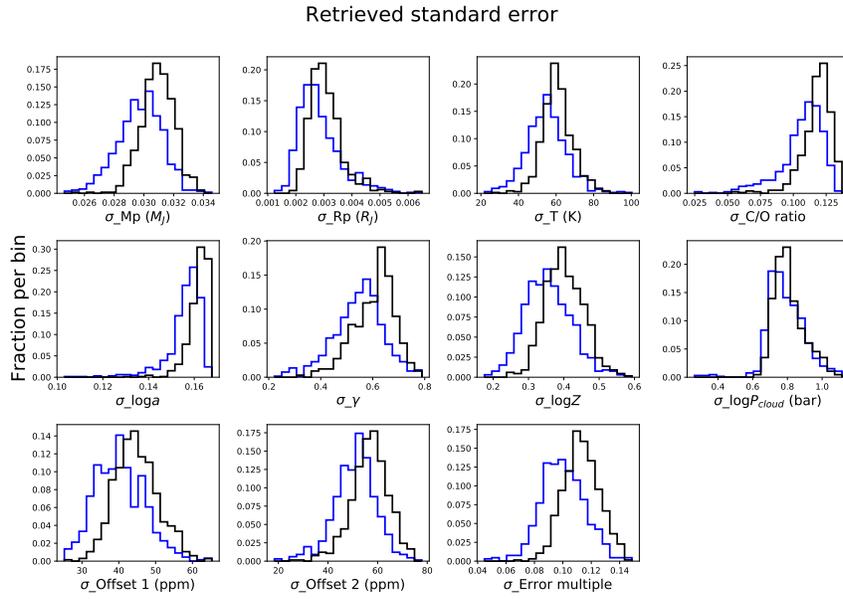


Figure A.1: Same as Figure 2.8, but for the hot Jupiter case including offsets. This figure now includes panels for both instrumental offsets in addition to the original set of retrieved parameters.

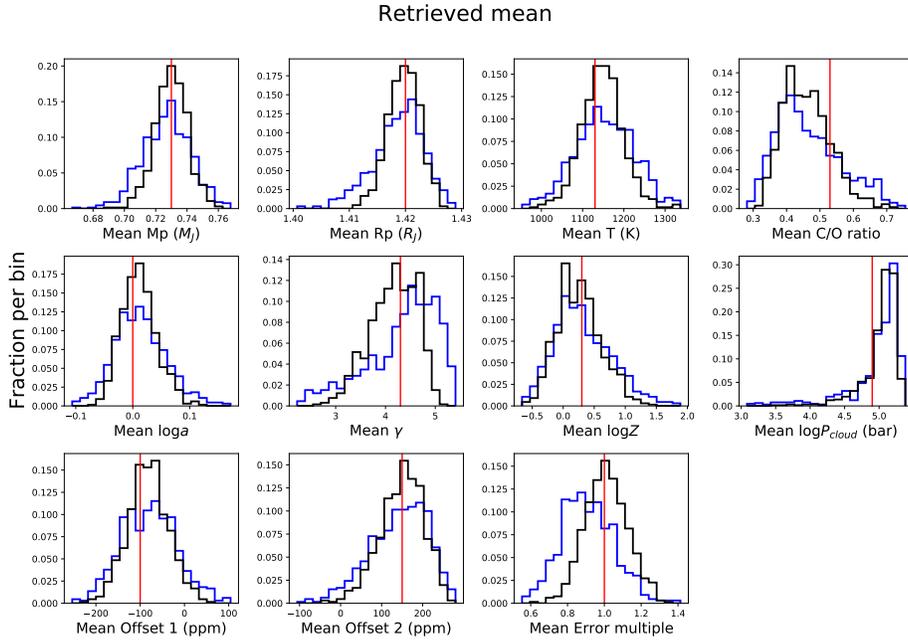


Figure A.2: Same as Figure 2.9, but for the hot Jupiter case including offsets. This figure now includes panels for both instrumental offsets in addition to the original set of retrieved parameters.

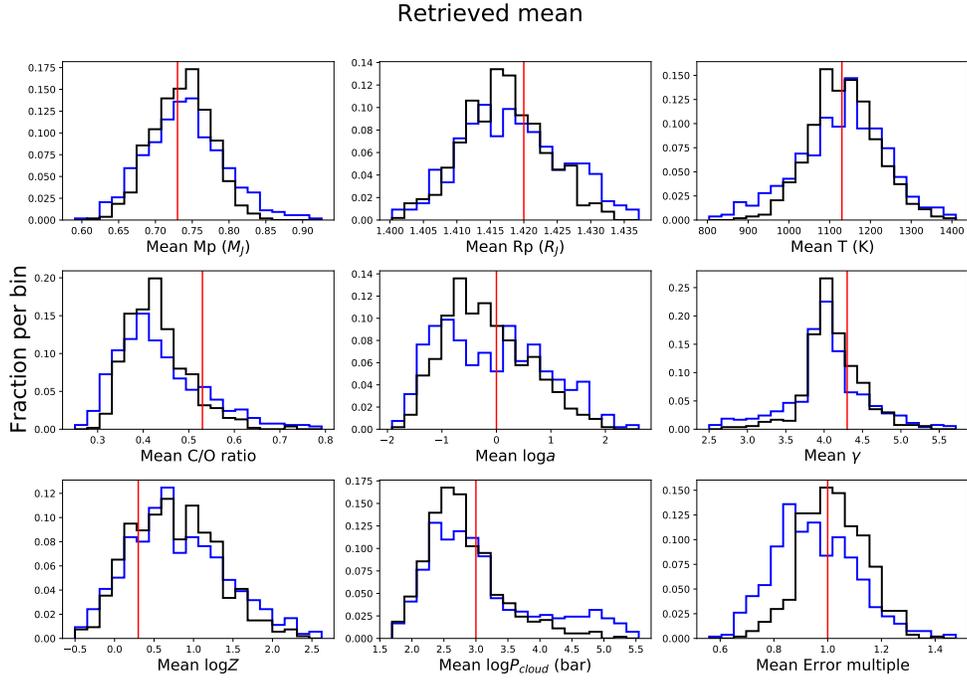


Figure A.3: Same as Figure 2.8, but for the cloudy hot Jupiter case.

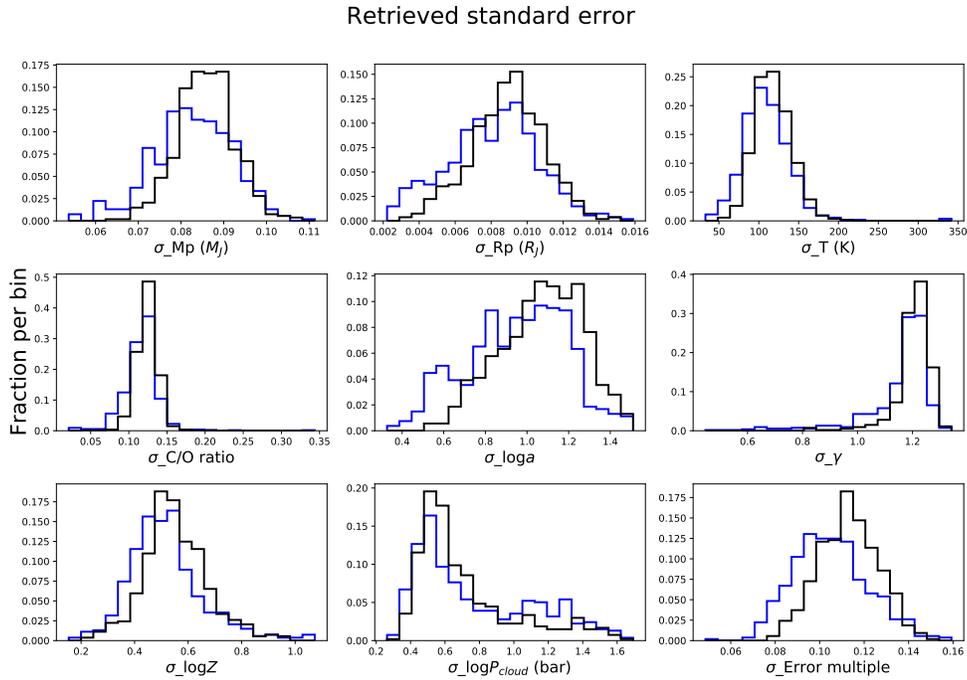


Figure A.4: Same as Figure 2.9, but for the cloudy hot Jupiter case.

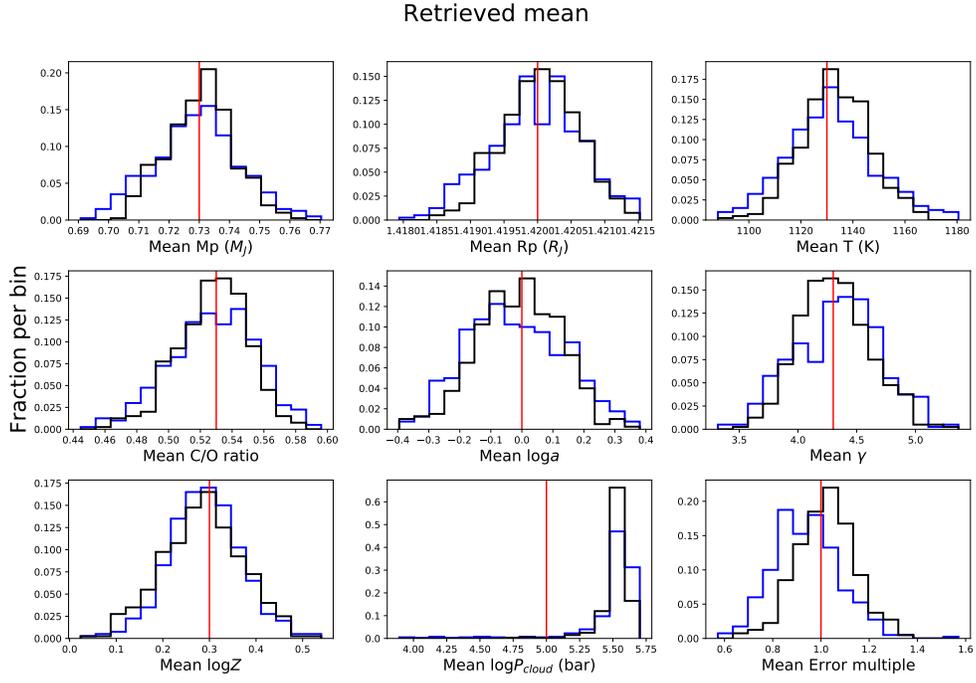


Figure A.5: Same as Figure 2.8, but for the high-precision hot Jupiter case.

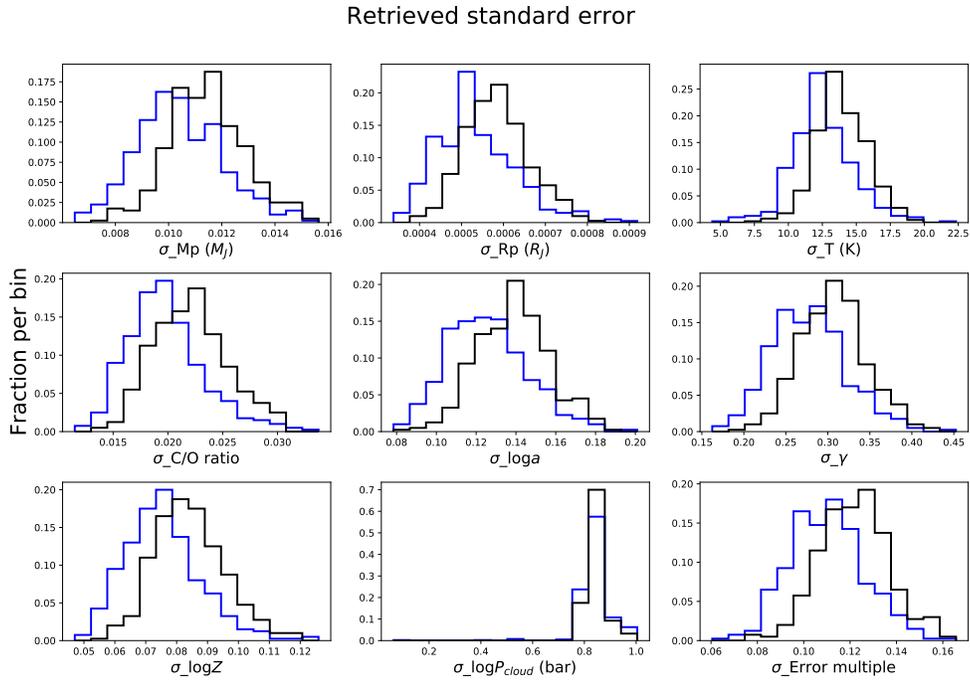


Figure A.6: Same as Figure 2.9, but for the high-precision hot Jupiter case.

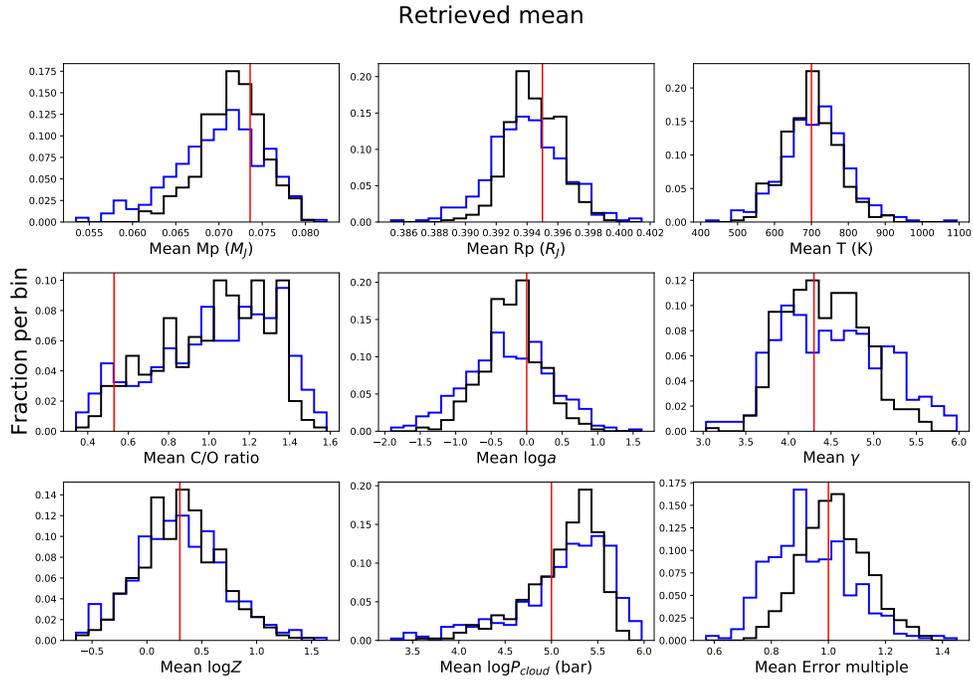


Figure A.7: Same as Figure 2.8, but for the warm Neptune case.

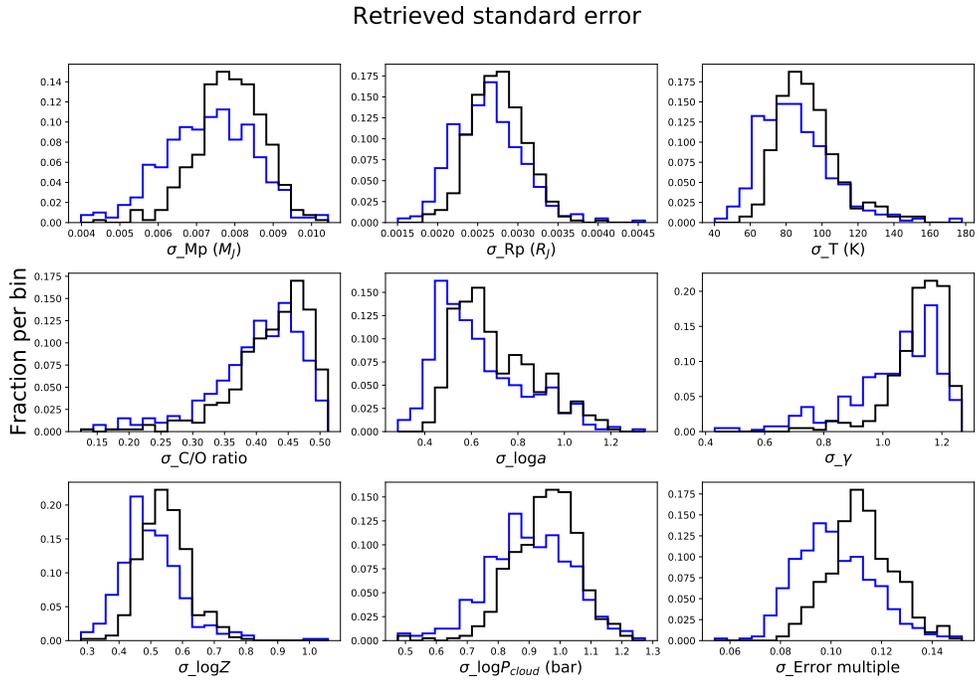


Figure A.8: Same as Figure 2.9, but for the warm Neptune case.

Appendix B: Facilities & Software used in this Thesis

B.1 Facilities

The modelling in Chapter 2 & 3 was performed in part on the Yorp and Astra clusters administered by the Center for Theory and Computation, part of the Department of Astronomy at the University of Maryland.

B.2 Software

1. astropy [[172](#), [173](#), [174](#)]
2. dynesty [[175](#)]
3. emcee [[176](#)]
4. helios [[35](#), [36](#), [37](#)]
5. matplotlib [[177](#)]
6. nestle [[102](#)]
7. numpy [[178](#)]
8. platon [[38](#), [39](#)]
9. pandexo [[125](#)]

10. `scipy` [\[179\]](#)

Bibliography

- [1] Jegug Ih and Eliza M. R. Kempton. Understanding the Effects of Systematics in Exoplanetary Atmospheric Retrievals. *AJ*, 162(6):237, December 2021. doi: 10.3847/1538-3881/ac173b.
- [2] Thomas P. Greene, Taylor J. Bell, Elsa Ducrot, Achrène Dyrek, Pierre-Olivier Lagage, and Jonathan J. Fortney. Thermal emission from the Earth-sized exoplanet TRAPPIST-1 b using JWST. *Nature*, 618(7963):39–42, June 2023. doi: 10.1038/s41586-023-05951-7.
- [3] Laura Kreidberg. Exoplanet Atmosphere Measurements from Transmission Spectroscopy and Other Planet Star Combined Light Observations. In Hans J. Deeg and Juan Antonio Belmonte, editors, *Handbook of Exoplanets*, page 100. 2018. doi: 10.1007/978-3-319-55333-7_100.
- [4] Jacob Lustig-Yaeger, Kristin S. Sotzen, Kevin B. Stevenson, Rodrigo Luger, Erin M. May, L. C. Mayorga, Kathleen Mandt, and Noam R. Izenberg. Hierarchical Bayesian Atmospheric Retrieval Modeling for Population Studies of Exoplanet Atmospheres: A Case Study on the Habitable Zone. *AJ*, 163(3):140, March 2022. doi: 10.3847/1538-3881/ac5034.
- [5] Kevin J. Zahnle and David C. Catling. The Cosmic Shoreline: The Evidence that Escape Determines which Planets Have Atmospheres, and what this May Mean for Proxima Centauri B. *ApJ*, 843(2):122, July 2017. doi: 10.3847/1538-4357/aa7846.
- [6] Eliza M. R. Kempton, Jacob L. Bean, Dana R. Louie, Drake Deming, Daniel D. B. Koll, Megan Mansfield, Jessie L. Christiansen, Mercedes López-Morales, Mark R. Swain, Robert T. Zellem, Sarah Ballard, Thomas Barclay, Joanna K. Barstow, Natasha E. Batalha, Thomas G. Beatty, Zach Berta-Thompson, Jayne Birkby, Lars A. Buchhave, David Charbonneau, Nicolas B. Cowan, Ian Crossfield, Miguel de Val-Borro, René Doyon, Diana Dragomir, Eric Gaidos, Kevin Heng, Renyu Hu, Stephen R. Kane, Laura Kreidberg, Matthias Mallonn, Caroline V. Morley, Norio Narita, Valerio Nascimbeni, Enric Pallé, Elisa V. Quintana, Emily Rauscher, Sara Seager, Evgenya L. Shkolnik, David K. Sing, Alessandro Sozzetti, Keivan G. Stassun, Jeff A. Valenti, and Carolina von Essen. A Framework for Prioritizing the TESS Planetary Candidates Most Amenable to Atmospheric Characterization. *PASP*, 130(993):114401, November 2018. doi: 10.1088/1538-3873/aadf6f.
- [7] Xueying Guo, Ian J. M. Crossfield, Diana Dragomir, Molly R. Kosiarek, Joshua Lothringer, Thomas Mikal-Evans, Lee Rosenthal, Bjorn Benneke, Heather A. Knutson,

- Paul A. Dalba, Eliza M. R. Kempton, Gregory W. Henry, P. R. McCullough, Travis Barman, Sarah Blunt, Ashley Chontos, Jonathan Fortney, Benjamin J. Fulton, Lea Hirsch, Andrew W. Howard, Howard Isaacson, Jaymie Matthews, Teo Mocnik, Caroline Morley, Erik A. Petigura, and Lauren M. Weiss. Updated Parameters and a New Transmission Spectrum of HD 97658b. *arXiv e-prints*, art. arXiv:2004.03601, April 2020.
- [8] A. Wolszczan and D. A. Frail. A planetary system around the millisecond pulsar PSR1257 + 12. *Nature*, 355(6356):145–147, January 1992. doi: 10.1038/355145a0.
- [9] Michel Mayor and Didier Queloz. A Jupiter-mass companion to a solar-type star. *Nature*, 378(6555):355–359, November 1995. doi: 10.1038/378355a0.
- [10] R. L. Akeson, X. Chen, D. Ciardi, M. Crane, J. Good, M. Harbut, E. Jackson, S. R. Kane, A. C. Laity, S. Leifer, M. Lynn, D. L. McElroy, M. Papin, P. Plavchan, S. V. Ramírez, R. Rey, K. von Braun, M. Wittman, M. Abajian, B. Ali, C. Beichman, A. Beekley, G. B. Berriman, S. Berukoff, G. Bryden, B. Chan, S. Groom, C. Lau, A. N. Payne, M. Regelson, M. Saucedo, M. Schmitz, J. Stauffer, P. Wyatt, and A. Zhang. The NASA Exoplanet Archive: Data and Tools for Exoplanet Research. *PASP*, 125(930):989, August 2013. doi: 10.1086/672273.
- [11] Adam P. Showman, Jonathan J. Fortney, Yuan Lian, Mark S. Marley, Richard S. Freedman, Heather A. Knutson, and David Charbonneau. Atmospheric Circulation of Hot Jupiters: Coupled Radiative-Dynamical General Circulation Model Simulations of HD 189733b and HD 209458b. *ApJ*, 699(1):564–584, July 2009. doi: 10.1088/0004-637X/699/1/564.
- [12] David S. Amundsen, Isabelle Baraffe, Pascal Tremblin, James Manners, Wolfgang Hayek, Nathan J. Mayne, and David M. Acreman. Accuracy tests of radiation schemes used in hot Jupiter global circulation models. *A&A*, 564:A59, April 2014. doi: 10.1051/0004-6361/201323169.
- [13] L. S. Rothman, I. E. Gordon, A. Barbe, D. Chris Benner, P. F. Bernath, M. Birk, V. Boudon, L. R. Brown, A. Campargue, J. P. Champion, K. Chance, L. H. Coudert, V. Dana, V. M. Devi, S. Fally, J. M. Flaud, R. R. Gamache, A. Goldman, D. Jacquemart, I. Kleiner, N. Lacome, W. J. Lafferty, J. Y. Mandin, S. T. Massie, S. N. Mikhailenko, C. E. Miller, N. Moazzen-Ahmadi, O. V. Naumenko, A. V. Nikitin, J. Orphal, V. I. Perevalov, A. Perrin, A. Predoi-Cross, C. P. Rinsland, M. Rotger, M. Šimečková, M. A. H. Smith, K. Sung, S. A. Tashkun, J. Tennyson, R. A. Toth, A. C. Vandaele, and J. Vander Auwera. The HITRAN 2008 molecular spectroscopic database. *JQSRT*, 110(9-10):533–572, June 2009. doi: 10.1016/j.jqsrt.2009.02.013.
- [14] Jonathan Tennyson and Sergei N. Yurchenko. ExoMol: molecular line lists for exoplanet and other atmospheres. *MNRAS*, 425(1):21–33, September 2012. doi: 10.1111/j.1365-2966.2012.21440.x.
- [15] F.D. Stacey and P.M. Davis. High pressure equations of state with applications to the lower mantle and core. *Physics of the Earth and Planetary Interiors*, 142(3):137–184, 2004. ISSN 0031-9201. doi: <https://doi.org/10.1016/j.pepi.2004.02.003>. URL <https://www.sciencedirect.com/science/article/pii/S0031920104001049>.

- [16] Diana Valencia, Richard J. O’Connell, and Dimitar Sasselov. Internal structure of massive terrestrial planets. *Icarus*, 181(2):545–554, April 2006. doi: 10.1016/j.icarus.2005.11.021.
- [17] Carl Sagan, W. Reid Thompson, Robert Carlson, Donald Gurnett, and Charles Hord. A search for life on Earth from the Galileo spacecraft. *Nature*, 365(6448):715–721, October 1993. doi: 10.1038/365715a0.
- [18] Timothy M. Brown. Transmission Spectra as Diagnostics of Extrasolar Giant Planet Atmospheres. *ApJ*, 553(2):1006–1026, June 2001. doi: 10.1086/320950.
- [19] James B. Pollack, Olenka Hubickyj, Peter Bodenheimer, Jack J. Lissauer, Morris Podolak, and Yuval Greenzweig. Formation of the Giant Planets by Concurrent Accretion of Solids and Gas. *Icarus*, 124(1):62–85, November 1996. doi: 10.1006/icar.1996.0190.
- [20] Laura Schaefer, Katharina Lodders, and Bruce Fegley. Vaporization of the Earth: Application to Exoplanet Atmospheres. *ApJ*, 755(1):41, August 2012. doi: 10.1088/0004-637X/755/1/41.
- [21] Laura Schaefer, Robin D. Wordsworth, Zachory Berta-Thompson, and Dimitar Sasselov. Predictions of the Atmospheric Composition of GJ 1132b. *ApJ*, 829(2):63, October 2016. doi: 10.3847/0004-637X/829/2/63.
- [22] Xander Byrne, Oliver Shorttle, Sean Jordan, and Paul B. Rimmer. Atmospheres as a Window to Rocky Exoplanet Surfaces. *arXiv e-prints*, art. arXiv:2312.11133, December 2023. doi: 10.48550/arXiv.2312.11133.
- [23] Renyu Hu, Mario Damiano, Markus Scheucher, Edwin Kite, Sara Seager, and Heike Rauer. Unveiling Shrouded Oceans on Temperate sub-Neptunes via Transit Signatures of Solubility Equilibria versus Gas Thermochemistry. *ApJL*, 921(1):L8, November 2021. doi: 10.3847/2041-8213/ac1f92.
- [24] David J. Des Marais, Martin O. Harwit, Kenneth W. Jucks, James F. Kasting, Douglas N. C. Lin, Jonathan I. Lunine, Jean Schneider, Sara Seager, Wesley A. Traub, and Neville J. Woolf. Remote Sensing of Planetary Properties and Biosignatures on Extrasolar Terrestrial Planets. *Astrobiology*, 2(2):153–181, June 2002. doi: 10.1089/15311070260192246.
- [25] David Charbonneau, Timothy M. Brown, David W. Latham, and Michel Mayor. Detection of Planetary Transits Across a Sun-like Star. *ApJL*, 529(1):L45–L48, January 2000. doi: 10.1086/312457.
- [26] David Charbonneau, Timothy M. Brown, Robert W. Noyes, and Ronald L. Gilliland. Detection of an Extrasolar Planet Atmosphere. *ApJ*, 568(1):377–384, March 2002. doi: 10.1086/338770.
- [27] Andrew Collier Cameron, Keith Horne, Alan Penny, and David James. Probable detection of starlight reflected from the giant planet orbiting τ Boötis. *Nature*, 402(6763):751–755, December 1999. doi: 10.1038/45451.

- [28] J. J. Fortney, M. S. Marley, K. Lodders, D. Saumon, and R. Freedman. Comparative Planetary Atmospheres: Models of TrES-1 and HD 209458b. *ApJL*, 627(1):L69–L72, July 2005. doi: 10.1086/431952.
- [29] A. R. Belu, F. Selsis, J. C. Morales, I. Ribas, C. Cossou, and H. Rauer. Primary and secondary eclipse spectroscopy with JWST: exploring the exoplanet parameter space. *A&A*, 525:A83, January 2011. doi: 10.1051/0004-6361/201014995.
- [30] D. Deming, S. Seager, J. Winn, E. Miller-Ricci, M. Clampin, D. Lindler, T. Greene, D. Charbonneau, G. Laughlin, G. Ricker, D. Latham, and K. Ennico. Discovery and Characterization of Transiting Super Earths Using an All-Sky Transit Survey and Follow-up by the James Webb Space Telescope. *PASP*, 121(883):952, September 2009. doi: 10.1086/605913.
- [31] Megan Mansfield, Edwin S. Kite, Renyu Hu, Daniel D. B. Koll, Matej Malik, Jacob L. Bean, and Eliza M. R. Kempton. Identifying Atmospheres on Rocky Exoplanets through Inferred High Albedo. *ApJ*, 886(2):141, December 2019. doi: 10.3847/1538-4357/ab4c90.
- [32] Daniel D. B. Koll, Matej Malik, Megan Mansfield, Eliza M. R. Kempton, Edwin Kite, Dorian Abbot, and Jacob L. Bean. Identifying Candidate Atmospheres on Rocky M Dwarf Planets via Eclipse Photometry. *ApJ*, 886(2):140, December 2019. doi: 10.3847/1538-4357/ab4c91.
- [33] Sebastian Zieba, Laura Kreidberg, Elsa Ducrot, Michaël Gillon, Caroline Morley, Laura Schaefer, Patrick Tamburo, Daniel D. B. Koll, Xintong Lyu, Lorena Acuña, Eric Agol, Aishwarya R. Iyer, Renyu Hu, Andrew P. Lincowski, Victoria S. Meadows, Franck Selsis, Emeline Bolmont, Avi M. Mandell, and Gabrielle Suissa. No thick carbon dioxide atmosphere on the rocky exoplanet TRAPPIST-1 c. *Nature*, 620(7975):746–749, August 2023. doi: 10.1038/s41586-023-06232-z.
- [34] Nikku Madhusudhan. Exoplanetary Atmospheres: Key Insights, Challenges, and Prospects. *ARA&A*, 57:617–663, August 2019. doi: 10.1146/annurev-astro-081817-051846.
- [35] Matej Malik, Luc Grosheintz, João M. Mendonça, Simon L. Grimm, Baptiste Lavie, Daniel Kitzmann, Shang-Min Tsai, Adam Burrows, Laura Kreidberg, Megan Bedell, Jacob L. Bean, Kevin B. Stevenson, and Kevin Heng. HELIOS: An Open-source, GPU-accelerated Radiative Transfer Code for Self-consistent Exoplanetary Atmospheres. *AJ*, 153(2):56, February 2017. doi: 10.3847/1538-3881/153/2/56.
- [36] Matej Malik, Eliza M. R. Kempton, Daniel D. B. Koll, Megan Mansfield, Jacob L. Bean, and Edwin Kite. Analyzing Atmospheric Temperature Profiles and Spectra of M Dwarf Rocky Planets. *ApJ*, 886(2):142, December 2019. doi: 10.3847/1538-4357/ab4a05.
- [37] Matej Malik, Daniel Kitzmann, João M. Mendonça, Simon L. Grimm, Gabriel-Dominique Marleau, Esther F. Linder, Shang-Min Tsai, and Kevin Heng. Self-luminous and Irradiated Exoplanetary Atmospheres Explored with HELIOS. *AJ*, 157(5):170, May 2019. doi: 10.3847/1538-3881/ab1084.

- [38] Michael Zhang, Yayaati Chachan, Eliza M. R. Kempton, and Heather A. Knutson. Forward Modeling and Retrievals with PLATON, a Fast Open-source Tool. *PASP*, 131(997): 034501, March 2019. doi: 10.1088/1538-3873/aaf5ad.
- [39] Michael Zhang, Yayaati Chachan, Eliza M. R. Kempton, Heather A. Knutson, and Wenjun (Happy) Chang. PLATON II: New Capabilities and a Comprehensive Retrieval on HD 189733b Transit and Eclipse Data. *ApJ*, 899(1):27, August 2020. doi: 10.3847/1538-4357/ab1e6.
- [40] Joachim W. Stock, Daniel Kitzmann, A. Beate C. Patzer, and Erwin Sedlmayr. FastChem: A computer program for efficient complex chemical equilibrium calculations in the neutral/ionized gas phase with applications to stellar and planetary atmospheres. *MNRAS*, 479(1):865–874, September 2018. doi: 10.1093/mnras/sty1531.
- [41] P. Woitke, Ch. Helling, G. H. Hunter, J. D. Millard, G. E. Turner, M. Worters, J. Blečić, and J. W. Stock. Equilibrium chemistry down to 100 K. Impact of silicates and phyllosilicates on the carbon to oxygen ratio. *A&A*, 614:A1, June 2018. doi: 10.1051/0004-6361/201732193.
- [42] Sarah D. Blumenthal, Avi M. Mandell, Eric Hébrard, Natasha E. Batalha, Patricio E. Cubillos, Sarah Rugheimer, and Hannah R. Wakeford. A Comparison of Simulated JWST Observations Derived from Equilibrium and Non-equilibrium Chemistry Models of Giant Exoplanets. *ApJ*, 853(2):138, February 2018. doi: 10.3847/1538-4357/aa9e51.
- [43] Claire Baxter, Jean-Michel Désert, Shang-Min Tsai, Kamen O. Todorov, Jacob L. Bean, Drake Deming, Vivien Parmentier, Jonathan J. Fortney, Michael Line, Daniel Thorngren, Raymond T. Pierrehumbert, Adam Burrows, and Adam P. Showman. Evidence for disequilibrium chemistry from vertical mixing in hot Jupiter atmospheres. A comprehensive survey of transiting close-in gas giant exoplanets with warm-Spitzer/IRAC. *A&A*, 648: A127, April 2021. doi: 10.1051/0004-6361/202039708.
- [44] Q. Changeat, B. Edwards, I. P. Waldmann, and G. Tinetti. Toward a More Complex Description of Chemical Profiles in Exoplanet Retrievals: A Two-layer Parameterization. *ApJ*, 886(1):39, November 2019. doi: 10.3847/1538-4357/ab4a14.
- [45] N. Madhusudhan and S. Seager. A Temperature and Abundance Retrieval Method for Exoplanet Atmospheres. *ApJ*, 707(1):24–39, December 2009. doi: 10.1088/0004-637X/707/1/24.
- [46] Michael R. Line, Aaron S. Wolf, Xi Zhang, Heather Knutson, Joshua A. Kammer, Elias Ellison, Pieter Deroo, Dave Crisp, and Yuk L. Yung. A Systematic Retrieval Analysis of Secondary Eclipse Spectra. I. A Comparison of Atmospheric Retrieval Techniques. *ApJ*, 775(2):137, October 2013. doi: 10.1088/0004-637X/775/2/137.
- [47] Syukuro Manabe, Joseph Smagorinsky, and Robert F. Strickler. Simulated Climatology of a General Circulation Model with a Hydrologic CYCLE1. *Monthly Weather Review*, 93(12):769, January 1965. doi: 10.1175/1520-0493(1965)093<0769:SCOAGC>2.3.CO;2.

- [48] Tristan Guillot. Condensation of Methane, Ammonia, and Water and the Inhibition of Convection in Giant Planets. *Science*, 269(5231):1697–1699, September 1995. doi: 10.1126/science.7569896.
- [49] R. J. Graham, Tim Lichtenberg, Ryan Boukrouche, and Raymond T. Pierrehumbert. A Multispecies Pseudoadiabat for Simulating Condensable-rich Exoplanet Atmospheres. *PSJ*, 2(5):207, October 2021. doi: 10.3847/PSJ/ac214c.
- [50] Antonio Claret and Guillermo Torres. The Dependence of Convective Core Overshooting on Stellar Mass: Additional Binary Systems and Improved Calibration. *ApJ*, 859(2):100, June 2018. doi: 10.3847/1538-4357/aabd35.
- [51] Emily A. Whittaker, Matej Malik, Jegug Ih, Eliza M. R. Kempton, Megan Mansfield, Jacob L. Bean, Edwin S. Kite, Daniel D. B. Koll, Timothy W. Cronin, and Renyu Hu. The Detectability of Rocky Planet Surface and Atmosphere Composition with the JWST: The Case of LHS 3844b. *AJ*, 164(6):258, December 2022. doi: 10.3847/1538-3881/ac9ab3.
- [52] Martin Turbet, Thomas J. Fauchez, Jeremy Leconte, Emeline Bolmont, Guillaume Chaverot, Francois Forget, Ehouarn Millour, Franck Selsis, Benjamin Charnay, Elsa Ducrot, Michaël Gillon, Alice Maurel, and Geronimo L. Villanueva. Water condensation zones around main sequence stars. *A&A*, 679:A126, November 2023. doi: 10.1051/0004-6361/202347539.
- [53] Daniel D. B. Koll. A Scaling for Atmospheric Heat Redistribution on Tidally Locked Rocky Planets. *ApJ*, 924(2):134, January 2022. doi: 10.3847/1538-4357/ac3b48.
- [54] A. Burrows, I. Hubeny, and D. Sudarsky. A Theoretical Interpretation of the Measurements of the Secondary Eclipses of TrES-1 and HD 209458b. *ApJL*, 625(2):L135–L138, June 2005. doi: 10.1086/431242.
- [55] A. Burrows, I. Hubeny, J. Budaj, H. A. Knutson, and D. Charbonneau. Theoretical Spectral Models of the Planet HD 209458b with a Thermal Inversion and Water Emission Bands. *ApJL*, 668(2):L171–L174, October 2007. doi: 10.1086/522834.
- [56] Bjoern Benneke and Sara Seager. Atmospheric Retrieval for Super-Earths: Uniquely Constraining the Atmospheric Composition with Transmission Spectroscopy. *ApJ*, 753(2):100, July 2012. doi: 10.1088/0004-637X/753/2/100.
- [57] Matthew C. Nixon, Luis Welbanks, Peter McGill, and Eliza M. R. Kempton. Methods for Incorporating Model Uncertainty into Exoplanet Atmospheric Analysis. *ApJ*, 966(2):156, May 2024. doi: 10.3847/1538-4357/ad354e.
- [58] Sean N. Raymond, Thomas Quinn, and Jonathan I. Lunine. Making other earths: dynamical simulations of terrestrial planet formation and water delivery. *Icarus*, 168(1):1–17, March 2004. doi: 10.1016/j.icarus.2003.11.019.
- [59] Linda T. Elkins-Tanton and Sara Seager. Ranges of Atmospheric Mass and Composition of Super-Earth Exoplanets. *ApJ*, 685(2):1237–1246, October 2008. doi: 10.1086/591433.

- [60] Feng Tian and Shigeru Ida. Water contents of Earth-mass planets around M dwarfs. *Nature Geoscience*, 8(3):177–180, March 2015. doi: 10.1038/ngeo2372.
- [61] Edwin S. Kite and Megan N. Barnett. Exoplanet secondary atmosphere loss and revival. *Proceedings of the National Academy of Science*, 117:18264–18271, July 2020. doi: 10.1073/pnas.200617711710.48550/arXiv.2006.02589.
- [62] K. Garcia-Sage, A. Gloer, J. J. Drake, G. Gronoff, and O. Cohen. On the Magnetic Protection of the Atmosphere of Proxima Centauri b. *ApJL*, 844(1):L13, July 2017. doi: 10.3847/2041-8213/aa7eca.
- [63] George D. McDonald, Laura Kreidberg, and Eric Lopez. The Sub-Neptune Desert and Its Dependence on Stellar Type: Controlled by Lifetime X-Ray Irradiation. *ApJ*, 876(1):22, May 2019. doi: 10.3847/1538-4357/ab1095.
- [64] Clive D Rodgers. *Inverse Methods for Atmospheric Sounding*. WORLD SCIENTIFIC, 2000. doi: 10.1142/3171. URL <https://www.worldscientific.com/doi/abs/10.1142/3171>.
- [65] M. R. Line, A. S. Wolf, X. Zhang, H. Knutson, J. A. Kammer, E. Ellison, P. Deroo, D. Crisp, and Y. L. Yung. A Systematic Retrieval Analysis of Secondary Eclipse Spectra. I. A Comparison of Atmospheric Retrieval Techniques. *ApJ*, 775:137, October 2013. doi: 10.1088/0004-637X/775/2/137.
- [66] I. P. Waldmann, G. Tinetti, M. Rocchetto, E. J. Barton, S. N. Yurchenko, and J. Tennyson. Tau-REx I: A Next Generation Retrieval Code for Exoplanetary Atmospheres. *ApJ*, 802(2):107, April 2015. doi: 10.1088/0004-637X/802/2/107.
- [67] Joseph Harrington. Atmospheric Retrievals from Exoplanet Observations and Simulations with BART. NASA Proposal id.16-XPR16-106, January 2016.
- [68] N. Madhusudhan. *Atmospheric Retrieval of Exoplanets*, page 104. 2018. doi: 10.1007/978-3-319-55333-7_104.
- [69] M. Zhang, Y. Chachan, E. M.-R. Kempton, and H. A. Knutson. Forward Modeling and Retrievals with PLATON, a Fast Open-source Tool. *PASP*, 131(3):034501, March 2019. doi: 10.1088/1538-3873/aaf5ad.
- [70] Björn Benneke, Heather A. Knutson, Joshua Lothringer, Ian J. M. Crossfield, Julianne I. Moses, Caroline Morley, Laura Kreidberg, Benjamin J. Fulton, Diana Dragomir, Andrew W. Howard, Ian Wong, Jean-Michel Désert, Peter R. McCullough, Eliza M. R. Kempton, Jonathan Fortney, Ronald Gilliland, Drake Deming, and Joshua Kammer. A sub-Neptune exoplanet with a low-metallicity methane-depleted atmosphere and Mie-scattering clouds. *Nature Astronomy*, 3:813–821, July 2019. doi: 10.1038/s41550-019-0800-5.
- [71] Joanna K. Barstow and Kevin Heng. Outstanding Challenges of Exoplanet Atmospheric Retrievals. *arXiv e-prints*, art. arXiv:2003.14311, March 2020.

- [72] Arazi Pinhas, Benjamin V. Rackham, Nikku Madhusudhan, and Dániel Apai. Retrieval of planetary and stellar properties in transmission spectroscopy with AURA. *MNRAS*, 480(4):5314–5331, November 2018. doi: 10.1093/mnras/sty2209.
- [73] A. Caldas, J. Leconte, F. Selsis, I. P. Waldmann, P. Bordé, M. Rocchetto, and B. Charnay. Effects of a fully 3D atmospheric structure on exoplanet transmission spectra: retrieval biases due to day-night temperature gradients. *A&A*, 623:A161, March 2019. doi: 10.1051/0004-6361/201834384.
- [74] Q. Changeat, B. Edwards, I. P. Waldmann, and G. Tinetti. Toward a More Complex Description of Chemical Profiles in Exoplanet Retrievals: A Two-layer Parameterization. *ApJ*, 886(1):39, November 2019. doi: 10.3847/1538-4357/ab4a14.
- [75] Joanna K. Barstow, Quentin Changeat, Ryan Garland, Michael R. Line, Marco Rocchetto, and Ingo P. Waldmann. A comparison of exoplanet spectroscopic retrieval tools. *MNRAS*, 493(4):4884–4909, April 2020. doi: 10.1093/mnras/staa548.
- [76] Brianna I. Lacy and Adam S. Burrows. JWST Transit Spectra I: Exploring Potential Biases and Opportunities in Retrievals of Tidally-locked Hot Jupiters with Clouds and Hazes. *arXiv e-prints*, art. arXiv:2006.06899, June 2020.
- [77] Jake Taylor, Vivien Parmentier, Patrick G. J. Irwin, Suzanne Aigrain, Graham K. H. Lee, and Joshua Krissansen-Totton. Understanding and mitigating biases when studying inhomogeneous emission spectra with JWST. *MNRAS*, 493(3):4342–4354, April 2020. doi: 10.1093/mnras/staa552.
- [78] Rene Andrae, Tim Schulze-Hartung, and Peter Melchior. Dos and don'ts of reduced chi-squared. *arXiv e-prints*, art. arXiv:1012.3754, December 2010.
- [79] Knicole D. Colón, Laura Kreidberg, Luis Welbanks, Michael R. Line, Nikku Madhusudhan, Thomas Beatty, Patrick Tamburo, Kevin B. Stevenson, Avi Mandell, Joseph E. Rodriguez, Thomas Barclay, Eric D. Lopez, Keivan G. Stassun, Daniel Angerhausen, Jonathan J. Fortney, David J. James, Joshua Pepper, John P. Ahlers, Peter Plavchan, Supachai Awiphan, Cliff Kotnik, Kim K. McLeod, Gabriel Murawski, Heena Chotani, Danny LeBrun, William Matzko, David Rea, Monica Vidaurri, Scott Webster, James K. Williams, Leafia Sheraden Cox, Nicole Tan, and Emily A. Gilbert. An Unusual Transmission Spectrum for the Sub-Saturn KELT-11b Suggestive of a Sub-Solar Water Abundance. *arXiv e-prints*, art. arXiv:2005.05153, May 2020.
- [80] Benjamin V. Rackham, Dániel Apai, and Mark S. Giampapa. The Transit Light Source Effect: False Spectral Features and Incorrect Densities for M-dwarf Transiting Planets. *ApJ*, 853(2):122, February 2018. doi: 10.3847/1538-4357/aaa08c.
- [81] Giovanni Bruno, Nikole K. Lewis, Munazza K. Alam, Mercedes López-Morales, Joanna K. Barstow, Hannah R. Wakeford, David K. Sing, Gregory W. Henry, Gilda E. Ballester, Vincent Bourrier, Lars A. Buchhave, Ofer Cohen, Thomas Mikal-Evans, Antonio García Muñoz, Panayotis Lavvas, and Jorge Sanz-Forcada. WASP-52b. The effect

- of star-spot correction on atmospheric retrievals. *MNRAS*, 491(4):5361–5375, February 2020. doi: 10.1093/mnras/stz3194.
- [82] A. Tsiaras, I. P. Waldmann, M. Rocchetto, R. Varley, G. Morello, M. Damiano, and G. Tinetti. A New Approach to Analyzing HST Spatial Scans: The Transmission Spectrum of HD 209458 b. *ApJ*, 832(2):202, December 2016. doi: 10.3847/0004-637X/832/2/202.
- [83] K. B. Stevenson and J. Fowler. Analyzing Eight Years of Transiting Exoplanet Observations Using WFC3’s Spatial Scan Monitor. Space Telescope WFC Instrument Science Report, September 2019.
- [84] Kai Hou Yip, Quentin Changeat, Billy Edwards, Mario Morvan, Katy L. Chubb, Angelos Tsiaras, Ingo P. Waldmann, and Giovanna Tinetti. On The Compatibility of Ground-based and Space-based Data: WASP-96 b, An Example. *arXiv e-prints*, art. arXiv:2009.10438, September 2020.
- [85] T. M. Evans, S. Aigrain, N. Gibson, J. K. Barstow, D. S. Amundsen, P. Tremblin, and P. Mourier. A uniform analysis of HD 209458b Spitzer/IRAC light curves with Gaussian process models. *MNRAS*, 451:680–694, July 2015. doi: 10.1093/mnras/stv910.
- [86] D. Deming, A. Wilkins, P. McCullough, A. Burrows, J. J. Fortney, E. Agol, I. Dobbs-Dixon, N. Madhusudhan, N. Crouzet, J.-M. Desert, R. L. Gilliland, K. Haynes, H. A. Knutson, M. Line, Z. Magic, A. M. Mandell, S. Ranjan, D. Charbonneau, M. Clampin, S. Seager, and A. P. Showman. Infrared Transmission Spectroscopy of the Exoplanets HD 209458b and XO-1b Using the Wide Field Camera-3 on the Hubble Space Telescope. *ApJ*, 774:95, September 2013. doi: 10.1088/0004-637X/774/2/95.
- [87] H. A. Knutson, D. Charbonneau, R. W. Noyes, T. M. Brown, and R. L. Gilliland. Using Stellar Limb-Darkening to Refine the Properties of HD 209458b. *ApJ*, 655:564–575, January 2007. doi: 10.1086/510111.
- [88] Drake Deming, Pedro V. Sada, Brian Jackson, Steven W. Peterson, Eric Agol, Heather A. Knutson, Donald E. Jennings, Flynn Haase, and Kevin Bays. Kepler and Ground-based Transits of the Exo-Neptune HAT-P-11b. *ApJ*, 740(1):33, Oct 2011. doi: 10.1088/0004-637X/740/1/33.
- [89] Jason A. Dittmann, Laird M. Close, Elizabeth M. Green, Louis J. Scuderi, and Jared R. Males. Follow-up Observations of the Neptune Mass Transiting Extrasolar Planet HAT-P-11b. *ApJ*, 699(1):L48–L51, Jul 2009. doi: 10.1088/0004-637X/699/1/L48.
- [90] Joshua D. Lothringer, Björn Benneke, Ian J. M. Crossfield, Gregory W. Henry, Caroline Morley, Diana Dragomir, Travis Barman, Heather Knutson, Eliza Kempton, Jonathan Fortney, Peter McCullough, and Andrew W. Howard. An HST/STIS Optical Transmission Spectrum of Warm Neptune GJ 436b. *AJ*, 155(2):66, February 2018. doi: 10.3847/1538-3881/aaa008.
- [91] M. Asplund, N. Grevesse, A. J. Sauval, and P. Scott. The Chemical Composition of the Sun. *ARA&A*, 47:481–522, September 2009. doi: 10.1146/annurev.astro.46.060407.145222.

- [92] Laura Kreidberg, Jacob L. Bean, Jean-Michel Désert, Björn Benneke, Drake Deming, Kevin B. Stevenson, Sara Seager, Zachory Berta-Thompson, Andreas Seifahrt, and Derek Homeier. Clouds in the atmosphere of the super-earth exoplanet gj1214b. *Nature*, 505(7481):69–72, Jan 2014. ISSN 1476-4687. doi: 10.1038/nature12888. URL <http://dx.doi.org/10.1038/nature12888>.
- [93] David K. Sing, Jonathan J. Fortney, Nikolay Nikolov, Hannah R. Wakeford, Tiffany Kataria, Thomas M. Evans, Suzanne Aigrain, Gilda E. Ballester, Adam S. Burrows, Drake Deming, Jean-Michel Désert, Neale P. Gibson, Gregory W. Henry, Catherine M. Huitson, Heather A. Knutson, Alain Lecavelier Des Etangs, Frederic Pont, Adam P. Showman, Alfred Vidal-Madjar, Michael H. Williamson, and Paul A. Wilson. A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion. *Nature*, 529(7584):59–62, January 2016. doi: 10.1038/nature16068.
- [94] Drake Deming and Sara Seager. Illusion and Reality in the Atmospheres of Exoplanets. *arXiv e-prints*, art. arXiv:1701.00493, Jan 2017.
- [95] M. S. Marley, A. S. Ackerman, J. N. Cuzzi, and D. Kitzmann. *Clouds and Hazes in Exoplanet Atmospheres*, page 367. 2013. doi: 10.2458/azu_uapress_9780816530595-ch15.
- [96] M. et al. Gennaro. WFC3 Data Handbook, Version 4.0. Baltimore: STScI, 2018.
- [97] Emily Garhart, Drake Deming, Avi Mandell, Heather A. Knutson, Nicole Wallack, Adam Burrows, Jonathan J. Fortney, Callie Hood, Christopher Seay, David K. Sing, Björn Benneke, Jonathan D. Fraine, Tiffany Kataria, Nikole Lewis, Nikku Madhusudhan, Peter McCullough, Kevin B. Stevenson, and Hannah Wakeford. Statistical Characterization of Hot Jupiter Atmospheres Using Spitzer’s Secondary Eclipses. *AJ*, 159(4):137, April 2020. doi: 10.3847/1538-3881/ab6cff.
- [98] Michael R. Line, Gautam Vasisht, Pin Chen, D. Angerhausen, and Yuk L. Yung. Thermochemical and Photochemical Kinetics in Cooler Hydrogen-dominated Extrasolar Planets: A Methane-poor GJ436b? *ApJ*, 738(1):32, September 2011. doi: 10.1088/0004-637X/738/1/32.
- [99] Eliza Miller-Ricci Kempton, Kevin Zahnle, and Jonathan J. Fortney. The Atmospheric Chemistry of GJ 1214b: Photochemistry and Clouds. *ApJ*, 745(1):3, January 2012. doi: 10.1088/0004-637X/745/1/3.
- [100] Jonathan J. Fortney, Channon Visscher, Mark S. Marley, Callie E. Hood, Michael R. Line, Daniel P. Thorngren, Richard S. Freedman, and Roxana Lupu. Beyond Equilibrium Temperature: How the Atmosphere/Interior Connection Affects the Onset of Methane, Ammonia, and Clouds in Warm Transiting Giant Planets. *AJ*, 160(6):288, December 2020. doi: 10.3847/1538-3881/abc5bd.
- [101] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: The MCMC Hammer. *PASP*, 125(925):306, March 2013. doi: 10.1086/670067.

- [102] F. Feroz and M. P. Hobson. Multimodal nested sampling: an efficient and robust alternative to markov chain monte carlo methods for astronomical data analyses. *Monthly Notices of the Royal Astronomical Society*, 384(2):449–463, Jan 2008. ISSN 1365-2966. doi: 10.1111/j.1365-2966.2007.12353.x. URL <http://dx.doi.org/10.1111/j.1365-2966.2007.12353.x>.
- [103] E. M. May, T. Gardner, E. Rauscher, and J. D. Monnier. MOPSS. II. Extreme Optical Scattering Slope for the Inflated Super-Neptune HATS-8b. *AJ*, 159(1):7, January 2020. doi: 10.3847/1538-3881/ab5361.
- [104] Adam D. Cobb, Michael D. Himes, Frank Soboczenski, Simone Zorzan, Molly D. O’Beirne, Atılım Güneş Baydin, Yarin Gal, Shawn D. Domagal-Goldman, Giada N. Arney, Daniel Angerhausen, and II 2018 NASA FDL Astrobiology Team. An Ensemble of Bayesian Neural Networks for Exoplanetary Atmospheric Retrieval. *AJ*, 158(1):33, July 2019. doi: 10.3847/1538-3881/ab2390.
- [105] D. Sivia and J. Skilling. *Data Analysis: A Bayesian Tutorial*. Oxford University Press, 1996. ISBN 9780198568322. URL <https://global.oup.com/academic/product/data-analysis-9780198568322?cc=us&lang=en&#>.
- [106] Ryan J. MacDonald, Jayesh M. Goyal, and Nikole K. Lewis. Why is it so Cold in Here?: Explaining the Cold Temperatures Retrieved from Transmission Spectra of Exoplanet Atmospheres. *arXiv e-prints*, art. arXiv:2003.11548, March 2020.
- [107] H. A. Knutson, D. Dragomir, L. Kreidberg, E. M.-R. Kempton, P. R. McCullough, J. J. Fortney, J. L. Bean, M. Gillon, D. Homeier, and A. W. Howard. Hubble Space Telescope Near-IR Transmission Spectroscopy of the Super-Earth HD 97658b. *ApJ*, 794:155, October 2014. doi: 10.1088/0004-637X/794/2/155.
- [108] Joanna K. Barstow, Quentin Changeat, Ryan Garland, Michael R. Line, Marco Rocchetto, and Ingo P. Waldmann. A comparison of exoplanet spectroscopic retrieval tools. *MNRAS*, 493(4):4884–4909, April 2020. doi: 10.1093/mnras/staa548.
- [109] Anjali A. A. Piette, Nikku Madhusudhan, and Avi M. Mandell. HyDRo: atmospheric retrieval of rocky exoplanets in thermal emission. *MNRAS*, 511(2):2565–2584, April 2022. doi: 10.1093/mnras/stab3612.
- [110] Luis Welbanks and Nikku Madhusudhan. Aurora: A Generalized Retrieval Framework for Exoplanetary Transmission Spectra. *ApJ*, 913(2):114, June 2021. doi: 10.3847/1538-4357/abee94.
- [111] F. Forget and J. Leconte. Possible climates on terrestrial exoplanets. *Philosophical Transactions of the Royal Society of London Series A*, 372(2014):20130084–20130084, March 2014. doi: 10.1098/rsta.2013.0084.
- [112] Björn Benneke and Sara Seager. How to Distinguish between Cloudy Mini-Neptunes and Water/Volatile-dominated Super-Earths. *ApJ*, 778(2):153, December 2013. doi: 10.1088/0004-637X/778/2/153.

- [113] A. Caldas, J. Leconte, F. Selsis, I. P. Waldmann, P. Bordé, M. Rocchetto, and B. Charnay. Effects of a fully 3D atmospheric structure on exoplanet transmission spectra: retrieval biases due to day-night temperature gradients. *A&A*, 623:A161, March 2019. doi: 10.1051/0004-6361/201834384.
- [114] Prajwal Niraula, Julien de Wit, Iouli E. Gordon, Robert J. Hargreaves, Clara Sousa-Silva, and Roman V. Kochanov. The impending opacity challenge in exoplanet atmospheric characterization. *Nature Astronomy*, September 2022. doi: 10.1038/s41550-022-01773-1.
- [115] Jake Taylor. Impact of variable photospheric radius on exoplanet atmospheric retrievals. *MNRAS*, 513(1):L20–L24, June 2022. doi: 10.1093/mnrasl/slac028.
- [116] Luis Welbanks and Nikku Madhusudhan. On Atmospheric Retrievals of Exoplanets with Inhomogeneous Terminators. *ApJ*, 933(1):79, July 2022. doi: 10.3847/1538-4357/ac6df1.
- [117] Jacob L. Bean, Eliza M. R. Kempton, Guangwei Fu, Peter Gao, Jegug Ih, Tiffany Kataria, Matej Malik, Megan Mansfield, Vivien Parmentier, Emily Rauscher, Michael Roman, Kevin Stevenson, and Jake Taylor. Unlocking the Mysteries of the Archetype Sub-Neptune GJ1214b with a Full-Orbit Phase Curve. JWST Proposal. Cycle 1, ID. #1803, March 2021.
- [118] Renyu Hu and Mario Damiano. Deep Characterization of the Atmosphere of a Temperate Sub-Neptune. JWST Proposal. Cycle 1, ID. #2372, March 2021.
- [119] Olivia Lim, Loic Albert, Etienne Artigau, Bjorn Benneke, Nicolas B. Cowan, Rene Doyon, and David Lafreniere. Atmospheric reconnaissance of the TRAPPIST-1 planets. JWST Proposal. Cycle 1, ID. #2589, March 2021.
- [120] Kevin Stevenson, Jacob Lustig-Yaeger, Munazza Alam, Natasha Batalha, Mercedes Lopez-Morales, Joshua D. Lothringer, Ryan J. MacDonald, Erin M. May, Sarah E. Moran, Sarah Peacock, Zafar Rustamkulov, David K. Sing, Kristin Showalter Sotzen, Jeff A. Valenti, and Hannah Wakeford. Tell Me How I'm Supposed To Breathe With No Air: Measuring the Prevalence and Diversity of M-Dwarf Planet Atmospheres. JWST Proposal. Cycle 1, ID. #1981, March 2021.
- [121] J. Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160, 1982. doi: <https://doi.org/10.1111/j.2517-6161.1982.tb01195.x>. URL <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1982.tb01195.x>.
- [122] Mark R. Swain, Raissa Estrela, Gael M. Roudier, Christophe Sotin, Paul B. Rimmer, Adriana Valio, Robert West, Kyle Pearson, Noah Huber-Feely, and Robert T. Zellem. Detection of an Atmosphere on a Rocky Exoplanet. *AJ*, 161(5):213, May 2021. doi: 10.3847/1538-3881/abe879.
- [123] E. M. May, Ryan J. MacDonald, Katherine A. Bennett, Sarah E. Moran, Hannah R. Wakeford, Sarah Peacock, Jacob Lustig-Yaeger, Alicia N. Highland, Kevin B. Stevenson, David K. Sing, L. C. Mayorga, Natasha E. Batalha, James Kirk, Mercedes López-Morales, Jeff A. Valenti, Munazza K. Alam, Lili Alderson, Guangwei Fu, Junellie Gonzalez-Quiles,

- Joshua D. Lothringer, Zafar Rustamkulov, and Kristin S. Sotzen. Double Trouble: Two Transits of the Super-Earth GJ 1132 b Observed with JWST NIRSpec G395H. *ApJL*, 959 (1):L9, December 2023. doi: 10.3847/2041-8213/ad054f.
- [124] Zachory K. Berta-Thompson, Jonathan Irwin, David Charbonneau, Elisabeth R. Newton, Jason A. Dittmann, Nicola Astudillo-Defru, Xavier Bonfils, Michaël Gillon, Emmanuël Jehin, Antony A. Stark, Brian Stalder, Francois Bouchy, Xavier Delfosse, Thierry Forveille, Christophe Lovis, Michel Mayor, Vasco Neves, Francesco Pepe, Nuno C. Santos, Stéphane Udry, and Anaël Wünsche. A rocky planet transiting a nearby low-mass star. *Nature*, 527 (7577):204–207, November 2015. doi: 10.1038/nature15762.
- [125] Natasha E. Batalha, Avi Mandell, Klaus Pontoppidan, Kevin B. Stevenson, Nikole K. Lewis, Jason Kalirai, Nick Earl, Thomas Greene, Loïc Albert, and Louise D. Nielsen. Pan-Exo: A Community Tool for Transiting Exoplanet Science with JWST & HST. *PASP*, 129(976):064501, June 2017. doi: 10.1088/1538-3873/aa65b0.
- [126] Anjali A. A. Piette, Peter Gao, Kara Brugman, Anat Shahar, Tim Lichtenberg, Francesca Miozzi, and Peter Driscoll. Rocky Planet or Water World? Observability of Low-density Lava World Atmospheres. *ApJ*, 954(1):29, September 2023. doi: 10.3847/1538-4357/acdef2.
- [127] Ryan J. MacDonald and Nikole K. Lewis. TRIDENT: A Rapid 3D Radiative-transfer Model for Exoplanet Transmission Spectra. *ApJ*, 929(1):20, April 2022. doi: 10.3847/1538-4357/ac47fe.
- [128] Matthew C. Nixon and Nikku Madhusudhan. Aura-3D: A Three-dimensional Atmospheric Retrieval Framework for Exoplanet Transmission Spectra. *ApJ*, 935(2):73, August 2022. doi: 10.3847/1538-4357/ac7c09.
- [129] The JWST Transiting Exoplanet Community Early Release Science Team, Eva-Maria Ahrer, Lili Alderson, Natalie M. Batalha, Natasha E. Batalha, Jacob L. Bean, Thomas G. Beatty, Taylor J. Bell, Björn Benneke, Zachory K. Berta-Thompson, Aarynn L. Carter, Ian J. M. Crossfield, Néstor Espinoza, Adina D. Feinstein, Jonathan J. Fortney, Neale P. Gibson, Jayesh M. Goyal, Eliza M. R. Kempton, James Kirk, Laura Kreidberg, Mercedes López-Morales, Michael R. Line, Joshua D. Lothringer, Sarah E. Moran, Sagnick Mukherjee, Kazumasa Ohno, Vivien Parmentier, Caroline Piaulet, Zafar Rustamkulov, Everett Schlawin, David K. Sing, Kevin B. Stevenson, Hannah R. Wakeford, Natalie H. Allen, Stephan M. Birkmann, Jonathan Brande, Nicolas Crouzet, Patricio E. Cubillos, Mario Damiano, Jean-Michel Désert, Peter Gao, Joseph Harrington, Renyu Hu, Sarah Kendrew, Heather A. Knutson, Pierre-Olivier Lagage, Jérémy Leconte, Monika Lendl, Ryan J. MacDonald, E. M. May, Yamila Miguel, Karan Molaverdikhani, Julianne I. Moses, Catriona Anne Murray, Molly Nehring, Nikolay K. Nikolov, D. J. M. Petit dit de la Roche, Michael Radica, Pierre-Alexis Roy, Keivan G. Stassun, Jake Taylor, William C. Waalkes, Patcharapol Wachiraphan, Luis Welbanks, Peter J. Wheatley, Keshav Aggarwal, Munazza K. Alam, Agnibha Banerjee, Joanna K. Barstow, Jasmina Blečić, S. L. Casewell, Quentin Changeat, K. L. Chubb, Knicole D. Colón, Louis-Philippe Coulombe, Tansu Daylan, Miguel de Val-Borro, Leen Decin, Leonardo A. Dos Santos, Laura Flagg, Kevin

- France, Guangwei Fu, A. García Muñoz, John E. Gizis, Ana Glidden, David Grant, Kevin Heng, Thomas Henning, Yu-Cian Hong, Julie Inglis, Nicolas Iro, Tiffany Kataria, Thaddeus D. Komacek, Jessica E. Krick, Elspeth K. H. Lee, Nikole K. Lewis, Jorge Lillo-Box, Jacob Lustig-Yaeger, Luigi Mancini, Avi M. Mandell, Megan Mansfield, Mark S. Marley, Thomas Mikal-Evans, Giuseppe Morello, Matthew C. Nixon, Kevin Ortiz Ceballos, Anjali A. A. Piette, Diana Powell, Benjamin V. Rackham, Lakeisha Ramos-Rosado, Emily Rauscher, Seth Redfield, Laura K. Rogers, Michael T. Roman, Gael M. Roudier, Nicholas Scarsdale, Evgenya L. Shkolnik, John Southworth, Jessica J. Spake, Maria E Steinrueck, Xianyu Tan, Johanna K. Teske, Pascal Tremblin, Shang-Min Tsai, Gregory S. Tucker, Jake D. Turner, Jeff A. Valenti, Olivia Venot, Ingo P. Waldmann, Nicole L. Wallack, Xi Zhang, and Sebastian Zieba. Identification of carbon dioxide in an exoplanet atmosphere. *arXiv e-prints*, art. arXiv:2208.11692, August 2022.
- [130] Surya T. Tokdar and Robert E. Kass. Importance sampling: a review. *WIREs Computational Statistics*, 2(1):54–60, 2010. doi: <https://doi.org/10.1002/wics.56>. URL <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wics.56>.
- [131] Willie Neiswanger and Eric Xing. Post-Inference Prior Swapping. *arXiv e-prints*, art. arXiv:1606.00787, June 2016. doi: 10.48550/arXiv.1606.00787.
- [132] Michaël Gillon, Emmanuël Jehin, Susan M. Lederer, Laetitia Delrez, Julien de Wit, Artem Burdanov, Valérie Van Grootel, Adam J. Burgasser, Amaury H. M. J. Triaud, Cyrielle Opitom, Brice-Olivier Demory, Devendra K. Sahu, Daniella Bardalez Gagliuffi, Pierre Magain, and Didier Queloz. Temperate Earth-sized planets transiting a nearby ultracool dwarf star. *Nature*, 533(7602):221–224, May 2016. doi: 10.1038/nature17448.
- [133] Michaël Gillon, Amaury H. M. J. Triaud, Brice-Olivier Demory, Emmanuël Jehin, Eric Agol, Katherine M. Deck, Susan M. Lederer, Julien de Wit, Artem Burdanov, James G. Ingalls, Emeline Bolmont, Jeremy Leconte, Sean N. Raymond, Franck Selsis, Martin Turbet, Khalid Barkaoui, Adam Burgasser, Matthew R. Burleigh, Sean J. Carey, Aleksander Chaushev, Chris M. Copperwheat, Laetitia Delrez, Catarina S. Fernandes, Daniel L. Holdsworth, Enrico J. Kotze, Valérie Van Grootel, Yaseen Almlucky, Zouhair Benkhaldoun, Pierre Magain, and Didier Queloz. Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1. *Nature*, 542(7642):456–460, February 2017. doi: 10.1038/nature21360.
- [134] Martin Turbet, Emeline Bolmont, Vincent Bourrier, Brice-Olivier Demory, Jérémy Leconte, James Owen, and Eric T. Wolf. A Review of Possible Planetary Atmospheres in the TRAPPIST-1 System. *SSRv*, 216(5):100, July 2020. doi: 10.1007/s11214-020-00719-1.
- [135] Robin Wordsworth and Laura Kreidberg. Atmospheres of Rocky Exoplanets. *ARA&A*, 60: 159–201, August 2022. doi: 10.1146/annurev-astro-052920-125632.
- [136] Julien de Wit, Hannah R. Wakeford, Nikole K. Lewis, Laetitia Delrez, Michaël Gillon, Frank Selsis, Jérémy Leconte, Brice-Olivier Demory, Emeline Bolmont, Vincent Bourrier, Adam J. Burgasser, Simon Grimm, Emmanuël Jehin, Susan M. Lederer, James E. Owen,

- Vlada Stamenković, and Amaury H. M. J. Triaud. Atmospheric reconnaissance of the habitable-zone Earth-sized planets orbiting TRAPPIST-1. *Nature Astronomy*, 2:214–219, March 2018. doi: 10.1038/s41550-017-0374-z.
- [137] Hannah Diamond-Lowe, Zachory Berta-Thompson, David Charbonneau, and Eliza M. R. Kempton. Ground-based Optical Transmission Spectroscopy of the Small, Rocky Exoplanet GJ 1132b. *AJ*, 156(2):42, August 2018. doi: 10.3847/1538-3881/aac6dd.
- [138] Hannah Diamond-Lowe, David Charbonneau, Matej Malik, Eliza M. R. Kempton, and Yuri Beletsky. Optical Transmission Spectroscopy of the Terrestrial Exoplanet LHS 3844b from 13 Ground-based Transit Observations. *AJ*, 160(4):188, October 2020. doi: 10.3847/1538-3881/abaf4f.
- [139] Lorenzo V. Mugnai, Darius Modirrousta-Galian, Billy Edwards, Quentin Changeat, Jeroen Bouwman, Giuseppe Morello, Ahmed Al-Refaie, Robin Baeyens, Michelle Fabienne Bieger, Doriann Blain, Amélie Gressier, Gloria Guilluy, Yassin Jaziri, Flavien Kiefer, Mario Morvan, William Pluriel, Mathilde Poveda, Nour Skaf, Niall Whiteford, Sam Wright, Kai Hou Yip, Tiziano Zingales, Benjamin Charnay, Pierre Drossart, Jérémy Leconte, Olivia Venot, Ingo Waldmann, and Jean-Philippe Beaulieu. ARES. V. No Evidence For Molecular Absorption in the HST WFC3 Spectrum of GJ 1132 b. *AJ*, 161(6):284, June 2021. doi: 10.3847/1538-3881/abf3c3.
- [140] Jessica E. Libby-Roberts, Zachory K. Berta-Thompson, Hannah Diamond-Lowe, Michael A. Gully-Santiago, Jonathan M. Irwin, Eliza M. R. Kempton, Benjamin V. Rackham, David Charbonneau, Jean-Michel Désert, Jason A. Dittmann, Ryan Hofmann, Caroline V. Morley, and Elisabeth R. Newton. The Featureless HST/WFC3 Transmission Spectrum of the Rocky Exoplanet GJ 1132b: No Evidence for a Cloud-free Primordial Atmosphere and Constraints on Starspot Contamination. *AJ*, 164(2):59, August 2022. doi: 10.3847/1538-3881/ac75de.
- [141] J. Lustig-Yaeger, G. Fu, E. M. May, K. N. Ortiz Ceballos, S. E. Moran, S. Peacock, K. B. Stevenson, M. López-Morales, R. J. MacDonald, L. C. Mayorga, D. K. Sing, K. S. Sotzen, J. A. Valenti, J. Adams, M. K. Alam, N. E. Batalha, K. A. Bennett, J. Gonzalez-Quiles, J. Kirk, E. Kruse, J. D. Lothringer, Z. Rustamkulov, and H. R. Wakeford. A JWST transmission spectrum of a nearby Earth-sized exoplanet. *arXiv e-prints*, art. arXiv:2301.04191, January 2023. doi: 10.48550/arXiv.2301.04191.
- [142] John Southworth, Luigi Mancini, Nikku Madhusudhan, Paul Mollière, Simona Ciceri, and Thomas Henning. Detection of the Atmosphere of the 1.6 M_{\oplus} Exoplanet GJ 1132 b. *AJ*, 153(4):191, April 2017. doi: 10.3847/1538-3881/aa6477.
- [143] Sarah E. Moran, Kevin B. Stevenson, David K. Sing, Ryan J. MacDonald, James Kirk, Jacob Lustig-Yaeger, Sarah Peacock, L. C. Mayorga, Katherine A. Bennett, Mercedes López-Morales, E. M. May, Zafar Rustamkulov, Jeff A. Valenti, Jéa I. Adams Redai, Munazza K. Alam, Natasha E. Batalha, Guangwei Fu, Junellie Gonzalez-Quiles, Alicia N. Highland, Ethan Kruse, Joshua D. Lothringer, Kevin N. Ortiz Ceballos, Kristin S. Sotzen, and Hannah R. Wakeford. High Tide or Riptide on the Cosmic Shoreline? A

Water-Rich Atmosphere or Stellar Contamination for the Warm Super-Earth GJ-486b from JWST Observations. *arXiv e-prints*, art. arXiv:2305.00868, May 2023. doi: 10.48550/arXiv.2305.00868.

- [144] Laura Kreidberg, Daniel D. B. Koll, Caroline Morley, Renyu Hu, Laura Schaefer, Drake Deming, Kevin B. Stevenson, Jason Dittmann, Andrew Vanderburg, David Berardo, Xueying Guo, Keivan Stassun, Ian Crossfield, David Charbonneau, David W. Latham, Abraham Loeb, George Ricker, Sara Seager, and Roland Vanderspek. Absence of a thick atmosphere on the terrestrial exoplanet LHS 3844b. *Nature*, 573(7772):87–90, August 2019. doi: 10.1038/s41586-019-1497-4.
- [145] Ian J. M. Crossfield, Matej Malik, Michelle L. Hill, Stephen R. Kane, Bradford Foley, Alex S. Polanski, David Coria, Jonathan Brande, Yanzhe Zhang, Katherine Wienke, Laura Kreidberg, Nicolas B. Cowan, Diana Dragomir, Varoujan Gorjian, Thomas Mikal-Evans, Björn Benneke, Jessie L. Christiansen, Drake Deming, and Farisa Y. Morales. GJ 1252b: A Hot Terrestrial Super-Earth with No Atmosphere. *ApJL*, 937(1):L17, September 2022. doi: 10.3847/2041-8213/ac886b.
- [146] J. Krissansen-Totton and J. J. Fortney. Predictions for Observable Atmospheres of Trappist-1 Planets from a Fully Coupled Atmosphere-Interior Evolution Model. *ApJ*, 933(1):115, July 2022. doi: 10.3847/1538-4357/ac69cb.
- [147] Renyu Hu, Bethany L. Ehlmann, and Sara Seager. Theoretical Spectra of Terrestrial Exoplanet Surfaces. *ApJ*, 752(1):7, June 2012. doi: 10.1088/0004-637X/752/1/7.
- [148] Eric Agol, Caroline Dorn, Simon L. Grimm, Martin Turbet, Elsa Ducrot, Laetitia Delrez, Michaël Gillon, Brice-Olivier Demory, Artem Burdanov, Khalid Barkaoui, Zouhair Benkhaldoun, Emeline Bolmont, Adam Burgasser, Sean Carey, Julien de Wit, Daniel Fabrycky, Daniel Foreman-Mackey, Jonas Haldemann, David M. Hernandez, James Ingalls, Emmanuel Jehin, Zachary Langford, Jérémy Leconte, Susan M. Lederer, Rodrigo Luger, Renu Malhotra, Victoria S. Meadows, Brett M. Morris, Francisco J. Pozuelos, Didier Queloz, Sean N. Raymond, Franck Selsis, Marko Sestovic, Amaury H. M. J. Triaud, and Valerie Van Grootel. Refining the Transit-timing and Photometric Analysis of TRAPPIST-1: Masses, Radii, Densities, Dynamics, and Ephemerides. *PSJ*, 2(1):1, February 2021. doi: 10.3847/PSJ/abd022.
- [149] Aishwarya R. Iyer, Michael R. Line, Philip S. Muirhead, Jonathan J. Fortney, and Ehsan Gharib-Nezhad. The SPHINX M-dwarf Spectral Grid. I. Benchmarking New Model Atmospheres to Derive Fundamental M-dwarf Properties. *ApJ*, 944(1):41, February 2023. doi: 10.3847/1538-4357/acabc2.
- [150] Andrew P. Lincowski, Victoria S. Meadows, David Crisp, Tyler D. Robinson, Rodrigo Luger, Jacob Lustig-Yaeger, and Giada N. Arney. Evolved Climates and Observational Discriminants for the TRAPPIST-1 Planetary System. *ApJ*, 867(1):76, November 2018. doi: 10.3847/1538-4357/aae36a.

- [151] Renyu Hu, Luke Peterson, and Eric T. Wolf. O₂- and CO-rich Atmospheres for Potentially Habitable Environments on TRAPPIST-1 Planets. *ApJ*, 888(2):122, January 2020. doi: 10.3847/1538-4357/ab5f07.
- [152] Julien de Wit, Hannah R. Wakeford, Michaël Gillon, Nikole K. Lewis, Jeff A. Valenti, Brice-Olivier Demory, Adam J. Burgasser, Artem Burdanov, Laetitia Delrez, Emmanuël Jehin, Susan M. Lederer, Didier Queloz, Amaury H. M. J. Triaud, and Valérie Van Grootel. A combined transmission spectrum of the Earth-sized exoplanets TRAPPIST-1 b and c. *Nature*, 537(7618):69–72, September 2016. doi: 10.1038/nature18641.
- [153] Linda T. Elkins-Tanton. Magma Oceans in the Inner Solar System. *Annual Review of Earth and Planetary Sciences*, 40(1):113–139, May 2012. doi: 10.1146/annurev-earth-042711-105503.
- [154] Olivia Lim, Loic Albert, Etienne Artigau, Bjorn Benneke, Nicolas B. Cowan, Rene Doyon, and David Lafreniere. Atmospheric reconnaissance of the TRAPPIST-1 planets. JWST Proposal. Cycle 1, ID. #2589, March 2021.
- [155] Eliza Miller-Ricci, Sara Seager, and Dimitar Sasselov. The Atmospheric Signatures of Super-Earths: How to Distinguish Between Hydrogen-Rich and Hydrogen-Poor Atmospheres. *ApJ*, 690(2):1056–1067, January 2009. doi: 10.1088/0004-637X/690/2/1056.
- [156] J. K. Barstow and P. G. J. Irwin. Habitable worlds with JWST: transit spectroscopy of the TRAPPIST-1 system? *MNRAS*, 461(1):L92–L96, September 2016. doi: 10.1093/mnras/slw109.
- [157] E. Ducrot, M. Gillon, L. Delrez, E. Agol, P. Rimmer, M. Turbet, M. N. Günther, B. O. Demory, A. H. M. J. Triaud, E. Bolmont, A. Burgasser, S. J. Carey, J. G. Ingalls, E. Jehin, J. Leconte, S. M. Lederer, D. Queloz, S. N. Raymond, F. Selsis, V. Van Grootel, and J. de Wit. TRAPPIST-1: Global results of the Spitzer Exploration Science Program Red Worlds. *A&A*, 640:A112, August 2020. doi: 10.1051/0004-6361/201937392.
- [158] Benjamin V. Rackham, Dániel Apai, and Mark S. Giampapa. The Transit Light Source Effect: False Spectral Features and Incorrect Densities for M-dwarf Transiting Planets. *ApJ*, 853(2):122, February 2018. doi: 10.3847/1538-4357/aaa08c.
- [159] Benjamin V. Rackham and Julien de Wit. Towards robust corrections for stellar contamination in JWST exoplanet transmission spectra. *arXiv e-prints*, art. arXiv:2303.15418, March 2023. doi: 10.48550/arXiv.2303.15418.
- [160] Rostom Mbarek and Eliza M. R. Kempton. Clouds in Super-Earth Atmospheres: Chemical Equilibrium Calculations. *ApJ*, 827(2):121, August 2016. doi: 10.3847/0004-637X/827/2/121.
- [161] Laura Kreidberg, Eric Agol, Emeline Bolmont, Michael Gillon, Renyu Hu, Daniel Koll, Avi Mandell, Victoria Suzanne Meadows, Caroline Morley, Laura Schaefer, Franck Selsis, and Julien de Wit. Hot Take on a Cool World: Does Trappist-1c Have an Atmosphere? JWST Proposal. Cycle 1, ID. #2304, March 2021.

- [162] Megan Mansfield, Jacob L. Bean, Eliza M. R. Kempton, Edwin S. Kite, Daniel Koll, and Matej Malik. Constraining the Atmosphere of the Terrestrial Exoplanet Gl486b. JWST Proposal. Cycle 1, ID. #1743, March 2021.
- [163] Jonathan I. Lunine and Jacob L. Bean. Extrasolar Planet Science with JWST. JWST Proposal. Cycle 1, ID. #1274, July 2017.
- [164] Laura Kreidberg, Renyu Hu, Edwin S. Kite, Daniel Koll, Matej Malik, Caroline Morley, Laura Schaefer, and Emily Whittaker. A Search for Signatures of Volcanism and Geodynamics on the Hot Rocky Exoplanet LHS 3844b. JWST Proposal. Cycle 1, ID. #1846, March 2021.
- [165] Olivia Lim, Björn Benneke, René Doyon, Ryan J. MacDonald, Caroline Piaulet, Étienne Artigau, Louis-Philippe Coulombe, Michael Radica, Alexandrine L’Heureux, Loïc Albert, Benjamin V. Rackham, Julien de Wit, Salma Salhi, Pierre-Alexis Roy, Laura Flagg, Marylou Fournier-Tondreau, Jake Taylor, Neil J. Cook, David Lafrenière, Nicolas B. Cowan, Lisa Kaltenegger, Jason F. Rowe, Néstor Espinoza, Lisa Dang, and Antoine Darveau-Bernier. Atmospheric Reconnaissance of TRAPPIST-1 b with JWST/NIRISS: Evidence for Strong Stellar Contamination in the Transmission Spectra. *ApJL*, 955(1):L22, September 2023. doi: 10.3847/2041-8213/acf7c4.
- [166] Nicole L. Wallack, Natasha E. Batalha, Lili Alderson, Nicholas Scarsdale, Jea I. Adams Redai, Artyom Aguichine, Munazza K. Alam, Peter Gao, Angie Wolfgang, Natalie M. Batalha, James Kirk, Mercedes López-Morales, Sarah E. Moran, Johanna Teske, Hannah R. Wakeford, and Nicholas F. Wogan. JWST COMPASS: A NIRSpec/G395H Transmission Spectrum of the Sub-Neptune TOI-836c. *arXiv e-prints*, art. arXiv:2404.01264, April 2024. doi: 10.48550/arXiv.2404.01264.
- [167] Natasha E. Batalha, Angie Wolfgang, Johanna Teske, Munazza K. Alam, Lili Alderson, Natalie M. Batalha, Mercedes López-Morales, and Hannah R. Wakeford. Importance of Sample Selection in Exoplanet-atmosphere Population Studies. *AJ*, 165(1):14, January 2023. doi: 10.3847/1538-3881/ac9f45.
- [168] Aomawa L. Shields, Sarah Ballard, and John Asher Johnson. The habitability of planets orbiting M-dwarf stars. *PhR*, 663:1, December 2016. doi: 10.1016/j.physrep.2016.10.003.
- [169] E. Magaudda, B. Stelzer, K. R. Covey, St. Raetz, S. P. Matt, and A. Scholz. Relation of X-ray activity and rotation in M dwarfs and predicted time-evolution of the X-ray luminosity. *A&A*, 638:A20, June 2020. doi: 10.1051/0004-6361/201937408.
- [170] Xintong Lyu, Daniel D. B. Koll, Nicolas B. Cowan, Renyu Hu, Laura Kreidberg, and Brian E. J. Rose. Super-Earth LHS3844b is Tidally Locked. *ApJ*, 964(2):152, April 2024. doi: 10.3847/1538-4357/ad2077.
- [171] B. Hapke. Bidirectional reflectance spectroscopy. I - Theory. *J. Geophys. Res.*, 86:3039–3054, April 1981. doi: 10.1029/JB086iB04p03039.

- [172] Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom, E. Bray, T. Aldcroft, M. Davis, A. Ginsburg, A. M. Price-Whelan, W. E. Kerzendorf, A. Conley, N. Crighton, K. Barbary, D. Muna, H. Ferguson, F. Grollier, M. M. Parikh, P. H. Nair, H. M. Unther, C. Deil, J. Woillez, S. Conseil, R. Kramer, J. E. H. Turner, L. Singer, R. Fox, B. A. Weaver, V. Zabalza, Z. I. Edwards, K. Azalee Bostroem, D. J. Burke, A. R. Casey, S. M. Crawford, N. Dencheva, J. Ely, T. Jenness, K. Labrie, P. L. Lim, F. Pierfederici, A. Pontzen, A. Ptak, B. Refsdal, M. Servillat, and O. Streicher. Astropy: A community Python package for astronomy. *A&A*, 558:A33, October 2013. doi: 10.1051/0004-6361/201322068.
- [173] Astropy Collaboration, A. M. Price-Whelan, B. M. Sipőcz, H. M. Günther, P. L. Lim, S. M. Crawford, S. Conseil, D. L. Shupe, M. W. Craig, N. Dencheva, A. Ginsburg, J. T. VanderPlas, L. D. Bradley, D. Pérez-Suárez, M. de Val-Borro, T. L. Aldcroft, K. L. Cruz, T. P. Robitaille, E. J. Tollerud, C. Ardelean, T. Babej, Y. P. Bach, M. Bachetti, A. V. Bakanov, S. P. Bamford, G. Barentsen, P. Barmby, A. Baumbach, K. L. Berry, F. Biscani, M. Boquien, K. A. Bostroem, L. G. Bouma, G. B. Brammer, E. M. Bray, H. Breytenbach, H. Buddelmeijer, D. J. Burke, G. Calderone, J. L. Cano Rodríguez, M. Cara, J. V. M. Cardoso, S. Cheedella, Y. Copin, L. Corrales, D. Crichton, D. D’Avella, C. Deil, É. Depagne, J. P. Dietrich, A. Donath, M. Droettboom, N. Earl, T. Erben, S. Fabbro, L. A. Ferreira, T. Finethy, R. T. Fox, L. H. Garrison, S. L. J. Gibbons, D. A. Goldstein, R. Gommers, J. P. Greco, P. Greenfield, A. M. Groener, F. Grollier, A. Hagen, P. Hirst, D. Homeier, A. J. Horton, G. Hosseinzadeh, L. Hu, J. S. Hunkeler, Ž. Ivezić, A. Jain, T. Jenness, G. Kanarek, S. Kendrew, N. S. Kern, W. E. Kerzendorf, A. Khvalko, J. King, D. Kirkby, A. M. Kulkarni, A. Kumar, A. Lee, D. Lenz, S. P. Littlefair, Z. Ma, D. M. Macleod, M. Mastroiello, C. McCully, S. Montagnac, B. M. Morris, M. Mueller, S. J. Mumford, D. Muna, N. A. Murphy, S. Nelson, G. H. Nguyen, J. P. Ninan, M. Nöthe, S. Ogaz, S. Oh, J. K. Parejko, N. Parley, S. Pascual, R. Patil, A. A. Patil, A. L. Plunkett, J. X. Prochaska, T. Rastogi, V. Reddy Janga, J. Sabater, P. Sakurikar, M. Seifert, L. E. Sherbert, H. Sherwood-Taylor, A. Y. Shih, J. Sick, M. T. Silbiger, S. Singanamalla, L. P. Singer, P. H. Sladen, K. A. Sooley, S. Sornarajah, O. Streicher, P. Teuben, S. W. Thomas, G. R. Tremblay, J. E. H. Turner, V. Terrón, M. H. van Kerkwijk, A. de la Vega, L. L. Watkins, B. A. Weaver, J. B. Whitmore, J. Woillez, V. Zabalza, and Astropy Contributors. The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *AJ*, 156(3):123, September 2018. doi: 10.3847/1538-3881/aabc4f.
- [174] Astropy Collaboration, Adrian M. Price-Whelan, Pey Lian Lim, Nicholas Earl, Nathaniel Starkman, Larry Bradley, David L. Shupe, Aarya A. Patil, Lia Corrales, C. E. Brasseur, Maximilian Nöthe, Axel Donath, Erik Tollerud, Brett M. Morris, Adam Ginsburg, Eero Vaher, Benjamin A. Weaver, James Tocknell, William Jamieson, Marten H. van Kerkwijk, Thomas P. Robitaille, Bruce Merry, Matteo Bachetti, H. Moritz Günther, Thomas L. Aldcroft, Jaime A. Alvarado-Montes, Anne M. Archibald, Attila Bódi, Shreyas Bapat, Geert Barentsen, Juanjo Bazán, Manish Biswas, M’ed’eric Boquien, D. J. Burke, Daria Cara, Mihai Cara, Kyle E. Conroy, Simon Conseil, Matthew W. Craig, Robert M. Cross, Kelle L. Cruz, Francesco D’Eugenio, Nadia Dencheva, Hadrien A. R. Devillepoix, J’org P. Dietrich, Arthur Davis Eigenbrot, Thomas Erben, Leonardo Ferreira, Daniel

- Foreman-Mackey, Ryan Fox, Nabil Frejj, Suyog Garg, Robel Geda, Lauren Glattly, Yash Gondhalekar, Karl D. Gordon, David Grant, Perry Greenfield, Austen M. Groener, Steve Guest, Sebastian Gurovich, Rasmus Handberg, Akeem Hart, Zac Hatfield-Dodds, Derek Homeier, Griffin Hosseinzadeh, Tim Jenness, Craig K. Jones, Prajwel Joseph, J. Bryce Kalmbach, Emir Karamehmetoglu, Mikolaj Kaluszy'nski, Michael S. P. Kelley, Nicholas Kern, Wolfgang E. Kerzendorf, Eric W. Koch, Shankar Kulumani, Antony Lee, Chun Ly, Zhiyuan Ma, Conor MacBride, Jakob M. Maljaars, Demitri Muna, N. A. Murphy, Henrik Norman, Richard O'Steen, Kyle A. Oman, Camilla Pacifici, Sergio Pascual, J. Pascual-Granado, Rohit R. Patil, Gabriel I. Perren, Timothy E. Pickering, Tanuj Rastogi, Benjamin R. Roulston, Daniel F. Ryan, Eli S. Rykoff, Jose Sabater, Parikshit Sakurikar, Jes'us Salgado, Aniket Sanghi, Nicholas Saunders, Volodymyr Savchenko, Ludwig Schwaradt, Michael Seifert-Eckert, Albert Y. Shih, Anany Shrey Jain, Gyanendra Shukla, Jonathan Sick, Chris Simpson, Sudheesh Singanamalla, Leo P. Singer, Jaladh Singhal, Manodeep Sinha, Brigitta M. SipHocz, Lee R. Spitler, David Stansby, Ole Streicher, Jani Sumak, John D. Swinbank, Dan S. Taranu, Nikita Tewary, Grant R. Tremblay, Miguel de Val-Borro, Samuel J. Van Kooten, Zlatan Vasovi'c, Shresth Verma, Jos'e Vin'icius de Miranda Cardoso, Peter K. G. Williams, Tom J. Wilson, Benjamin Winkel, W. M. Wood-Vasey, Rui Xue, Peter Yoachim, Chen Zhang, Andrea Zonca, and Astropy Project Contributors. The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package. *ApJ*, 935(2):167, August 2022. doi: 10.3847/1538-4357/ac7c74.
- [175] Joshua S. Speagle. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *MNRAS*, 493(3):3132–3158, April 2020. doi: 10.1093/mnras/staa278.
- [176] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: The MCMC Hammer. *PASP*, 125(925):306, March 2013. doi: 10.1086/670067.
- [177] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.
- [178] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL <https://doi.org/10.1038/s41586-020-2649-2>.
- [179] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H.

Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.