SAPIENZA
UNIVERSITÀ DI ROMA

Characterization of the Atmospheres of Extrasolar Planets with Ariel

Scienze Matematiche Fisiche e Naturali

Doctor of Philosophy in Astronomy, Astrophysics, and Space Science (XXXVI cycle)

Andrea Bocchieri
ID number 1714815

Advisor
Prof. Enzo Pascale

Academic Year 2022/2023

Thesis defended on 21 December 2023
in front of the National Board of Examiners composed by:

Prof. Stefano Borgani, University of Trieste (chairman)
Prof. Salvatore Capozziello, University of Napoli Federico II
Prof. Francesco Rosario Ferraro, University of Bologna

External Referee commission composed by:

Prof. Giorgio Savini, University College London
Prof. Ingo Waldmann, University College London

Characterization of the Atmospheres of Extrasolar Planets with Ariel
Ph.D. thesis. Sapienza University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: andrea.bocchieri@uniroma1.it

Abstract

Extrasolar planet (exoplanet) science has rapidly expanded over the past decades, with thousands of exoplanets discovered using techniques such as the transit method. As the field progressed, the focus shifted from measuring fundamental planetary parameters to understanding the nature of these faraway worlds. The James Webb Space Telescope is a monumental step in this direction. However, studying select targets needs complementing with unbiased surveys of the diverse exoplanets we can currently observe. This will be a legacy of the Ariel space mission, which will perform the first spectroscopic survey of the atmospheres of exoplanets, enabling comparative planetology studies on a Galactic scale. This thesis presents my work over three years of Ph.D. research to prepare exoplanet atmospheric characterization with Ariel. Preparing the Ariel mission requires extensive studies, from optimizing the mission design and observation strategy to developing representative simulators and data analysis tools. At a high level, this thesis is a three-stage progression: it begins by framing the scientific context and motivation for Ariel, then evaluates the expected performance of the mission, and finally develops tools to reduce and analyze simulated Ariel data. Characterizing the atmospheres of exoplanets with Ariel requires the combined expertise and dedication of numerous scientists, engineers, and institutions. While advanced instrumentation enables the mission, it is the people involved who drive its success. Ariel is more than flying hardware: it is a culmination of human ingenuity, teamwork, and vision. My work fits into this broader effort of individuals collaborating to maximize Ariel's scientific potential and ensure its legacy, on the path of understanding these alien worlds.

Dedicated to
my late grandfathers.

Antonino,
who taught me to be curious,
and Giovanni Battista,
who taught me to be patient.

Preface

The field of exoplanetary science has seen tremendous growth in recent years, fueled by continuous advancements in observational techniques and technology. As we set our sights on understanding the nature and formation of these distant worlds, it is essential to develop a comprehensive and systematic approach to studying their atmospheres. In this Ph.D. thesis, I present my contributions to the characterization of extrasolar planetary atmospheres with the Ariel space mission.

In approaching this thesis, I have chosen a three-tiered structure of logos, ethos, and pathos to present my work. This structure is inspired by the Aristotelian modes of persuasion, which I believe provide an effective framework for presenting the technical, mission-oriented, and scientific impact aspects of the research. However, elements of each mode are inherently interwoven throughout the thesis.

"Part I: Logos - Scientific Framework, Technical Contributions, and Development" begins by providing a concise overview of the scientific context and the objectives of the Ariel space mission. Then, it delves into the technical aspects of my research, detailing the software learning and contributions, as well as the development of a physical optics propagation simulator. By discussing these general and technical aspects, this part emphasizes the logical underpinnings and methodological foundations of my research, its *Logos*.

"Part II: Ethos - Mission Performance, Calibration, and Data Reduction Strategies" focuses on examining the various aspects of the Ariel mission's performance, as well as my contribution to developing the processes and techniques necessary to analyze the collected data. Through a detailed discussion of the noise budget, mission validation, and data detrending, this part presents the rigorous standards and requirements in my doctoral work: its *Ethos*.

"Part III: Pathos - Scientific Impact and Future Prospects" illustrates the scientific impact and future prospects of the study. Through a discussion of publications, collaborations, and proposed observations, this part emphasizes the broader implications of my work for our understanding of extrasolar planets and their atmospheres. It highlights the potential for interdisciplinary collaboration and the development of innovative research methodologies. Most of all, it conveys the central theme in my research: the exploration of the unknown, its *Pathos*.

* * *

The boundless reaches of outer space have often been likened to a Cosmic Ocean, and just like a seafarer traversing the open waters, the scientist who seeks to explore them needs a suitable vessel and a skilled crew. In my voyage of exploration, I have had the most suitable vessel, "Sapienza" University of Rome, the best captain, Enzo, and the best admiral,

Giovanna. As the captain's attendant, I have kept a logbook of our expeditions, which I recount in this thesis.

But first, I want to thank two notable crew members: Lorenzo, for instructing and mentoring me as the former attendant, and Jiří, for inspiring me to always gaze beyond the ship's windows. Other friends and crew members that are always invited guests for a good beer and cheer are Angelos, Luke, Quentin, Billy, Howard, William, Yassin, and Anastasia. Whatever the storm, this crew will bring you to the stars.

Some days at sea, however, exact a heavy toll. They say that in those days a dark mood befalls the mariners and their hearts start to covet just one thing: a home. These borrowed rhymes put it better than I ever could. *Day after day, day after day,\ We stuck, nor breath nor motion;\ As idle as a painted ship\ Upon a painted ocean* [S.T. Coleridge, The Rime of the Ancient Mariner].

When the wind fails to blow and the quest seems doomed, the sailors without a home are lost in despair. I, however, have a home, and my home is with Caterina. Rephrasing from another master of Literature, J.R.R. Tolkien, without her "I would not do half of what I do half as well as I should like; and I would like to do less than half of what I do half as well as it deserves."

Contents

Part I

Logos
Scientific Framework, Technical Contributions, and Development

Chapter 1

Introduction

This chapter provides a brief introduction to the field of exoplanetary science and the Ariel mission. It sets the stage for the entire thesis, outlining the scope and objectives of my doctoral work and briefly introducing the computational tools used throughout.

1.1 A Primer to Exoplanetary Science

The idea that planets are not unique to the solar system, but could be found orbiting other stars as well, was proposed by a few scientists and philosophers like Giordano Bruno in the past[1]. This was a radical statement, as it challenged the axioms of Aristotelian natural philosophy[2], which dominated the world cultural stage at the time. Due to a lack of evidence to support either argument, the debate died out eventually. It was not until 1992 that the first planets outside the solar system, or *exoplanets*, were discovered around an unlikely host, a pulsar (Wolszczan and Frail, 1992). Soon after, Mayor and Queloz (1995) struck gold with the discovery of *51 Pegasi b*, the first exoplanet found orbiting a main-sequence star, which was awarded the Nobel Prize in Physics in 2019. The observations now support a scenario where on average each star hosts at least one planet (Howard et al., 2010; Cassan et al., 2012), and therefore there are more than $\sim$ 100 billion planets in the Milky Way alone. We are the first generation to know with absolute certainty that other planetary systems besides our own exist. This profound acknowledgment has sparked our most vivid dreams that, if we continue to look, someday we may find evidence of life elsewhere in the Universe. At present, a major scientific goal for astronomers worldwide is "Understanding the connections between stars and the worlds that orbit them" (National Academies of Sciences, 2021) and "[...] the conditions for planet formation and the emergence of life" (ESA, 2015). To achieve this

[1] Excerpt from a dialogue between two imaginary characters in (Bruno, 1584) I translated.

> <Elpino> Thus, there are innumerable Suns and Earths that orbit around those Suns just as these seven orbit this Sun close to us.
> <Filoteo> Indeed.
> <Elpino> Then, how is it possible that we cannot see these Earths orbiting their Suns [...]?
> <Filoteo> That is because we see the Suns, which are the largest bodies, but we cannot see the Earths which are much smaller and are thus invisible.

[2] Notably the idea that sublunary elements occupied or strove to return to their natural places, that is, the elemental spheres, at the center of the cosmos (Knox, 2019).

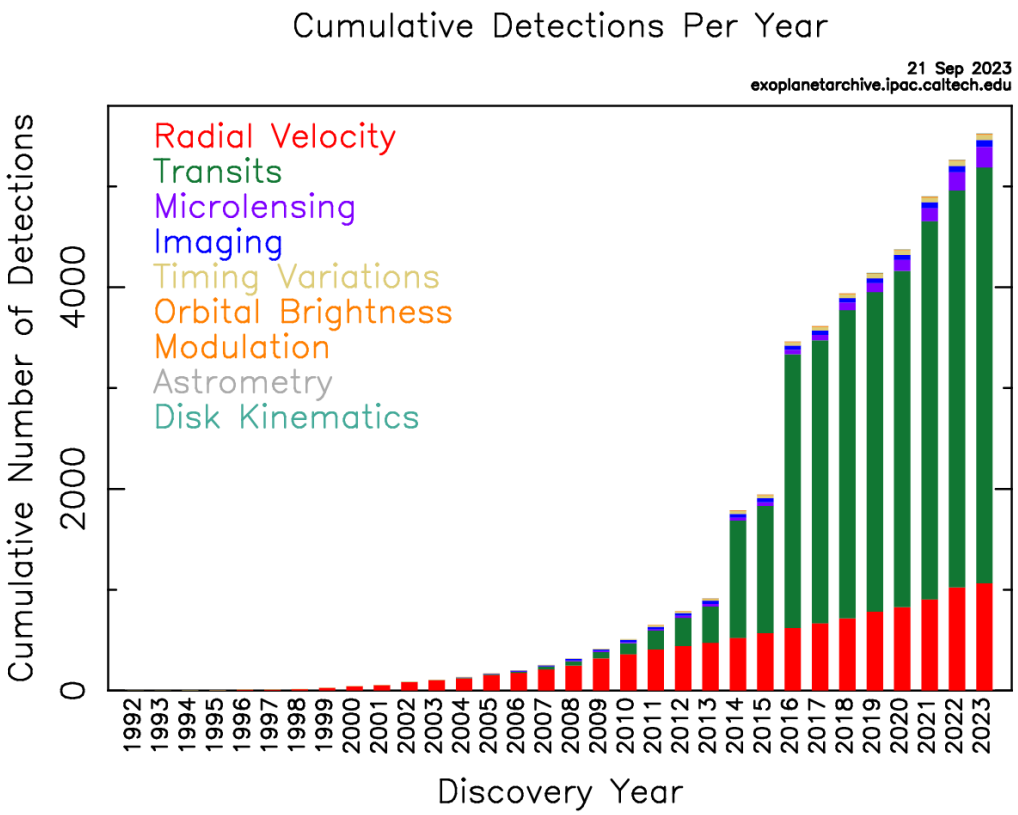


Figure 1.1. Cumulative histogram of confirmed exoplanet discoveries over the years, colored according to the method used for their detection. The discovery of *51 Pegasi b* in 1995 is the turning point in this timeline. The two large spikes in 2014 and 2016 correspond to the data releases of the NASA Kepler space mission (Koch et al., 2001). From the NASA Exoplanet Archive.

goal, we need to discover and characterize a large sample of exoplanets. The cumulative discoveries of exoplanets are represented in Figure 1.1 on a yearly basis. In total, more than 5000 exoplanets have been revealed using different detection techniques including the transit method, which yielded the largest number of discoveries to date. However, this number is tiny when compared to the billions of planets expected in our Galaxy, and the sample is not complete, i.e. statistically unbiased. In this respect, the Kepler (Koch et al., 2001) space mission, discontinued in 2018, was a stepping stone in the field. Launched to search a portion of the Milky Way galaxy for Earth-sized planets orbiting stars outside our solar system, Kepler provided a large and homogenous sample of exoplanets, which enabled the first statistical studies of the exoplanetary population (Batalha, 2014). Another pioneering mission was the CoRoT space telescope (Baglin et al., 2006; Barge and CoRoT Exoplanet Science Team, 2009), which was launched in 2007 and operated until 2013. CoRoT was the first space mission dedicated to the search for exoplanets using the transit method, opening the way for more advanced probes such as Kepler.

The number of confirmed exoplanets will increase significantly in the coming years, thanks to existing and future space missions such as TESS (Ricker et al., 2015), CHEOPS (Fortier et al., 2014; Cessa et al., 2017), Gaia (Gaia Collaboration et al., 2016), and PLATO (Rauer, 2013), as well as ground-based instrumentation such as HARPS (Mayor et al., 2003),

WASP (Pollacco et al., 2006), OGLE (Udalski et al., 2015), HATNet (Bakos, 2018), KELT (Pepper et al., 2018), and NGTS (Wheatley et al., 2018). However, as will be discussed in the next sections, the emphasis in the field is expanding from the discovery of exoplanets to their characterization. This shift in focus is driven by the need to understand the physical and chemical properties of exoplanets, in relation to their host stars and the environment in which they formed and evolved. This understanding is essential to advancing our knowledge of the formation and evolution of planetary systems, and ultimately to answering the question of what is the true nature of (exo)planets.

1.1.1 Exoplanet Demographics in 2023

For centuries, the terrestrial and gas giant planets in the solar system have been our only templates for planetary bodies in the Galaxy, and we have been able to study them in detail (Galanti et al., 2019; Cao et al., 2020). As a consequence, our understanding of planetary systems was predominantly heliocentric. Current observations of exoplanets, however, suggest that the processes of formation and evolution of planetary systems produce a gamut of outcomes (e.g. Turrini et al., 2015; Zhu and Dong, 2021), with a diverse range of parameters in terms of masses, sizes, orbits, and types of planets. These parameters often differ from those of the planets in the solar system, calling for a Copernican revolution. For instance, *51 Pegasi b* was the first discovered Hot Jupiter, a Jupiter-sized planet with an orbital period of less than ~10 days, and served as a prototype for a new class of planets[3]. Moreover, observations indicate that super-Earths like *Proxima Centauri b*, rocky exoplanets with a mass between ~2 and 10 Earth masses, absent from the solar system, are among the most common in our Galaxy.

1.1.2 Observational Constraints

Four or five key diagrams can represent the basic properties of exoplanets (Pudritz et al., 2018). However, before discussing them, it is crucial to consider the observational biases that may affect the interpretation and generalization of the data. For instance, the solar system seems to be unique in the Milky Way because our current technical capabilities cannot detect an exoplanetary system with Earth-sized planets in 1 AU orbits around a G-type star and large gas giants in decade-long outer orbits (Howard et al., 2012). Moreover, transiting planets – planets that cross the disk of their host star with respect to our line of sight – provide the most informative data set on the properties of exoplanets orbiting close to their host star (within 1 AU). However, the transit method is inherently biased towards detecting short-period and large planets, due to both geometric and detection factors (Kipping and Sandford, 2016).

Another notable observational bias is the one that affects the shape of the distribution of the masses of the exoplanet host stars. The peak emission of G-type stars is at visible wavelengths, matching the peak sensitivity of charge-coupled device (CCD) detectors, used e.g. on the focal plane of Kepler (Koch et al., 2001), which reported the highest number of exoplanet discoveries so far. Thus, the observed mass distribution of exoplanet host stars is

[3] In a famously ignored paper, Struve (1952) actually asserted Doppler measurements had become good enough to detect planets – if they existed – at least as massive as Jupiter with orbital periods as short as a few days, which no physical law forbade. However, the discoverers of *51 Pegasi b* were completely uninfluenced by this paper, and the first Hot Jupiter was a genuine surprise.

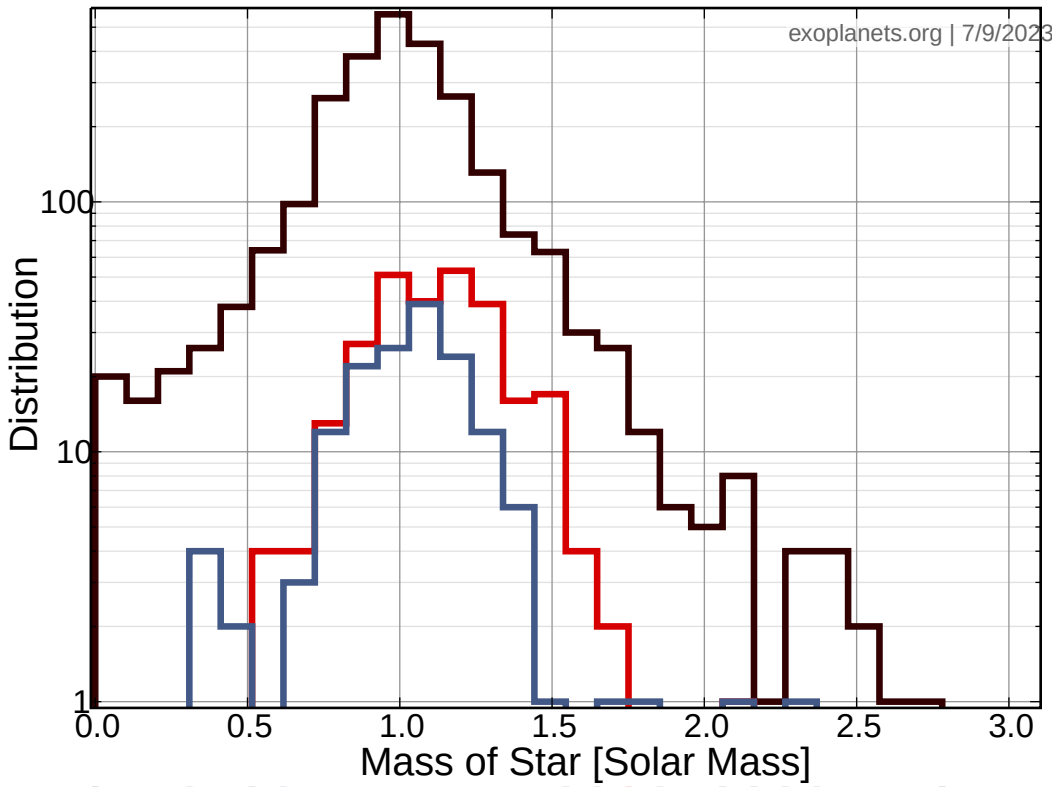


Figure 1.2. Observed mass distribution of exoplanet stellar hosts, from the Exoplanet Orbit Database. Black: full data set of planets. Blue: data set restricted to exoplanets found in peer-reviewed literature before 2005. Red: data set restricted to stars with hot Jupiters as planetary companions. The red and blue distributions are similar because Doppler radial velocity measurements of G-type stars provided the first exoplanet discoveries, essentially hot Jupiters (Deeg and Belmonte, 2018).

primarily centered around $\sim 1 M_{\odot}$ (see Figure 1.2). Detecting planets around more common and cooler M-type stars requires detectors sensitive to infrared wavelengths.

Interestingly, one detection method, microlensing – the method of detecting exoplanets by observing the effects of the gravitational field of a planet on the passing light of a distant background star, is largely unaffected by these observational biases. In microlensing, host stars are essentially selected by nature, independent of the observer. Consequently, the properties of these stars, such as their spectral type or activity, are not significantly biased. As a result, both planets and their host stars should be detected in a proportion reflecting their true occurrence within the Galaxy's disk (Jørgensen and Hundertmark, 2018). However, microlensing events, though unbiased, are rare and challenging to validate using other methods, which limits their statistical significance and the potential for detailed characterization.

In discussing basic exoplanet properties, the mass-period (or semimajor axis) diagram (see Figure 1.3) is a crucial initial reference. Here, planets are divided into three or four populations: Jovian planets, hot Jupiters, super-Earths, and mini-Neptunes. Jovian planets cluster around Earth's orbital period, while hot Jupiters are slightly less massive, boasting orbital periods of less than roughly 10 days. Super-Earths and mini-Neptunes populate the

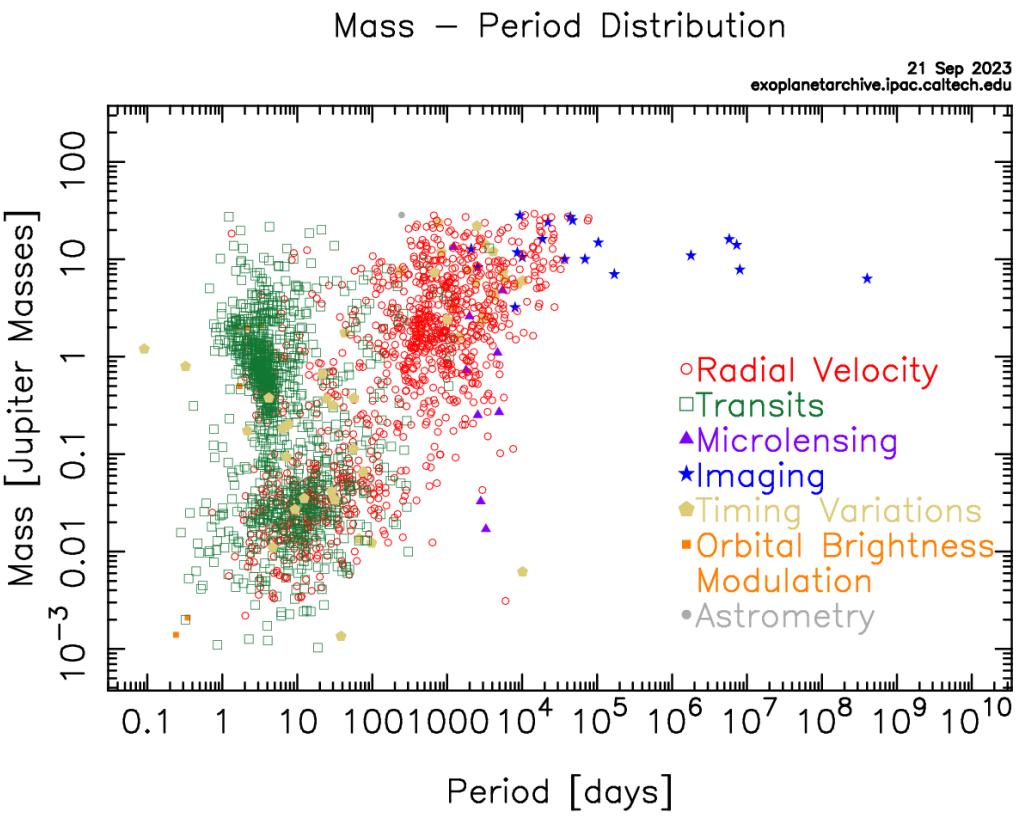


Figure 1.3. Mass-period distribution of the discovered exoplanets, colored according to the method used for their detection. From the NASA Exoplanet Archive.

lower-mass and lower-period sectors of the diagram. The distinction between super-Earths and mini-Neptunes lies in their composition and mean density. Super-Earths, rocky planets with outgassed atmospheres, exhibit a mean density greater than water. Mini-Neptunes, in contrast, contain a substantial primordial gaseous envelope, resulting in a mean density less than that of water (Kaltenegger, 2017). The existence and location of hot Jupiters in the diagram suggest a significant movement of these massive planets during their formation within disks (Turrini et al., 2015), delivering planets formed at different times, under different conditions, and at different distances from their host star to an optimal location for e.g. transit observation (Turrini et al., 2018). Finally, observational bias works to under-represent small-radius and long-period planets (Welsh and Orosz, 2018). Thus, the lower right-hand corner of Figure 1.3 is sparsely populated.

The mass-radius distribution of exoplanets (see Figure 1.4) is the second key diagram to discuss. Mass and radius are fundamental properties for exoplanet characterization, yet for a substantial portion of exoplanets, one or the other remains unknown. That is because they are derived from different measurement methods: radial velocity measurements are used to derive the mass[4] (Wright, 2018), while the radius is primarily determined using the transit method. Various published mass-radius relationships, contingent upon planet type, frequently enable radius prediction (Ulmer-Moll et al., 2019). These relationships facilitate

[4]More precisely, mass · sin(i), where i represents the inclination of the plane of the orbit with respect to the plane of the sky (i = 0 represents a face-on orbit with no radial component).

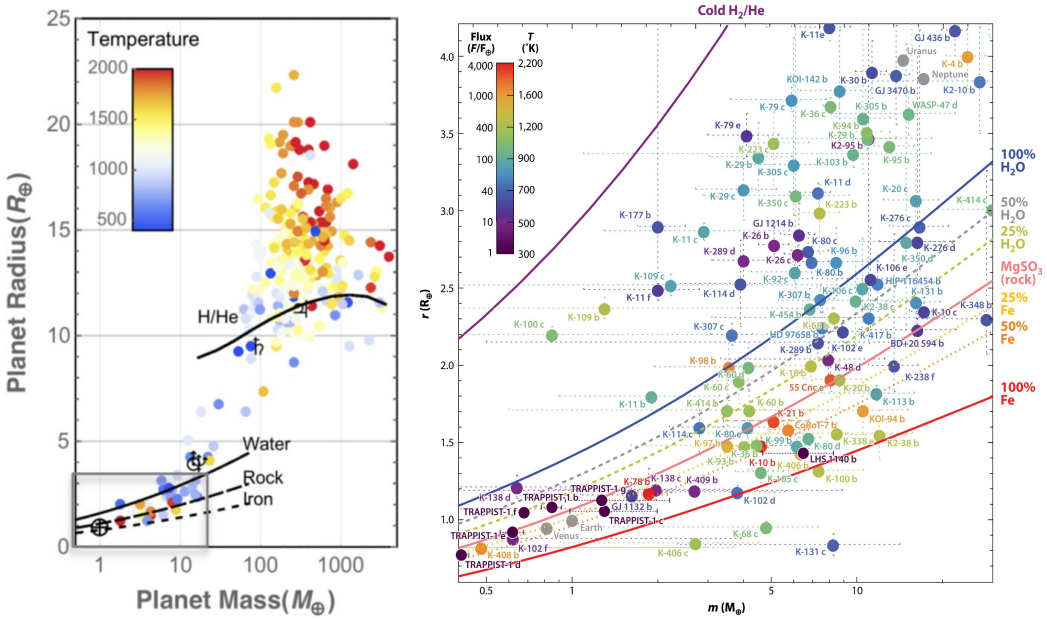


Figure 1.4. *Left panel*: Masses and radii of known transiting exoplanets. Black lines show mass-radius relations for a variety of internal compositions: the models cannot fully capture the variety of cases and break the degeneracies in the interpretation of the bulk composition (Tinetti et al., 2018). *Right panel*: zoom into the lower mass regime (roughly) indicated as a grey rectangle on the left. Coloured lines show mass-radius relations for a variety of internal compositions (Kaltenegger, 2017).

the first-time exploration of planetary structure and composition (Howard et al., 2013). A crucial factor, especially for low-mass planets, is the overall elemental abundances of the materials accreted onto forming planets, which largely determines the radius of a planet for a given mass (Pudritz et al., 2018). For instance, the radius can be radically altered if the planet is rocky, has substantial water content, or retains a hydrogen-helium atmosphere. Inferences about planetary structures rely on the compressibility of potential compositions within a planet's pressure and temperature range, which in turn depends on the Equation of State (EoS) for the relevant matter composition (Seager et al., 2007; Swift et al., 2012). The minimum conceivable radius for a given mass corresponds to a planet composed purely of iron. The mass-radius relationships for silicate planets intersect with those of carbon and water planets due to similar EOS (Seager et al., 2007). These intersections represent degeneracies, where planets of different compositions could end up having similar radii and masses.

The third key diagram to consider is the planet-metallicity relation (Petigura et al., 2018) (see Figure 1.5). This relation suggests that massive planets are more likely to be found around stars with solar or higher metallicity. Specifically, within a limited range of stellar masses ($0.7 - 1.2\ M_\odot$), the probability of a star hosting a giant planet scales as the square of the number of iron atoms, represented as $P_{planet} \propto N_{Fe}^2$ (Fischer and Valenti, 2005). Less massive, longer-period planets show a similar, but significantly weaker correlation. This suggests that low-mass planets can form in any disk, while only high metallicity disks or sufficiently massive disks can generate giant planets within the disk's lifespan (Hasegawa and Pudritz, 2014).

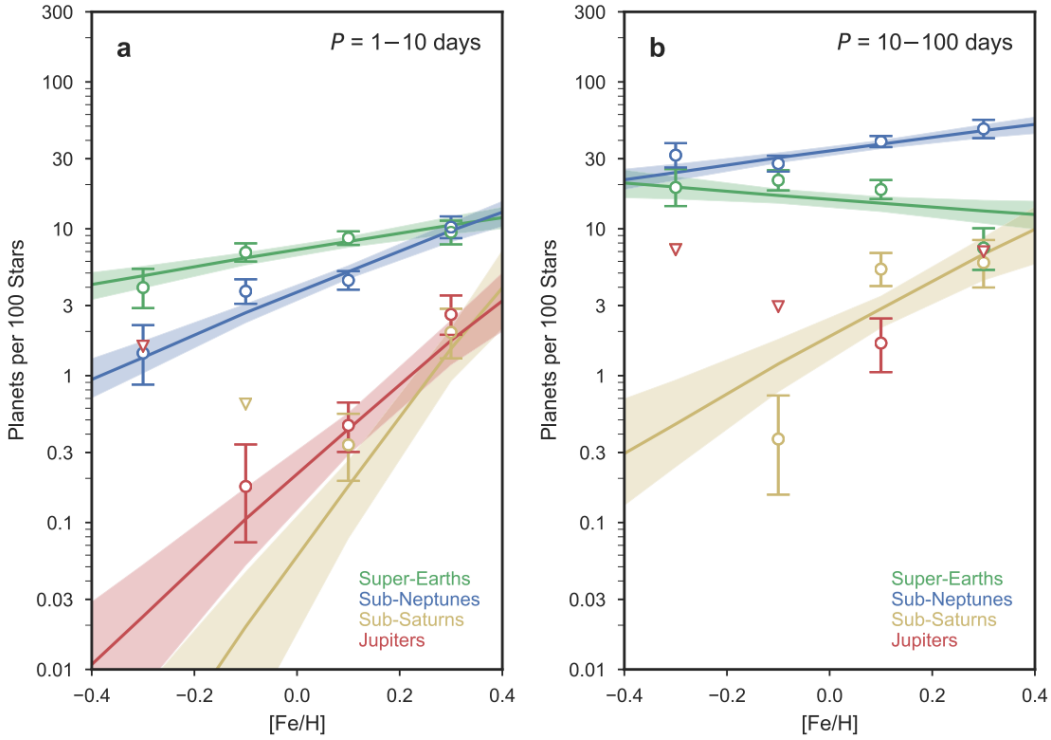


Figure 1.5. Planet-metallicity relations for short (left panel) and long (right panel) period exoplanets. Data points show the number of planets per 100 stars for bins of host star metallicity from −0.4 to 0.4 in logarithmic iron to hydrogen proportion. Triangles represent upper limits (90%). The solid lines and bands show the best-fitting model and the 1σ credible range of models, respectively. The strength of the planet occurrence-metallicity correlation increases with planet size and is stronger for shorter periods. Figure credits: Petigura et al. (2018).

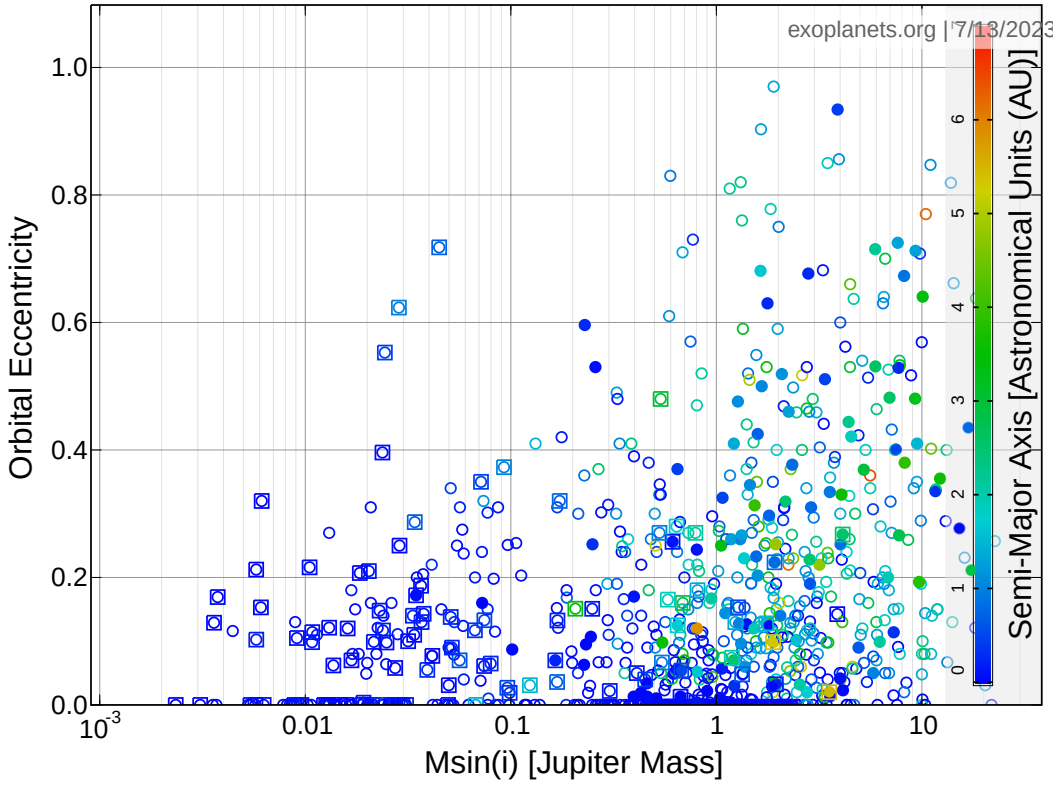


Figure 1.6. The mass-orbital eccentricity distribution of planets, plotted using data from the Exoplanet Orbit Database, is color-coded according to the planet's semi-major axis in astronomical units (AU). Square markers emphasize planetary systems with a multiplicity greater than three. Filled dots signify exoplanets discovered in peer-reviewed literature prior to 2005. Initial measurements predominantly populate the high mass, high eccentricity parameter space. In contrast, highly multiplexed planetary systems tend to occupy the lower mass, lower eccentricity region due to their lower gravitational stability and increased susceptibility to disruption.

The fourth important diagram to discuss to consider is the distribution of planetary eccentricities, which reveals substantial eccentricities for a significant number of massive exoplanets (see Figure 1.6). In particular, planets with mass $\gtrsim 4M_J$ exhibit an eccentricity distribution akin to that of binary stars. In contrast, planets with $\lesssim 4M_J$ are less eccentric than both binary stars and more massive planets (Ribas and Miralda-Escudé, 2007). The median value of this eccentricity is very high ≈ 0.25 (Pudritz et al., 2018). The eccentricity of single massive planets can be attributed to planet-planet scattering interactions following the dispersal of the gas disk (Chatterjee et al., 2008). Based on these trends, Ribas and Miralda-Escudé (2007) posited that there are two populations of gaseous planets, each arising from a distinct formation mechanism: a low-mass population formed by *gas accretion* onto a rock-ice core within a circumstellar disk, and a high-mass population formed directly via pre-stellar cloud *fragmentation*. Planets of the first population are theorized to form in initially circular orbits that become more eccentric over time. In contrast, planets of the second population exhibit higher mean orbital eccentricity due to the extensive radial migration required to move from their initial orbits at $\gtrsim 30\ AU$ to their significantly smaller observed orbits (Ribas and Miralda-Escudé, 2007). Additionally, population studies of

exoplanets' orbital characteristics within multi-planet systems reveal an inverse correlation between average planetary orbital eccentricity and system multiplicity (Jurić and Tremaine, 2008). In Turrini et al. (2020), they identified an anti-correlation between the Normalized Angular Momentum Deficit (NAMD) of planetary systems, a dimensionless measure of the difference between an idealized circular-orbit system's angular momentum and that of a real planetary system, and the multiplicity of these systems. This finding explains the previously observed inverse relationship between eccentricity and multiplicity. Planetary systems with exceptionally high multiplicity (6 or more) are rare but not impossible, as observations have shown (Tuomi, 2012; Gillon et al., 2017; Shallue and Vanderburg, 2018). Such systems necessitate very low eccentricities for maintaining stability and averting gravitational disruption. A prominent example is the TRAPPIST-1 system (Gillon et al., 2017), hosting seven Earth-sized exoplanets orbiting an ultra-cool star. Notably, all seven TRAPPIST-1 planets constitute a single resonant chain, despite the likelihood of instability typically associated with resonant systems. This unique arrangement is interpreted as the result of a comparatively slow migration of the whole system to its observed orbital configuration, resulting in a long-lived resonant system (Grimm et al., 2018).

1.1.3 The Radius Gap

The field of exoplanets has made significant advances in recent years, thanks to the availability of fresh data and measurements using different instruments. This has enabled the validation of observed trends and the discovery of entirely new features in the parameters space of exoplanets. For instance, the pioneering study of Owen and Wu (2013) initially observed a bimodality in the planetary radii using the Kepler exoplanet population. Still, it could not confirm this feature due to insufficient data and observational biases. However, the California-Kepler Survey in 2017 confirmed this bimodality using updated Kepler statistics and recent spectroscopic observations of stellar companions, along with Gaia (Gaia Collaboration et al., 2016) observations to refine the stellar types and radii (Fulton et al., 2017; Fulton and Petigura, 2018). The resultant distribution shows that, after accounting for selection effects, planets fall into two groups: those with radii of approximately $\sim 2.4\ R_\oplus$ and those with $\sim 1.3\ R_\oplus$ (see Figure 1.7). In between these groups, planets with radii from $1.5 - 2.4\ R_\oplus$ are relatively uncommon, forming a "radius gap".

The distribution peaks correlate with the orbit environment, specifically the host star's level of irradiation. The $1.3\ R_\oplus$ peak corresponds to planets that typically receive irradiation $S_{inc} > 100\ S_\odot$, while the $2.4\ R_\oplus$ peak planets usually orbit in less irradiated environments. There are two potential mechanisms to explain this distribution. The first is mass loss due to photoevaporation, a mechanism that was proposed in a study by Owen and Wu (2013). In this scenario, Extreme Ultra-Violet (EUV) and X-ray radiation heat the atmosphere, causing a mass outflow. This outflow is dependent on Extreme Ultra-Violet (XUV) time-integrated flux, and hence on stellar mass, scaling as $\sim M_*^{-3}$. This process decreases significantly after the first 100 Myr of the stellar lifetime (Ribas et al., 2005). The peak at lower radii is composed of rocky cores that were unable to retain their primordial hydrogen-helium atmosphere, which was completely stripped off by photoevaporation. The peak at slightly higher radii, on the other hand, consists of larger planets with thicker atmospheres that have preserved some of their original gaseous envelopes. This retention of gas significantly inflates their radii, shifting these planets to the larger peak of the radius distribution (see Figure 1.8, 1.9).

An alternative explanation is core-powered mass loss, a mechanism dependent on both

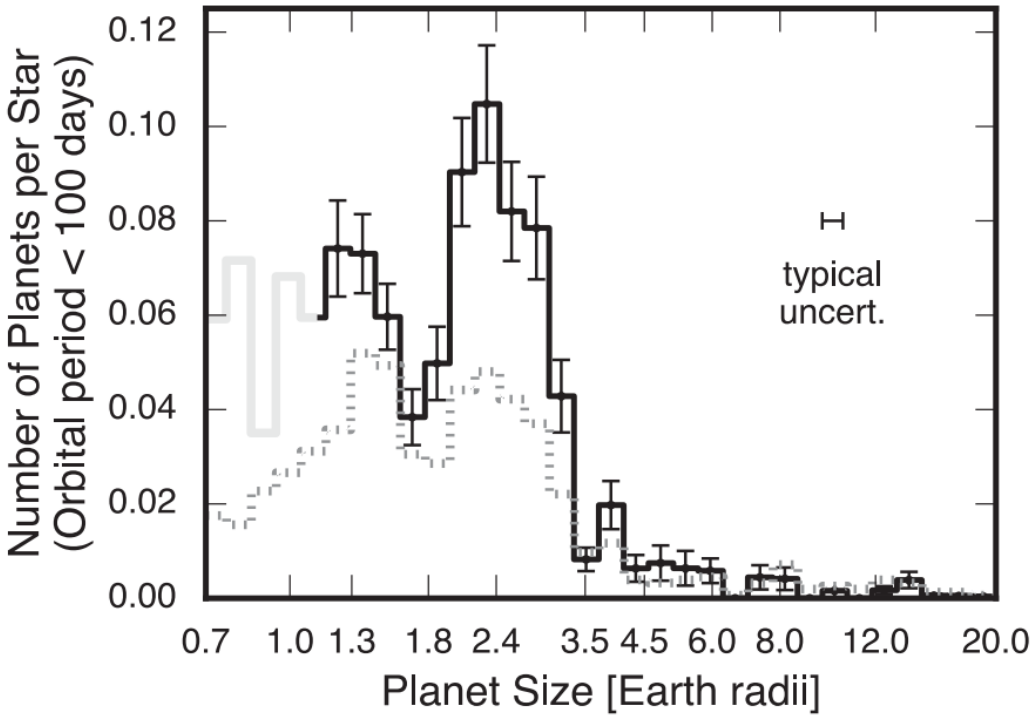


Figure 1.7. Histogram illustrating the distribution of planet radii for planets with orbital periods less than 100 days, adjusted for selection effects (completeness). The typical uncertainties are represented by bars. The median radius uncertainty is indicated in the upper right corner of the plot. A light gray thick line emphasizes the region of the histogram representing radii smaller than $\sim 1.14\ R_{\oplus}$, a range that is subject to low completeness due to observational bias. A dotted gray line overlays the histogram, representing the uncorrected version of the same distribution. A distinct bimodality is visible in the planet size distribution, along with the presence of the radius gap. Figure credits: Fulton and Petigura (2018).

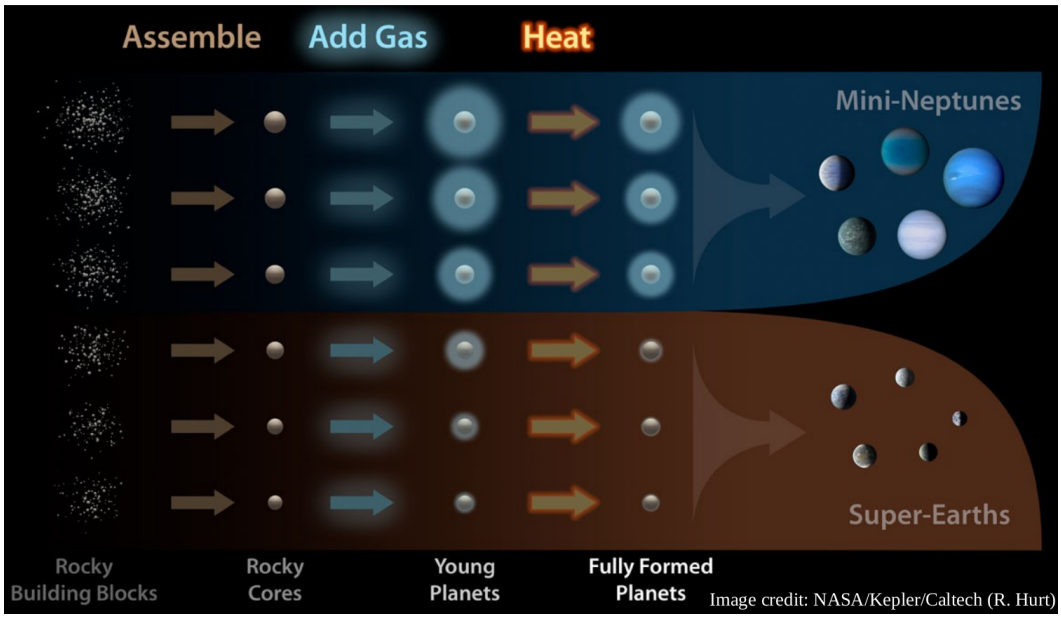


Figure 1.8. Infographic of the formation and evolution mechanism behind the bi-modal distribution of exoplanet radii. Exoplanets mostly fall into two classes, super-Earths and mini-Neptunes. These classes are the end product of various evolutionary processes including accretion of material from the proto-planetary disk, gas accretion to form the primordial hydrogen-helium envelope, and heating by the host star. The further advancement of our understanding of these processes relies on future ground-based and space-based exoplanet observation facilities. Credits: NASA/Kepler/Caltech (R. Hurt).

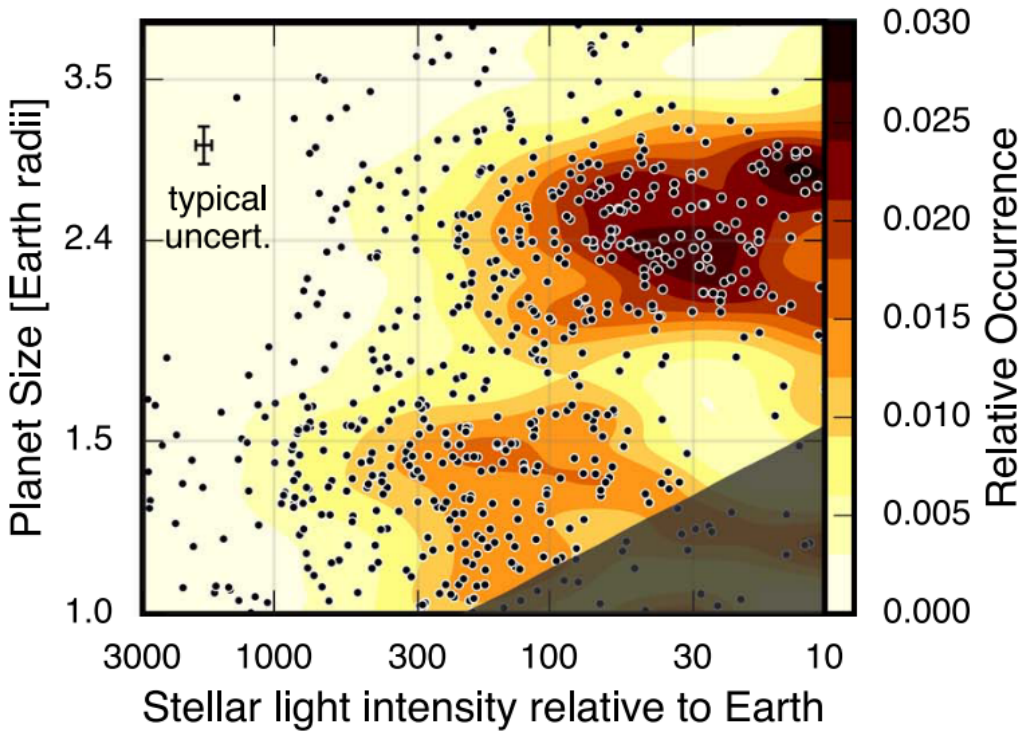


Figure 1.9. Two-dimensional distribution of planet size (on the vertical axis) and insolation flux (on the horizontal axis), with median uncertainty depicted in the upper left corner. The distribution shows two distinct peaks corresponding to planets receiving $S_{inc} > 100\ S_{\odot}$ and $S_{inc} < 200\ S_{\odot}$, as discussed in the text. Possible explanations for the bimodal nature of the distribution involve either mass loss by photoevaporation or core-powered mass loss. Credits: Fulton and Petigura (2018).

stellar irradiation and planetary properties. This scenario, first proposed in Ginzburg et al. (2018), postulates that the luminosity from a cooling rocky core produces a scarcity of intermediate-sized planets, irrespective of the high-energy flux incident upon them. Both peaks of the distribution naturally emerge from this process, which can fully erode light envelopes while preserving heavier ones. With core-powered mass loss, the results align with a dependency on mass, whereby the peak occurrence shifts towards higher irradiation levels as stellar mass increases. To differentiate between these two mechanisms - photoevaporation and core-powered mass loss - it could be useful to observe planets orbiting stars of different types. This is because photoevaporation is driven by the high-energy tail of stellar radiation, while core-powered mass loss depends on the planet's equilibrium temperature, which determines both its cooling and mass loss rates (Ginzburg et al., 2018).

A notable feature of the mass-period diagram, shown in Figure 1.3, is the dearth of exoplanets with Neptune-like masses and radii on short orbital periods (below 2–4 days). This feature is known as the "Neptune desert" (Mazeh et al., 2016; Szabó M et al., 2023), and cannot be explained by observational biases since many Neptunian planets with longer orbital periods have been detected. The exact origin of this desert is currently unresolved, and there is active debate in the literature about its possible causes, including photoevaporation. A parallelism with the so-called brown-dwarf desert suggests different mechanisms of formation and evolution for hot Jupiters and short-period super-Earths. By investigating the shape of the desert's boundary in the mass-period or radius-period diagram, different formation-evolution scenarios can be inferred, dependent on the stellar environment. In this respect, systems with multiple planets (each receiving different levels of irradiation), where at least one planet is inside the desert, can provide further insight into the origin of the desert (Kubyshkina et al., 2022b). For further reading on the subject, see Chapter 8.

1.1.4 Beyond Mass and Radius

From most of the current exoplanet observations, we could infer only a limited set of observable properties related to the orbits and the fundamental planetary parameters. These observations have highlighted an unprecedented diversity among exoplanets, posing significant challenges to our understanding of their true nature. We may question, for instance, "What are exoplanets made of?". However, answering this is more complicated than it might initially seem. Indeed, when only the mass and radius (and therefore the average density) information is available, intrinsic degeneracies between different internal compositions prevent us from reaching a definitive answer (e.g. Swift et al., 2012; Lopez and Fortney, 2014; Deming and Seager, 2017). For instance, according to a density estimate alone, the super-Earth GJ 1214 b could either have a rocky interior with a primordial hydrogen-helium atmosphere or a water interior and a steam atmosphere (Valencia et al., 2013). This is evidenced in Figure 1.10, which illustrates the degeneracy between different compositions for a given mass and radius for this planet.

Additional observational data are required to complement the basic mass and radius information and enable a more accurate characterization of the composition. This information can be obtained from the spectroscopic study of the atmosphere - the link between the interior of the planet and the stellar environment in which it has formed and evolved - and its composition. In this respect, we have yet to understand if the chemistry of exoplanets is intrinsically linked to their formation environment or if the nature of the host star directs the physics and chemistry of the planets' origin and evolution (Tinetti et al., 2018).

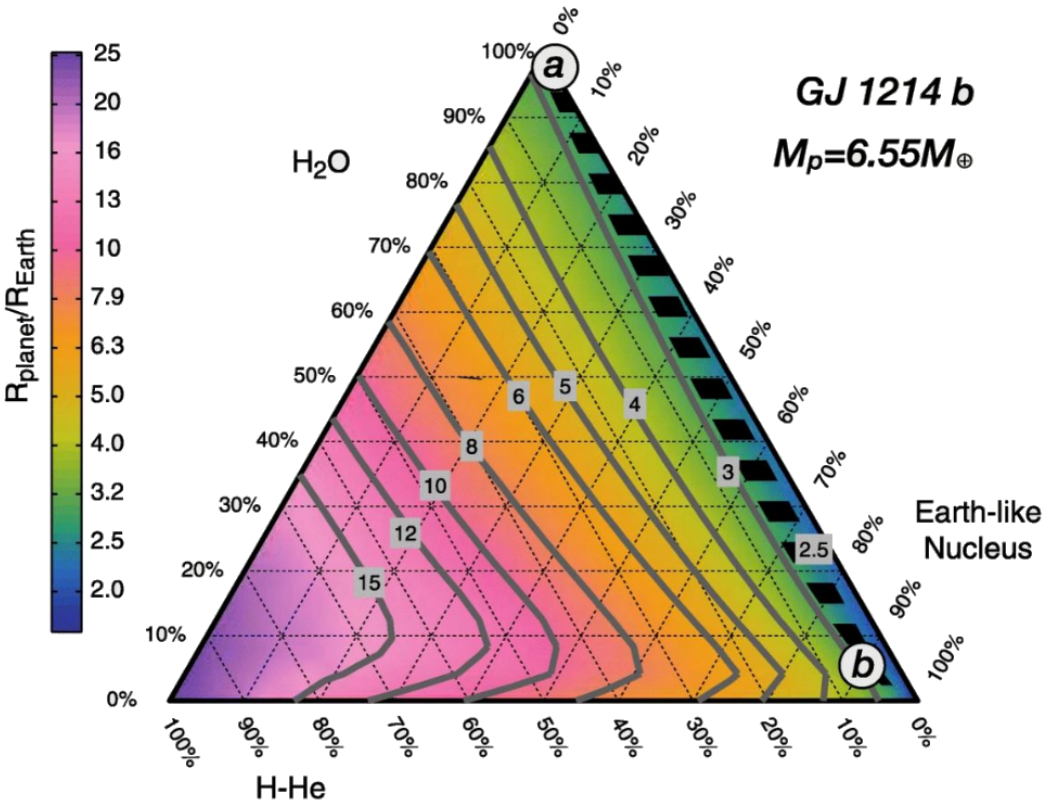


Figure 1.10. Ternary diagram of the exoplanet GJ 1214b. The composition χ is related to the radius for a specific planetary mass in terms of Earth-like nucleus fraction, water/ices fraction, and H/He fraction to total mass. Each vertex corresponds to 100% and the opposite side to 0% of a particular component. The color bar shows the radius in Earth radii and the gray lines are the isoradius curves labeled in Earth radii. Note that isoradius curves correspond to isodensity curves since the mass is fixed. Credits: Valencia et al. (2013).

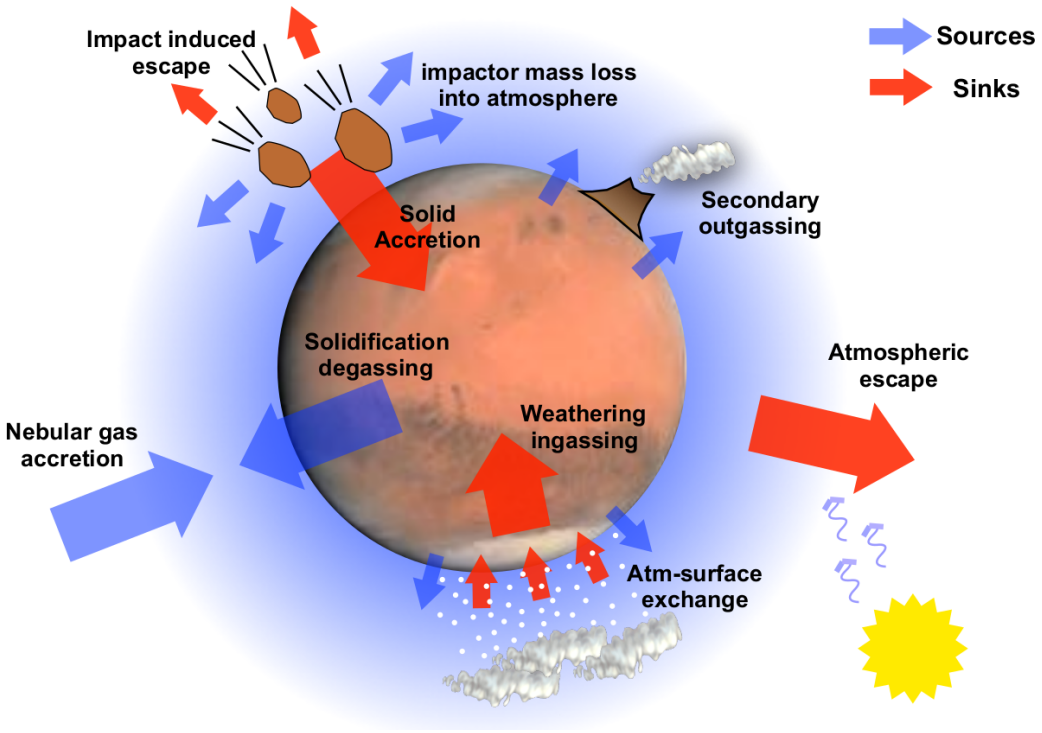


Figure 1.11. Various mechanisms responsible for adding (blue arrows) or removing (red arrows) gas from the atmosphere of a planet. The former includes accretion of gas from the surrounding solar nebula, degassing of solids and condensed species from the planetary surface, and possibly secondary outgassing from active volcanoes. The latter include atmospheric escape driven by stellar irradiation, in-gassing through atmospheric-surface thermal exchange, and impactor-induced escape from the atmosphere. Credits: Tinetti et al. (2021).

Additionally, the composition and structure of planetary atmospheres are influenced by a variety of physical processes, hard to disentangle, such as accretion, degassing, outgassing, escape, in-gassing, and impacts (see Figure 1.11). Therefore, it becomes clear that all these aspects are intertwined, and a comprehensive, multi-faceted effort is needed to answer these questions.

While the task of isolating the relative impact of all these processes on a single planet poses a considerable challenge, we can potentially decipher these influences by expanding observations to a larger and more diverse sample of planets. Through these expanded observations, we could extract general correlations or "trends" between e.g. atmospheric chemistry, planetary size, density, temperature, stellar type, and metallicity (Changeat et al., 2020a). Importantly, this information could shed insight into the key mechanisms governing planetary evolution at different time scales. Planets, unlike stars, do not occupy a well-behaved parameter space, exhibiting instead a broad and complex array of characteristics. As a result, the rigorous validation of our theoretical models and the understanding of the most critical physical parameters necessitate spectroscopic observations of a statistically significant sample of exoplanets (Tinetti et al., 2018). This task involves observing hundreds of planets, often repeatedly or over extended periods. Achieving such an extensive level of observation requires the use of dedicated space-based instruments, such as the Ariel mission

(see Section 1.2 for details).

1.1.5 Transit spectroscopy

The study of exoplanetary atmospheres is a relatively new field of research, with the first detection of an exoplanet atmosphere occurring in 2001 (Charbonneau et al., 2002). Since then, a growing number of atmospheres have been characterized by ground- and space-based observatories, using different techniques, including transmission and emission spectroscopy of transiting planets. Transiting planets are those that cross the disk of their host star with respect to our line of sight. During a primary transit (henceforth, *transit*), the planet blocks a fraction of the stellar light, causing a dip in the observed flux. Four different phases of a transit can be distinguished:

1. *First contact* – T_1, when the planet first overlaps with the stellar disk;
2. *Second contact* – T_2, when the planet starts being fully inside the stellar disk;
3. *Third contact* – T_3, when the planet starts exiting the stellar disk;
4. *Fourth contact* – T_4, when the planet is fully outside the stellar disk.

Therefore, besides the orbital period P, we can measure four observables which characterize the duration and profile of the transit (see Figure 1.12), assuming a uniformly bright stellar disk:

1. the interval between first and last contact, T_{14}, also referred to as "transit duration";
2. the interval between second and third contact, T_{23};
3. the Out-of-Transit (OOT) flux, F_{OOT}, measured outside the transit, when the planet is not blocking any light from the star;
4. the In-Transit (IT) flux, F_{IT}, measured during T_{23}.

During a transit, the observed flux from the star is reduced from the baseline value F_{OOT} to the in-transit value F_{IT}, by an amount $\delta F = F_{OOT} - F_{IT}$. The transit depth is defined as the ratio between the flux decrement and the baseline flux, $\delta F / F_{OOT}$. To first order, the transit depth measured from a mono-chromatic light curve, δ, is given by the ratio of the projected area of the planet to the projected area of the star:

$$\delta = \left(\frac{R_p}{R_*}\right)^2 \tag{1.1}$$

Assuming that the stellar radius is known, the transit depth can be used to measure the radius of the planet. The atmosphere of the planet, if present, induces a chromatic absorption, increasing the measured radius of the planet at some specific wavelengths corresponding to the absorption features of atomic or molecular species present in the atmosphere (Seager and Sasselov, 2000; Brown, 2001; Tinetti et al., 2013). By measuring the transit depth at different wavelengths, one can reconstruct the atmospheric spectrum of the planet. The wavelength-dependent transit depth, $\delta(\lambda)$, is then given by:

$$\delta(\lambda) = \frac{\pi R_p^2 + A(\lambda)}{\pi R_*^2} \tag{1.2}$$

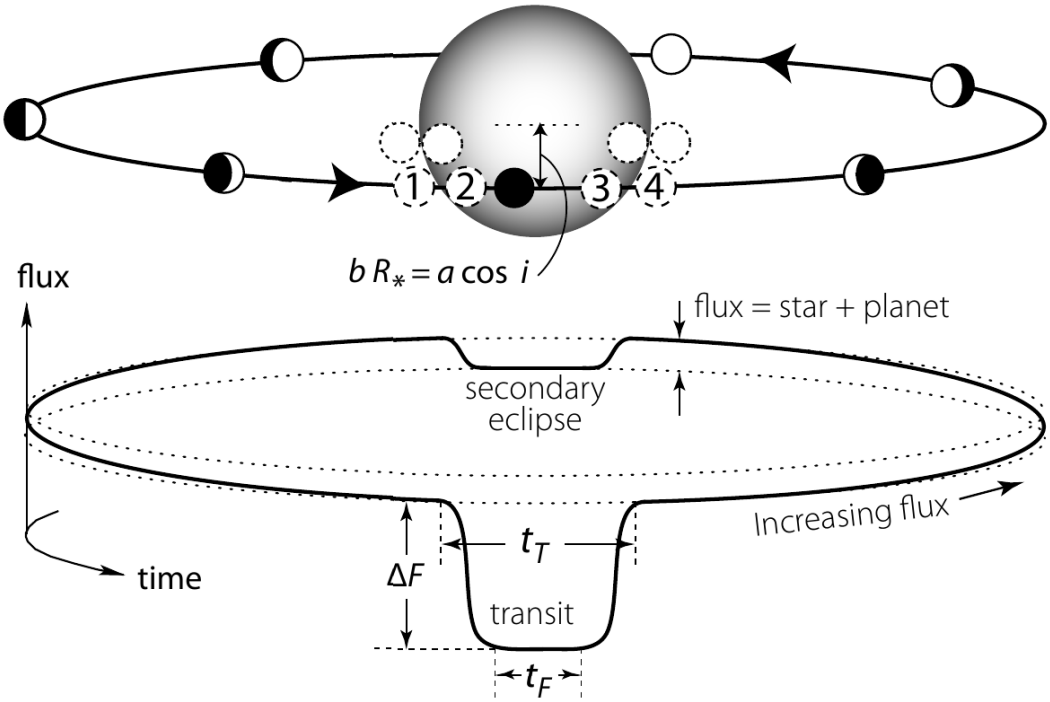


Figure 1.12. Schematic of a transit of an orbiting planet (Perryman, 2018). The total transit duration is between first and fourth contact, while the ingress/egress duration is timed between second and third contact. The planet blocks a fraction of the star light, causing a dip in the observed flux. After the fourth contact, the planet's brighter day-side progressively comes into view, and the total flux rises. The flux drops again during the secondary eclipse as the planet passes behind the star.

where R_p is the radius at which the planet becomes opaque at all wavelengths, and $A(\lambda)$ is the atmospheric contribution. Equation 1.2 has a unique solution provided we know R_p accurately. For a terrestrial planet, R_p usually coincides with the radius at the surface. For a gaseous planet, R_p may correspond to a pressure $p_0 \sim 1 - 10$ bar. $A(\lambda)$ can be calculated by integrating the absorptivity of the atmosphere from R_p to the top of the atmosphere (e.g. Sing, 2018):

$$A(\lambda) = \int_{R_p}^{z_{max}} [1 - \mathbb{T}(z, \lambda)]\, 2\pi z dz \tag{1.3}$$

where the transmittance $\mathbb{T}(z, \lambda)$ is the fraction of the radiation transmitted through a given atmospheric layer, and z is the radial coordinate direction, with $z = 0$ at R_p. $\mathbb{T}(z, \lambda)$ is related to the optical depth of the atmosphere at altitude z, $\tau(z, \lambda)$, by the Beer-Bouguer-Lambert law:

$$I(z, \lambda) = I_0(\lambda)\mathbb{T}(z, \lambda) = I_0(\lambda)e^{-\tau(z,\lambda)} \tag{1.4}$$

where $I_0(\lambda)$ is the intensity of the incoming stellar radiation and $I(z, \lambda)$ is the intensity after the radiation has been filtered through the atmosphere. τ accounts for the quantum interaction between the stellar photons and the atmospheric absorbers, which depends on the absorption coefficient and mixing ratio of each species.

The characteristic length scale of the atmosphere is the scale height, H, defined as the height at which the pressure decreases by a factor e in hydrostatic equilibrium:

$$H = \frac{k_B T}{\mu g} \tag{1.5}$$

where k_B is the Boltzmann constant, T is the temperature, μ is the mean molecular weight, and g is the surface gravity. At zero order approximation the atmosphere is shaped as an annulus around the planet with a radial height of N scale heights, so the amplitude of the atmospheric absorption is:

$$A \sim N \times 2R_p H \tag{1.6}$$

A typical value of N in the infrared is 5. This formulation highlights that the atmospheric signature is stronger for hot planets, light atmospheres, and low-gravity objects. "Hot Jupiters" are therefore amongst the most suitable targets for transit spectroscopy.

Figure 1.13 shows the spectro-photometric light curves of WASP-39b, a hot Jupiter orbiting a G-type star at a distance of about 0.05 AU (Rustamkulov et al., 2023). The observations cover the entire 0.5 to 5.5 μm JWST/PRISM wavelength range at 20–300 in resolving power, and were taken as part of the JWST Transiting Exoplanet Community Early Release Science Program. The light curves were created by binning the data into wide wavelength channels, before removing known systematics. This beautifully demonstrates the unprecedented stability and precision of this observing mode.

Figure 1.14 illustrates the transmission spectrum obtained from the dame data, processed with the FIREFLy (Rustamkulov et al., 2022) data reduction pipeline. The PICASO 3.0 grid (Mukherjee et al., 2023) was used to fit the data and retrieve the atmospheric parameters, and found evidence for the presence various atmospheric species, including Na, H_2O, CO_2, CO, and CH_4. An unexpected absorption feature at 4.05 μm was also detected, which was attributed to SO_2. Extensive photochemical modelling (Tsai et al., 2023b) has shown that SO_2 is a robust outcome of the chemical environment in WASP-39b's atmosphere.

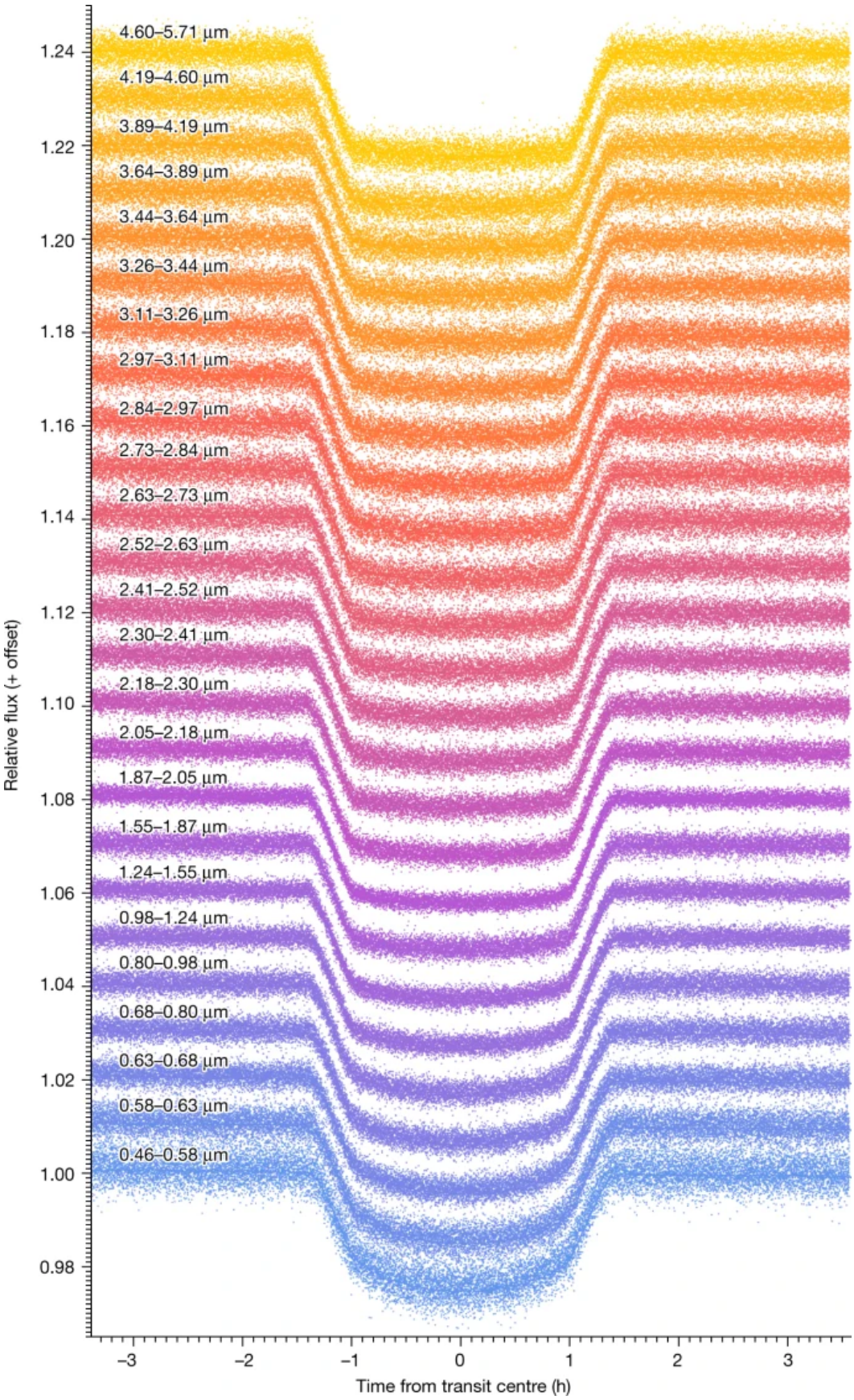


Figure 1.13. Normalized spectrophotometric light curves for the JWST-PRISM transit of WASP-39b (Rustamkulov et al., 2023). The light curves were created by summing over wide wavelength channels (wavelength ranges indicated on the plot). Overplotted on each light curve are their best-fit models, which include a transit model and detector systematics. Light curve systematics have not been removed from the data.

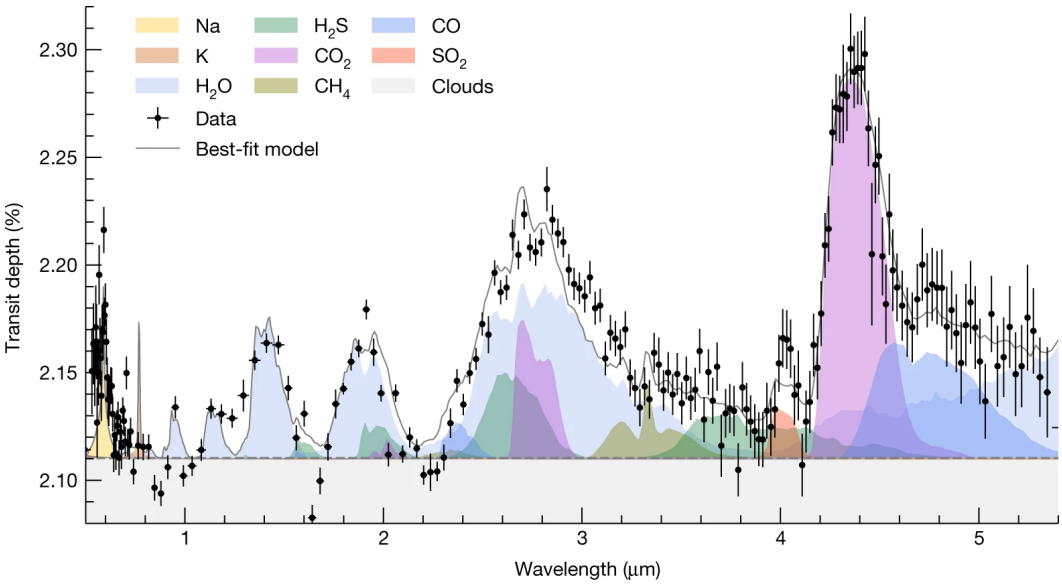


Figure 1.14. The JWST-PRISM transmission spectrum of WASP-39b with key contributions to the atmospheric spectrum (Rustamkulov et al., 2023). The black points with error bars correspond to the measured FIREFLy transit depths of the spectrophotometric light curves at different wavelengths. The best-fitting model spectrum from the PICASO 3.0 grid is shown as the grey line and the coloured regions correspond to the chemical opacity contributions at specific wavelengths.

Follow-up observations with the JWST/MIRI instrument are scheduled[5] to unambiguously confirm this detection using the SO_2 7.5 μm and 8.8 μm features. This remarkable result demonstrates the power of transit spectroscopy to lead the charge in the study of exoplanet atmospheres. It also highlights the ability of JWST to provide transformative insights into the nature of exoplanets.

As opposed to a transit, the secondary eclipse (henceforth, *eclipse*) occurs when the planet passes behind the star. The planet is then hidden from view, and the observed flux drops from the combined flux of the star and the planet to the flux of the star alone. The secondary eclipse is usually much shallower than the transit, and therefore more challenging to observe. Before and after the eclipse, the signal from the planet is the combination of thermally-emitted radiation and reflected stellar light (Charbonneau et al., 2005; Deming et al., 2005). Then, the contrast ratio between the planet and the star is given by:

$$\delta(\lambda) = \frac{R_p^2}{R_*^2}\frac{F_p(\lambda)}{F_*(\lambda)} + A_l(\lambda) \tag{1.7}$$

where $F_p(\lambda)$ and $F_*(\lambda)$ are the fluxes emitted by the planet and the star, respectively. $A_l(\lambda)$ is the reflected light contribution, which can be expressed as a function of the geometric albedo $p(\lambda)$:

$$A_l(\lambda) = \frac{R_p^2}{a^2}p(\lambda) \tag{1.8}$$

The geometric albedo is determined by the optical properties of the atmosphere and the aerosols (condensate clouds/photochemical hazes) cover. Eclipse spectra are sensitive to

[5] https://www.stsci.edu/jwst/phase2-public/2783.pdf

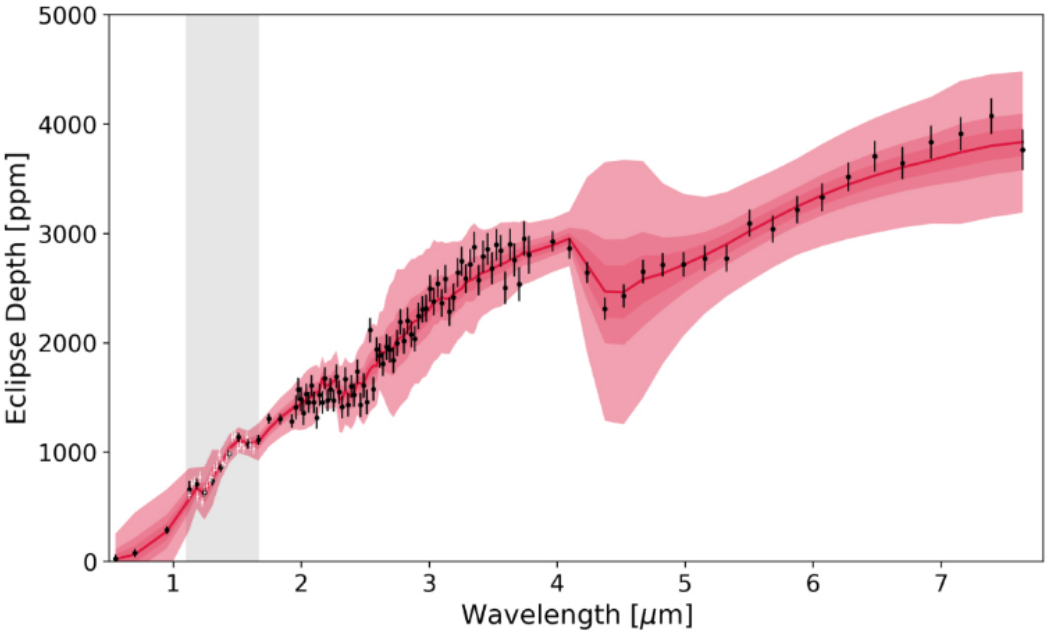


Figure 1.15. Illustration of WASP-76 b eclipse spectra with HST and Ariel (Edwards et al., 2020). The gray box indicates the wavelength range covered by the HST/WFC3 G141 grism. The data points from HST are shown in white. The black points show the simulated Ariel observations of the same planet, where Gaussian scatter was added. The Ariel spectra are for a single observation at the native resolution of the instrumentation (i.e., Tier 3). The red areas indicate the 1–3σ uncertainties from the Ariel data.

the bulk temperature of the atmosphere, its vertical structure, and global energy budget. This is important to reveal possible temperature inversions – where an upper layer of the atmosphere is warmer than the layer beneath it, strong thermal emitters, and the presence of reflective aerosols. Additionally, the acquisition of eclipse spectra allows us to gain information on atmospheric molecules (Cartier et al., 2017), although their spectral features may be more difficult to detect than in transmission. Recent detections of thermal inversions in Hot Jupiters and ultra-Hot Jupiters[6] have linked it to the presence of optical molecular absorbers e.g. TiO, VO, AlO, as well as ionic species such as FeI and H- (von Essen et al., 2019; Edwards et al., 2020; Yan et al., 2020; Changeat, 2022).

Figure 1.15 reports a measured eclipse spectrum acquired using the HST/WFC3 G141 grism (Edwards et al., 2020). The spectrum covers the wavelength range 1.1 – 1.7 μm, and it was used to infer the possible presence of TiO and an atmospheric thermal inversion, along with a significant amount of H_2O. The extent probed via HST should be compared to the Ariel wavelength range, which covers 0.5 – 7.8 μm. The figure shows the simulated Ariel observations of the same planet, for a single observation. Compared to HST, Ariel will be able to probe the emission of the planet at longer wavelengths, where the thermal emission is more prominent. This will allow us to better constrain the atmospheric temperature structure and place far more stringent constraints on the presence of molecular constituents.

[6]Hot Jupiters with day-side temperatures higher than 2200 K (Parmentier et al., 2018).

1.1.6 Atmospheric Retrieval

Retrieval codes are essential for correctly interpreting spectroscopic observations of exoplanet atmospheres. The idea being to infer the atmospheric properties from comparing the observed spectrum to a forward model, iteratively adjusting the model parameters until the best fit is found (see Figure 1.16). At each step, the forward model is computed, and the likelihood of the model is calculated. The likelihood is then used to update the posterior distribution of the model parameters. Upon convergence, the posterior distribution is sampled to obtain the best-fit model and the uncertainties on the model parameters. See Appendix A for a brief introduction to Bayesian statistics which explains in more detail how posterior distributions are derived and how we can use them to interpret the observations.

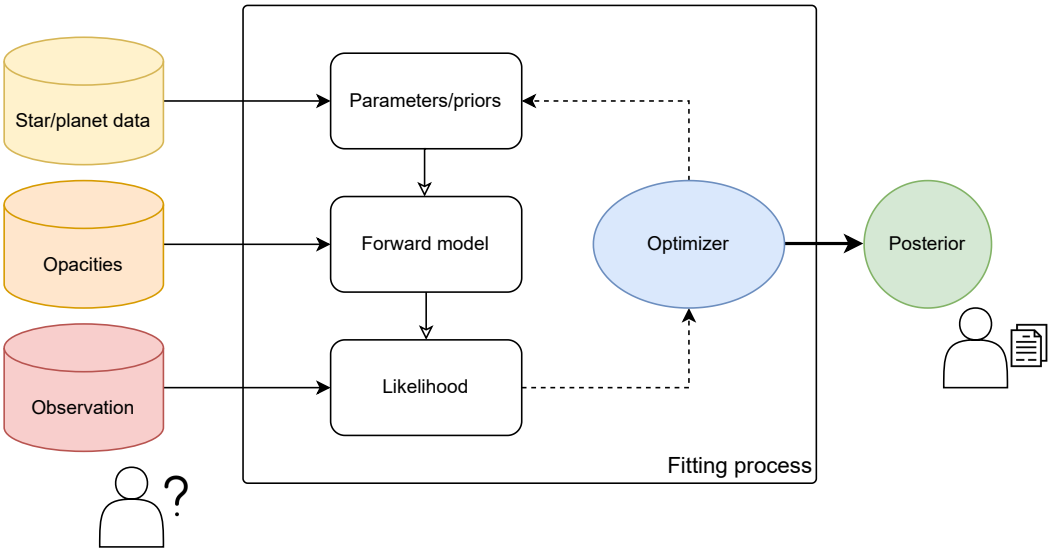


Figure 1.16. Schematic representation of the atmospheric retrieval process. The forward model is computed from the input parameters and opacities data. The forward model is then compared to the observed spectrum, and the likelihood is calculated. The optimizer adjusts the input parameters to maximize the likelihood, i.e. obtain the best fit. The process is repeated iteratively until convergence. The end products are the best-fit model for the atmospheric spectrum and the uncertainties on the parameters, obtained from the posterior distribution.

To date, there exist a number of open-source retrieval tools (e.g. Irwin et al., 2008; Line et al., 2013; Benneke, 2015; Waldmann et al., 2015b; Gandhi and Madhusudhan, 2017; MacDonald and Madhusudhan, 2017; Zhang et al., 2019; Mollière et al., 2019; Kitzmann et al., 2020; Harrington et al., 2020; Al-Refaie et al., 2021), each with its own strengths. Following an intense effort to compare different retrieval tools, Barstow et al. (2020) demonstrated the robustness of the retrieval results obtained by different codes. However, the comparison also highlighted the need for a unified framework that can be used to perform retrievals with different forward atmospheric models and retrieval algorithms. As will be discussed later, the atmospheric retrieval framework TauREx 3 is the first attempt to address this need.

1.1.7 Key Scientific Questions

While recent observations are beginning to reveal the diversity of exoplanets, we still lack a comprehensive understanding of their true nature. Further observations with dedicated space-based instruments are required to provide pivotal insights into a number of fundamental questions about exoplanets:

- *What are exoplanets made of?* The study of the composition and elemental abundances of planetary atmospheres – the window into the interior structure and bulk composition – can constrain the nature of the planet's interior.
- *How do planetary systems form and evolve?* The planetary population is a fossil record of the formation and evolution of planetary systems. The stellar environment, the disk composition, and the disk dispersal mechanisms all influence the formation and evolution of planetary systems. Therefore, by studying a diverse sample of exoplanets, we can decipher the key processes that determine the formation pathways and migration histories of exoplanets.
- *What processes shape planetary atmospheres?* A planet's atmosphere is shaped by a variety of physical processes, including photochemistry, transport-induced quenching, surface interactions, clouds, and hazes. Each of these processes leaves a unique, wavelength-dependent imprint on the atmospheric spectrum. Therefore, multi-wavelength observations are required to disentangle the roles of these processes in sculpting atmospheric chemistry.
- *How do exoplanets and their atmospheres evolve over time?* The study of Earth's climate is an example of how the evolution of a planet's atmosphere can be studied over time. The study of exoplanets can provide a broader perspective on the evolution of planetary atmospheres, including the role of stellar activity, the evolution of the host star, and the evolution of the planet itself. Single observations of exoplanets can provide a snapshot of their atmospheric properties at a particular time. However, repeated observations over time can provide insights into the planet's weather, global circulation, spatial variability, and seasonal changes.

1.2 The Ariel Contribution to Exoplanetary Characterization

The upcoming Ariel mission will conduct the first chemical survey of a large and diverse sample of exoplanets by performing simultaneous spectro-photometric observations across the visible to infrared wavelength range. As the first dedicated exoplanet atmosphere characterization mission, Ariel will provide key insights into the nature, formation, and evolution of planets beyond our solar system.

1.2.1 An Overview of the Ariel Mission

The Atmospheric Remote-Sensing Infrared Exoplanet Large-survey (Ariel) is a Medium Class science mission of the ESA's Cosmic Vision science programme (Figure 1.17). More in-depth information about the mission can be found in the Ariel Definition Study Report[7],

[7] https://sci.esa.int/web/ariel/-/ariel-definition-study-report-red-book

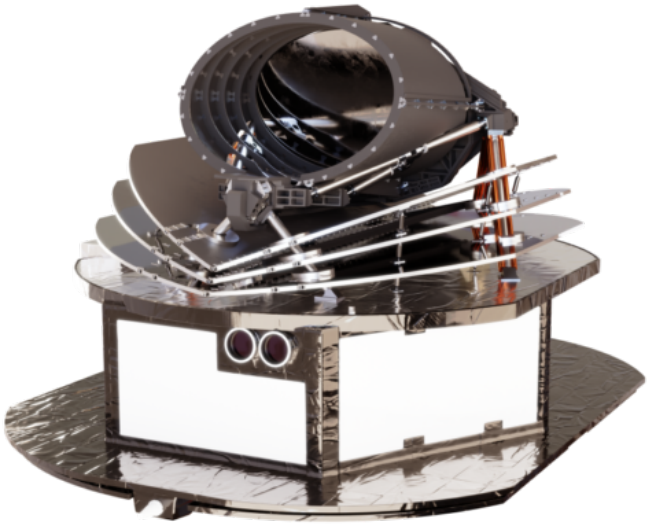


Figure 1.17. Artist impression of ESA's Ariel exoplanet satellite. Credit: Airbus.

commonly referred to as the "Red Book". Ariel will be the first space observatory fully devoted to studying the atmospheres of a large and diverse population of known exoplanets orbiting different types of stars (Pascale et al., 2018; Tinetti et al., 2018).

The mission launch is planned in 2029, with an Ariane 6.2 vehicle[8]. After reaching the operational orbit, a large halo orbit around the L2, Ariel will begin its commissioning phase[9] and instrument performance verification phase[10]. Nominal science phase operations will start six months after launch, and the nominal mission will last 4 years[11]. Upon completion of the nominal mission phase, Ariel will enter the extended mission phase[12], foreseeing an additional 2 years of operations. During the nominal mission phase, Ariel will survey approximately 1000 exoplanets, ranging from Jupiter- and Neptune-size gas giants down to sub-Neptunes and super-Earths, orbiting host stars with a range of spectral types – from hot F- to cooler K- and M-types, metallicity, and activity level.

This large and diverse sample of targets is essential for extending comparative planetology to the Galaxy scale and uncovering the processes that shape planetary formation and evolution. Most observations will consist in transit and/or eclipse spectroscopy of the atmospheres of warm and hot exoplanets to take advantage of their well-mixed atmospheres which should present minimal sequestration of heavy elements. These atmospheres are thus expected to be more representative of the planetary bulk composition than those of colder planets. Thanks to this approach, Ariel will be able to provide key insights into the observed planetary population to address the following fundamental questions:

- What are exoplanets made of?

[8] https://www.esa.int/Enabling_Support/Space_Transportation/Launch_vehicles/Ariane_6

[9] R-MIS-170

[10] R-MIS-190

[11] R-MIS-010

[12] G-MIS-030

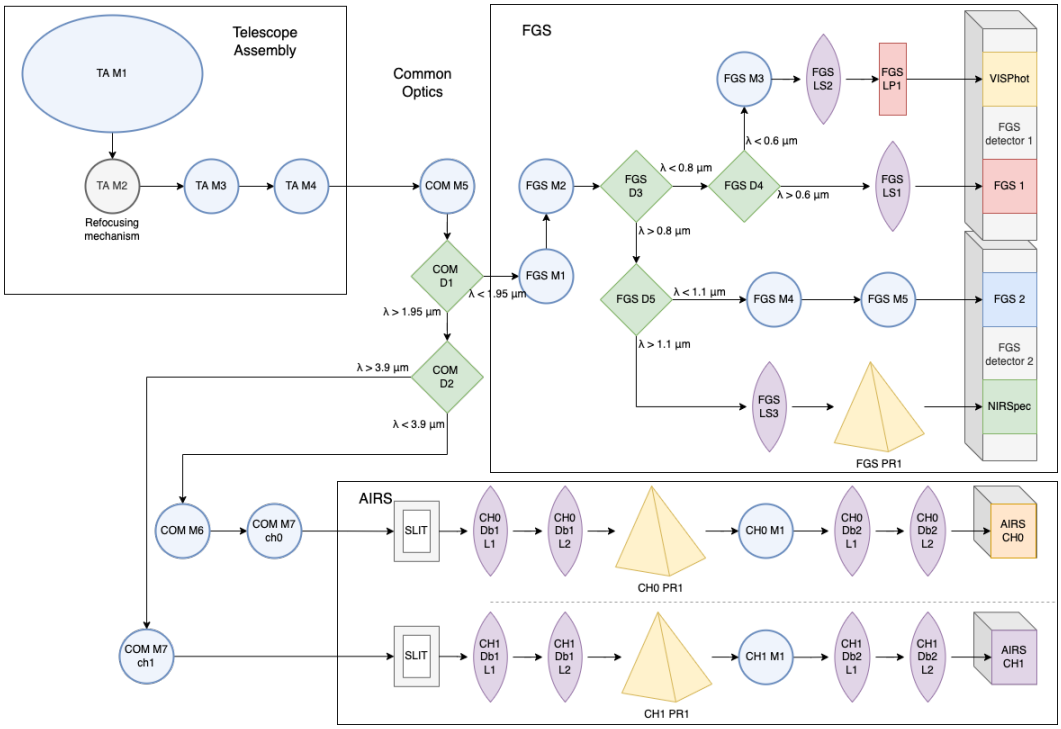


Figure 1.18. The optical layout assumed representative of the Ariel payload. The components of the system are represented as follows: mirrors as light blue circles, lenses as purple ovals, dichroics as green diamonds, prisms as yellow pyramids, filters as red rectangles, and detectors as gray cubes. Additionally, two slits (field stops) are shown as gray frames. The incoming light is initially collected by the primary mirror in the Telescope Assembly and then reflected onto the secondary mirror, which includes a refocusing mechanism. After propagating through the Telescope Assembly, the light enters the Common Optics, which splits the flux into two instruments: the Fine Guidance System (FGS) and Ariel Infra-Red Spectrometer (AIRS). Within FGS, the light is divided into four channels on two separate detectors. The AIRS instrument encompasses two channels, each with a separate detector in the current Ariel design.

- How do planets and planetary systems form?
- How do planets and their atmospheres evolve?

The observations will cover a wide spectral range from 0.5 to 7.8 µm using two instrument modules integrated into a single optical path (see Figure 1.18). This allows simultaneous and co-aligned observations by two science instruments: the Fine Guidance System (FGS) and Ariel Infra-Red Spectrometer (AIRS). FGS mounts three photometers and a low-resolution spectrometer, while AIRS has two infrared-optimized spectrometer channels (see Table 1.1 for details). With this optimized instrumental setup, Ariel will measure atmospheric signals from the planet at the level of 20–100 ppm relative to the host star, depending on the target brightness. This level of photometric precision is required to probe the atmospheric composition and thermodynamical structure of transiting planets. Additionally, the L2 orbit provides excellent thermal stability and maximum field of regard, ensuring high observational efficiency on the time-constrained observations of transits.

The simultaneous spectral coverage from visible to infrared wavelengths is another critical aspect of the Ariel mission. Robust relative calibration afforded by the simultaneous

Table 1.1. Ariel required spectral coverage and resolving power by instrument/channel.

Instrument	Channel	Spectral Range [μm]	Spectral Resolution
FGS	VISPhot	0.5 – 0.6	Integrated band
	FGS-1	0.6 – 0.8	Integrated band
	FGS-2	0.8 – 1.1	Integrated band
	NIRSpec	1.1 – 1.95	~20
AIRS	Ch0	1.95 – 3.9	~100
	Ch1	3.9 – 7.8	~30

observations is required to extract tiny signals from differences between measurements, avoiding potential offsets between detector channels which would compromise the quality of the data. From the shorter wavelengths, we can monitor stellar activity, detect the existence of haze and/or clouds in the planetary atmospheres, and provide an estimate of the planetary albedo. The longer wavelengths are key for detecting spectral features of the expected major atmospheric constituents, such as H_2O, CO_2, CH_4, NH_3, HCN, H_2S, TiO, and VO (Tinetti et al., 2013; Encrenaz et al., 2015), allowing to infer their abundances.

Transit spectroscopy observations require no significant angular resolution and detailed performance studies show that a telescope collecting area of 0.64 m^2 is sufficient to achieve the necessary observations on all the Ariel targets within the mission lifetime (see Chapter 3 for details). The Ariel telescope (Pace et al., 2022; Chioetto, 2022) is an off-axis 1.1 × 0.73 meter all-Aluminium Cassegrain[13] (M1, M2) with an elliptical primary mirror (see Figure 1.19). The Cassegrain telescope provides diffraction-limited performance beyond 3 μm. Behind M2, a refocusing mechanism (M2M) with three degrees of freedom (focus, tip, and tilt) is responsible for the in-flight focus adjustment after launch/cool-down and to compensate for any long term drifts in structural stability. The Cassegrain feeds the optical beam to the common optics, which include a re-collimating off-axis parabola (M3), two plane fold mirrors (M4, M5), and a dichroic beam splitter (D1) that splits the beam into two paths leading to the FGS and AIRS instruments. The Payload Module (PLM) is passively cooled to ~55 K by isolation from the Spacecraft (S/C) bus via a series of V-Groove radiators to limit thermal infrared background; the AIRS detectors are the only items that require cooling to <42 K, achieved via an active Joule-Thomson (JT) cooler with Neon as the working fluid.

1.2.2 The Ariel Observing Strategy

The primary science objectives of the Ariel mission, summarized above, necessitate the collection of a large and diverse sample of atmospheric spectra or photometric light curves of known exoplanets covering a wide parameter space in terms of planetary parameters and host-stars properties. In contrast, other science objectives require the very deep knowledge of a select sub-sample of objects. The total sample size of approximately 1000 targets calls for a well-defined observing strategy to maximize the scientific return of the mission.

After each observation, the resulting spectrum from each spectrometer – taken always at

[13] Parabolic primary and hyperbolic secondary.

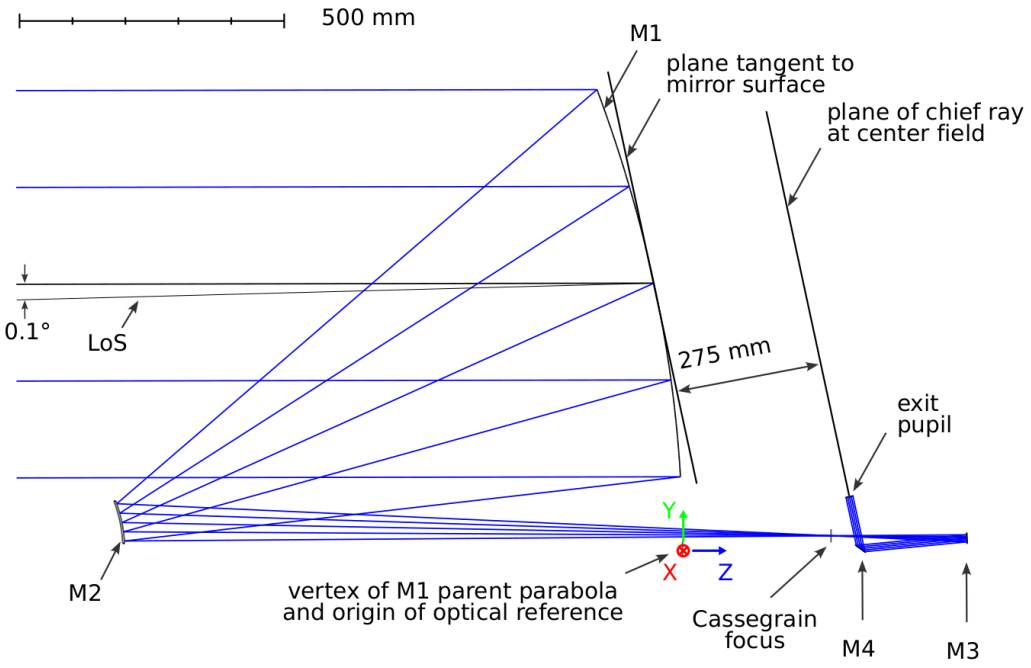


Figure 1.19. Scale drawing of the telescope and common optics on the Y-Z optical axis plane. The 0.1°offset is exaggerated for clarity.

the focal plane native resolution – is binned during data analysis in bins of different widths to optimize the Signal-to-Noise Ratio (SNR). Therefore, by implementing different binning options, the mission adopts a four-tiered approach tailored to address both questions on the general population of exoplanets and in-depth studies of individual targets. The four tiers are as follows (see Table 1.2 for a quick-look summary):

1. Tier 1 – shallow *Reconnaissance Survey* of the full sample of 1000 exoplanet atmospheres. It is designed to reach a SNR >7 when measurements are binned into 7+ photometric bands. Among its main purposes is to provide a rapid and broad characterization of planets to identify and flag featureless spectra, detect key molecular features and potentially provide a preliminary classification of planets through color-color diagrams (or other metrics) to reveal hidden properties/trends in the population, and guide decisions about priorities for future observations in higher Tiers. It will also address questions concerning the fraction of planets with clouds, the fraction of small planets that have retained their primordial gaseous envelope, and will characterize the albedo, bulk temperature, and energy balance for a sub-sample of planets.

2. Tier 2 – *Deep Survey* observations at higher spectral resolution and SNR of about half of the Tier 1 sample. The goal is to uncover the atmospheric structure and composition, as well as to search for potential correlations between atmospheric chemistry and fundamental parameters such as planetary radius, density, temperature, stellar type, and metallicity. The Tier 2 observations will bridge single planets and population-level studies, providing detailed information to fill the gaps in our understanding of the exoplanet population.

Table 1.2. Summary of survey tiers and detailed science objectives. From the Red Book.

Tier Name	Observational Strategy	Science case
Tier 1 *Reconnaissance survey*	Low Spectral Resolution observation of ~ 1000 planets in the VIS and IR, with SNR ~ 7	• What fraction of planets are covered by clouds? • What fraction of small planets have still retained H/He? • Classification through color-color diagrams? • Constraining/removing degeneracies in the interpretation of mass-radius diagrams • Albedo, bulk temperature, and energy balance for sub-sample
Tier 2 *Deep survey*	Higher Spectral Resolution observations of a sub-sample in the VIS-IR	• Main atmospheric component for small planets • Chemical abundances of trace gases • Atmospheric thermal structure (vertical/horizontal) • Cloud characterization • Elemental composition (gaseous planets)
Tier 3 *Benchmark planets*	High SNR observations in 1–2 events, re-observed over time	• Very detailed knowledge of the planetary chemistry dynamics • Weather and temporal variability
Tier 4 *Phase-curves & bespoke observations*	Phase-curves & bespoke observations	• Very detailed knowledge of the planetary chemistry dynamics • Targets of special interests • Weather, spatial, and temporal variability

3. Tier 3 – observations of *Benchmark Planets*, a small subsample of planets for which the maximum Ariel resolving power and SNR can be reached in one or two observations. Repeat observations of these planets will shed light on the temporal variability of their atmospheres, due to variations in the cloud coverage or patterns in the global circulation. These observations will in short enable extremely detailed and time-resolved studies of atmospheric chemistry and dynamics in key exoplanets.

4. Tier 4 – *Phase-curves & Bespoke Observations* will be allocated to targets of special interest, identified based on observations by previous missions as well as Ariel in its Tier 1. These targets may require tailored observations rather than the standard Tier-based observational strategy. Short-period exoplanets orbiting bright host stars are the most favorable for phase-curve observations in this Tier to probe the spatial and temporal variability of their atmospheres.

The interdependence among the tiers means that the scientific value of the data collected by Ariel will grow over time. The planets observed in each tier will give us deeper insight into the nature of the planets in preceding tiers. Tier 1 will provide the base for selecting Tier 2 planets, which in turn will inform the selection of Tier 3 targets. Tier 4 observations will benefit from the insight gained by the planetary populations of the other Tiers and will in turn provide a better understanding of their atmospheric behaviour. At the end of the nominal mission, the wealth of information supplied by the four-Tier approach will guide the decision-making process for the selection of the new observational sample.

1.3 Software Ecosystem

The Ariel mission is a multifaceted project that requires the integration of many different tools, each with a specific purpose, into a single ecosystem. From early-on in the mission stages, Ariel has developed advanced simulators to assess the performance of the mission, prepare the observations, and analyze the data. The software ecosystem is shown in Figure 1.20.

The figure is not meant to be exhaustive, and its main purpose is to show the interdependence of each component and underlying flow of information, activities, and networks of people. My research has led me to the center of this flow, where I have been able to contribute to the development of the software ecosystem in several ways, leading the actual development or contributing to it in a significant way[14]. Additionally, I have taken a proactive approach in the utilization of the software in critical aspects of mission preparation. This section aims to give a high-level introduction to this software, providing the reader with a foundation to be engaged in the following chapters with hands-on analyses.

The software ecosystem is composed of four main building blocks: the generic tools, the Ariel customization, the Ariel products, and the main finality of each product. Among the generic tools are the radiometric simulator ExoRad 2.0, the atmospheric retrieval framework TauREx 3, the time-domain simulator of an exoplanet observation ExoSim 2, and the physical optics propagation code PAOS. These tools are described in the following sections and chapters. The customization layer takes the current knowledge of the Ariel payload specifications (e.g., telescope, instruments, detectors, etc.) and the mission design (e.g.,

[14]With TauREx 3 being the exception, as it pre-dates my involvement in the Ariel mission.

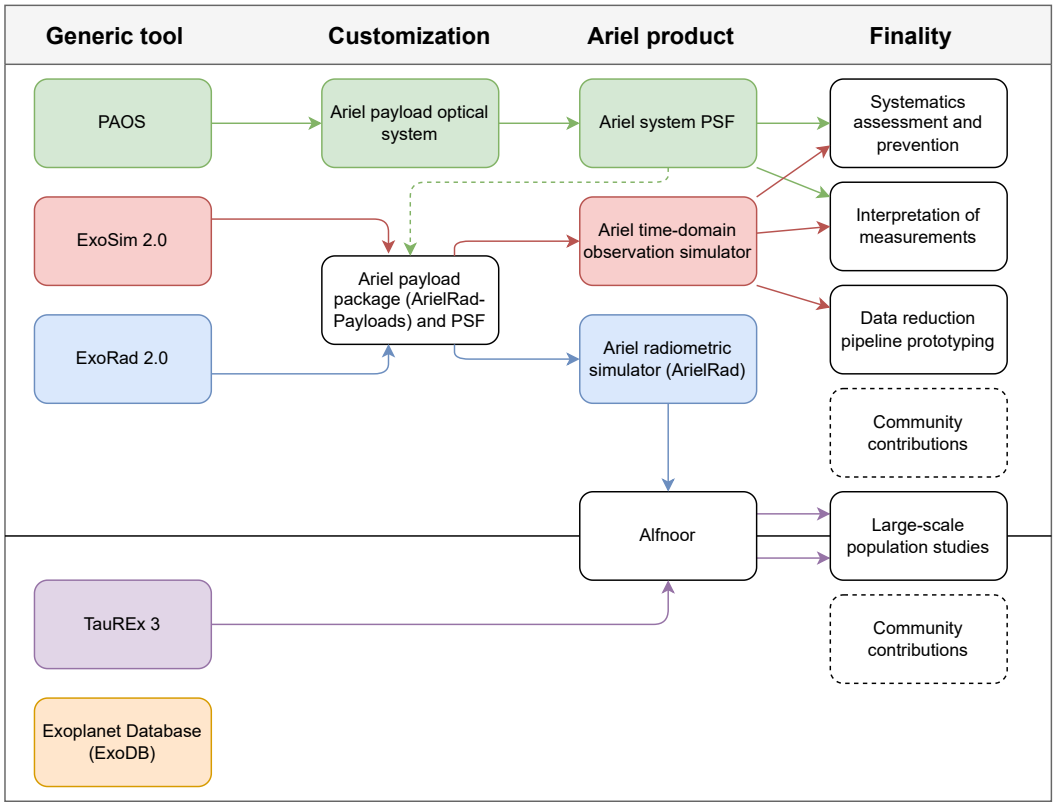


Figure 1.20. Illustration of the Ariel software ecosystem. The building blocks are generic tools, which can be customized for the Ariel application. The customized tools are then used to prepare the mission and the data interpretation, as well as to organize community contributions.

orbit, observing strategy, etc.) to adapt the generic tools and obtain the Ariel products. Major Ariel products are the system Point Spread Function (PSF) – which is fed back to the customization layer, the Ariel radiometric simulator ArielRad, ExoSim 2 adapted to Ariel, and Alfnoor – a tool for studies of synthetic populations of atmospheres observed by Ariel.

Thanks to these products, and in close interaction with all involved teams, we can assess e.g. the role of known systematics to improve the overall resilience of the mission design and devise strategies for their correction in post-processing. The goals being the optimization of the overall scientific output of the mission and the correct interpretation of measurements. These goals require (i) a thorough understanding of the instrument, (ii) the timely prototyping of the data reduction pipeline, and (iii) large-scale studies of thousands of representative atmospheres. All these aspects are covered by the Ariel software ecosystem, which will continue to evolve with the mission to best serve its scientific needs and organize community contributions.

1.3.1 ArielRad

The Ariel consortium has developed a radiometric simulator named ArielRad (Mugnai et al., 2020) to support the mission design and the scientific preparation of the mission. ArielRad enables the extensive modeling required for:

1. verifying performance requirements,
2. optimizing the payload design,
3. assessing the feasibility of observing Ariel's target list,
4. estimating required telescope time, and
5. thoroughly budgeting noise sources.

ArielRad is an adapted version of the generic point source radiometric simulator ExoRad 2.0 (Mugnai et al., 2023)[15]. This simulator enables the evaluation of the payload science performance by propagating signals through a detailed, physically-motivated model of the Ariel payload to estimate the noise performances achievable on each target. By facilitating detailed sensitivity analyses, ArielRad is an indispensable tool for maximizing the science return of Ariel.

The ArielRad software, implemented in Python, takes as input a configuration file specifying the Ariel payload parameters, a file defining the astrophysical parameters of potential exoplanet targets, and a file detailing the mission configuration (see Figure 1.21). Each simulation starts with the generation of the source signal from the target star. ArielRad currently supports black body sources, Phoenix stellar models[16] (Baraffe et al., 2015), or custom files that describe the source's spectral energy density vs. wavelength. The signal is then propagated through the modeled Ariel telescope and instruments to the detector focal planes. This includes accounting for each optical component's wavelength-dependent transmission and the dispersion of the prism spectrometers. Background signals from zodiacal light and thermal instrument emission are added, and the total signal is dispersed onto the appropriate detector pixels depending on the channel (photometer or spectrometer).

[15] ExoRad 2.0 is released under the BSD 3-Clause license and it is available at https://pypi.org/project/exorad/.

[16] https://phoenix.ens-lyon.fr/Grids/BT-Settl/CIFIST2011_2015/FITS/

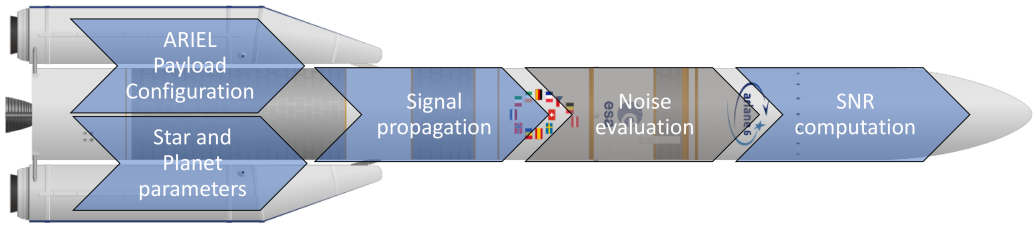


Figure 1.21. Illustration of the ArielRad workflow (Mugnai et al., 2020). ArielRad takes two input files: a payload configuration file and a candidate planet list. Then, it propagates the target host star signal through the payload before evaluating the noise. Finally, ArielRad estimates the transit or eclipse observation and calculates the resultant Signal-to-Noise Ratio (SNR).

With the input signal modeled, ArielRad estimates the noise variance on these signals in the spectral bin, considering photon noise from the star, zodiacal background, instrument emission contribution, detector noise (readout noise and read noise), dark current, and gain variations. Pointing jitter noise is imported from external time-domain simulations with ExoSim (Sarkar et al., 2021) and now ExoSim 2 (see Section 1.3.3 below). Additional margins are included to account for uncertainties in the current noise estimates and instrument performances, and a noise floor is implemented to prevent over-optimistic estimates (i.e., noise that integrates down indefinitely with time). The noise model[17] implemented in ArielRad is summarized in the below equation.

$$\frac{Var(S)}{S^2} = \underbrace{\sigma_G^2 \times \frac{1}{T}}_{\text{gain terms}} + \underbrace{g_\gamma \frac{1+X}{k\eta_s\, QE\, N_0} \times \frac{1}{\epsilon T}}_{\text{photon noise term}} + \underbrace{\frac{N_{pix} I_D + N_{pix}\sigma_{add}^2/(\epsilon\Delta t)}{(k\eta_s\, QE\, N_0)^2} \times \frac{1}{\epsilon T}}_{\text{dark current and additive noise}} + \underbrace{p_0^2}_{\text{payload noise floor}} \tag{1.9}$$

In this equation, S represents the time-averaged signal with efficiency ϵ. σ_G denotes electronic gain noise. g_γ is a photon noise amplification factor arising from detector ramp fitting, taking a value of 1 for Correlated Double Sampling (CDS). X is a margin term for photon noise. N_{pix} gives the number of pixels used in aperture photometry, and k is the aperture correction factor. I_D symbolizes the dark current per detector pixel. QE denotes the quantum efficiency. N_0 gives the incident photons reaching the focal plane in each photometric channel or spectral bin. Δt represents the detector exposure time. Note that the expected value of S equals $k\eta_s\, QE\, N_0$, where η_s is an astrophysical source efficiency set to unity here. Additive noise σ_{add} accounts for detector read noise and other additive terms beyond read noise. The noise floor p_0 is the limit below which noise does not integrate down with time, interpretable as low-frequency Brownian noise.

Using this noise model, ArielRad calculates the atmospheric signal and resulting Signal-to-Noise Ratio (SNR) achievable for a given exoplanet target in both transit and eclipse. Each simulated observation spans 2.5 times the transit duration between first and fourth contact points (T_{14}). This approach enables out-of-transit data collection alongside in-transit measurements for light curve fitting and transit depth determination. By iterating through all prospective targets over multiple tiers of spectral resolution, ArielRad can robustly predict Ariel's performance across its three-tier observing strategy (see Section 1.2.2). Finally, ArielRad estimates the number of observations required to achieve the desired SNR for

[17] ARIEL-SAP-PL-TN-004

each planet (e.g. SNR = 7) and from this the number of transit observations and the total observing time required per target.

ArielRad enables verification that the Ariel payload meets the mission science requirements. By modeling the noise contributions in detail, it facilitates thorough sensitivity analyses to identify optimal design solutions that maximize SNRs achieved. Computationally efficient to run for thousands of targets, ArielRad permits assessment of whether the proposed Ariel target list can actually be observed to the required sensitivity within the mission lifetime (e.g. Edwards et al., 2019; Edwards and Tinetti, 2022). This is crucial for verifying the feasibility of Ariel's observing plan.

ArielRad has undergone extensive validation, with its results cross-checked against two other simulators, including the first version of ExoSim (Sarkar et al., 2021), which has been validated against real astrophysical observations. The comparison shows excellent agreement between the simulators. While ExoSim provides unparalleled fidelity, ArielRad offers sufficiently accurate performance estimates with massively improved computational efficiency. The development of ArielRad has been critical in supporting the payload design convergence and performance verification activities during Ariel's Phase B1 study, leading up to mission adoption. Recently, it was used in the payload performance analysis for the Preliminary Design Review (PDR) (see Chapter 3 for details). As Ariel moves towards the Critical Design Review (CDR), ArielRad will continue to be an essential tool for finalizing the payload design and optimizing the expected scientific return.

1.3.2 TauREx 3

Tau Retrieval for Exoplanets (TauREx 3) (Al-Refaie et al., 2021) provides a fully Bayesian inverse atmospheric retrieval framework for exoplanetary atmosphere modeling and retrievals, written in Python. The framework builds forward models, simulates instruments, and performs retrievals, and provides a rich library of classes for building additional programs and using new atmospheric models[18]. TauREx 3 is very flexible, allowing users to mix and match atmospheric models and retrieval algorithms, and is also designed to be easily extensible, allowing users to add their own forward models and retrieval algorithms. This is achieved by using a modular design (see Figure 1.22) that allows the user to easily add new components to the framework. This flexibility is enhanced by a plugin-based system (Al-Refaie et al., 2020), inspired by Flask extensions[19], allowing the framework to be extended with new functionality without the need to modify the core code. Anyone can host and develop taurex plugins, and they are designed to be fully interoperable with each other. This allows TauREx 3 to potentially become a hub for the exoplanet atmospheric retrieval community. Table 1.3 reports a list of plugins available on Python Package Index (PyPI); the list is not exhaustive, as new plugins are continuously being developed and released[20].

Compared to its previous incarnations (Waldmann et al., 2015a,b), TauREx 3 is a complete rewrite and redesign of the codebase, with a focus on performance, flexibility, and ease of use. At its core, TauREx 3 is a line-by-line radiative transfer code embedded in a Bayesian framework. Rather than a monolithic block of code, the new framework represents different atmospheric properties like temperature profiles and chemical abundances as

[18] A propedeutical introduction to TauREx 3 can be found in Quentin Changeat's Ph.D. thesis.
[19] https://flask.palletsprojects.com/en/2.3.x/extensions/
[20] The main GitHub repository for TauREx 3 and its plugins is https://github.com/ucl-exoplanets.

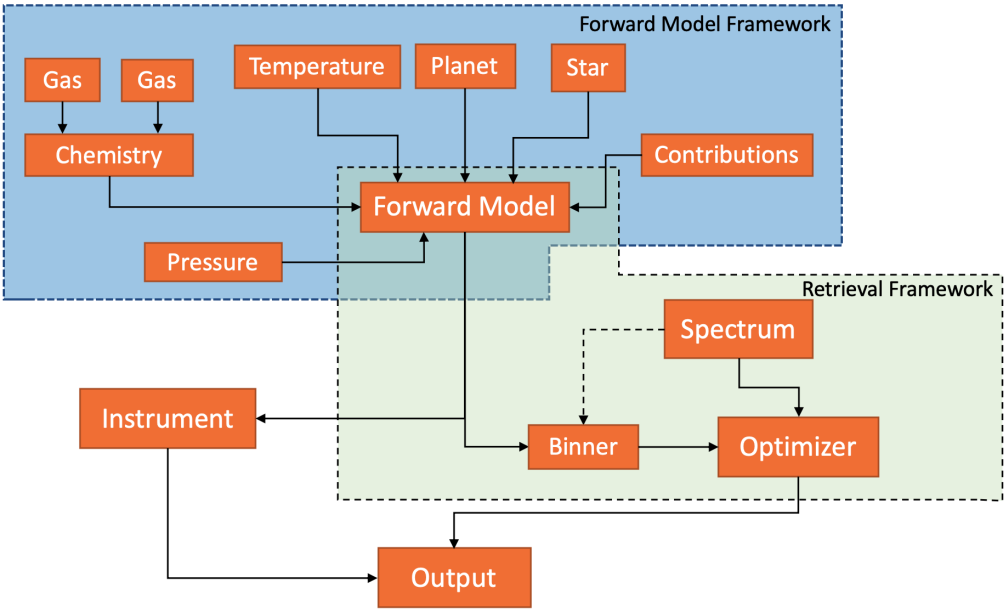


Figure 1.22. The overall structure of TauREx 3 (Al-Refaie et al., 2021). Highlighted is the two-framework structure of the complete framework. Each box describes a class from Table 1, solid arrows flowing out describe outputs, and solid arrows flowing in describe inputs. A dashed arrow describes the creation of an object.

Table 1.3. A non-exhaustive list of TauREx 3 plugins available on PyPI.

Plugin name	Description	PyPi
taurex_ace	ACE (Agúndez et al., 2012, 2020) equilibrium chemistry	link
taurex_fastchem	FastChem (Stock et al., 2018, 2022) equilibrium chemistry	link
taurex_ggchem	GGChem (Woitke et al., 2018) equilibrium chemistry	link
taurex_cuda	CUDA-acceleration of forward models	link
taurex_ultranest	Ultranest (Buchner, 2021b) sampler for the retrieval	link
taurex_petitrad	petitRADTRANS (Mollière et al., 2019) forward models and opacity formats	link

interchangeable building blocks. Each module exposes a simple interface specifying what inputs it requires and outputs it generates. The modules can be connected together into a full forward model pipeline without tightly coupled interdependencies. This approach offers great flexibility: users are able to develop new models and components into the existing framework with minimal coding effort. Another breakthrough is a system for dynamic detection and fitting of model parameters, which eliminates much of the complexity required in previous versions to manually define parameter lists and handle their sampling. This is further helped by the high-level input file, designed to allow users to easily enable or disable the fitting of any parameter without editing core code. These capabilities make TauREx 3 an extremely powerful tool for the exoplanet community to investigate new data sets and test new theoretical advances in exoplanet atmosphere modeling by the community.

Out-of-the-box, TauREx 3 contains a suite of state-of-the-art models relevant for exoplanet characterization. For primary transit spectroscopy, a 1D plane-parallel forward model computes transmission spectra using efficient numerical integration. Emission spectra can be generated with a 1D hemispherically-integrated thermal emission model. Multiple temperature profile parameterizations are available, including analytic two-stream solutions and flexible N-point profiles. Equilibrium and free chemistry routines provide atmospheric compositions, with configurable mixing ratio profiles for individual species. Opacities are handled using cross-sections with high-performance interpolation, allowing high spectral resolution. Contribution modules represent processes including molecular absorption, collision-induced absorption (CIA), Rayleigh scattering, and gray clouds. TauREx 3 supports a number of formats for wavelength-dependent opacities including ExoMol[21] (Tennyson and Yurchenko, 2012), HITRAN[22] (Gordon et al., 2017), and HITEMP[23] (Rothman et al., 2010). Also supported are opacity tables for fast computation of molecular absorption (e.g. Yurchenko et al., 2018). A flexible binner resamples model spectra onto arbitrary observational grids. By importing noise modeling in TauREx 3, users can simulate observations and estimate the retrievability of atmospheric parameters given a range of telescope and instrument configurations. This extensive suite of modern components enables TauREx 3 to synthesize high-fidelity atmospheric spectra for solar system and exoplanet applications spanning a wide phase space.

The software employs state-of-the-art Bayesian retrieval techniques to infer model parameters and uncertainties from simulated/observational data. Nesting samplers – including MultiNest (Feroz et al., 2009; Buchner, 2021a) and Nestle (Barbary, 2021) – are supported to rigorously map posterior distributions efficiently. The dynamic fitting system auto-selects an appropriate set of parameters to retrieve based on the chosen forward model configuration. Uniform priors are currently implemented, but the modular design and available plugins readily allow the addition of more complex priors (Al-Refaie et al., 2020). Native TauREx 3 can thus conduct retrievals to constrain properties like chemical abundances, temperature structure, planetary radii, and cloud properties given transmission or emission spectra. The versatile retrieval functionality provides a robust statistical framework for analyzing spectroscopic data sets. Moreover, the flexibility of the plugin-enhanced TauREx 3 has enabled to cross-compare state-of-the-art chemical equilibrium codes (Al-Refaie et al., 2022a).

Thanks to these capabilities, in recent years TauREx 3 has played a key role in enabling

[21] https://www.exomol.com/
[22] https://hitran.org/
[23] https://hitran.org/hitemp/

the detailed analysis of spectroscopic data sets to constrain the properties of exoplanets. An example is the study presented in (Changeat et al., 2022) where they used TauREx 3 to re-analyze archival data of the atmospheres of 25 exoplanets with observations taken with the HST/WFC3. Another is the study (Al-Refaie et al., 2022b) which demonstrated the viability of total disequilibrium chemistry[24] retrievals with simulated JWST observations of HD189733b, implemented thanks to a TauREx 3 plugin. Figure 1.23 showcases that the full chemical scheme matches the simulated observations, with recovered free parameters close to the chosen true value.

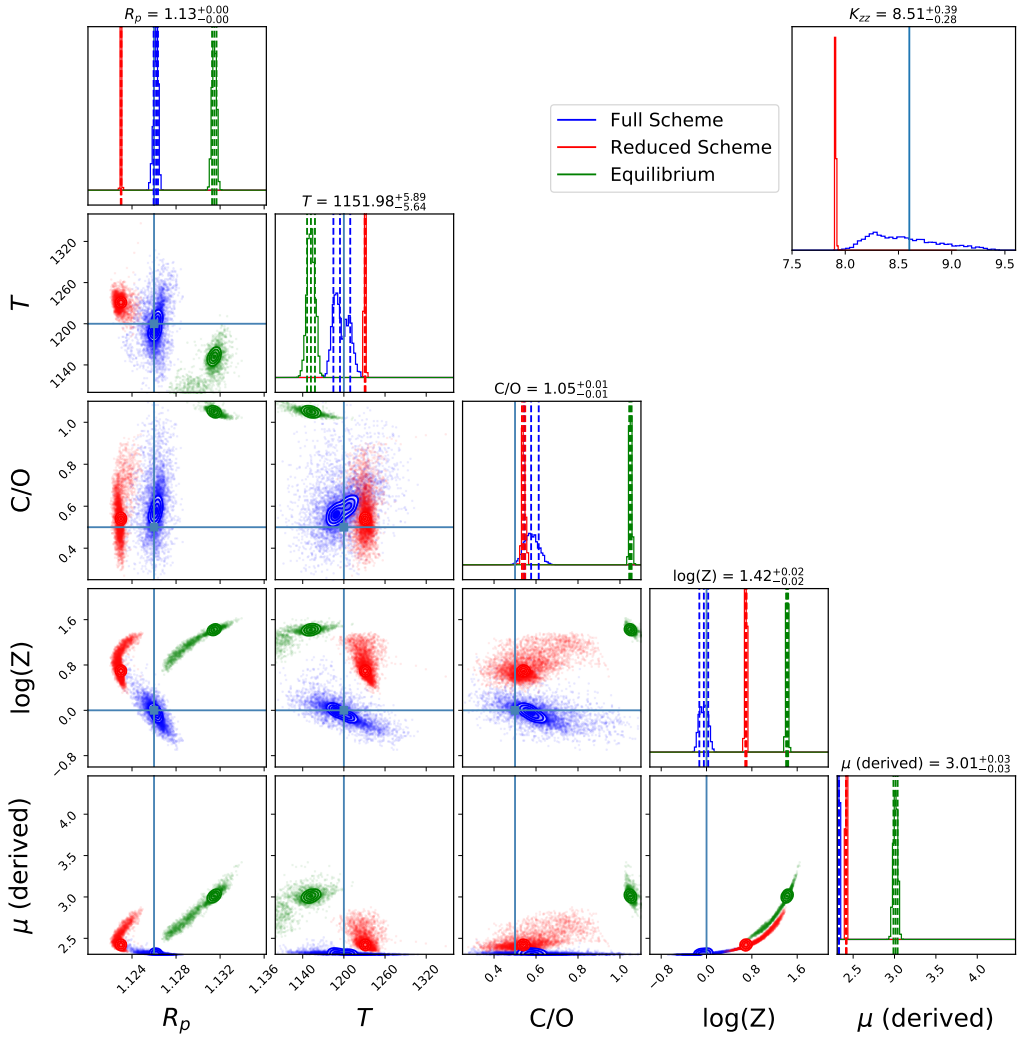


Figure 1.23. Posteriors from the retrieval of a simulated JWST spectrum of HD 189733 b (Al-Refaie et al., 2022b) using full and reduced chemical network of Venot et al. (2020a) and equilibrium chemistry (Agúndez et al., 2012). Blue and red are the posteriors using the full and reduced schemes respectevly and green are the posteriors using equilibrium chemistry.

In summary, TauREx 3 represents a major leap forward in publicly available software

[24] As opposed to equilibrium chemistry, where the detailed molecular abundances depend solely on the temperature, pressure, and elemental composition.

for exoplanet atmosphere characterization. The flexible, high-performance design facilitates leveraging cutting-edge theoretical models. Dynamic fitting and modular components enable extensive customization. These capabilities make Taurex an extremely powerful tool for the exoplanet community to investigate new data sets. By lowering barriers to rapidly prototype and evaluate new retrieval approaches, TauREx 3 helps accelerate understanding of these faraway worlds. It is publicly released under an open-source license to maximize its value to the scientific community. Ongoing development will further enhance TauREx 3's capabilities through new forward model physics, support for more general prior shapes, and expanded spectral generation options. By providing a state-of-the-art, shared platform for atmospheric retrievals, TauREx 3 aims to be an enabling resource helping unlock insights from current and future exoplanet observations.

1.3.3 ExoSim 2

ExoSim 2 (Mugnai et al., 2022) is a time-domain simulator for exoplanet observations, designed to simulate the entire observation process, from the generation of the astrophysical signal to the final detector readout. It can capture temporal effects, such as correlated noise and systematics on the light curve. The simulator produces spectral images like those produced by an actual observation, aiming to:

1. inform the design of the instrument;
2. assess the impact of systematics on the measurements;
3. evaluate the compliance with scientific objectives;
4. prepare the data reduction pipeline against realistic data sets;
5. assist in the selection of targets and observing strategies;
6. interpret the measurements;
7. assess the confidence level of retrieved quantities.

The first version of ExoSim (Sarkar et al., 2021) was developed for the Ariel space mission, then adapted to the James Webb Space Telescope (JWST) and presented to the community as JexoSim (Sarkar et al., 2020; Sarkar and Madhusudhan, 2021).

ExoSim 2, a fully refactored version, represents a major advancement in exoplanet observation simulation tools, developed to address the evolving needs of the exoplanet community. Built entirely in Python, ExoSim 2 employs an object-oriented architecture focused on flexibility, extensibility, and ease of use compared to its previous version. The code is open-source and designed to simulate high-fidelity focal planes alongside photometric and spectroscopic timelines for diverse optical systems studying exoplanets. A huge effort has been put in place to systematically document the code and include readily executable examples to facilitate adoption by new users. This effectively enables community-driven code growth as needed moving forward.

A key innovation in ExoSim 2 is its modular framework based on encapsulated Task classes. Different physical models and simulation steps are implemented as swappable Tasks. This allows straightforward customization and expansion of ExoSim 2's capabilities by users writing their own Tasks, without needing to modify the core code. Tasks automatically share

data through a common `Signal` class representing a general photometric or spectroscopic data cube. As shown in Figure 1.24, ExoSim 2 parses the input parameters description and data, then executes through a three-stage pipeline: 1) focal planes creation, 2) sub-exposures generation, and 3) final pixel reads with noise. This produces focal plane, sub-exposure, and Non-Destructive Reads (NDRs) data products. This structure optimizes resource usage by enabling selective recomputation of only updated components.

ExoSim 2 incorporates detailed models of astrophysical scenes and instrumental effects. Point Spread Functions (PSFs) can be generated analytically (Airy or Gaussian) or supplied by the user. PSFs from PAOS (see Chapter 2 for details), are natively supported. Pixel-level effects include intra-pixel sensitivity variations, diffusion, and multiple sampling modes. Sophisticated noise models account for contributions from source shot noise, dark current, readout noise, and zodiacal light. The flexible Task framework allows integrating additional components like background stars through custom extensions. This enables ExoSim 2 to produce realistic synthetic observations for a wide variety of instruments and targets, beyond its initial use case.

Concerning the astrophysical scene, ExoSim 2 includes a module for loading and applying different astronomical signals to the target source. This expands upon the "Astroscene" functionality of the previous code version (Sarkar et al., 2021), allowing users to write their own `Tasks` that describe the target astronomical signal. The signal is convolved with the spectral instrument line shape before being applied as a perturbation to the data. Figure 1.25 illustrates an example transmission spectrum simulated with ExoSim 2. The high-resolution input spectrum is observed by Ariel in the FGS-NIRSpec and AIRS spectroscopic channels, and the output spectrum is obtained after post-processing the science images produced by ExoSim 2. This functionality enables ExoSim 2 to provide realistic data sets to assess the data reduction pipeline's capability to recover spectral features of a given amplitude post-processing.

The development of ExoSim 2 has been driven by the evolving needs of the Ariel space mission. As a pioneering spectroscopic survey of exoplanet atmospheres, Ariel requires detailed simulations of payload performance and related systematics to develop its observation strategy and data analysis pipeline. ExoSim 2 models the specific instrumental characteristics of Ariel across its multiple detector channels. The simulator has undergone extensive testing and validation against ArielRad (see above for details), showing excellent consistency in estimating photon conversion efficiency, saturation times, and signal generation. The software has also been independently validated for instantaneous read-out and jitter simulation, as well as for astronomical signal representation, attesting to its robustness and versatility. This gives confidence ExoSim 2 will provide Ariel with representative simulations to inform all mission activities in the coming years.

Looking beyond Ariel, ExoSim 2's flexibility enables adaptation to other exoplanet studies. Its generic optical payload and noise frameworks allow the modeling of any combination of imagers and spectrographs. User customization through Task extensions provides the capability to tailor ExoSim 2 to different satellite, balloon-borne, and ground-based instruments. Already ExoSim 2 simulations are being generated for the Exoplanet Climate Infrared TElescope (EXCITE) mission (Tucker et al., 2018), and its ability to simulate the physics encapsulated in user-defined models opens avenues for other observatories, such as the JWST, to be modeled within its framework. This demonstrates the significant value of ExoSim 2's versatile architecture in reliably serving varied scientific instrumentation.

In conclusion, ExoSim 2 represents a state-of-the-art tool for synthesizing observations

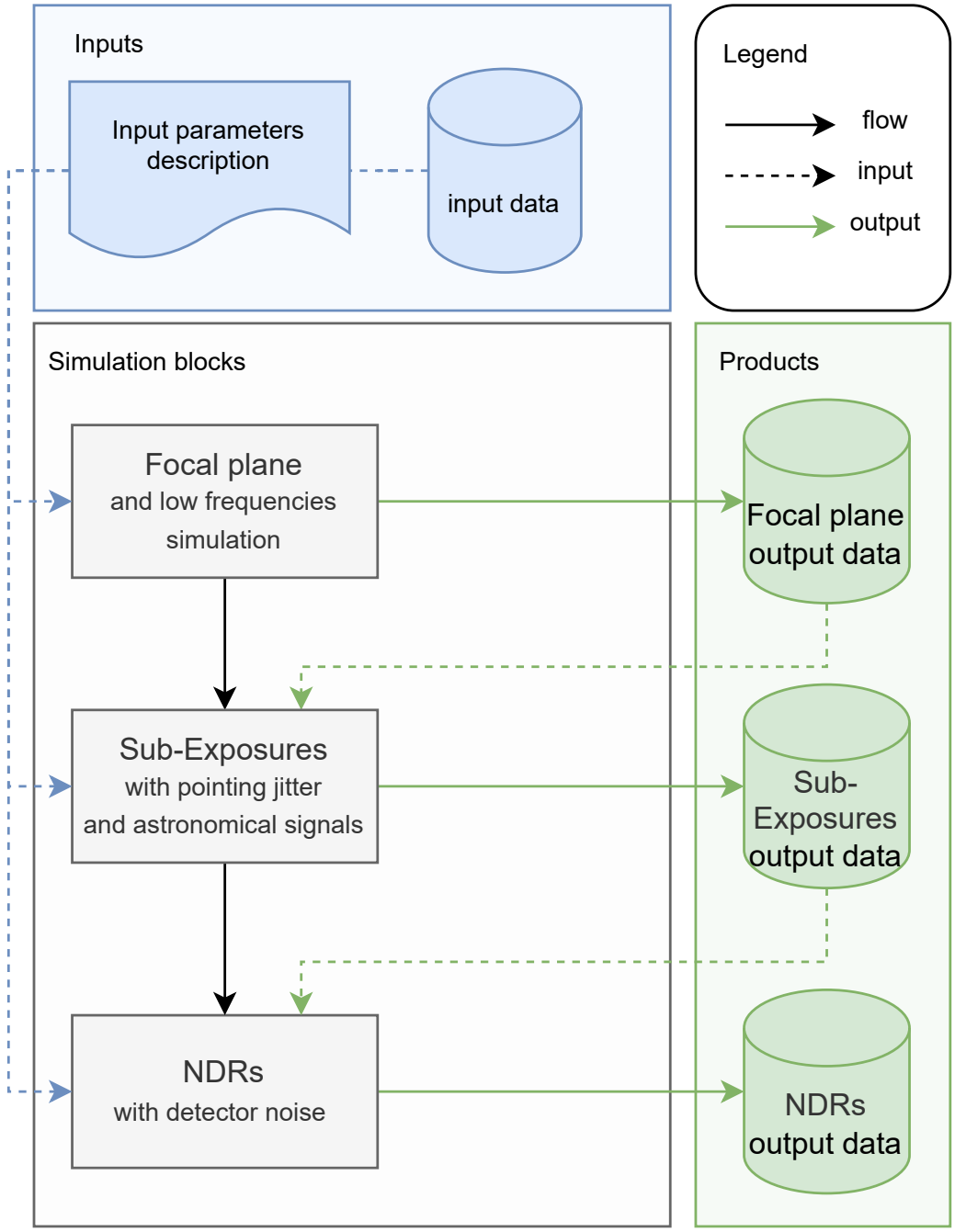


Figure 1.24. ExoSim 2 operates via a three-step workflow. The initial step involves the creation of focal planes based on source and instrument specifications. Subsequently, the Sub-Exposure step simulates high-frequency sampled focal planes, incorporating astronomical signals, pointing jitter, and detector read-out modes. Finally, the Non-Destructive Read (NDR) step generates NDRs from the Sub-Exposure, accounting for detector noise. Figure from Mugnai et al., 2023 (in preparation).

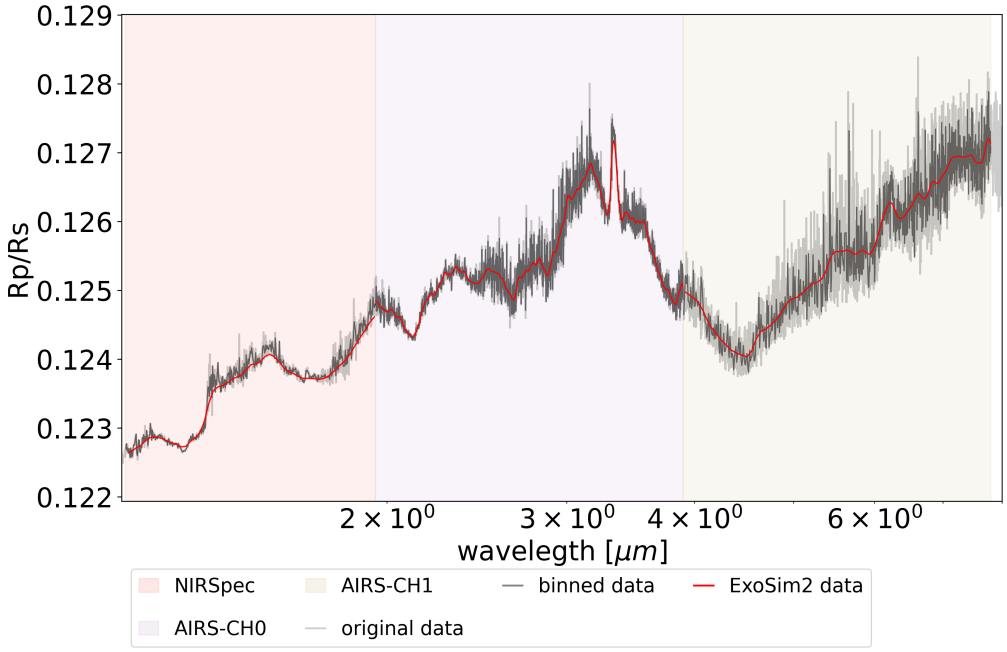


Figure 1.25. Transmission spectrum simulated in ExoSim 2. The gray curve represents the original high-resolution spectrum used as input for the astronomical signal simulation. The black curve corresponds to the same spectrum binned down to the focal plane resolution at the pixel level. The red curve represents the spectrum obtained as output from ExoSim 2 data, post-processing. The red spectrum is modulated by the instrument line shape. The colored background indicates the spectral range of the spectrometers. Figure from Mugnai et al., 2023 (in preparation).

to enable exoplanet atmosphere characterization studies. For future missions like Ariel, it provides critical support for mission design and planning. In the era of JWST and extremely large telescopes, models like those encapsulated in ExoSim 2 will be essential to understand the biases of estimators and instrumental effects to accurately interpret the data. The open-source nature of ExoSim 2 will foster growth through community contributions. As we push the boundaries of exoplanet science with new discoveries and high-fidelity observations, robust and adaptable simulators like ExoSim 2 will be ever more indispensable tools for understanding these faraway worlds.

Chapter 2

The Physical Optics Simulator

This chapter presents the development, validation, and application of a Physical Optics Propagation (POP) code called Physical Optics Simulator (PAOS). The original motivation behind the development of PAOS is to assess the optical performance of the Ariel space mission (see Section 1.2 for details). Reliable estimates of the Ariel system Point Spread Function (PSF) are critical to estimate saturation times and photometric performance. During different mission phases, the PSF needs to be (i) estimated from manufacturing models, (ii) extrapolated from ground measurements, and (iii) measured in-flight. For each of these phases, PAOS can be used to estimate the representative system PSF in the Fresnel approximation, accounting for the impact of diffraction, aberrations, and related systematics. This is essential to interpret the measurements and ensure that Ariel achieves a sufficiently compact PSF to operate as a light-bucket (see Section 2.3 for details) and a sufficiently sampled PSF to mitigate line-of-sight jitter (see Section 4.1 for details), as per design. Additionally, tight integration with the mission simulators such as ArielRad and ExoSim 2 (see Section 1.3.1 and Section 1.3.3 for details) has been achieved to provide the representative PSF at the detector level. Beyond Ariel, PAOS can be used to simulate the optical performance of a wide range of optical systems, including non-axial symmetric ones, as long as the Fresnel approximation remains valid.

* * *

Optical system design has witnessed significant advancements in recent years, necessitating efficient and reliable tools to simulate and optimize complex systems (Smith, 2000). Ray-tracing and POP are the two primary methods for modeling the propagation of electromagnetic fields through optical systems. Ray-tracing is often employed during the design phase due to its speed, flexibility, and efficiency in determining basic properties such as optical magnification, aberrations, and vignetting. POP provides a comprehensive understanding of beam propagation by directly calculating changes in the electromagnetic wavefront (Goodman and Weare, 2010). POP is particularly useful for predicting diffraction effects and modeling the propagation of coherently interfering optical wavefronts. Yet, it may require supplementary input from direct measurements or a ray-tracing model for comprehensive analysis including aberration variations, especially in the Fresnel approximation.

Commercial ray-tracing codes like `Zemax` and `Code V` provide tools for performing POP calculations and optimization, offering advanced capabilities in aberration reduction

and optical system enhancement. However, these programs often come with substantial costs and steep learning curves, which may not be justifiable for every application. Furthermore, accessibility to their source code could be limited.

To overcome these limitations, we present the PAOS, a reliable, user-friendly, and open-source POP code that integrates an implementation of Fourier optics. It employs the Fresnel approximation for efficient and accurate optical system simulations. By including a flexible configuration file and paraxial ray-tracing, PAOS seamlessly facilitates the study of various optical systems, including non-axial symmetric ones, as long as the Fresnel approximation remains valid.

Initially developed to evaluate the optical performance of the Ariel space mission (Tinetti et al., 2018, 2021), PAOS has proven its value in assessing the impact of diffraction, aberrations, and related systematics on Ariel's optical performance. By offering a general-purpose tool capable of simulating the optical performance of diverse optical systems, PAOS fills a crucial gap in the field and makes advanced physical optics research more accessible. This chapter presents the development, validation, and application of PAOS and its limitations, showcasing its potential to advance optical system design and analysis for a wide range of scientific and engineering applications.

2.1 Methods

2.1.1 PAOS: an overview

PAOS, is an end-to-end POP simulator developed using a full Python 3 stack. The motivation behind the development of PAOS is that of assessing the optical performance of Ariel. However, PAOS is not specific to Ariel and it is general enough to be employed as a library for implementing custom propagation routines. The code will soon be openly distributed under the BSD 3-Clause License.

2.1.1.1 Initial setup

The PAOS software package is equipped with an installer, comprehensive documentation in HTML and PDF formats, and a collection of example scripts that showcase the core functionalities of the code. To compile the documentation from the source code, users can follow the instructions in the `README.md` file and use the `make` command (see Listing 2.1).

```
$ cd PAOS/docs/
$ make html
$ make latexpdf
```

Listing 2.1. Compiling the PAOS documentation.

This process generates an HTML file in the `build/html` directory and a PDF file in the `build/latex` directory within `PAOS/docs/`. The documentation provides a detailed explanation of the code and the API, accompanied by example scripts that serve as a valuable resource for new users.

PAOS necessitates a Python 3 installation and the `numpy` package as minimum requirements. The installation can be compiled from the source (see Listing 2.2).

```
$ cd PAOS/
$ pip install .
```

Listing 2.2. Installing PAOS from source.

This action automatically downloads and installs the requisite dependencies. After installation, users can initiate PAOS from any location on the operating system using the command `paos` (see Listing 2.3).

```
$ paos --help
usage: paos [-h] -c CONF [-o OUTPUT] [-lo] [-wl WAVELENGTHS] [-wlg
    WL_GRID] [-wfe WFE] [-keys STORE_KEYS] [-n N_JOBS] [-p] [-d] [-l]
```

Listing 2.3. PAOS help message.

To execute an input file `input.ini`, save the output, generate a plot of the results, and employ 2 parallel cores, the command in Listing 2.4 is used:

```
$ paos -c input.ini -o output.h5 -p -n 2
```

Listing 2.4. Example PAOS command.

In this case, `output.h5` represents a `HDF5` file containing the simulation output, which by default includes the wavefront amplitude, spatial sampling along the x and y axes, and wavelength.

Upon installation, users also gain access to the `paosgui` command, which initiates a GUI for PAOS (see Listing 2.5).

```
$ paosgui --help
usage: paosgui [-h] [-c CONF] [-d] [-l] [-o OUTPUT]
```

Listing 2.5. PAOS GUI help message.

The GUI, acting as a wrapper for the `paos` command, provides users with a streamlined and user-friendly interface to access input files and effectively utilize the most important functionalities of the code.

2.1.1.2 Software architecture

PAOS boasts a modular software architecture, composed of four primary modules: `core`, `classes`, `util`, and `gui`. The `core` module manages the central POP propagation loop, employing the Python classes `ABCD`, `WFO`, and `Zernike` from the `classes` module, along with the `Material` class from the `util` module (Table 2.1). These classes serve as the fundamental building blocks of the code. In addition to directing the main propagation process, the `core` module encompasses advanced plotting routines and output file writing capabilities, streamlining data visualization and storage. Lastly, the `gui` module implements the GUI for PAOS.

Table 2.1. The main Python classes in PAOS.

Class	Main purpose	Module
`ABCD`	Implement ABCD matrix theory	`classes`
`WFO`	Implement wavefront propagation	`classes`
`Zernike`	Handle Zernike polynomials	`classes`
`Material`	Handle optical glasses	`util`

2.1.1.3 Simulation workflow

Figure 2.1 presents an overview of the PAOS code and the standard simulation workflow. The user can launch the code (i) from the terminal using the `paos` command, (ii) from a Python script using the `pipeline` routine in the `core` module, or (iii) by manually executing the steps in a Jupyter notebook.

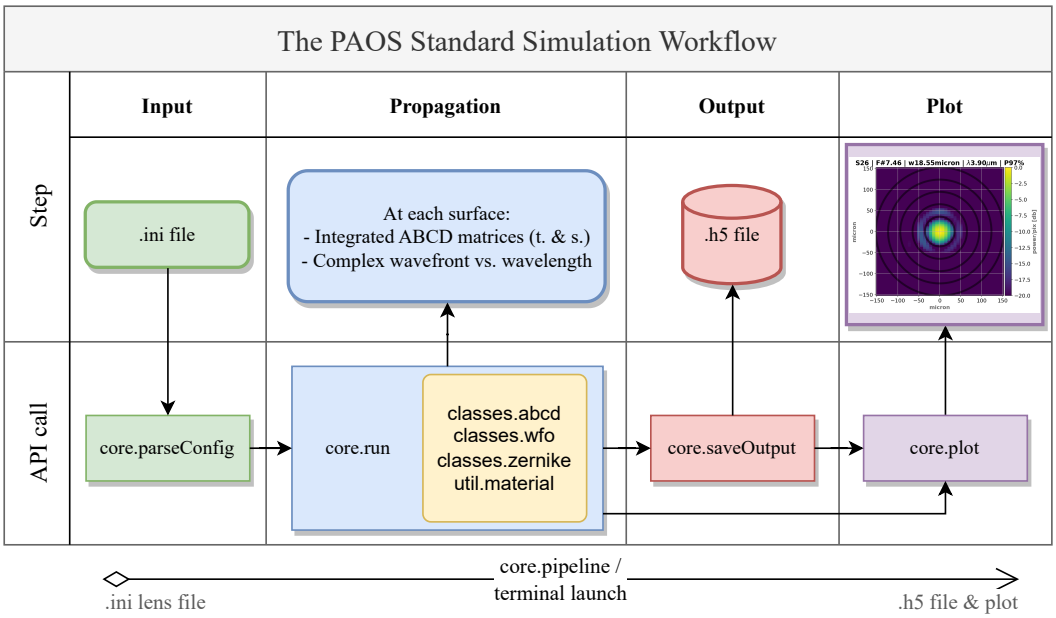


Figure 2.1. Schematic representation of the PAOS standard simulation workflow, illustrating each step and associated API call. The standard input is a .ini file containing the initial beam properties, ambient pressure and temperature, wavelengths, fields, and the sequence of optical surfaces from the input to the output. The simulation can be launched from a shell or a Python script. PAOS automatically parses the input file and runs the propagation, accounting for refractive index changes. It outputs the sagittal and tangential ABCD matrices, along with the complex wavefront vs wavelength at each required plane. The output is saved in a .h5 file, with visualization tools provided for PSF representation. Additionally, PAOS can function as a library, enabling custom propagation routines and surface types.

The simulation begins by parsing the user-defined input, a .ini file that contains the simulation configuration (Section 2.1.2.1). It then proceeds with wavefront propagation through the optical system.

Wavefront propagation is performed using the Paraxial theory described in Lawrence et al. (1992), that is, Fourier optics in the Fresnel approximation (Section 2.1.2.3). PAOS

initially propagates a surrogate beam profile using an analytically-traced on-axis Gaussian beam (Appendix B.1.2). From the properties of the surrogate beam, PAOS then automatically selects the algorithm for propagating the actual wavefront in near and far-field regions and the corresponding computational grid sampling (Appendix B.1.7).

The propagation accounts for any aberration that can be described by Zernike polynomials (Section 2.1.2.4). Such aberrations can occur due to e.g. imperfections in mirror manufacturing and the presence of refractive elements at non-normal angles, such as prisms. However, the current implementation of PAOS lacks the capability to independently evaluate wavefront aberrations. Instead, the user is required to input wavefront aberrations in the .ini file, utilizing Zernike polynomial coefficients. To address this limitation, a future version of PAOS may include a dedicated ray-tracing module.

Moreover, PAOS features a built-in refractive index database and calculates refractive indices using the Sellmeier equation (Section 2.1.2.5). Deviations from the refractive index at standard temperature and pressure are accounted for by specifying ambient pressure and temperature in the .ini file.

Throughout propagation, PAOS computes the complex wavefront and updates the ABCD matrices in the sagittal and tangential directions at each surface (Section 2.1.2.6). PAOS can save this data in an `HDF5` file at each point in the optical system or only at the output. Additionally, PAOS provides routines to plot the PSFs at the principal and intermediate focal planes. Parallelization using the `joblib` module in Python can be activated to speed up computation.

Finally, PAOS implements a logger based on the `logging` module in Python, allowing users to save output parameters, debug the code, and monitor the simulation progress. The log file is saved in the same directory as the output file and contains the relevant information for reproducibility (e.g. code version). This information is also stored in the `HDF5` file under the `info` group.

2.1.2 PAOS: implementation

2.1.2.1 Input system

PAOS utilizes the standard .ini file format for its input system. The configuration file is a single .ini file that serves as the main source of input parameters and contains the description of the optical chain, which defines the optical system from the input to the output. The input structure is generic, flexible, and intuitive and can be applied to configure and customize a wide range of optical simulations.

For instance, let us consider the optical layout in Figure 2.2, exported from Zemax. This layout represents a simple periscope, ray-traced from input to output. The collimated input rays (top left) are double-folded by a sequence of two plane mirrors at an angle of 45° before being brought to a focus (bottom right) by a thin lens. In between the mirrors, a generic surface can be customized to impose a wavefront aberration.

This periscope layout can be propaedeutic to illustrating the input structure in PAOS. To this aim, let us refer the reader to the input .ini file provided in Listing 2.6. This input file showcases how to set up all the various components to model the same periscope optical system, from input to output, including its off-axis geometry. Therefore, this learning material also demonstrates the potentiality of PAOS in modelling advanced optical systems.

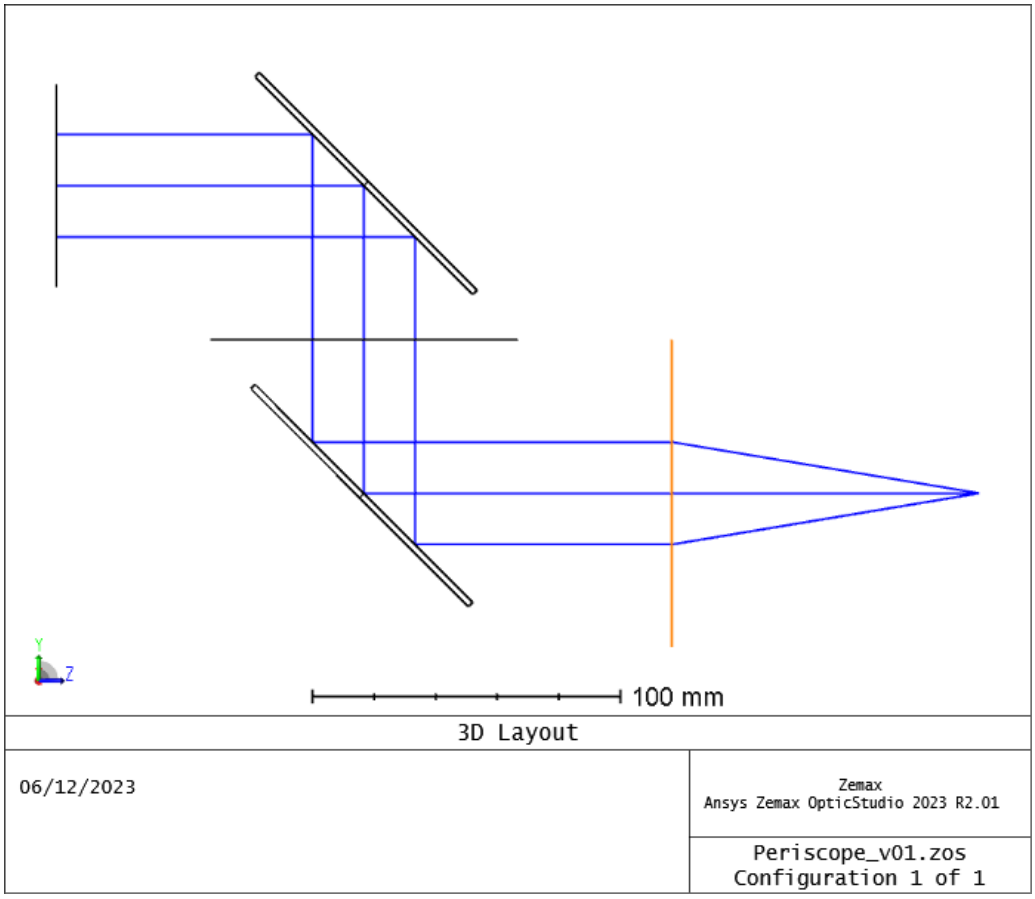


Figure 2.2. Zemax 3D layout of a simple periscope to be modeled using PAOS.

```
[general]
project = Periscope
version = 1.0
grid_size = 512
zoom = 4
lens_unit = m
tambient = 20.0  # C
pambient = 1.0  # atm

[wavelengths]
w1 = 1.0
w2 = 3.0

[fields]
f1 = 0.0,0.0  # Slope: Sagittal, Tangential
f2 = 0.0,0.001

[lens_01]
surfacetype = INIT  # Input Beam Initialization
aperture = elliptical aperture,0.050,0.050,0.0,0.0  ## Aperture Type, X-Half Width,
     Y-Half Width, X-Decenter, Y-Decenter

[lens_02]
surfacetype = Standard  # Move to M1
thickness = 0.100

[lens_03]
surfacetype = Coordinate Break  # LOS tilt 1
par3 = -45.0

[lens_04]
surfacetype = Standard  # M1
material = MIRROR
save = True
stop = True
aperture = elliptical aperture,0.050,0.033,0.0,0.0

[lens_05]
surfacetype = Coordinate Break  # LOS tilt 2
par3 = -45.0

[lens_06]
surfacetype = Standard  # Move to Zernike surface
thickness = -0.050

[lens_07]
surfacetype = Zernike  # Z1
zindex = 0,1,2,3,4,5
z = 0.0,0.0,0.0,0.05,0.0,0.0
par1 = 3.00  ## Wavelength
par2 = standard  ## Ordering
par3 = False  ## Normalization
par5 = x  ## Origin for Angle Measurement

[lens_08]
surfacetype = Standard  # Move to M2
thickness = -0.050
```

```
57
58 [lens_09]
59 surfacetype = Coordinate Break  # LOS tilt 3
60 par3 = 45.0
61
62 [lens_10]
63 surfacetype = Standard  # M2
64 material = MIRROR
65 aperture = elliptical aperture,0.050,0.050,0.0,0.0
66
67 [lens_11]
68 surfacetype = Coordinate Break  # LOS tilt 4
69 par3 = 45.0
70
71 [lens_12]
72 surfacetype = Standard  # Move to L1
73 thickness = 0.100
74
75 [lens_13]
76 surfacetype = Paraxial Lens  # L1
77 thickness = 0.100
78 par1 = 0.100  ## Focal Length
79
80 [lens_14]
81 surfacetype = Standard  # Image Plane
82 save = True
```

Listing 2.6. Example input file for PAOS. Remove comments to use.

The .ini file is organized into various sections, each denoted by a header enclosed in square brackets. The `[general]` section contains general information about the simulation, including the project name, version, grid size, and zoom level. This section includes parameters for the ambient temperature (`tambient`) and pressure (`pambient`). To ensure consistency, PAOS assumes meters, micrometers, and degrees, respectively, for the unit of measurement of lens parameters, wavelengths, and angles.

In the `[wavelengths]` section, the input wavelengths used in the simulation are specified. Each wavelength is assigned a unique identifier, such as `w1` and `w2` in the example provided. The `[fields]` section contains information about the *field slopes* in the sagittal and tangential directions. Field slopes represent the tangent of the angle between the rays and the optical axis. In the PAOS code, these field slopes are set equal to the angles themselves under the paraxial approximation. The field slopes are identified with labels like `f1` and `f2` in the provided example.

The optical chain description begins with the `[lens_01]` header, which represents the initial surface and is always of type `INIT`. Following this, subsequent headers are used to define each optical surface in sequential order. There are several supported surface types, including `Coordinate Break`, `Standard`, `ABCD`, `Zernike`, and `Paraxial Lens`. Each optical surface is defined by a set of parameters, such as surface type, curvature, material, shape and position of the aperture, and thickness. The thickness parameter specifies the propagation distance to the next surface. Additional parameters can be utilized to place an aperture stop, ignore the surface, or save the current wavefront properties. The user can refer to the documentation for a comprehensive list of parameters and their description.

In Listing 2.6, the input beam is initialized on-axis at `[lens_01]` and has a radius

of 50 mm; the next surface, [lens_02], propagates the collimated beam by 100 mm to the first fold mirror, M1. Before M1, [lens_03] tilts the beam about the sagittal axis clockwise by 45°. M1 is defined at [lens_04] with an elliptical aperture that vignettes the incoming beam and planar geometry. M1 also normalizes the wavefront to unit energy. [lens_05] and [lens_06] perform another beam tilt of 45° and a propagation of 50 mm, respectively. [lens_07] is a Zernike surface and imposes a wavefront error of 50 nm RMS astigmatism on the beam. Then, [lens_08] propagates the beam to M2 by 50 mm. M2 is preceded by [lens_09], which tilts the beam about the sagittal axis counter-clockwise by 45°. [lens_10] is M2, the second fold mirror, plane and circular. After M2, [lens_11] performs the last tilt that restores the original propagation direction. The next surface is [lens_12], from where the beam is propagated to the paraxial lens at [lens_13], described by a focal length of 100 mm. Finally, the lens brings the beam to its focus, at [lens_14]. When using this input file, the wavefront properties are saved in the HDF5 file at M1 and the image plane.

This example demonstrates all essential components in PAOS. It, along with other examples provided in the documentation, serves as a valuable starting point for users to understand and explore the capabilities of the software. It also showcases the standard naming conventions and syntax adopted by PAOS, similar to those used in commercial optical computer-aided design (CAD) software like Zemax. This approach minimizes the learning curve and allows users to quickly leverage their existing knowledge and experience in optical design when working with PAOS.

The input file is parsed using the core.parseConfig method, which utilizes the ConfigParser class from the Python standard library. This method parses the input file and stores the data in a dictionary. It returns a tuple consisting of simulation parameters and the optical chain list. Each entry in the list is a dictionary that represents an optical surface in the .ini file. The dictionary contains the parameters and ABCD matrices of the optical surface, estimated at the given wavelength. It should be noted that this information is relevant only for diffractive components, as their inclusion results in wavelength-dependent optical paths. The output of this method is then used for the POP simulation.

2.1.2.2 GUI

PAOS's generic input system is complemented by a GUI, which can be accessed through the paosgui command in the terminal. The GUI serves as a user-friendly interface, providing an intuitive way for users to interact with the essential functionality of the PAOS code – the POP simulation. One notable advantage of the GUI is that it eliminates the requirement for users to be versed in Python or any programming language. This accessibility makes PAOS suitable for optical engineers, scientists, and even students who may not be programming experts. By providing a convenient and user-friendly GUI, PAOS (i) extends its user base, (ii) enables users to focus on the scientific analysis while seamlessly interacting with the code and input file, and (iii) enhances usability and appeal compared to other open-source POP codes.

The GUI is implemented using the PySimpleGUI[1] module in Python. This module aims to bridge the gap between software developers and end-users by offering high-level access to the Tkinter library[2] (and other backends), which is built-in in Python. With

[1] https://www.pysimplegui.org/en/latest/, insert-version-here
[2] https://docs.python.org/3/library/tk.html

this module, creating GUI elements such as windows, buttons, and input fields becomes straightforward, requiring only a few lines of code. PAOS provides a GUI that emphasizes simplicity, featuring a minimal number of elements and a straightforward interface.

To illustrate the GUI, let us refer to Figure 2.3, which presents a series of screenshots displaying its various tabs. Users can either load an existing input file (we use the same one from Section 2.1.2.1) or create a new one using the GUI as an editor. The GUI loads instantly and the landing tab is `General` (screenshots 1 and 2), which contains two different sections corresponding to the `[general]` and `[wavelengths]` sections in the input file. Users have the flexibility to insert wavelengths for the simulation either manually or by pasting a column of values. The second tab, labeled as `Fields` (screenshot 3), corresponds to the `[fields]` section. The next tab, `Lens Data` (screenshot 4), displays a table that represents the optical chain from input to output. Each column corresponds to a keyword that can be selected from a dropdown menu or entered in an input field, such as `SurfaceType` or `Thickness`. The figure only shows part of the table, which comprises many more columns, including `Aperture` (with its own dropdown menu in a collapsible window) and the columns from `par1` to `par8`. This table allows users to create and edit the input file section `[lens_01]` and following in an interactive and intuitive way, with event-driven programming (e.g. keyboard navigation) and dynamic headers (e.g. `par1` to `par4` change to A_x, B_x, C_x, and D_x for an ABCD surface). This feature improves the user experience and facilitates efficient editing of the input file. The user can access a special window (`Zernike Setup`) for viewing and editing the Zernike coefficients (screenshot 5) from the `Lens Data` tab by selecting 'Zernike' as the `SurfaceType`. In it, the user can enter coefficients individually under 'Z' or paste a column of values for convenience. The window shows the index as well as the radial and azimuthal order of the coefficients, according to the chosen ordering. This window enables more convenient and precise editing of Zernike coefficients than pasting them in the input file. The GUI saves any modification by the user in a temporary .ini file that corresponds to the current status of the GUI. The user can click 'Save' to finalize the changes and optionally overwrite the previous file. The GUI can be exited at any time, but the temporary .ini file will be deleted.

The user can run a diagnostic ray-tracing (optional) and a POP simulation of the optical system in the `Launcher` tab (screenshot 6) after modifying the simulation parameters or loading the input file. The wavelength and field for the simulation can be selected from dropdown menus. A progress bar indicates the simulation status. The results can be saved to an `HDF5` file and the squared amplitude of the wavefront can be plotted at the chosen surface (with the chosen scale), and saved to disk. This plot can be displayed in a dedicated frame that shows its cross-sectional and 2D views.

This description covers the main GUI functionalities, but there is another tab where the user can run a predefined set of Monte Carlo (MC) simulations, including running the POP simulation for different realizations of the optical system Wavefront Error (WFE). We refer to the documentation for more details on this tab. The GUI is designed to be extensible and can accommodate additional functionality based on user feedback. However, users who want more flexibility and customization may prefer to interact with the Python code directly. This allows them to perform a wider range of tasks, such as implementing new optical surface types or different MC simulations.

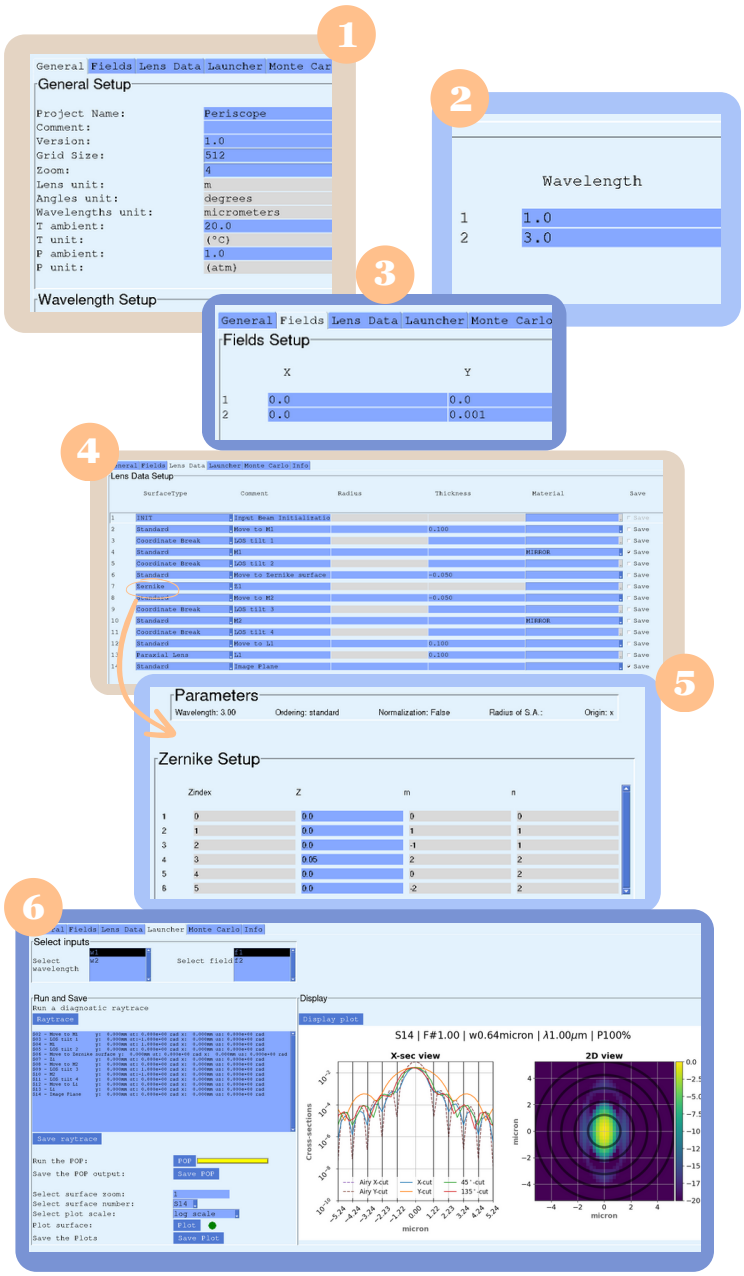


Figure 2.3. Screenshots of the GUI tabs. (1) and (2) show the `General` tab, which contains sections for the `[general]` and `[wavelengths]` parameters in the input file. (3) shows the `Fields` tab, which corresponds to the `[fields]` section. (4) shows the `Lens Data` tab, which displays a table that represents the optical chain from input to output. (5) shows the `Zernike Setup` window, which can be accessed from the `Lens Data` tab by selecting 'Zernike' as the `SurfaceType`. This window allows users to view and edit the Zernike coefficients. (6) shows the `Launcher` tab, which allows users to run a diagnostic ray-tracing and the POP simulation.

2.1.2.3 Physical Optics Propagation

PAOS estimates the complex-valued wavefront at any distance along the optical axis using the scalar Fresnel diffraction theory (Goodman, 2005). The Fresnel approximation requires i) aperture sizes to be significantly larger than the wavelength, ii) modest numerical apertures and iii) thin optical elements. Because the Fresnel and the paraxial approximations are equivalent, PAOS can estimate the wavefront propagation through optical layouts that can be described paraxially. The propagation of the wavefront is implemented using near-field and far-field propagators, required to guarantee adequate sampling of the fields while maintaining computational efficiency. This is a well-consolidated algorithm, fully described by Lawrence (Lawrence et al., 1992) in 1992. The near- and far-field propagators are combined to estimate wavefronts traveling from near-to-near, near-to-far, far-to-far, and far-to-near fields. The propagators to use are selected by studying a pilot beam, an ideal Gaussian beam whose propagation through an arbitrary optical system can be computed analytically in paraxial approximation. The pilot beam can be propagated through on-axis as well as off-axis optical layouts. In practice, the *xyz* reference frame used to estimate the field has the *z*-axis parallel to the direction of propagation of the pilot beam, and the paraxial quantities are estimated in this reference frame at all positions along the optical axis.

The paraxial behavior of an arbitrary optical system can be conveniently studied using a matrix method, often called the ABCD method. A ray is represented by a ray vector $\mathbf{v} = (y, u)$, where y is the ray height and u is its slope. In this formalism, the propagation through an optical element is described by a matrix P

$$\begin{pmatrix} y' \\ u' \end{pmatrix} = \mathbf{v}' = P\,\mathbf{v} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} y \\ u \end{pmatrix} \tag{2.1}$$

The matrix P can be factorized in a number of different ways; PAOS adopts the following

$$P = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\psi & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & n_1/n_2 \end{pmatrix} \begin{pmatrix} M & 0 \\ 0 & 1/M \end{pmatrix} \tag{2.2}$$

From left to right, the matrices represent a translation T, a lens with optical power ψ, a boundary between two media with n_1 and n_2, and an optical magnification M. These four variables can be obtained from the elements of the matrix P

$$\begin{aligned} M &= \tfrac{AD-BC}{D} \\ n_1/n_2 &= MD \\ T &= \tfrac{B}{D} \\ \psi &= -\tfrac{C}{M} \end{aligned} \tag{2.3}$$

The first three steps (power, medium change, magnification) are fast to execute because they correspond to a geometrical transformation of the wavefront and only the translation requires the application of the Fresnel theory (see Lawrence, 1992, section IV Lawrence et al. (1992)). With this formalism, PAOS can model any optical system involving reflective flat or powered mirrors, thin lenses as well as thick dioptres, prisms, etc.

The propagation of optical waves through any well-behaved system can be separated into geometric aberration calculations and propagation in homogeneous media (Lawrence et al., 1992). Geometric aberrations are discussed in Section 2.1.2.4; this section mainly describes the assumptions and limitations in the treatment of wavefront propagation in PAOS.

PAOS employs the Fresnel approximation for analyzing wave propagation, which is a more feasible alternative than the Rayleigh-Sommerfeld solution for complex systems. The Rayleigh-Sommerfeld solution is a fundamental approach in wave optics that provides a rigorous analysis of wave propagation, but it is impractical for most applications because of its high mathematical and computational demands. The Fresnel approximation is derived from the Rayleigh-Sommerfeld solution by assuming paraxial conditions, where the angles involved are small, and the distance between planes is much larger than the wavelength. The conditions can be summarized as (Lawrence et al., 1992):

1. aperture sizes significantly larger than the wavelength;
2. modest numerical apertures;
3. thin optical elements.

The Fresnel approximation has some advantages and disadvantages that affect its applicability and accuracy. A major advantage of this approximation is that it enables the use of the Fast Fourier Transform (FFT) in numerical computation, which greatly reduces the computational time and memory requirements. As a result, most optical systems can be efficiently modeled and analyzed using this method. However, this approximation cannot model systems that violate paraxial conditions. Moreover, aberrations cannot be independently evaluated and need to be input separately by the user.

Therefore, PAOS can treat all types of systems to the accuracy of small angle, scalar Fresnel diffraction theory, but it may not be suitable for systems that require higher accuracy or involve non-paraxial conditions. In these scenarios, the Rayleigh-Sommerfeld solution or other advanced methods might be more appropriate for analyzing wave propagation. These topics are well-known in the literature (e.g. Goodman, 2005) and were briefly summarized here for completeness. For more details on the code implementation of wavefront propagation in PAOS, as well as practical examples, the reader is referred to the Appendix B.1 and the code documentation. Next, we present a distinctive feature of PAOS that enables it to model off-axis optical systems.

Unlike other publicly available POP codes such as PROPER, PAOS can model off-axis optical systems, where the optical axis does not coincide with the mechanical axis. This is achieved by combining Fresnel diffraction theory and paraxial ray-tracing. Fresnel diffraction theory requires propagation along the optical axis, but this axis is not fixed in an off-axis system. Therefore, PAOS uses paraxial ray-tracing to estimate the location and projected shape of the apertures (and obscurations) that may affect the wavefront with respect to the optical axis, before propagating the wavefront (example in Appendix B.1.11). This information allows the translation of physical apertures into ABCD matrices for an on-axis optical chain, which enables propagation of an off-axis optical system with minimal phase aberrations. In some cases, such as the prism, PAOS also uses ray-tracing to estimate the appropriate ABCD matrices that account for the refraction and dispersion effects. For more information on ABCD matrices implemented in PAOS, we refer the reader to Section 2.1.2.6.

The Ariel Telescope, which has an off-axis Cassegrain design, is the main system that inspired the development of PAOS. Thanks to the combination of Fresnel diffraction and paraxial ray-tracing, PAOS was able to accurately simulate the wavefront propagation through the telescope and provide useful information on optical performance and tolerances. However, a detailed description of the analyses done for the Ariel Telescope is beyond the scope of this thesis and will be presented elsewhere.

2.1.2.3.1 Propagation loop The propagation loop is the main function of PAOS that performs the POP simulation over each element of the optical system according to the configuration file which must be already parsed (Section 2.1.2.1). This section describes the sequential flow diagram to execute the entire POP simulation, implemented in the `core.run` method.

First, PAOS initializes the beam at the center of the aperture. Then, it initializes the propagation ABCD matrix. Once the initialization is complete, the simulation run is completed in a single *for* loop over all the optical elements. For each element (step), PAOS applies sequentially different actions according to a series of if-clauses, which are all evaluated before going to the next step, as shown in Figure 2.4. The if-clauses and actions

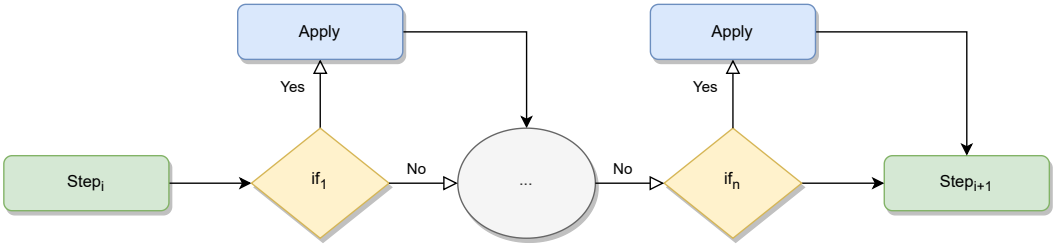


Figure 2.4. Sequential flow diagram that is repeated over all the optical elements in the system.

are in the specific order below.

1. Check coordinate break; if true, transform the wavefront coordinates.
2. Check aperture; if true, apply (and save) the given mask to the beam.
3. Check stop; if true, normalize the current wavefront to unit energy.
4. Check aberration; if true, add a phase term represented by a Zernike expansion to the wavefront.
5. (*) Save the wavefront properties.
6. Check magnification; if true, rescale the sampling and the relevant beam parameters.
7. Check medium change; if true, adjust the relevant beam parameters.
8. Check lens; if true, apply a paraxial lens phase term to the wavefront.
9. Check propagation by thickness; if true, select the appropriate propagation primitive and propagate the wavefront by a given distance.
10. (*) Update beam vector and ABCD matrix (and save it).

Note that actions with an asterisk are not in if-clauses and are always applied.

To illustrate a POP simulation using PAOS, we show the wavefront propagation of the optical system of the Hubble Space Telescope (HST) in Figure 2.5. The HST input file uses open-source information and is included in the PAOS package as an example. We describe this input file in more detail in Section 2.2. The figure displays the wavefront intensity at 1 μm for different locations in the optical path: the input pupil (left), the secondary mirror (center), and the Point Spread Function (PSF) at the telescope's focal plane (right).

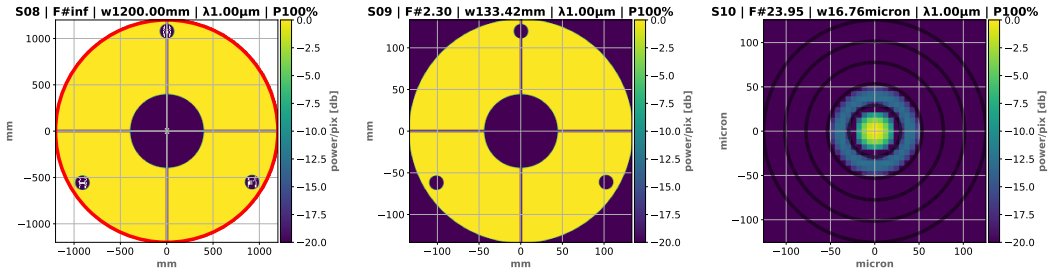


Figure 2.5. Wavefront propagation of the optical system of the Hubble Space Telescope (HST) at 1 µm. The panels show the squared amplitude of the wavefront at different positions, and the titles report information on the current surface number, beam focal ratio, Gaussian beam width, beam wavelength, and the fraction of total transmitted power. *Left panel*: the input pupil, which is uniformly illuminated and apodized by the primary mirror circular pads (1, 2, and 3), the circular obscuration of the secondary mirror, and the rectangular vanes (vertical and horizontal) that block part of the incoming light. *Center panel*: the beam at the secondary mirror (M2). *Right panel*: the Point Spread Function (PSF) at the HST telescope focus. Black rings are positioned at the first five dark rings for a diffraction-limited Airy PSF for reference. The color scale shows the power in each over-sampled pixel in decibels (dB).

2.1.2.4 Aberrations

One of the most important applications of optical modeling methods is to properly account for aberrations. Aberrations can arise in different ways, such as inherent optical configuration design and manufacturing errors (e.g. Lawrence et al., 1992; Welford, 1986; Mah, 2013). In this section, we describe how PAOS handles aberrations.

PAOS implements low-spatial-frequency aberrations (geometric aberrations), which can result from various factors, including gravity, mirror manufacturing, mounting, and alignment errors, and can affect the shape and large-scale spatial features of the PSF in distinctive ways. In contrast, mid- or high-spatial frequency errors (surface roughness) scatter light to larger angles, reducing the Strehl ratio without significantly altering the PSF shape. These aberrations can be statistically characterized relative to the spatial scales of interest using a parameterized Power Spectral Density (PSD) specification (e.g. Church and Takacs, 1991). PAOS does not support these aberrations natively at present; however, the user can add custom aberration maps when using PAOS as a library. We may include this capability in the next version of PAOS based on user feedback.

2.1.2.4.1 Zernike polynomials PAOS models a geometric aberration as a superposition of Zernike polynomials, up to a specified radial order. This is implemented in the Python class `Zernike`. The function describing an arbitrary wavefront in polar coordinates $W(\rho, \theta)$ can be expanded in terms of a sequence of Zernike polynomials (Lakshminarayanan and Fleck, 2011) as:

$$W(\rho, \theta) = \sum_{n,m} C_n^m Z_n^m(\rho, \theta) \tag{2.4}$$

where C_n^m are the coefficients of the Zernike polynomial $Z_n^m(\rho, \theta)$, and n and m are the radial and azimuthal numbers, respectively. The radial coordinate is normalized so that it is equal to 0 at the center of the pupil and equal to 1 at the edge of the pupil. Let x_{pup} and y_{pup} be the physical coordinates of a pupil, and let a and b be, respectively, the semi-major and

semi-minor axes of an elliptical pupil. Then:

$$\rho^2 = \frac{x_{pup}^2}{a^2} + \frac{y_{pup}^2}{b^2} \tag{2.5}$$

This formalism captures both circular and elliptical pupils. The first three terms in Equation 2.4 describe trivial aberrations (Piston, Tip, and Tilt) and are usually neglected. Note that, due to orthogonality, the Zernike coefficients are independent of the number of polynomials in the sequence.

PAOS can generate ortho-normal polynomials and orthogonal polynomials; orthonormal polynomials are normalized to RMS = 1, and each coefficient of the expansion represents the RMS Wavefront Error (WFE) of that specific mode. Orthogonal polynomials are generated as described in (Lakshminarayanan and Fleck, 2011):

$$Z_n^m = \begin{cases} R_n^m(\rho)\, cos(m\phi) & m \geq 0 \\ R_n^{-m}(\rho)\, cos(m\phi) & m < 0 \\ 0 & n - m \text{ is odd} \end{cases} \tag{2.6}$$

The radial polynomial is estimated using the Jacobi polynomial expression as in (Lakshminarayanan and Fleck, 2011, Eq. 14) and is normalized to 1 at the unit radial coordinate, that is, $R_n^m(\rho = 1) = 1$. Equation 2.6 can be expressed as:

$$\left\langle [Z_n^m(\rho, \phi)]^2 \right\rangle = 2\frac{n+1}{1+\delta_{m0}} \tag{2.7}$$

where δ_{m0} is the Kronecker delta ($\delta_{m0} = 0$ for m $\neq$0) and the average operator $\langle\rangle$ is intended over the pupil. Note that when a set of Zernike coefficients, D_n^m, are obtained from an interferometric measurement of a mirror, they can be scaled to match the desired RMS of the WFE at the pupil. This is useful for modeling realistic mirror aberrations.

The `Zernike` class supports four different orderings: ANSI, Noll, Fringe, or Standard (e.g. Born et al., 1999). For any ordering, it can convert n and m into the corresponding coefficient index and vice versa. This is particularly useful for collaborative work, since different teams often use different orderings, and doing a manual conversion is a cumbersome task.

PAOS can convert the Zernike coefficients between any of the ANSI, Noll, Fringe, or Standard orderings, ensuring compatibility and interoperability.

To illustrate how PAOS can model different types of aberrations, we present two examples below. Listing 2.7 shows an example of using PAOS as a library to simulate a 100-nm RMS wavefront aberration of defocus (4th coefficient in ANSI ordering) for an optical system with an input pupil similar to Ariel's. Figure 2.6 reports simulated PSFs at the Ariel Telescope exit pupil for a set of large wavefront aberrations: primary astigmatism, defocus, coma, and trefoil[3]. This figure illustrates that each optical aberration affects the final PSF shape and optical quality for the Ariel application in a distinctive way.

[3]The interested reader can verify that as the input parameters, aberration, and the implemented $f_\#$ are comparable to those used in Listing 2.7, the defocused PSF is similar.

```python
import numpy as np
from paos.classes.wfo import WFO
wfo = WFO(beam_diameter=1.1, wl=0.55e-6, grid_size=512, zoom=4)
wfo.aperture(xc=0.0, yc=0.0, hx=0.55, hy=0.365, shape="elliptical")
wfo.make_stop()
Z_index = np.array([0, 1, 2, 3,      4, 5])
Z_value = np.array([0, 0, 0, 0, 100e-9, 0])  # defocus: 100 nm RMS WFE
wfo.zernikes(index=Z_index, Z=Z_value, ordering="ansi", normalize="False"
    , radius=wfo.wz, origin="x")
wfo.lens(lens_fl=12)
wfo.propagate(dz=12)
psf = wfo.amplitude ** 2  # aberrated PSF at image plane
```

Listing 2.7. Zernike defocus aberration. We expand the wavefront function in terms of Zernike polynomials with ANSI ordering. The input pupil is elliptical, so we use orthogonal polynomials by setting *normalize* to false. The first three terms are trivial aberrations and are neglected. The fourth term is the defocus coefficient, set to 100 nm RMS. The radius parameter is the current Gaussian beam width, and the origin parameter is the x-axis. We propagate the wavefront to the image plane and calculate the PSF.

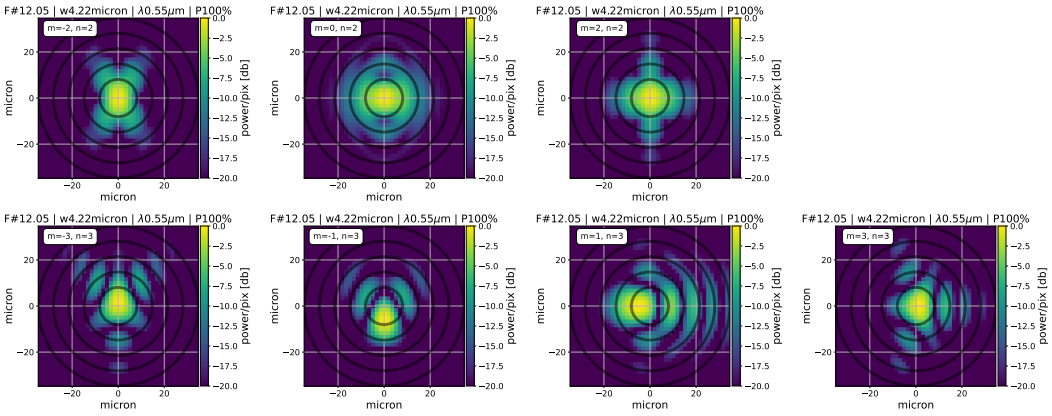


Figure 2.6. PSFs at the Ariel Telescope exit pupil simulated for different aberrations and same 100-nm Root Mean Square (RMS) Wavefront Error (WFE). The PSFs are reported in ascending ANSI index (Born et al., 1999). From the top left: oblique astigmatism, defocus, vertical astigmatism, vertical trefoil, vertical coma, horizontal coma, and oblique trefoil. The color scale represents the power per pixel in decibels (dB), normalized to the maximum value in the array. For reference, five black rings show the locations of the first zeros of the corresponding Airy PSF.

2.1.2.5 Refractive elements

PAOS can simulate both reflective and refractive elements, such as mirrors, dichroics, and prisms. This section describes how it accounts for the dispersion, temperature, and pressure effects on the refractive index of optical materials. This is another feature that distinguishes PAOS from other POP codes.

The dispersion of light can be explained by applying the electromagnetic theory to the molecular structure of matter. The properties of optical materials depend on wavelength through the index of refraction. The chromatic variation in the paraxial image formation properties of a system is generally of most interest (Lawrence et al., 1992). For instance, the wavefront propagating through the Ariel optical system is transmitted through dichroics and prisms. At each limiting surface, there is a change in the refractive index due to the different optical materials. Only for mirrors, the refractive index is constant between the mirror space and the media before.

In PAOS, the optical materials need to be specified in the input file (see Section 2.1.2.1) by using the *material* keyword for each surface. Note that blanks are treated as air. The catalog of supported optical materials includes:

- CAF2 (calcium fluoride) (Bass et al., 2009);
- SAPPHIRE (mainly aluminum oxide) (Rogatto et al., 1993);
- ZNSE (zinc selenide) (Bass et al., 2009);
- BK7 (borosilicate crown glass) [Adept-infrared catalog];
- SF6 (a dense-flint glass) [Schott catalog];
- SF11 (a dense-flint glass) [Adept-infrared catalog];
- BAF2 (barium fluoride) (Bass et al., 2009).

"MIRROR" is a valid key that can be entered for a mirror, although this name is not present in the catalog. We acknowledge that this catalog is largely Ariel-oriented and we may expand it in a future PAOS release upon user feedback.

Once the optical materials are defined, PAOS can compute their refractive index as a function of wavelength, temperature, and pressure. For more details, please refer to Appendix B.3 and the code documentation.

Note that the current PAOS implementation is not intended to perform thermal analysis, which must be done externally, and the results of which must be input by the user. As such, it allows one to specify only one pressure and temperature for the optical system under consideration; in addition, PAOS does not implement changes in the size of glasses with temperature from the value specified in the input file.

2.1.2.6 ABCD matrix method

In the past sections, we have introduced some of the peculiar features of PAOS that distinguish it from other POP codes, such as in Section 2.1.2.3, where we discussed how PAOS can model off-axis optical systems using paraxial ray-tracing and ABCD matrices. Here, we discuss how the ABCD matrix method is implemented in PAOS and what are its advantages and limitations.

PAOS implements the Paraxial theory described in Lawrence et al. (1992), which applies to the infinitesimal thread-like region around the optical axis where all the angles involved are small enough to be approximated by their sines and tangents (Smith, 2000). This approximation allows for estimating the image position and sizes in a real optical system by comparing them with the ideal paraxial image. Also, a well-corrected optical system will have only minor deviations from these first-order expressions, and the analysis will be greatly simplified compared to the exact trigonometric relations.

ABCD matrix theory was developed to break down the paraxial behavior of an optical system into a series of elementary transformations, which can be easily implemented in a computer program. Using ABCD matrices, the propagation of a paraxial beam is performed by representing each optical surface as a 2-by-2 matrix that transforms the beam parameters – height and slope of the rays – from one surface to the next. The ABCD matrix (less colloquially, ray transfer matrix) is a characteristic of each optical element. This method can be used to describe the paraxial behavior, the properties of the Gaussian beam, and the overall diffraction behavior of any given optical system described by a sequence of optical surfaces. In general, any optical system can be considered a black box described by an effective ABCD matrix:

$$\text{Optical system} \leftarrow \begin{pmatrix} A & B \\ C & D \end{pmatrix}_{eff} \tag{2.8}$$

This black box and its matrix may be factored into a set of four elementary non-commuting operations or *primitives* (Lawrence et al., 1992):

1. magnification change;
2. change of refractive index;
3. divergence (“thin lens”);
4. translation of distance (“thickness”).

Assuming that the apertures in the intervening surfaces are sufficiently large, we can neglect the diffraction propagation effects in all steps except the single propagation step, which is the most computationally intensive one. We can estimate the parameters of the effective ABCD matrix as follows:

$$\begin{aligned} M &= \tfrac{AD-BC}{D} \\ n_1/n_2 &= MD \\ T &= \tfrac{B}{D} \\ \psi &= -\tfrac{C}{M} \end{aligned} \tag{2.9}$$

where T, ψ, n_1/n_2, and M are the effective thickness, power, refractive index ratio, and magnification of the black box system, respectively. These parameters are not the same as the thickness, power, refractive index ratio, and magnification of the *actual* optical system or its components. With these definitions, we can obtain the effective focal length as $f_{eff} = \frac{1}{\psi M}$.

Using the Paraxial theory and the ABCD matrix method, PAOS can analyze the paraxial behavior of any optical system (or subsystem) as a black box, as long as it has the effective ABCD matrix of that system (or subsystem).

This is a flexible feature of PAOS that distinguishes it from other POP codes and enables the modeling of optical systems that may not have a simple analytical description or a

standard optical design software representation. It also allows for using placeholder ABCD matrices in early design stages, when the exact parameters of the optical elements are not yet known. The input file format and syntax for specifying the ABCD matrix are explained in Section 2.1.2.1.

In the following section, we demonstrate using the Python class `ABCD` in PAOS to estimate the appropriate ABCD matrices for a prism. The reader is referred to the Appendix B.2 and the code documentation for further examples of how to implement different ABCD operators using this class as a library.

2.1.3 Prism

The prism changes the slope and the magnification of an incoming beam.

Following (Taché, 1987), after some algebra, the ABCD matrix for the tangential transfer can be rewritten as:

$$P_t = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{2.10}$$

where

$$\begin{aligned} A &= \frac{cos(\theta_2)cos(\theta_4)}{cos(\theta_1)cos(\theta_3)} \\ B &= \frac{L}{n}\frac{cos(\theta_1)cos(\theta_4)}{cos(\theta_2)cos(\theta_3)} \\ C &= 0 \\ D &= 1/A \end{aligned} \tag{2.11}$$

where n is the refractive index of the prism, L is the geometrical path length in the prism, and the angles $\theta_{1..4}$ are as described in Figure 2.7, copied from (Taché, 1987).

The sagittal transfer, instead, is simply a propagation by the thickness $t = L/n$.

Listing 2.8 lists the steps of the computation of the tangential and sagittal ABCD matrices for an optical component with tilted flat interfaces such as a prism.

```
import numpy as np
from paos.classes.abcd import ABCD
thickness = 20  # mm
n = 1.5
theta_1 = np.deg2rad(60)  # modify angles as necessary
theta_2 = np.deg2rad(-30)
theta_3 = np.deg2rad(-20)
theta_4 = np.deg2rad(45)
A = np.cos(theta_2) * np.cos(theta_4) / (np.cos(theta_1) * np.cos(theta_3
    ))
B = np.cos(theta_1) * np.cos(theta_4) / (np.cos(theta_2) * np.cos(theta_3
    )) * thickness / n
C = 0.0
D = 1.0 / A
abcdt = ABCD()
abcds = ABCD()
abcdt.ABCD = np.array([[A, B], [C, D]])
abcds.ABCD = np.array([[1, thickness / n], [0, 1]])
(At, Bt), (Ct, Dt) = abcdt.ABCD  # tangential
(As, Bs), (Cs, Ds) = abcds.ABCD  # sagittal
```

Listing 2.8. Computing the tangential and sagittal matrices for a prism with given parameters using the `ABCD` class. Two instances of ABCD class are created and assigned the tangential and sagittal matrices, respectively. The components of the matrices are calculated using Equation 2.11; they are assigned and extracted from the ABCD instances.

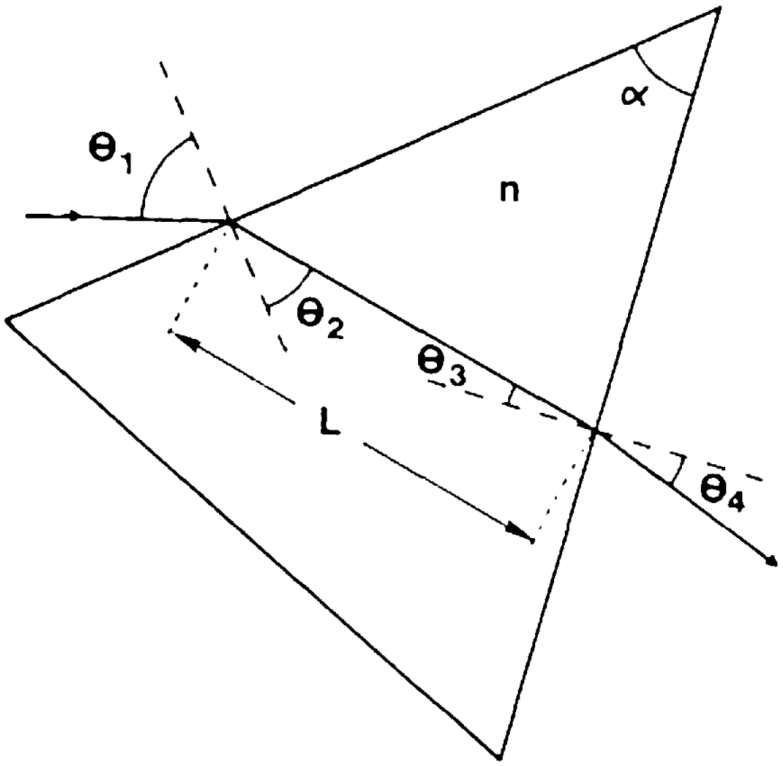


Figure 2.7. Propagation of a ray through a prism of refractive index n, apex angle α, and geometrical path length in the prism L. From Taché (1987, Figure 2).

2.2 Benchmark

We benchmarked PAOS against PROPER on the HST optical system, because PROPER is not designed to handle Ariel's, which (i) is off-axis and (ii) involves more complex elements than simple thin lenses (Krist, 2007), such as dichroics.

The description of the HST system used is the one provided in the "Hubble_simple.py" file in the PROPER package[4]. This description was translated into an input file for PAOS, "Hubble_simple.ini", included in the package under the "lens data" directory, for reproducibility. All simulation inputs have been matched (e.g., wavelength, grid size, `zoom`). We added a line in the PROPER HST routine to set the pixel subsampling factor used to antialias the edges of shapes. We set this value to 101 from the default 11 to improve the accuracy of the result to compare with the exact treatment given in PAOS.

We compared the resulting PSFs at the focal plane of the telescope, both in the central region and in the outer wings. The first benchmark is reported in Figure 2.8, showing the results for the PSFs at 1 μm. The plots in the upper row show the central region of the HST PSFs as computed with PAOS and PROPER, and their difference. No significant residuals were found, with sporadic outlier pixels showing deviations by less than 0.1 dB in regions corresponding to the PSF zeros due to small numerical errors. The middle and lower rows show a detailed view of the slices of the PSFs along the horizontal and vertical axes, and

[4]The PROPER source code and documentation can be downloaded at https://proper-library.sourceforge.net/

their differences. The signal curves show an almost perfect overlap, with negligible residuals, all corresponding to values lower than -50 dB from the PSF maximum in the far wings.

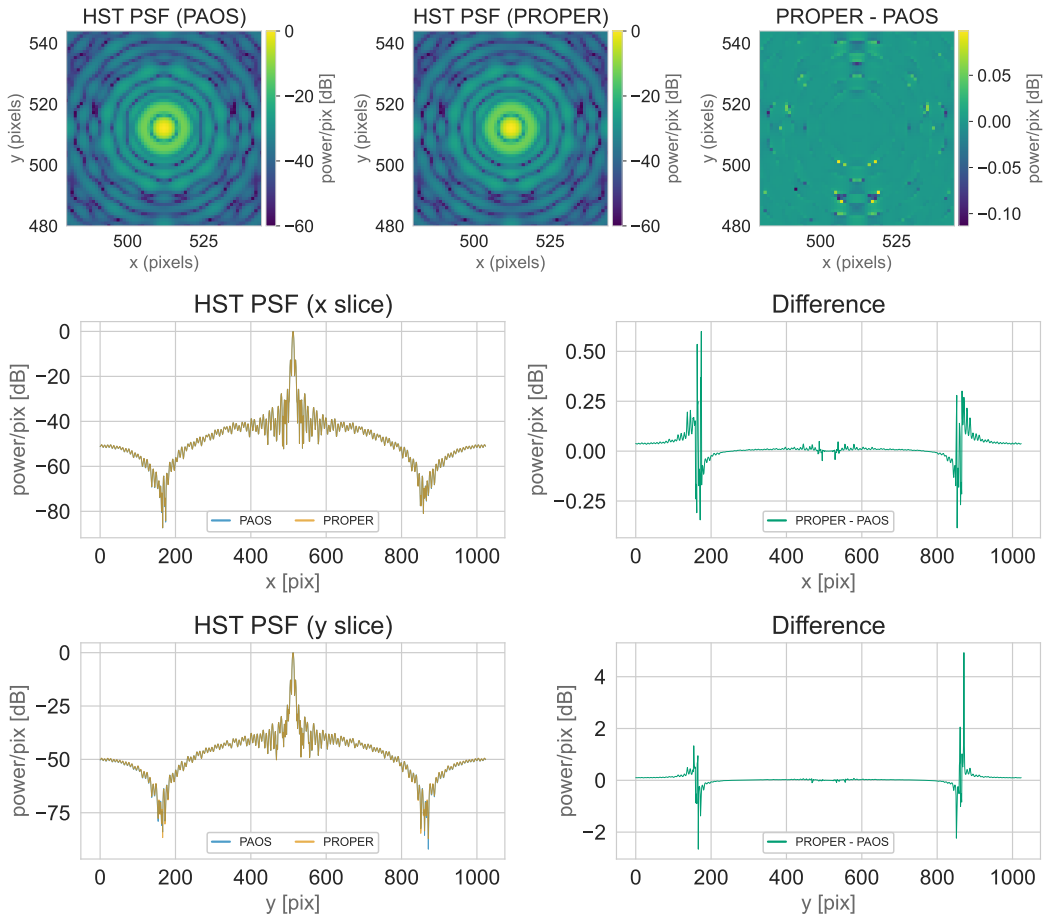


Figure 2.8. Top row: the central region of the HST PSF at 1 μm as estimated with PAOS (left) and PROPER (center) and normalized to the maximum value in the array. The axes are in oversampled pixels. The color scale represents the power per pixel in decibels (dB), with a lower cut-off at -60 dB for better visualization. The right panel reports the difference between the PSF computed with PROPER and PAOS in the same physical units. Middle and bottom rows: comparison between PSF slices along the x and y axis, respectively. The left column reports the slice values for both codes, whilst the right column reports their difference. The units are the same (power per pixel in dB) to highlight even the smallest discrepancies. As can be observed, these differences are negligible for powers $\gtrsim$-50 dB for the HST application.

Figure 2.9 reports the second benchmark; in this case, we simulate an aberrated PSF by HST described by a superposition of Zernike polynomials. At M2, we added 100 nm RMS (WFE) each for defocus, vertical astigmatism, and oblique astigmatism, totaling $\sigma \approx 173.2$ nm WFE. The simulation is performed at $\lambda = 1.0$ μm; therefore, using the Ruze formula (Ross, 2009), the Strehl ratio is $S = \exp(-2\pi\sigma/\lambda) \approx 0.3$. Consequently, the PSF is highly aberrated and the main lobe is spread over more pixels. Thus, we can validate the PAOS implementation of optical aberrations, and we have a larger region of high signal. The latter is especially useful for investigating aliasing errors, which tend to occur more severely where the distribution has the highest amplitude because the amplitudes of the signal and

the error add rather than the intensities (Lawrence et al., 1992).

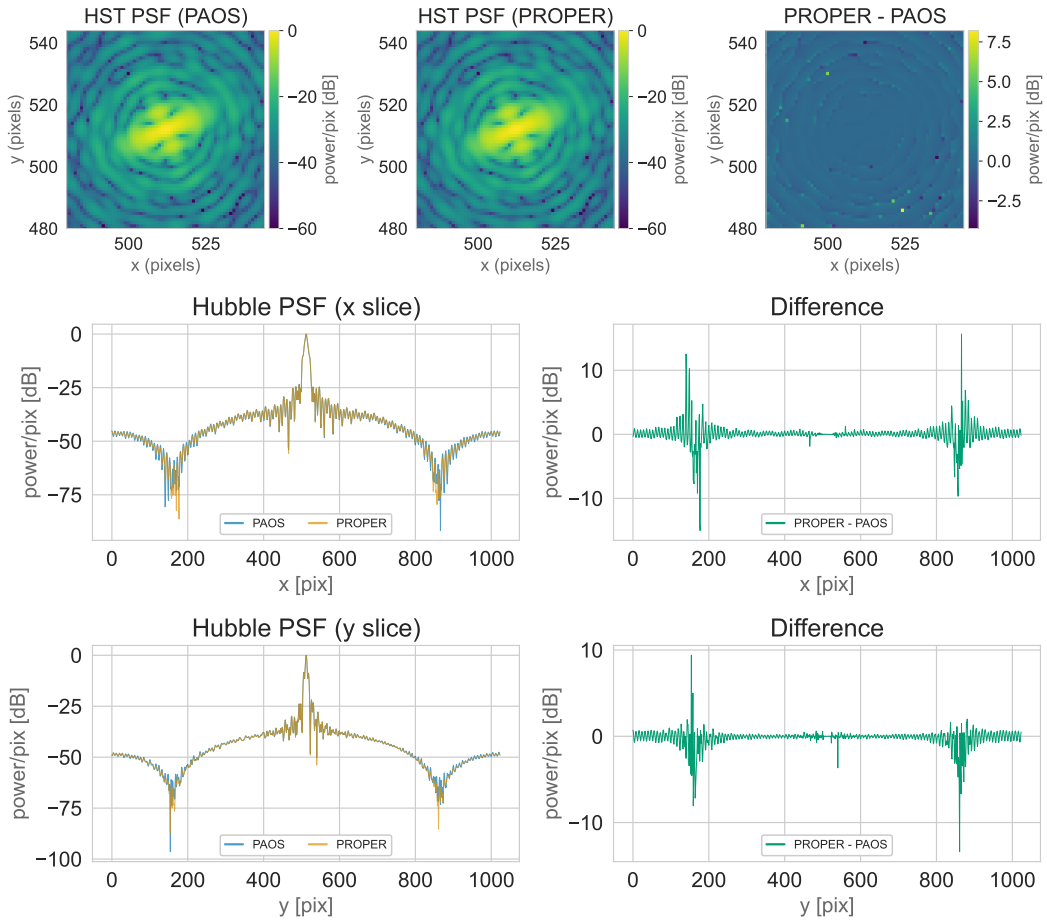


Figure 2.9. Same as Figure 2.8, but adding an optical aberration using Zernike polynomials: 100 nm RMS (WFE) for each of three low-order coefficients in the Zernike expansion: defocus and primary astigmatism (vertical and oblique), corresponding to the coefficients 4, 5, and 6 in the Noll ordering, respectively. The difference between the large-scale features of the PSFs is negligible. Locally, slightly "hotter" and "colder" pixels can be identified in the PSF wings, although, for powers $\gtrsim$-50 dB, this happens only sporadically. These minute numerical differences may be caused by the different treatment of aperture edges (exact for PAOS, sub-pixelled for PROPER), causing tiny aliasing errors.

We find that the differences between the aberrated PSFs are negligible and reach peaks of few dB only in the far wings. However, even in the central region, there is an increase in "hot" and "cold" pixels compared to the unaberrated case. These discrepancies are probably due to the different treatment of the edges of apertures and vanes in the optical system, causing small aliasing errors when not exact. However, they are so tiny that they can be safely neglected for the HST application.

2.3 From Radius of Encircled Energy to Wavefront

As an example PAOS application, this section aims to convey the work I led to convert a high-level optical requirement of the Ariel mission into information that can be used by the telescope manufacturer. The analysis started from developing a mathematical framework from first principles in optics and Zernike polynomials to link aberrations and PSF. Then, we sampled from solutions that were meeting a given criterion and compiled a table with compliant aberration realizations. To validate this table with an independent framework, we used PAOS and obtained results comparable with the expectations.

* * *

The optical performance requirements of the Ariel Payload and its subsystems are provided in terms of Encircled (or Enclosed) Energy (EE)[5], i.e. power within specified elliptical apertures at surfaces along the optical path. This is because Ariel is designed to be diffraction-limited at 3 μm, and in an aberrated system the energy in the Point Spread Function (PSF) spreads out. This is particularly relevant at FGS wavelengths: here the payload design relies on geometrical aberrations to broaden the PSF relative to the scale of the detector pixels. Otherwise, diffraction-limited optics would result in sub-sampled PSFs at FGS wavelengths, given the focal numbers implemented (see Section 4.1 for details).

Transmission spectroscopy observations do not require significant angular resolution: imaging quality is not critical, and the PSF can be aberrated. However, photons need to be collected in a sufficiently compact region of the focal plane. Translating this PSF "compactness" requirement into compliant aberration families is not trivial and a possible attempt is provided here. Additionally, the exact distribution of geometrical aberrations on the optical beam feeding the FGS instrument will be known later in the project.

From the manufacturer point of view, it is crucial to know the maximum allowable amplitude of aberrations from manufacturing the large all-Aluminum primary mirror, which poses fabrication challenges (Chioetto, 2022). To balance scientific requirements with manufacturing feasibility, the mirror fabrication process should be provided with optical surface shape requirements as relaxed as possible.

A measure of the degree of aberration of an optical system is given by the Wavefront Error (WFE). The WFE is defined in the geometric optics approximation as the optical path difference between the actual wavefront and an ideal one. Being easy to model with optical analysis software and measured precisely with interferometers, the WFE is a very commonly used performance parameter for optical systems.

A Wavefront Error (WFE) Root Mean Square (RMS) specification provides a useful budgeting tool as it can be directly allocated by setting an upper limit to optical surface errors. However, families of WFE functions meeting the RMS may still produce non-compliant PSFs[6]. This is because no one-to-one relation exists between the WFE RMS and PSF shape. Reducing WFE to its RMS discards information on the distribution of spatial frequencies, preventing unambiguous PSF (and therefore EE) determination. In other words, different WFE functions may result in the same RMS but different PSF shapes, and therefore different EEs. Therefore, the optical requirement is provided in terms of EE rather than WFE RMS, as verying EE ensures compliance regardless of the PSF shape.

[5] Perhaps it could be more precisely described as "enclosed" energy.
[6] ARIEL-SAP-PL-TN-005

In this section, we detail the analysis[7] done to investigate the shape of complex wavefronts sampled at the Telescope Exit Pupil (TEP) that are compatible with the EE requirement presented in Section 2.3.1, i.e. result in a PSF that is not too large. Somewhat serendipitously, later we also study the constraints to implement a PSF that is not too small (see Section 4.1 for details). Together, these analyses represent examples of practical application of the PAOS software described above to realistic scenarios requiring detailed knowledge of optical performance.

As shown in Section 2.3.2, we derive a mathematical relation between wavefront shape and requirements in terms of coefficients of Zernike polynomials, a sequence of polynomials that are orthogonal on the unit disk. We provide 1000 simulated wavefronts that comply with the requirements, sampled at the TEP[8]. The mathematical constraints on wavefront shape detailed below, and the realizations provided can be used to assess the performance of the Ariel subsystems, including Common Optics, AIRS and FGS. Notably, this preliminary attempt has led to further optical tolerancing analyses, described elsewhere[9].

2.3.1 Requirement at Telescope Exit Pupil

This section reports the top-level optical performance requirement for achieving the Ariel science goals. The allocated budget specifies an EE for each instrument channel that the telescope and payload optics must comply with. Table 2.2 summarizes the EE requirements at the TEP as elliptical apertures capturing 83.8% of the optical power.

2.3.2 Methodology

This section outlines the approach to derive validated WFE constraints from the top-level EE requirements. The WFE is estimated at the TEP and represented it in terms of Zernike polynomials. Clearly, there is a continuum of WFEs that are compatible with a given rEE and EE. Therefore, the overall approach consists of three steps:

1. Develop a mathematical and statistical framework linking EE and aberrations;
2. Generate aberration realizations compatible with EE requirements;
3. Validate realizations using an independent physical optics model (PAOS) and an appropriate optical chain representation.

The first step requires an involved digression into the theory behind that is beyond the scope of this section. Here we provide a brief summary and gloss over much of the mathematics and technical details[10].

The analysis used a geometric optics model to establish a relationship between wavefront errors described by Zernike polynomials and the resulting EE. Geometric optics is suitable to describe low spatial-frequency aberrations. As shown below, we also account for diffraction effects and scattering from optical surface imperfections (e.g. surface roughness). At the very core, geometric optics is based on angles diffracted by refractive elements or deviated

[7] ARIEL-SAP-INST-TN-001

[8] We later realized that those wavefront are all over-performing, i.e., each gives a Radius of Encircled Energy (rEE) that is better (smaller) than required.

[9] ARIEL-INAF-PL-TN-015

[10] See ARIEL-SAP-INST-TN-001 for details.

Table 2.2. EE Requirements at the Telescope Exit Pupil (TEP).

Instrument channel	Wavelength [μm]	Ellipse axes (83.8% EE) Semi-major [arcsec]	Semi-minor [arcsec]
VISPhot	0.55	41	27
FGS-1	0.70	40	26
FGS-2	0.90	41	27
NIRSpec	1.00	42	28
	1.24	45	30
	1.48	48	32
	1.71	49	33
	1.95	53	35
CH0	1.95	53	35
	3.00	73	49
	3.90	91	61
CH1	3.90	91	61
	5.90	132	88
	7.80	169	113

[1] Angles are measured at the collimated beam from the telescope on the optical bench.

[2] The major axes of these elliptical apertures are orthogonal to the major axis of the telescope pupil.

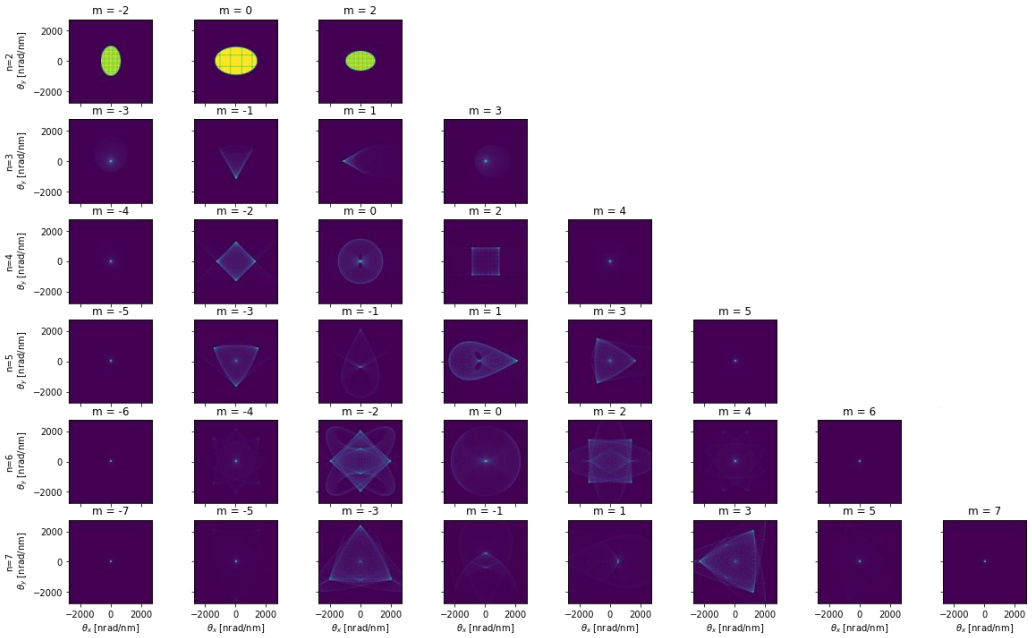


Figure 2.10. Probability Density Function (pdf) of the first 36 Zernike aberrations. The trivial terms (piston, tip, and tilt) are not shown and are not used in the analysis. n is the radial order and m is the azimuthal order of the Zernike polynomials. Note the Standard ordering.

by reflective elements. To simplify the discussion, with no loss of generality, we assume that aberrations are introduced by reflection off a fictitious surface located at the TEP. For reflective optical elements the Surface Form Error is $SFE = WFE/2$. The SFE is related to the Power Spectral Density (PSD) of the Surface Deviation (SD) – the difference from the ideal surface of the mirror – by the Parseval theorem:

$$\mathrm{SFE}^2 = \int_{Aperture} 2\pi f \ \mathrm{PSD}(f)\, df \tag{2.12}$$

where the integral is over the aperture described by the mirror. Moreover, there is a relation between the variance of the distribution of the incidence angles and the Power Spectral Density (PSD) of the SD:

$$Var(\theta) = 8\pi^3 \int_{Aperture} f^3 \ \mathrm{PSD}(f)\, df \tag{2.13}$$

The geometrical Point Spread Function, PSF_G, is given by the distribution of all reflected rays' angles, $\vec{\theta}_R = -2\,\vec{\theta}$, i.e. their Probability Density Function (pdf). Specifically, it is given by the convolution of the pdf of all individual reflection angles $\vec{\theta}_{R,k}$. Figure 2.10 shows the pdfs over the elliptical Ariel TEP for the first 36 Zernike aberrations, estimated numerically from the wavefront gradient. To account for diffraction, the geometrical PSF, PSF_G, can be convolved with the PSF expected from diffraction (Airy), PSF_D:

$$\mathrm{PSF}_{tot}(\vec{\theta}_R) = \mathrm{PSF}_G(\vec{\theta}_R) * \mathrm{PSF}_D(\vec{\theta}_R) \tag{2.14}$$

where the symbol $*$ denotes the convolution operation.

The residual Surface Form Error at mid and high frequencies, SFE_R, cannot be adequately described using a Zernike polynomial expansion and is instead described with a suitable PSD (e.g. Church et al., 1990):

$$\mathrm{PSD}_R(f) = \frac{w_0^4}{1 + (f/f_0)^\beta} \tag{2.15}$$

where w_0 is the power at low frequency in units of meters, f_0 is the cutoff frequency corresponding to the −3db point of the PSD, and β is the spectral index. The net effect of mid and high frequencies is that of removing power by scattering photons in all directions, but without significantly affecting the PSF shape. Therefore, we account for this effect by requiring that the EE in the target aperture be larger than required by the Total Integrated Scatter (TIS):

$$\mathrm{TIS}(\lambda) = 1 - exp\left[-\left(\frac{4\pi}{\lambda}\mathrm{SFE}_R\right)^2\right] \tag{2.16}$$

To obtain WFE constraints, first, we perform an *inverse sensitivity analysis*, where we evaluate the maximum level of aberrations compliant with the EE requirement assuming that the wavefront is affected by a single aberration. Thus, we obtain 36 Zernike coefficients c_k and discard the first three (trivial). These coefficients define a volume in the $N - 3$ dimensional space, where N is the number of Zernike coefficients, 36, and 3 are those discarded. The volume is bounded by the hyper-ellipsoid with semi-major axes equal to the coefficients.

Then, we consider a superposition of aberrations. In this case, we seek to sample from the set of all possible $\{\tilde{c}_k\}$ compliant with the EE requirement. Therefore, we consider the boundary of the hyper-ellipsoid, i.e. $\{\tilde{c}_k\}$ is a member of the set if

$$\sum_k \left(\frac{\tilde{c}_k}{c_k}\right)^2 \leq 1 \tag{2.17}$$

Using the Marsaglia (1972) method, we sample 1000 realizations from the boundary, attempting to combat the *curse of dimensionality*. Sampling from the volume would otherwise result in a disproportionate number of too optimistic realizations.

2.3.3 Results and POP validation

For each realization, and accounting for diffraction, we estimate the radius at which the EE allocation is met and compare this with the rEE allocation. Given that the geometrical PSF is independent of the wavelength, it is best studied at visible wavelengths, where diffraction effects are smallest. Therefore, the analysis is performed at the wavelength of the FGS–VISPhot channel (550 nm). We conservatively consider an allocation more demanding than the requirement (allowing some margin). Additionally, instead of the 83.8% EE requirement, we allocate 90% to account for TIS. At the end of the analysis, we obtained around 20% outliers violating the target allocation, which we may consider acceptable.

To validate the results of the geometrical treatment, we estimate the 550 nm VISPhot PSF for each $\{\tilde{c}_k\}$ realization using a Physical Optics Propagation (POP) model of the Ariel Telescope Assembly (TA) implemented with PAOS. The model consists of the Ariel TA optical design[11] up to the TEP, where an ideal thin lens is located. The lens has a 240 mm

[11] ARIEL-RAL-PL-ML-001

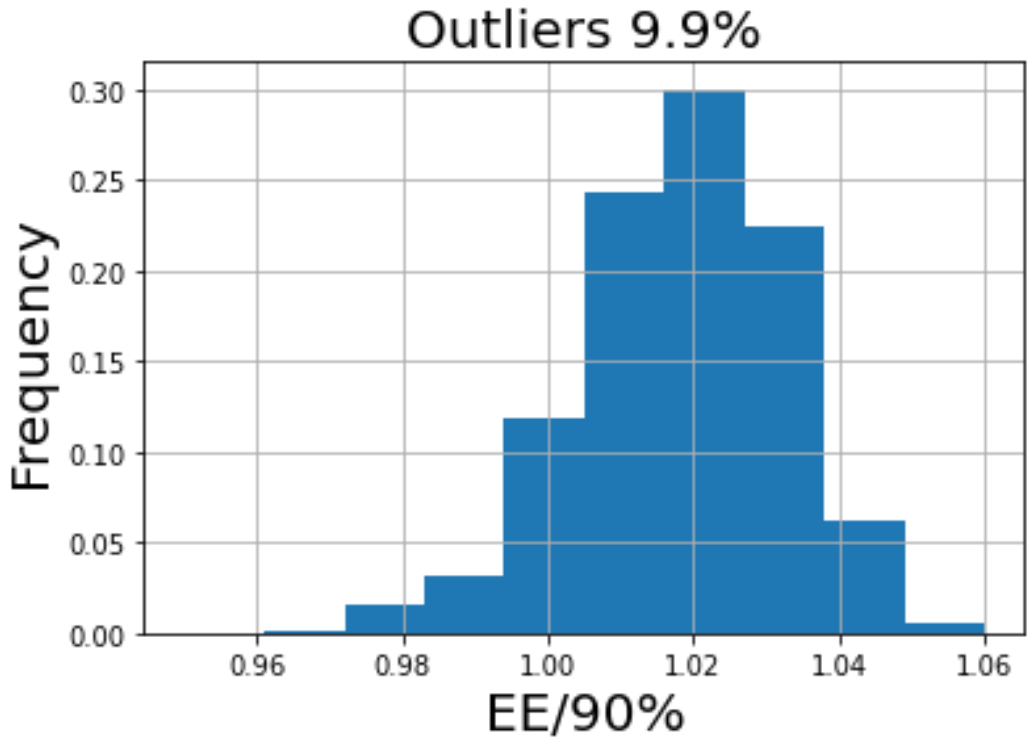


Figure 2.11. EE achieved in the allocated aperture, estimated using a PAOS physical optics propagation (POP) model of the Ariel Telescope Assembly (TA).

focal length to create an $f/12$ focal plane. The distribution of the EE estimated over the 1000 physical optics simulations is shown in Figure 2.11. The number of outliers is reduced from 21% to about 10%. The reason for this decrease has yet to be determined, but it is likely numerical (the POP simulation may be less precise than the geometrical treatment because of the finite resolutions of the grids used to sample the fields at the pupil and intermediate planes).

The obtained aberration realizations that are compatible with the EE allocation at 550 nm are shown in the example Table 2.3, along with the achieved EE within the specified elliptical aperture. These WFE limits represent the edge of the hyper-volume consistent with meeting the EE allocation.

Deriving WFE constraints tied directly to the EE requirements enables quantitative evaluation of the optical design. The whole set of aberrations realizations provides an effective way to test the optical performances of subsystems ahead of measurement of the SD of the Ariel TA. Additionally, it supports specifying the maximum allowable amplitude of aberrations from manufacturing the large all-Aluminum primary mirror. The goal, mentioned above, is to balance scientific requirements with manufacturing feasibility, without challenging the mirror fabrication process with overly stringent optical surface shape requirements.

The validation with PAOS gives high confidence that the geometrical limits define a bound on permissible aberrations. The integrated approach, starting from first principles in optics, followed by independent POP validation, provides an efficient process for cascading

Table 2.3. Example of WFE table of current pupil model.

				EE000	**EE001**	**EE002**	**EE003**	**EE004**	**...**	**EE999**
				86.7	87.5	87.6	87.8	87.9	...	94.6
Noll Index	**J**	**N**	**M**	**WFE000**	**WFE001**	**WFE002**	**WFE003**	**WFE004**	...	**WFE999**
6	4	2	2	77.2	-25.6	-16.2	3.6	32.6	...	14.8
4	5	2	0	70	-29.4	-15.4	-52.2	48.4	...	15
5	6	2	-2	-3.6	6.4	33	-3.4	21	...	25.4
10	7	3	3	-7.6	14	-2.8	5.6	-21.2	...	-54.2
...	...	...	...	...	...	...	...	...	...	...
35	36	7	-7	1.4	0.6	-4.6	-10.8	-12.4	...	-9.8

top-level science requirements down to detailed optical subsystem budgets. By combining these complementary approaches, the analysis produced realistic performance budgets validated through end-to-end modeling. Finally, it motivated further, large-scale analyses with commercial optics software[12] that yielded compatible results.

[12]Chiefly, the analysis with Zemax Opticstudio® presented in ARIEL-INAF-PL-TN-015.

Part II

Ethos
Mission Performance, Calibration, and Data Reduction Strategies

Chapter 3

Ariel's Payload PDR Performance Analysis

An important milestone in the development of the ESA's Ariel space mission was the Preliminary Design Review (PDR). The PDR reviewed the completeness and consistency of the preliminary mission design to ensure it meets the technical, scientific, and operational requirements within the constraints of cost, risk, and schedule. In May 2023 the ESA review board accepted that all the objectives had been completed, and confirmed the successful closure of the Ariel payload PDR. As a result of this major achievement, Ariel's payload critical technology is now considered at Technical Readiness Level 6, indicating that the mission can now proceed to payload Critical Design Review (CDR) and begin to manufacture its first prototype models. In support of these activities, I have contributed to investigating the overall scientific performance of the Ariel payload. The outcome of this study is summarized here and discussed in detail in the Performance Analysis Report[1]. Before delving into this study, I provide a brief overview of key aspects of the Ariel payload design, some already introduced in Section 1.2.1, to provide the reader with a self-consistent chapter.

* * *

Transit spectroscopy and multi-band photometry of exoplanet atmospheres have been so far conducted using general-purpose space telescopes and ground-based instruments. While these have achieved some success, measurements often suffer from limited and patchy spectral coverage and systematic errors from factors such as pointing jitter, thermal variations, opto-mechanical instabilities, and detector effects. Drawing upon lessons learned from past missions including Spitzer Space Telescope (SST), Hubble Space Telescope (HST), and ground-based observatories, the Ariel payload has been strategically designed to overcome these limitations. Ariel will obtain spectroscopic and photometric time series of transiting exoplanets with post-processing stability better than 20 to 100 ppm over a single transit observation, depending on the target brightness. Key aspects that allow Ariel to obtain its stable performance are:

1. Simultaneous observations by all photometric and spectroscopic channels;

[1] ARIEL-SAP-PL-AN-002

2. Continuous observations such that thermal and photometric stability are minimized;

3. A payload design that makes Ariel resilient to major sources of systematics, or allows their removal via post-processing.

The most significant noise and systematic uncertainties have been identified[2] and are summarized in Table 3.1, alongside mitigation approaches to minimize their impact on detection sensitivity and photometric stability.

Table 3.1. Sources of uncertainty and mitigation strategies implemented by Ariel.

Type of Uncertainty	Source	Mitigation Strategy
Detector noise	Dark current noise, Readout noise	Choice of low-noise detectors
	Gain stability	Calibration, post-processing data analysis, requirement of stable detectors, thermal stability of both cold and warm electronics
	Persistence[3]	Post-processing decorrelation. Continuously staring at a target for the whole duration of the observation.
Thermal noise	Emission from telescope, common optics, and all optical elements	Negligible impact by design
	Temperature fluctuations in time	Negligible impact by design
Astrophysical noise	Photon noise from the target	Fundamental noise limit, choice of aperture size (M1 diameter).
	Photon noise from local zodiacal light	Negligible over Ariel band
	Stellar variability with time	Multi-wavelength stellar monitoring, post-processing decorrelation

[2]ARIEL-RAL-PL-DD-001

[3]Later analysis (ARIEL-SAP-PL-TN-008) found persistence negligible for the Ariel application.

Pointing jitter	Relative Performance Error (RPE) and Performance Drift Error (PDE) effects on the position, Spectral Energy Distribution (SED), and detector intra/inter pixel response	Combination of instrument design, Attitude and Orbit Control System (AOCS) stability, and post-processing decorrelation
	Slit losses	Spectrometer input slit sufficiently large

Below I provide a targeted overview of salient aspects of the science performance of the Ariel payload, demonstrating the compliance of the payload design with the Ariel science requirements. The analysis discussed here uses radiometric and time-domain simulations together with data reduction techniques. The radiometric estimates are obtained using the Ariel Radiometric Model (ArielRad) (Mugnai et al., 2020), while non-stationary noise processes are evaluated using ExoSim 2, a new implementation of the time-domain simulator described in (Sarkar et al., 2021). ArielRad (see Section 1.3.1 for details) implements a detailed instrument model and performance is radiometrically estimated from a physically motivated noise model. ExoSim 2 (Mugnai et al., 2022) (see Section 1.3.3 for details) utilizes a more detailed instrument model and generates photometric and spectroscopic images in the time domain similar to those Ariel will provide during science operations. The timelines are processed with a data reduction pipeline that implements prototypical steps of ADaRP, the Ariel Data Reduction Pipeline[4], to provide estimates of the spectral and photometric light curves, of the noise processes and systematics, and to study the overall payload performance. Finally, I report on the capability of the Ariel payload in conducting observational campaigns to detect hundreds of exoplanetary atmospheres. An observational programme compliant with the Ariel science requirements is compatible with available observing time within the mission lifetime.

3.1 Model Parameters and Margins

To conduct these performance analyses, appropriate model parameters and margins are critical. This section details the key model parameters drawn from the Ariel Performance Parameters Database (PPD), listed in the table below as per the payload PDR snapshot[5], including both requirements-driven values and estimated performance.

[4]ARIEL-CRDF-GS-DD-001
[5]ARIEL-UCL-PLD-ML-001

Table 3.2. ArielRad model parameters utilized in the performance analysis for the payload PDR. Note that, when a requirement reference is given in the last column, then the value parameter is from requirements instead of the PPD.

Instrument Prescription	VISPhot	FGS-1	FGS-2	NIRSpec	AIRS Ch0	AIRS Ch1	Requirement
Telescope collecting area	0.63 m^2						R-TEL-10000
Wavelength coverage (µm)	0.50–0.60	0.60–0.80	0.8–1.1	1.1–1.95	1.95–3.90	3.90–7.80	R-PRD-0380 R-PRD-0390
Telescope temperature (K)	60	60	60	60	60	60	R-PRD-1310
Instrument optics temperature (K)	60	60	60	60	60	60	R-PRD-1310
Instrument enclosure temperature[6] (K)	60	60	60	60	55	55	
Detector temperature (K)	65	65	65	65	42	42	R-PRD-1310
FoM[7] (m^2)	0.130	0.130	0.165	0.165	0.132	0.132	R-PRD-0440 R-PRD-0441 R-PRD-0442
Emissivity of optical surfaces	0.03	0.03	0.03	0.03	0.03	0.03	
Slit width (mm)	N/A	N/A	N/A	N/A	1.2	0.94	
Slit image width (mm)	N/A	N/A	N/A	N/A	1.07	0.39	
Image Space F/#	33.67 50.63	18.70 28.12	24.83 37.34	19.66 29.57	15.5	7.7	
Plate Scale (mas/pix)	106	173	135	165 130	204 312	440 660	
Detector Pixel Size (µm)	18	18	18	18	18	18	
Detector Dark Current[8] (e-/s/pixel)	5	5	5	5	5	5	
Detector pixel linear well depth (ke-)	100	100	100	100	85	50	
Read Noise[9] (CDS/pixel) (e-)	20	20	20	20	22	23	

[6] Box light-tightening the instrument optics. Assumed 55 K for AIRS Ch0 and Ch1.
[7] Defined in Equation 3.1.
[8] Mean value
[9] Mean value

Excess additive noise[10] (CDS/pixel) (e-)	6	6	6	6	7	7	R-PRD-8811
Excess photon noise variance margin[11]	40%	40%	40%	40%	40%	40%	
Straylight[12]	1%	1%	1%	1%	1%	1%	
Noise Floor	20 ppm						R-PRD-8810
Gain noise	40 ppm $\sqrt{\text{hr}}$						R-PRD-8810
Jitter noise	20 ppm, constant at time scales T >1hr						
Observing time (transit or eclipse)	$2.5 \times T_{14}$						
Zodiacal light model	Average, using 2.5 multiplier in Zodiacal light expression						R-PRD-0500

Conservative assumptions and margins are also applied in the ArielRad model in Table 3.2 to account for uncertainties in this design phase. The most critical ones are as follows.

1. A photon noise variance margin of 18% (X = 0.4) is allocated to the target star noise, which dominates the budget as shown later. This margin accounts for uncertainties in quantum efficiency, optical efficiency, and observing efficiency.

2. The electronic gain noise is expected to be a pessimistic overestimate, as it was negligible for HST and SST. However, it does dominate the noise budget for bright targets.

3. The Photometric Conversion Efficiency (PCE) is potentially underestimated for AIRS bands by ~30%, based on JWST data indicating higher quantum efficiency from 2–5 μm.

4. A noise floor of 20 ppm is used in the analysis as a consequence of stability requirements at 10 hrs. This is considered a worst-case assumption by more than 30% compared to the stability of <15 ppm achieved by HST and SST. Given Ariel's optimization for exoplanets and L2 orbit, the 20 ppm noise floor is likely a significant overestimate of the true stability.

As Ariel detections are photon noise limited (Mugnai et al., 2020), neglecting gain noise, that may perhaps affect only a very small portion of the Ariel targets, the margins are carried mainly by a 40% excess noise variance and pessimistic QE requirements in AIRS. Overall, this represents conservative margins of ~20% for FGS bands and ~30% for AIRS bands on the noise model.

[10] Mean value

[11] Margin allocated

[12] The value used accounts for all sources of straylight (TIS, sky and ghosts, respectively R-PRD-0780/1/2) as discussed in ARIEL-SAP-PL-TN-003. It does not include self emission as this is accounted for separately.

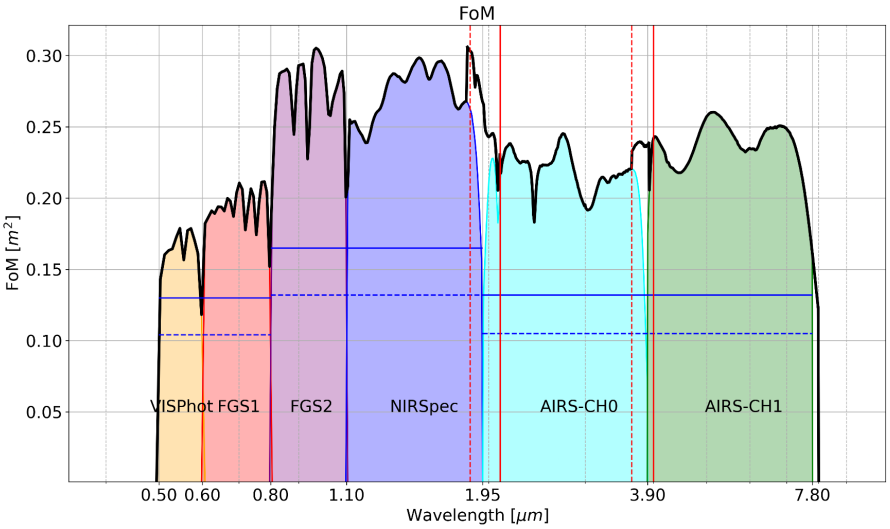


Figure 3.1. The payload end-to-end Figure of Merit (FoM), calculated as the product of the telescope collecting area, the detector quantum efficiency, and the optical efficiency. Horizontal solid lines in blue show science performance requirements: 0.130 m^2 from 0.5 µm to 0.8 (R-PRD-0441), 0.165 m^2 from 0.8 µm to 1.95 µm (R-PRD-0442) and 0.132 m^2 from 1.95 µm to 7.8 µm (R-PRD-0440). Blue horizontal dashed lines show the minimum allowed Figure of Merit (FoM) (R-PRD-0450). Vertical red lines represent the extended spectrometer wavelength ranges (R-PRD-0400). The solid black line is the total resulting FoM considering the overlap of adjacent spectroscopic channels.

3.2 Optical efficiency

The overall optical efficiency, named Figure of Merit (FoM), is defined as the product between the telescope collecting area, the optical efficiency, and the detector quantum efficiency[13]:

$$\mathrm{FoM}(\lambda) = \mathrm{A_{tel}} \cdot \mathrm{QE}(\lambda) \cdot \mathrm{TR}(\lambda) \tag{3.1}$$

where A_{tel} is the effective telescope collecting area, QE is the detector Quantum Efficiency (QE) and TR is the optical efficiency of the instrument.

The FoM is computed using parameters from Table 3.2 and is shown in Figure 3.1. Ariel's current best estimate is significantly above requirements and discussed in the Payload Throughput and FoM Analysis document[14].

3.3 Noise Budget

The overall noise budget is shown in Figure 3.2. It gives the best estimate of the complete instrument noise performance for the three sizing targets: GJ 1214[15], HD 219134[16], and

[13] R-PRD-0440
[14] ARIEL-SAP-PL-AN-001
[15] R-PERF-010
[16] R-PERF-020

HD 209458[17]. GJ 1214 is a faint target of spectral type M5V and $\text{mag}_k = 8.8$. HD 219134 has spectral type K3V, $\text{mag}_k = 3.25$, and represents the brightest target observable by Ariel. HD 209458 is a G0V star with $\text{mag}_k = 6.3$; it represents a typical bright target Ariel will observe. Figure 3.2 shows that excluding gain noise, the dominant noise source is the target photon shot noise.

The noise budget presented provides a cautious estimate of Ariel's post-processing photometric performance. The following considerations are worth making:

1. At the red end of AIRS-Ch1, the detector read noise becomes comparable to the photon noise. The analysis assumes that detectors are analyzed using the first and the last science frame of one exposure to form a Correlated Double Sampling (CDS) sample. However, as Ariel detectors are sampled up the ramp, read noise reduction strategies can be implemented when advantageous.

2. The noise evaluation assumes aperture photometry signal extraction. Optimal estimation methods could reduce noise components other than target photon noise, including read noise, by 40% compared to aperture photometry. Thus aperture photometry represents a worst-case assumption.

3. For the bright targets HD 209458 and HD 219134, gain noise dominates the budget of HD 209458 at wavelengths shorter than 1.95 μm and across the full spectrum, respectively. While Ariel is fully compliant with the science requirements as shown later, past observations by HST and SST achieved total noise within 15% of photon noise for similar bright stars. And that includes noise from both the instrument and of astronomical origin. Because Ariel implements very similar detector focal planes and acquisition electronics but has an optimized payload design operating in a more stable environment, it is likely that this noise source, which primarily affects bright targets, is significantly over-estimated. Thus, while detector gain noise is expected to be efficiently detrended, we retain it as a worst-case.

4. A 20 ppm noise floor is included, corresponding to an allocation of Brown noise with a $1/f$ knee at 10 hr. An additional noise floor of 20 ppm is added to account for the photometric noise arising from pointing jitter (see Section 3.4 for details).

3.4 Pointing stability

As part of the payload Preliminary Design Review (PDR), I contributed to a detailed study of photometric uncertainties arising from pointing jitter[18]. Below, I summarize the key findings and design solutions from that study relevant to this analysis.

The pointing stability of the telescope is quantified in terms of[19] Absolute Performance Error (APE), Performance Drift Error (PDE), Relative Performance Error (RPE), and Mean Performance Error (MPE)-RPE. The effect of the jitter on the observed time series is the introduction of noise, characterized by the power spectrum of the telescope pointing. It is correlated in time if the power spectrum is non-white. Since jitter randomly shifts the

[17] R-PERF-025
[18] ARIEL-CRDF-PL-AN-005
[19] ESA Pointing Error Engineering Handbook

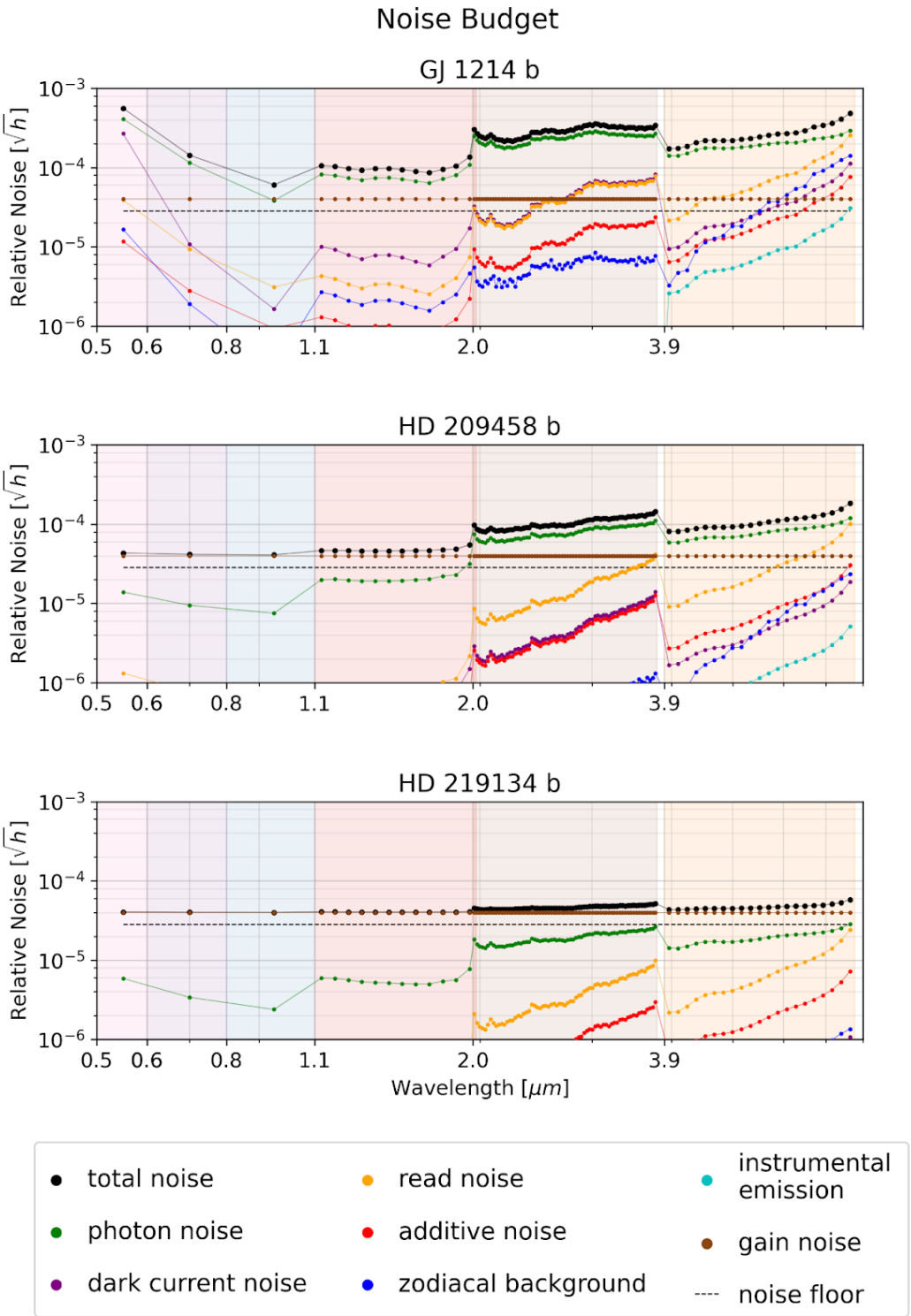


Figure 3.2. Noise budget for the three sizing planets with channel bands highlighted. The total noise is the sum in quadrature of all noise sources, where the target photon noise has been scaled up by 40%. The noise floor is the sum in quadrature of the true noise floor and jitter noise; it has no units as it does not integrate down with time. The noise floor is not included in the total noise and should be added in quadrature to the total noise, at the time scale of interest.

line of sight, it affects all illuminated pixels identically and is 100% correlated across all wavelengths and channels. The amplitude of the resultant photometric noise depends on the amount of displacement of the spectrum, the Point Spread Function (PSF) of the instruments vs. wavelength, the detector intra-pixel response, and the amplitude of the inter-pixel variations (Sarkar et al., 2021, Figure 4).

The Ariel spectrometers utilize a slit-less design. Field stops in the form of rectangular apertures are implemented to mitigate possible stray light contamination. Through Physical Optics Propagation (POP) simulations using Physical Optics Simulator (PAOS) (see Chapter 2 for details), the field stop sizes were characterized to determine sufficient widths that prevent photometric instabilities from field losses. This quantification prompted design changes to ensure field stop dimensions that maintain photometric stability. However, the detailed characterization and analysis are not discussed here due to length constraints.

For CDS, pointing jitter mainly affects the second Non-Destructive Read (NDR). Therefore, the relevant frequency is set by the exposure time, and pointing jitter induces photometric error on two different time scales:

- Jitter faster than the exposure rate effectively enlarges the monochromatic PSF, slowly changing its size with time. However, total energy collected per unit time does not incur significant losses, making this photometric error negligible compared to other experimental uncertainties on 1-hour timescales (vs. max 4-hour transit durations). The slight spectral blurring is inconsequential[20].

- Jitter slower than the exposure rate can be a non-negligible source of photometric errors. For spatial jitter, photometric uncertainties arise from the combined effect of a wobbling spectrum sampled in the presence of intra- and inter-pixel variations. For spectral jitter, uncertainties arise mostly from the modulation imposed in the dispersion direction. Compared to spatial jitter, spectral jitter can result in larger photometric instabilities, and if left uncorrected can severely impair the quality of the final science result. The effect of spectral jitter is most problematic in areas of steep flux gradients, whether due to stellar lines or variations in the continuum. Photometric variations caused by PSF shape variations induced by a time-varying RPE are comparatively much smaller.

Jitter can be mitigated by a combination of instrument design, Attitude and Orbit Control System (AOCS) stability, and post-processing analysis which takes advantage of the correlated nature of this systematic component[21]. Spatial-jitter can be made negligible by requiring that photometer and spectrometer signals are at least Nyquist sampled (see Section 4.1 for more details). This results in a monochromatic PSF with widths spanning two or more detector pixels, eliminating any photometric instability from intra-pixel responses. Inter-pixel variations do still have an impact, but this is rendered negligible by flat-fielding quantum-efficiency variations across pixels to the required level (~0.5%), in post-processing. Moreover, spectral jitter can be decorrelated in post-processing by using the spectral information alone, correlating each exposure with all other exposures to estimate the shift. However, even with these design and post-processing solutions, without proper decorrelation, jitter can be a significant source of photometric noise.

[20] ARIEL-SAP-PL-TN-003

[21] ARIEL-SAP-PL-TN-003 and ARIEL-CRDF-PL-AN-001

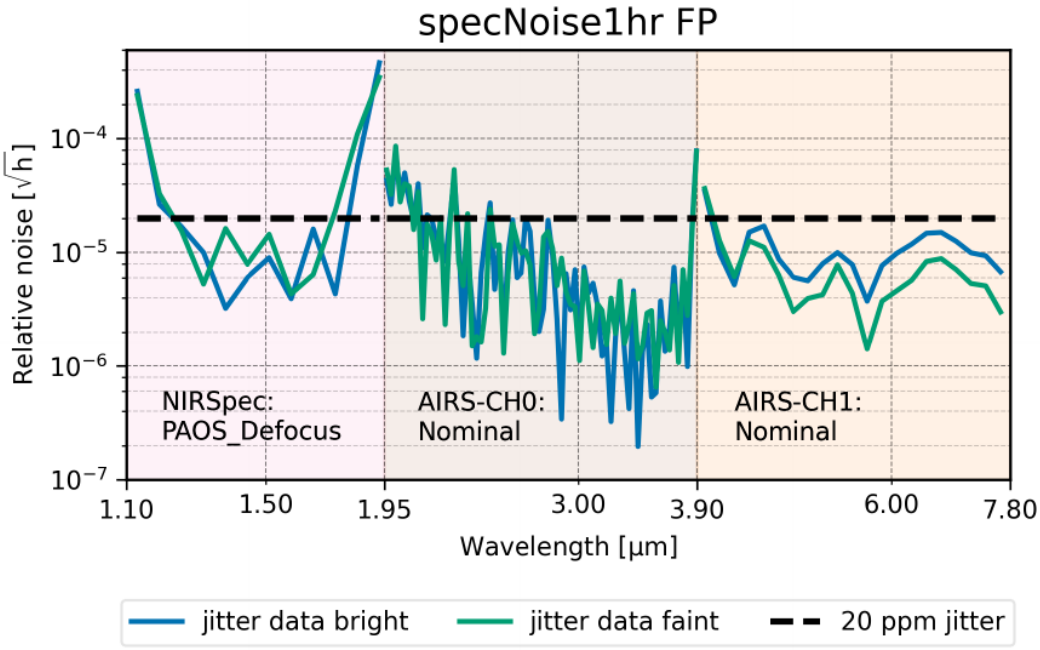


Figure 3.3. Simulated spectro-photometric relative noise at a 1-hour time scale pointing jitter (wiggly lines in the three panels) using simulations from Airbus. The black horizontal line corresponds to the 20ppm $\sqrt{\text{hr}}$ bright target pointing jitter detrending requirement (R-PRD-8927).

The photometric and spectroscopic lightcurves are simulated using AOCS inputs from Airbus[22] - specifically pitch and yaw deviations relative to the target vs. time. The AOCS models used for both bright and faint targets include several strong reaction wheel harmonics (5 Hz – 300 Hz) and time-varying RPE. The simulations are processed through ExoSim 2 and a dedicated simulator for reading out HgCdTe (MCT) detectors. Airy PSFs are used for AIRS channels. This is appropriate for the diffraction limited Ch1, but conservative for Ch0 because it is narrower than expected in the integrated payload, increasing the jitter noise. On the other hand, NIRSpec utilizes PAOS PSFs to capture the PSF broadening critical for jitter noise mitigation (see Section 4.1 for details).

The simulated lightcurves are jitter-noise-decorrelated, before estimating the uncertainty on the photometry. Decorrelation is a data processing step that, taking advantage of the Nyquist sampled nature of photometric and spectral images, can reduce jitter noise by close to two orders of magnitude in noise variance[23]. The decorrelation implemented in this analysis is only a possible solution, and it was later found suboptimal for the Ariel photometers. I present an optimal decorrelation algorithm in Section 4.2, along with the updated results.

The uncertainty on the photometry after decorrelation is proportional to the signal, and therefore the relative uncertainty is independent from the target brightness[24].

Figure 3.3 summarizes an example of the analysis of the estimated jitter noise for the spectroscopic channels. At 1 hr, the noise is generally < 20 ppm in NIRSpec/AIRS-Ch1

[22]ARIEL-ADST-SC-FI-000001
[23]ARIEL-CRDF-PL-AN-001
[24]ARIEL-CRDF-PL-AN-005

and across most of AIRS-Ch0. However, the Ch0 blue end exhibits excess jitter noise up to 50–70 ppm from 1.95-2.1 μm. While this marginally exceeds the bright target jitter requirement[25], it does not impact overall science capabilities as discussed later. Improved detrending using the optimal decorrelation algorithm, adapted to spectrometers, and more realistic PSFs are expected to resolve this small non-compliance.

While we expect to be able to de-trend jitter noise significantly better than 20 ppm, in the performance analyses presented it is conservatively assumed that the jitter noise contributes 20 ppm at all time scales from 1 hr and longer, i.e. it does not integrate down with time, and at all Ariel wavelengths.

3.5 Compliance with requirements

This final section demonstrates compliance of the Ariel science payload with the science requirements[26], showing the Mission Reference Sample (MRS) can be observed during the available mission lifetime at the specified tier spectral resolution and Signal-to-Noise Ratio (SNR). The Mission Reference Sample 2019 (MRS19) (Edwards et al., 2019) lists potential Ariel targets that are in number and type compliant with the science requirements[27] This target list is not the final one, nor the only possible one; it is used mainly to provide a testbed for the performance of the payload to validate the compliance with the science requirements consistently as the mission evolves.

MRS19 includes information on stellar and planet parameters, assigned observing tiers, and preferred transit or eclipse methods. Phase curve observations are not considered. 400, 550, and 50 targets are listed in the Reconnaissance, Deep, and Benchmark tiers, respectively. For Reconnaissance survey, we use this definition[28]: average SNR $\gtrsim 7$ for $\gtrsim 10$ averaged bands for $1.10<\lambda<7.80$ μm for all targets. This is slightly different than reported in Edwards et al. (2019).

The validation procedure is described in Table 3.3. For each target in the MRS19, we use

Table 3.3. Procedure utilized to validate the mission reference sample (MRS19).

Step	Description
1	Use ArielRad to calculate observing time needed per target to reach required SNR and resolution
2	Sum observing times over all MRS19 targets to get total time T_{OBS}
3	Estimate total science time available in mission lifetime T_{SCI}
4	Compare T_{OBS} and T_{SCI}
5	If $T_{OBS} < T_{SCI}$, payload design meets requirements for MRS19 sample

ArielRad to estimate the amount of observing time required to reach the SNR appropriate for the tier each target belongs to, at the spectral resolution of the tier. The target observing time is always an integer multiple of the minimum target observing time, that is 2.5 times the transit duration. The sum of all target observing time, T_{OBS}, is the total observing time

[25] R-PRD-8927
[26] R-PERF-130, ESA-ARIEL-EST-MIS-RS-001
[27] From R-PERF-130 to R-PERF-136, included.
[28] R-PERF-134

required for Ariel to comply the top level performance requirements[29]. T_{OBS} is compared with an estimate of the fraction of the mission lifetime available for science observations, T_{SCI}. T_{SCI} is evaluated as:

$$T_{SCI} = \eta_{OBS}\eta_{SE}\left(T_{MIS} - T_{CP}\right) \times \left[1 - \frac{T_{CAL}}{24 \cdot 7}\right] \tag{3.2}$$

with parameters T_{MIS} (mission life time, 4 years), T_{CP} (commissioning phase + performance verification + science demonstration, 180 days), η_{OBS} (observation efficiency, 85%), η_{SE} (target scheduling efficiency, 100%), and T_{CAL} (calibration time, 7 hours/week).

Given these parameters, T_{SCI} is estimated to be 25044 hours. The mission is considered compliant if the science implemented by the MRS19 can be executed within the mission lifetime, i.e if $T_{OBS} < T_{SCI}$, and not compliant otherwise.

Below we demonstrate the compliance of the Ariel science payload with Table 3.2 parameters with the science requirements on the Ariel core survey[30], i.e. the observation of a core sample of ≥ 500 exoplanet targets divided in at least three survey tiers. This analysis uses a modified version of the payload, where the overall Figure of Merit (FoM) from payload requirements is used (horizontal solid red lines in Figure 3.1), that is significantly worse than the current design estimate. As a stress-test, we study the largest number of MRS19 targets that can be observed in the deep survey tier[31]. Figure 3.4 shows that observations of 557 targets can be completed during the nominal mission lifetime, 57 targets more than the required 500. As a further stress test, we estimate that 428 planets can be observed if Ariel targets were all Benchmark targets. Figure 3.4 further shows that observing all MRS19, 1000 targets at the low resolution of the Reconnaissance survey would take less than the available time. Because targets have to be distributed among different tiers, we conclude that Ariel can easily observe targets in excess of 500, when tier-optimized.

As further evidence of this, we analyzed the case of MRS19, that is compliant with 1000 targets[32] optimally distributed in the three Ariel tiers (Reconnaissance tier, Deep tier, and Benchmark planets). We utilize the Current Best Estimate (CBE) FoM, but retaining simulations margins. Here, for optimization we adopt the definition of Edwards et al. (2019) for the Reconnaissance survey spectral data, binned in 4 spectral bins (1 in NIRSpec, 2 in AIRS-Ch0 and 1 in AIRS-Ch1). Additionally, we utilize the AIRS capability of averaging consecutive NDRs[33]. As shown in Table 3.4, 981 MRS19 targets observations can be completed in less than ~25000 hr, i.e. the available science observing time during the nominal mission, having deducted overheads and calibration events from the nominal mission lifetime. Finally, we obtain that 1753 hr of the extended mission lifetime are needed for completion of the full sample of 1000 planets, i.e. about 14% of the extended lifetime[34].

[29] R-PRD-0470
[30] R-PERF-130, R-PERF-133
[31] R-PERF-135
[32] Twice more than R-PERF-130.
[33] R-PRD-1967
[34] G-MIS-030

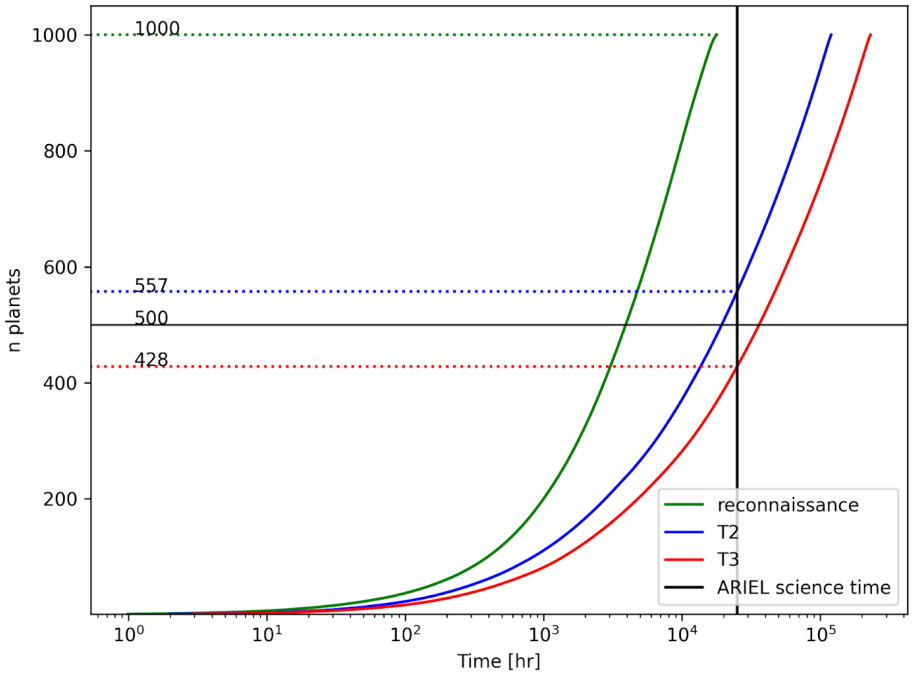


Figure 3.4. Cumulative number of targets observable vs. time from the start of science observations. The vertical line marks T_{SCI}, the maximum available science time within the nominal mission lifetime. All targets observed at Deep survey tier resolution (blue solid line). All targets observed at Benchmark resolution (red solid line). All targets observed at Reconnaissance resolution (green solid line). Figure of Merit (FoM) from requirements is assumed.

Table 3.4. Performance on the MRS19 sample, where 19 out of 1000 targets that require a large number of observations have been excluded.

Tier	Number of targets	Total observing time [h]	Relative observing time
Reconnaissance survey	391	7151	29 %
Deep survey	541	16247	65 %
Benchmark	49	1559	6 %
Total	981	**24958**	100 %
Available time		**25 044**	

Chapter 4

Data Reduction Strategies

This chapter bridges theory and practice, cementing the connection between my technical analyses and their practical application. Below I present a targeted selection of my work done for Ariel mission data simulations and subsequent data reduction. This work encompasses the start and end of the analysis cycle - from generating Point Spread Functions (PSFs) with PAOS (see Chapter 2 for details) to the final de-jittering steps. This end-to-end involvement afforded critical insights into the data being simulated and post-processing to extract exoplanet atmospheres' signals.

* * *

To establish context before delving into these analyses, here is a high-level summary of key points, sketched in Figure 4.1.

- Transit spectroscopy measurements are subject to sources of systematic uncertainty, which must be characterized and removed in post-processing, e.g. photometric noise from pointing jitter, *before* the data can be interpreted.
- The observation method of transit spectroscopy requires no significant angular resolution. Detailed performance studies show that a telescope collecting area of 0.64 m^2 is sufficient to achieve the necessary observations on all the Ariel targets within the mission lifetime. Since imaging quality is not a driver, Ariel is designed to be an efficient *light bucket*, collecting photons in a compact but adequately sampled PSF. This reduces costs and risks, ensuring that the instrument design is not above specification while retaining the necessary measurement quality.
- Among the main Ariel mission requirements is to provide critically sampled focal planes because Nyquist-sampled optical signals enable efficient rejection of these systematics. At small wavelengths, the Ariel payload design relies on geometrical aberrations to broaden the PSF relative to the scale of the detector pixels: diffraction-limited optics would otherwise result in sub-sampled PSFs at FGS wavelengths, given the image space f-numbers implemented.
- Ariel spectrographs operate in a slit-less mode for point source spectroscopy. To reduce background and straylight, AIRS implements field stops at the intermediate

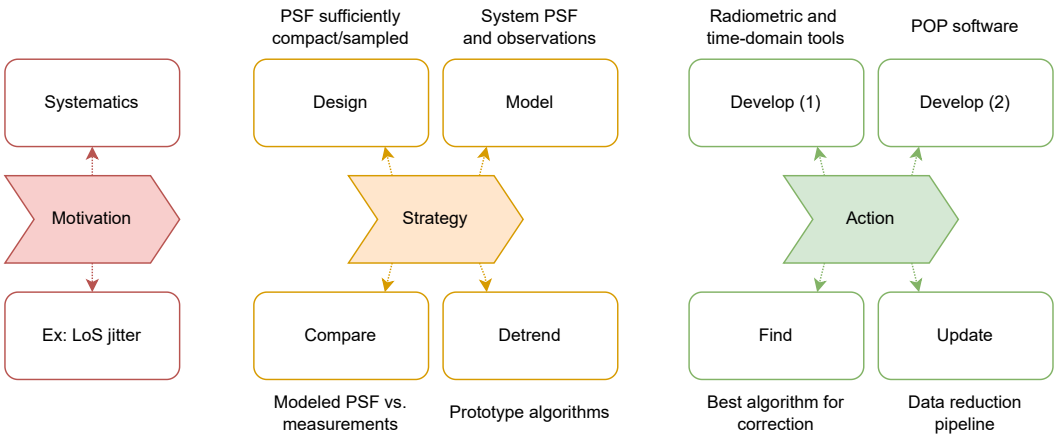


Figure 4.1. Schematic representation of the analysis cycle, from the definition of the scientific problem to a comprehensive strategy to address it and actionable steps needed to implement the strategy. Here, the problem is pointing jitter, which affects transmission spectroscopy measurements. The devised strategy includes designing the Point Spread Function (PSF) to be sufficiently compact and sufficiently sampled to mitigate jitter effects, and then modeling the system PSF and comparing it to measurements taken on-ground and in-flight. At the same time, realistic observations need to be simulated and algorithm prototyping and testing need to be carried out to ensure that the data analysis pipeline can effectively remove the systematic. To execute this strategy, a suite of tools is needed, including PAOS to generate PSFs, ArielRad to verify performance requirements, and ExoSim 2 to simulate observations. At the end of this process, the best algorithm for correcting the systematic effects can be selected based on representative simulations and the Ariel Data Reduction Pipeline (ADaRP) is updated accordingly. This process is iterative and requires the collaboration of multiple teams. All steps are documented and timely feedback is provided to the Consortium and at mission reviews.

image planes, at the input of Ch0 and Ch1. A key requirement is to implement sufficiently wide field stop apertures relative to the PSF dimensions[1] and Absolute Performance Error (APE) to prevent photometric instabilities from field losses in the presence of pointing jitter[2].

- Therefore, detailed studies of diffraction, aberrations, sampling, and related systematics must be carried out before the mission launch. These studies require the simulation of the propagation of the electromagnetic field through the Ariel optical chain, including diffractive elements, and the resulting PSFs vs. wavelength at the intermediate and focal planes. This is essential, especially at wavelengths where Ariel is not diffraction-limited (FGS), to properly capture the broadening of the PSF that is of fundamental importance for the rejection of jitter noise.

- The generation of simulated science frames requires knowledge of the illumination of the detector array by the Focal Plane Spectrum (FP). The FP depends on the stellar spectrum and takes into account the entire telescope/instrument optical chain. To simulate representative science frames, the PSFs generated with PAOS (see Chapter 2 for details) are combined with the time-domain simulation of an Ariel observation, from the astrophysical source to the focal planes, provided by ExoSim 2 (see Section 1.3.3 for details).

- Finally, the science frames need to be processed during data analysis and dedicated detrending modules need to be established to deal with systematics such as pointing jitter, that would otherwise interfere with the correct interpretation of the science data. Extensive testing and validation of the detrending modules are necessary to ensure that the atmospheric parameters are inferred within the required precision *and* accuracy, avoiding biases. These data analysis techniques will be eventually implemented in the Ariel Data Reduction Pipeline (ADaRP).

4.1 PSF sampling analysis

This section presents the analysis I conducted to evaluate Point Spread Function (PSF) sampling across the focal planes. Adequate PSF sampling is critical for enabling the effective mitigation of systematic uncertainties such as photometric noise from pointing jitter. This analysis was instrumental in establishing fully Nyquist-sampled PSFs for the detailed analyses of jitter noise mitigation, carried out for the payload Preliminary Design Review (PDR) (see Chapter 3 for details) and improved in subsequent work as shown later.

Rejection of photometric noise induced by pointing jitter is facilitated when the optical signals (PSFs) are critically sampled. At short wavelengths, the Ariel payload design relies on geometrical aberrations to broaden the PSF relative to the scale of the detector pixels: as mentioned above, diffraction-limited optics would otherwise result in sub-sampled PSFs at FGS wavelengths.

The exact distribution of geometrical aberrations on the optical beam feeding the FGS instrument will be known later in the project. As discussed in Section 2.3, we made a preliminary attempt at modeling what might be the distribution of aberrations compatible

[1] R-PRD-8390, R-PRD-8400

[2] This analysis is not reported in this thesis for length constraints.

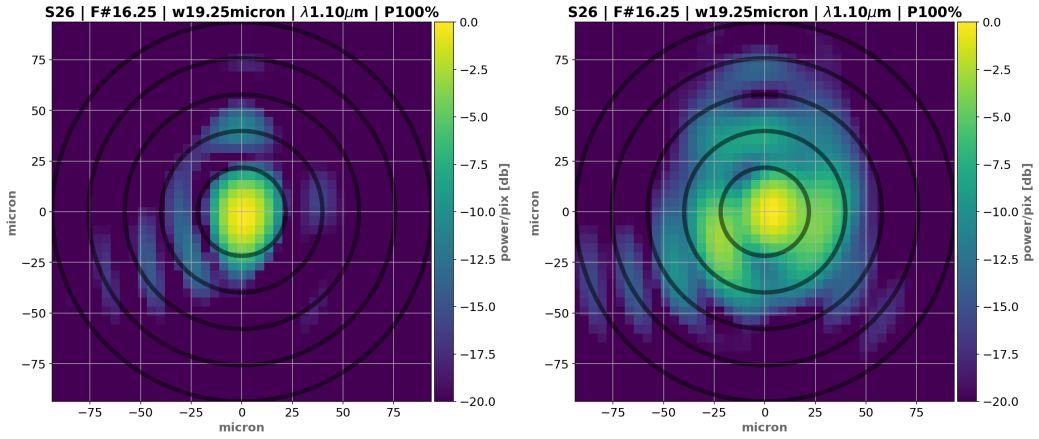


Figure 4.2. The Point Spread Function (PSF) used at the blue end of NIRSpec is critically sampled by 18 μm detector pixels, and it is shown in the right panel. It is obtained from the aberrated PSF shown on the left, to which we added 200nm of extra defocus. The PSFs used across the NIRSpec band are scaled to the appropriate wavelength.

with the Radius of Encircled Energy (rEE) requirements. That report[3] estimates 1000 possible realizations, all of which exceed performance requirements, with a smaller rEE. While in future work we will estimate aberrations realizations compatible with the rEE requirements, here we consider the possibility[4] of placing the instrument deliberately out of focus to recover the optimal operational conditions from the optical point of view.

For these simulations, we take the realization labeled EE001 (see Table 2.3) and add 200 nm of additional defocus aberration in the form of the appropriate Zernike term. This makes the PSF, shown in Figure 4.2, critically sampled. We adopt the following definition for Nyquist sampling compliance:

1. The Optical Transfer Function (OTF) is obtained by taking the modulus of the Fourier transform of the PSF.

2. The spatial frequencies (f_x, f_y) at which the OTF drops to 1/e of its peak value OTF(0,0) must lie within a circle centered at zero frequency.

3. The radius of this circle is defined as the Nyquist frequency f_N, calculated as $1/2\Delta_{pix}$, where Δ_{pix} is the detector pixel size (18 μm, see Table 3.2).

This definition is illustrated graphically in Figure 4.3.

It can be instructive to consider the diagram of Figure 4.4. The blue line relates the amount of defocus (x-axis) introduced to the size of the detector pixel (y-axis) that would ensure Nyquist sampling at a wavelength of 1.1 μm. With 200 nm of defocus, the PSF would be critically sampled by a pixel of about 20 μm. The 18 μm sampling provided by the NIRSpec detector, therefore, ensures Nyquist sampling with 1 – 18 μm/20 μm = 10% margin in our simulations.

It should be stressed that the PSFs estimated here provide a representative way to estimate the effect of jitter on the photometric timelines. This means that it is plausible to

[3] ARIEL-SAP-INST-TN-001
[4] Foreseen by R-PRD-8370.

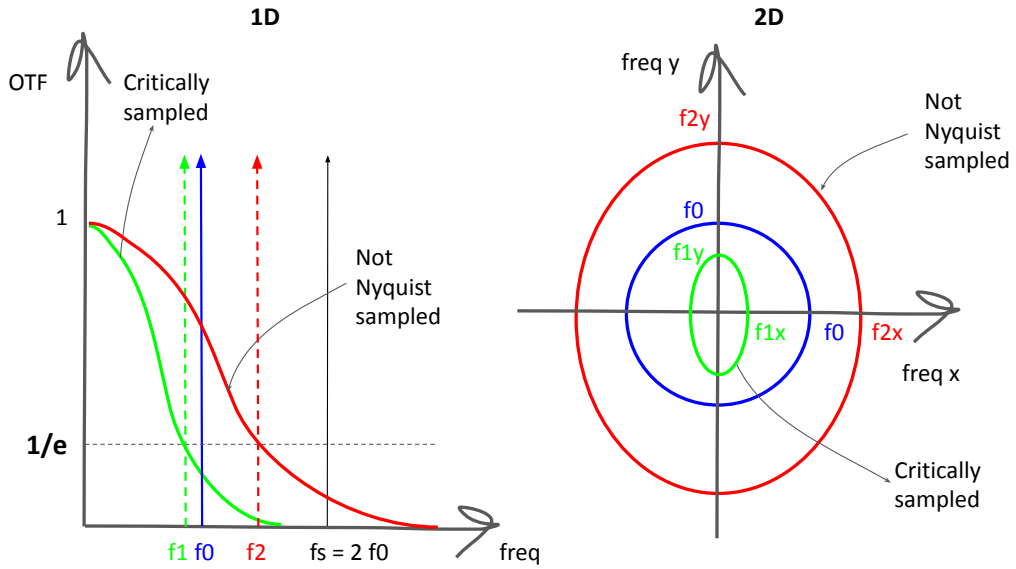


Figure 4.3. *Left panel*: A Nyquist sampling-compliant and a non-compliant example are shown for the 1-D case, in green and red respectively. Compliance ensures that aliasing from frequencies $f > f_0$ is contained. *Right panel*: Both spatial frequencies need to be enclosed within the circle with radius f_0.

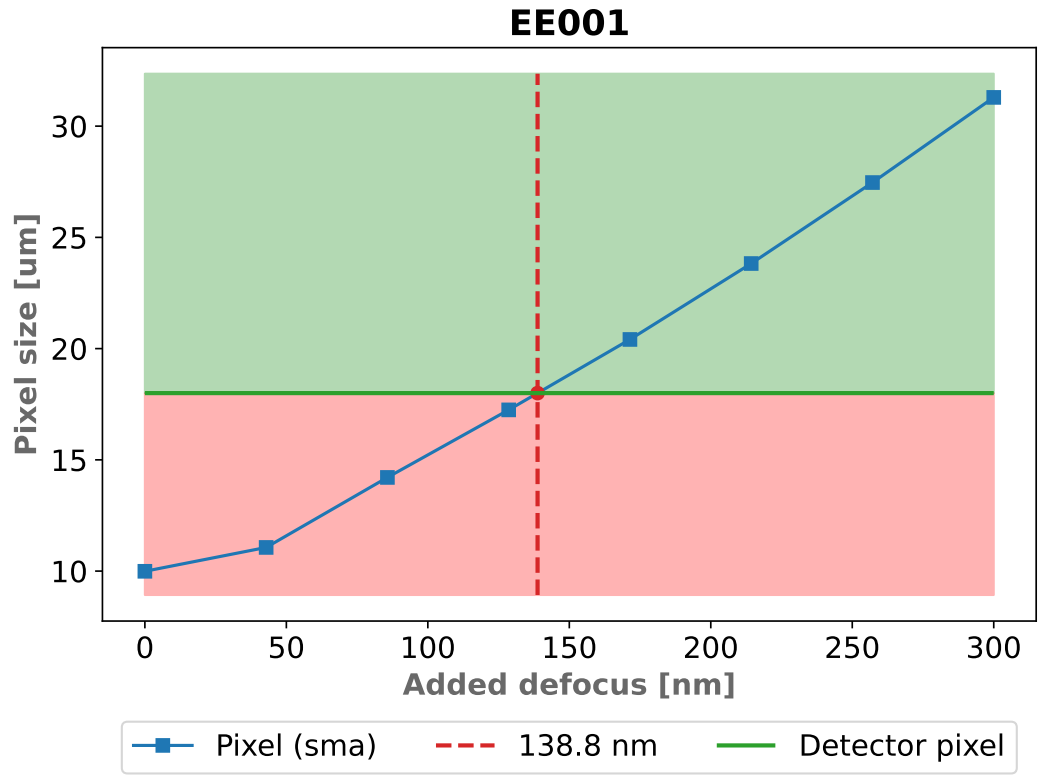


Figure 4.4. Nyquist sampling achieved by the NIRSpec PSF generated by PAOS with increasing defocus. The red shaded area indicates the sub-sampling condition, and the green shaded area indicates the critical sampling condition.

expect that PSFs of different shapes that are critically sampled and therefore comply with the rEE requirements to result in similar photometric performance. At any rate, these PSFs - with 200 nm defocus - have been used in the jitter performance simulations (see Section 3.4 and the next section for details).

4.2 Pointing Jitter Detrending

As part of the Ariel Simulators Software, Management, and Documentation (S2MD) working group, I have participated in key efforts to develop and apply simulation and data analysis tools to ensure Ariel can meet its exoplanet characterization goals. This section focuses on my contributions to the de-jittering analysis of simulated Ariel observations.

The S2MD collaboration encompasses a multidisciplinary team working to build an end-to-end modeling and analysis framework, including the ExoSim 2 and PAOS simulators, detector electronics modeling, and advanced data reduction techniques. Through weekly mini-workshops over multiple years, the team has constructed an ecosystem of tools while training junior members.

While the work featured here was led primarily by myself, Lorenzo V. Mugnai, and Andreas Papageorgiou, it reflects the efforts of the entire S2MD team. My contributions centered on developing and applying data analysis steps to quantitatively assess Ariel's capability to meet photometric stability requirements in the presence of pointing jitter disturbances. The following sections detail this jitter analysis process, methods, and results, representing a critical S2MD achievement on the path to ensuring Ariel's success and legacy.

4.2.1 Pointing jitter timeline models

The pointing jitter timelines are a critical input to assessing Ariel's capability to meet photometric stability requirements. Over several years, our team has compiled a dataset of timelines from various contributors. Early simulated timelines (2019) explored different aspects including Relative Performance Error (RPE), Mean Performance Error (MPE), Performance Drift Error (PDE), transients, and frequency content non-stationarity. A second set of timelines were provided in 2020. The most recent addition to the dataset comes from Airbus (2022), with different cases for bright and faint targets.

In this analysis, we utilize the Airbus timelines represented in Figure 4.5[5]. As evident in the raw timelines, these contain small spikes embedded within the jitter noise. However, these are negligible compared to the jitter noise itself. More impactful are temporal effects such as the RPE variation changing the envelope, and the short-duration "hiccup" in the middle of the timeline that will be discussed later. Other features include linear drift, easily removable by detrending, and a sinusoidal variability attributed to fuel sloshing, negligible for detrending and expected to be reduced. We process the timelines to remove linear drift and sinusoidal signals prior to analysis. This is better shown in terms of Power Spectral Density (PSD) in the figure. After correction, the only remaining feature is the time-varying RPE of the pointing timelines.

Ultimately, the various timelines provided by Airbus behave similarly for the purpose of our performance analyses. Therefore, we focused on a single example, `nom_bright_10h_1kz`[6],

[5] ARIEL-CRDF-PL-AN-005
[6] ARIEL-ADST-SC-FI-000001

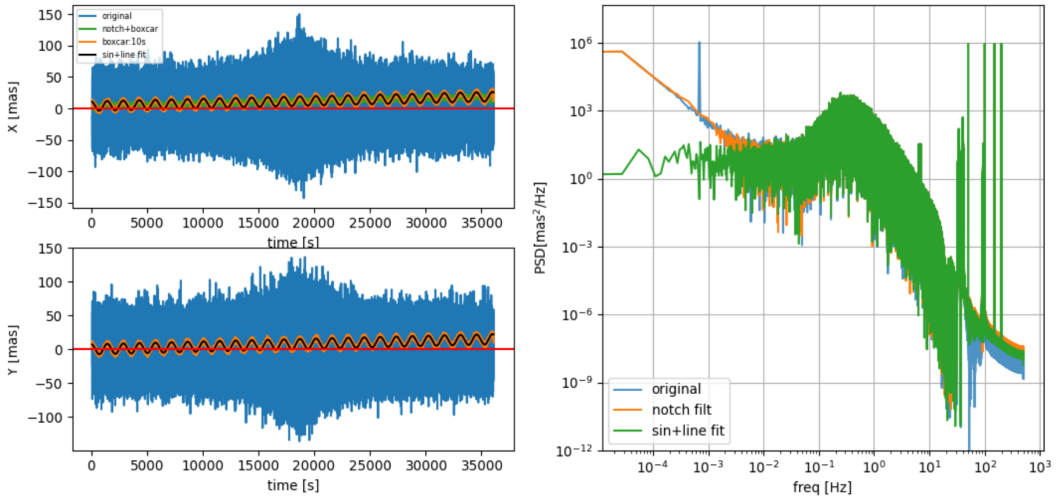


Figure 4.5. Summary of the latest (Mar 2022) AOCS simulations utilized in the analysis. *Left panels*: The raw pointing jitter timelines along x- and y-axes versus time (blue lines). A 10s boxcar average (orange lines) reveals underlying drift and strong sinusoidal components. Fitting the average data (black lines) models the drift and sinusoid. Right panel: Power spectral density (PSD) representation. The raw timeline PSD (blue) features strong harmonics around 7×10^{-4} Hz and 1/f noise below 1×10^{-2} Hz, causing the fitted sinusoidal and linear drift behavior, respectively. Notch filtering of the raw timeline (orange) rejects the sinusoidal component. Subtracting the fitted model yields the corrected PSD (green) without the sinusoidal and linear drift features.

that corresponds to the nominal Reaction Wheel (RW) case, bright stellar target, and 10 hours duration at the sample time of 1 kHz. The value of pointing indexes for this timeline is: 2-mas PDE, 65-mas MPE-RPE, and 60-mas RPE. As this example shows, the PDE, MPE, and RPE are small compared to requirements[7]. Ariel's resilience depends more on stability than absolute jitter - rapid temporal variations are more problematic than e.g. a stable, high-amplitude RPE.

4.2.2 Lessons learned from past analyses

A time-varying RPE has significant implications on how noise integrates down if the effect is not properly corrected. This was highlighted in previous analyses (see Section 3.4 for details) that motivated the present investigation in our detrending pipeline, as this is a known issue that has been successfully addressed in the past. Figure 4.6 illustrates the RPE time variability in the Airbus timelines. The envelope of the jitter noise exhibits a slow rise in Root Mean Square (RMS) toward the timeline center, followed by a decrease back to roughly the initial value. Around the time of peak RMS, the timelines present a "hiccup" feature that originates from variations in vibration frequencies modulated by the S/C Transfer Function. This rapid temporal variation induces photometric instability on timescales of a few minutes, posing a detrending challenge and a potential risk of data loss.

The coupling mechanism between the Line of Sight (LoS) jitter and the resulting photometric noise is as follows. In the presence of slow-varying RMS compared to the exposure rate (~3 s for bright targets) the signal is sampled by tiny amounts of different

[7] find-requirement.

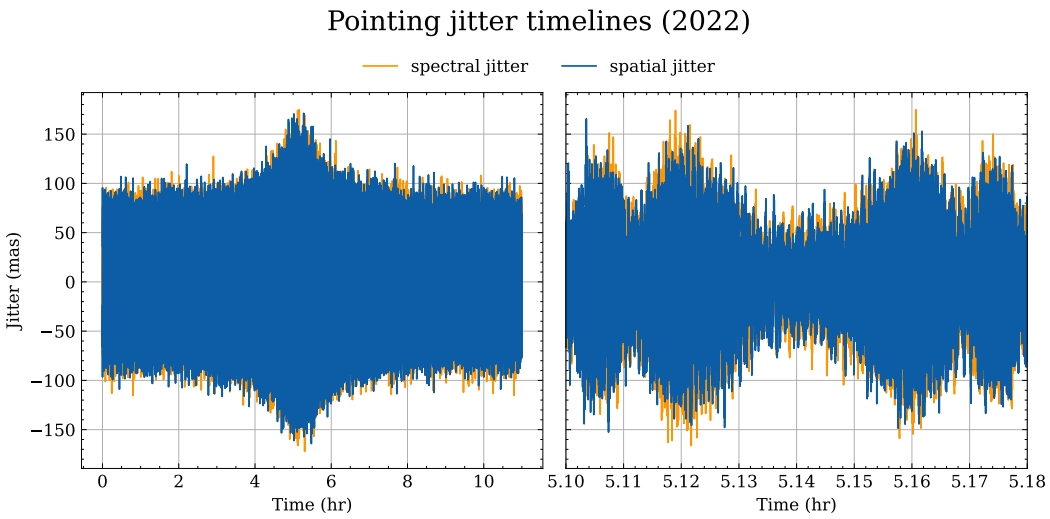


Figure 4.6. Spatial and temporal jitter timelines provided by Airbus and used in the analysis, after subtracting the linear drift and sinusoidal components. *Left*: The extended, 10-hour timelines illustrating the slow change in the jitter envelope. *Right*: A zoomed-in version of the left plot, showcasing the fast change in Relative Performance Error (RPE) around the center of the timeline.

pixels (a 1/10 of a pixel or even smaller) from exposure to exposure. This results in small variations in the position of the PSF. These shifts are not a problem in themselves, rather the photometric instability originates from the PSF distortions caused by variations in frequency content[8]. These frequency content variations arise from focal plane imperfections manifesting as inter-pixel Quantum Efficiency (QE) differences[9] and intra-pixel response variability. The intra-pixel response function (IPRF) is such that the pixel is generally more sensitive near the center and less at the edges (see Figure 4.7 for details).

This effect severely impacted SST/IRAC infrared observations[10]. However, as discussed above, Ariel is designed to provide Nyquist-sampled optical signals, minimizing the IPRF influence on photometric stability. This leaves inter-pixel QE variations, whose impact can be mitigated via flat fielding. The measured PSF depends on the combination of the optical PSF, with the jitter pdf, and the focal plane imperfections. Therefore, it becomes apparent that the measured PSF is distorted by the jitter, as well as the coupling between the jitter and focal plane imperfections. As a result, the measured PSF changes in shape with some skewness and kurtosis. Figure 4.8 illustrates these deformations of the PSF shape for a Gaussian, Nyquist-sampled PSF, jittered on a simple focal plane with randomized QE and constant IPRF. These subtle effects were not corrected by previous jitter detrending analyses, even those including flat fielding, and the resultant residual noise did not integrate down over time as an uncorrelated process. In hindsight, the limitation in accounting for PSF shape variations represents the missing step in our post-processing pipeline that prevented proper noise integration.

[8]Indeed, the jitter of the LoS would have no effect on an ideal detector: provided that the optical signal is spatially Nyquist-sampled, we can always reconstruct the signal shape with no loss of information.

[9]Each pixel has its own QE, i.e. the fraction of the counts that contributes to the photocurrent in the pixel.

[10]See IRAC Instrument Handbook, v.4.0, Figure 8.10

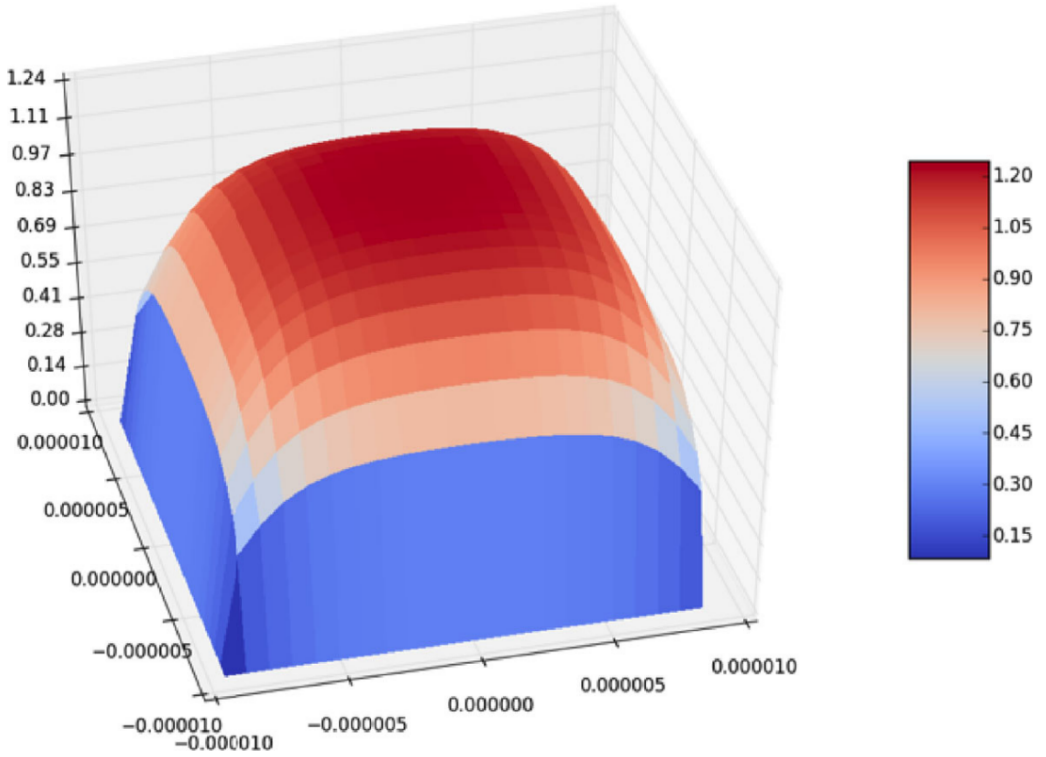


Figure 4.7. 2-D pixel response function used in ExoSim to simulate intra-pixel variation in responsivity (Sarkar et al., 2021). x and y axes are show distance in units of m; z axis shows the responsivity after normalizing the volume to unity.

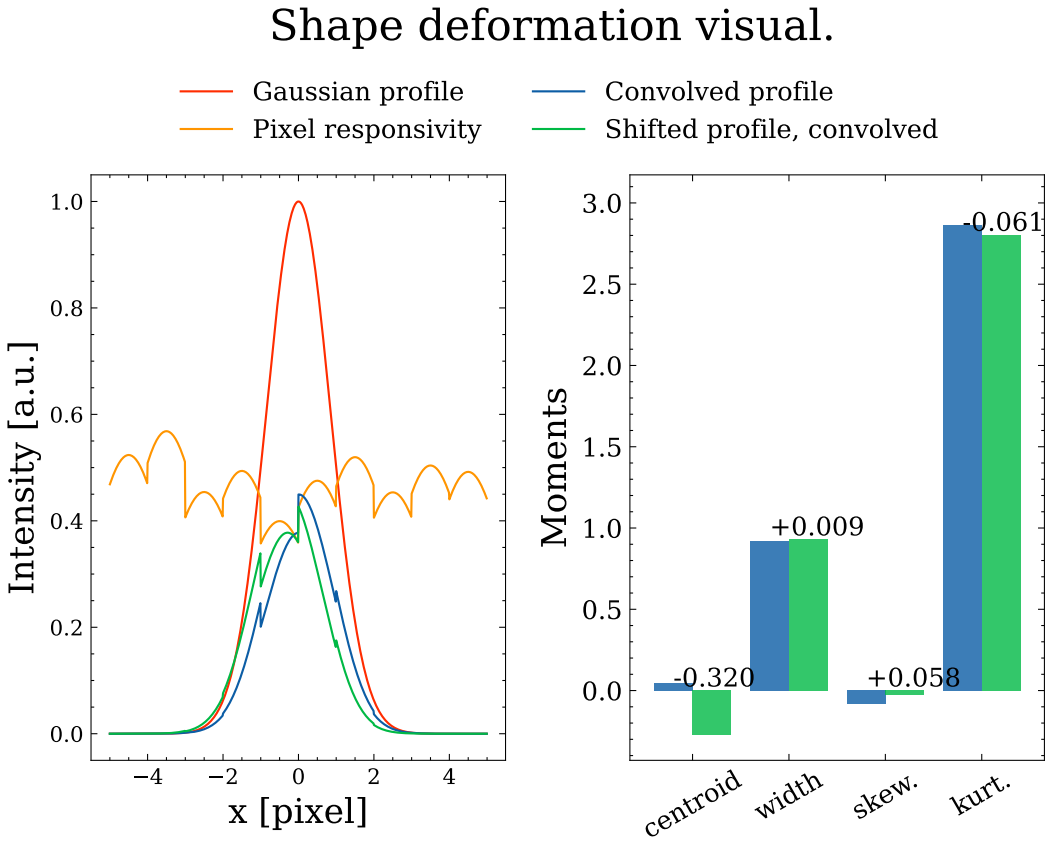


Figure 4.8. Visualization of the shape deformation of the PSF due to the combination of jitter movements and focal plane imperfections. *Left panel*: The PSF is represented with a Nyquist-sampled Gaussian profile (red curve). The illumination profile before and after the jitter (sub-pixel shift of -0.3 pixels) are obtained by convolving the Gaussian profile with the pixel IPRF and multiplying by the QE. The product of the QE and IPRF is shown in the orange pixel responsivity curve for reference. *Right panel*: bar plot of the first order moments of the illumination profile before and after the jitter shift, using the same colors. The variation in the moments with respect to the initial position is reported on top of the bars of the shifted profile.

4.2.3 Constant signals

This section presents the detrending analysis performed using constant timelines, as opposed to time-varying timelines containing planetary signals discussed later. Constant timelines are obtained through stable calibration source observations over time or out-of-transit monitoring of transiting exoplanet systems. We first focus on this simpler case to establish a foundation before addressing more complex signals. Moreover, we focus on photometric timelines; an extension of this work to spectroscopic timelines has already been initiated by the S2MD team, led by student Angèle Syty whom I co-supervised during her three-months internship at La Sapienza University of Rome. Her work has shown promising results, demonstrating that the main methods developed here apply across the Ariel wavelength band.

I produced the signal timelines using the state-of-the-art ExoSim 2, an end-to-end time-domain simulator of Ariel observations (see Section 1.3.3 for details). By implementing a detailed model of the payload design, expected noises, and systematics, it is capable of evaluating representative photometric and spectroscopic light curves. Given the pointing timelines, detector arrays, PSF vs. wavelength, and IPRF, ExoSim 2 calculates the detector array response to the target signal, modulated in position according to the AOCS pointing ("signal integration"). Using ExoSim 2, I simulated a 10-hour jittered timeline of the star HD 209458[11], observed out-of-transit. This simulation contains no astrophysical noise, e.g. variability from the star, and the stellar disk is flat, with no limb darkening. Additionally, to start from a simple case, the stellar flux follows a Planck model.

At specific times, the detector array is read to produce NDRs and when appropriate the array is reset. For implementation simplicity, the simulations used here ignore detector saturation and non-linearity. The detector charge increases linearly without limitation until the reset takes place. Even though the simulation can produce multiple NDRs before resetting, this work has focused on Correlated Double Sampling (CDS) mode, producing just two NDRs per exposure. Given the target brightness, we set the exposure time to 3 seconds between consecutive resets. The simulation also accounts for detector QE and its variation across the array, simulated with a standard deviation of 5% pixel-to-pixel[12]. Additionally, we do not consider detector persistence, which was extensively modeled and found to have a negligible impact on the Ariel application[13], even in the presence of pointing jumps[14].

4.2.3.1 Signal timeline

The output produced by ExoSim 2 for a photometric timeline is a data cube of two-dimensional frames representing the two spatial directions (x, y) indexed by the NDR time stamp. To obtain the "science frames", we apply the CDS and subtract the first frame of each exposure from the second (i.e., last minus first). Then, we perform aperture photometry on each frame to extract the photometric signal. That is, we sum the counts read by the photometer during the exposure within an aperture of specified sizes centered on the PSF. Given the Ariel elliptical primary, the PSF is elliptical, with the semi-major axis oriented as the semi-minor axis of the telescope aperture. Therefore, the photometric aperture is also elliptical.

[11] R-PERF-025
[12] In agreement with past analyses, e.g. ARIEL-CRDF-PL-AN-005
[13] ARIEL-SAP-PL-TN-008
[14] ARIEL-CRDF-PL-AN-005

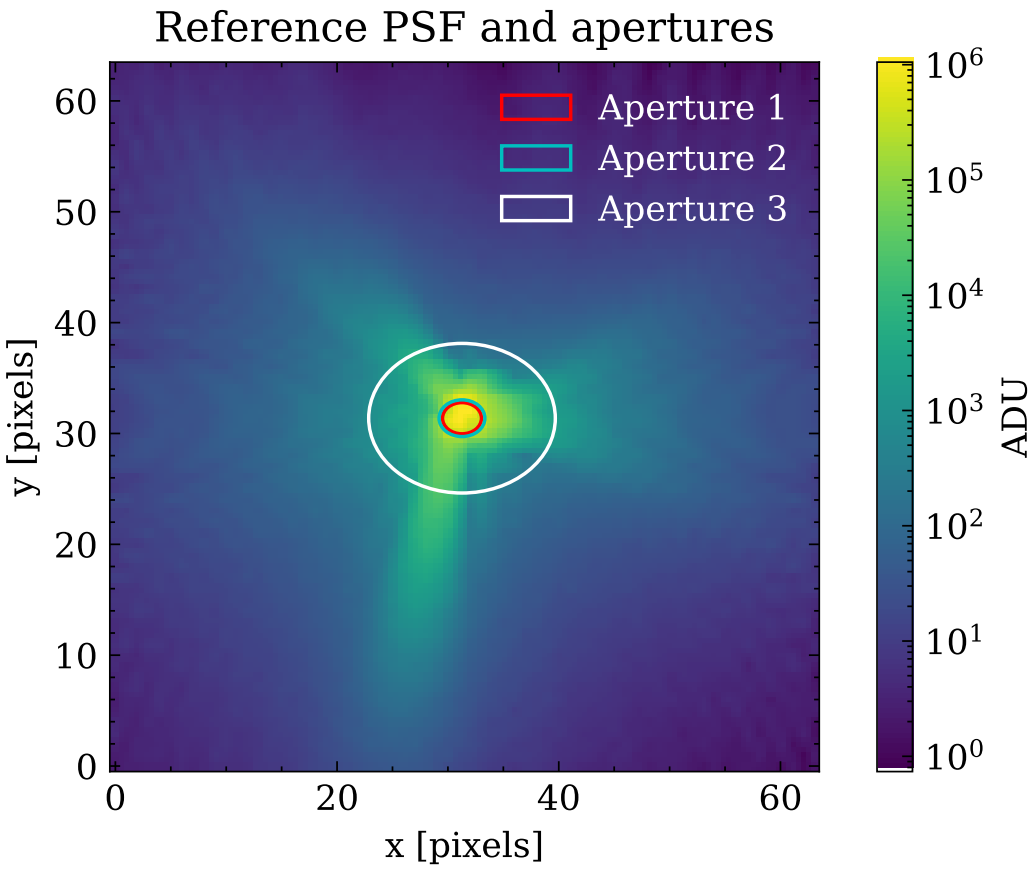


Figure 4.9. Aperture photometry for FGS-1. The image values are the median of the full image stack (reference PSF). The aperture is elliptical, with semi-major axis oriented as the semi-minor axis of the telescope aperture. Aperture 3 is the one used to compute the signal. The other apertures are sized according to the widths of the reference PSF (Aperture 1) and its Half Width at Half Maximum (HWHM) (Aperture 2) along each axis.

Figure 4.9 illustrates the aperture photometry process for the FGS-1 photometer. To have a good estimate of the PSF shape and consequently of the aperture, we start from the median of the image stack to obtain a reference PSF[15]. This choice is not strictly necessary for the analysis, as no transients are present in the simulation. However, it is a good practice to remove outliers and spurious signals that could affect the PSF shape estimate. To determine the aperture sizes, we define a Radius of Encircled Energy (rEE) criterion such that the aperture contains around 95% of the incoming power. The value of this rEE is to an extent arbitrary; however, a larger aperture would incorporate more background noise.

It should be noted that these science frames have considerable Signal-to-Noise Ratio (SNR) and the PSF is very stable, variations from one to the next are tiny. As a consequence, we need high-precision aperture photometry to capture tiny signal variations. Given that changing the aperture size between frames would introduce noise (the estimator gain would change), we center the aperture on the centroid, calculated with sub-pixel precision for each

[15]Here we have a constant illumination from the source and therefore we can take the median of the full image timeline. When dealing with signals containing a transit, we restrict the median operation to the frames collected out of transit.

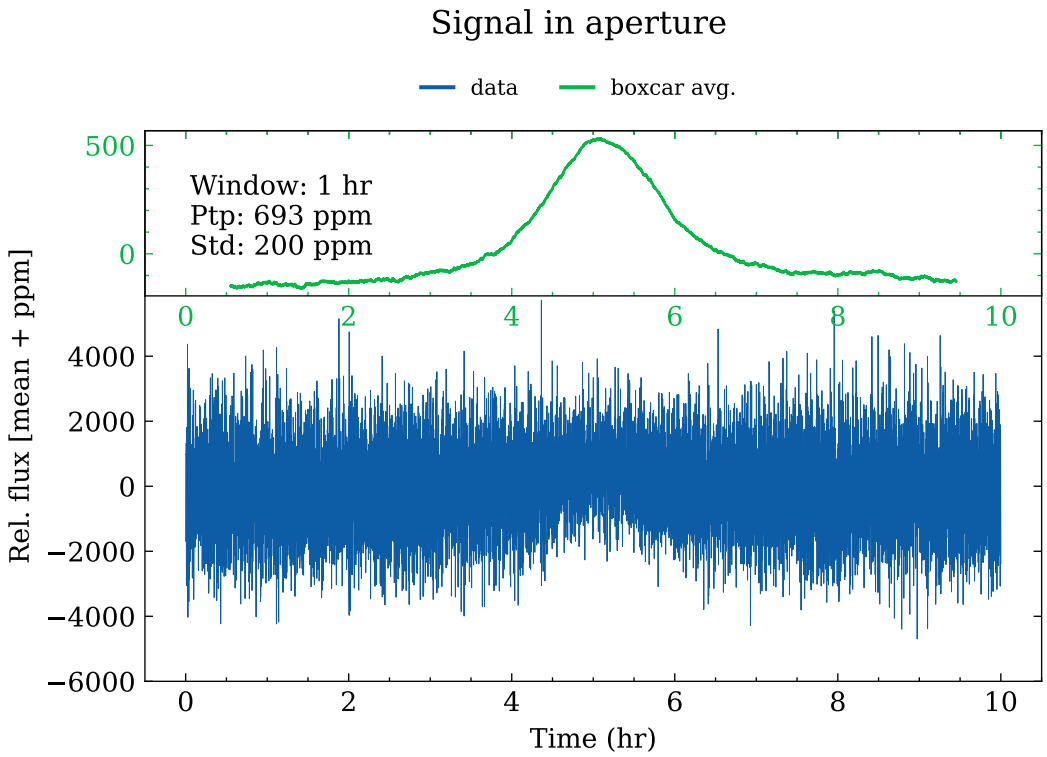


Figure 4.10. An example FGS-1 photometric timeline. *Bottom*: The raw signal timeline with only pointing jitter present - photon noise and read noise are not included here. The data is plotted in ppm, after subtracting the mean. *Top*: A low-pass filtered version of this timeline. The data were boxcar-averaged using a window of 1 hour to show temporal trends at the observation timescale.

frame, and apply the same aperture throughout. This should correspond to having a fixed PSF for optimal photometry and then correcting for gain variations. Finally, the photometric signal is normalized by the total counts in the first frame. This is done to work with values close to unity.

Figure 4.10 shows an example signal timeline for the FGS-1 photometer, in the presence of pointing jitter noise only. Photon and readout noise add a layer of complexity for jitter detrending and are discussed later. The other photometer channels (VISPhot and FGS-2) behave similarly and the main analysis findings apply to them as well. Therefore, in this section, we focus on the FGS-1 channel. In addition, unless specified differently, we do not apply flat fielding (i.e., divide the CDS data by the distribution of QE across the detector array, assuming perfect knowledge or within the required mission uncertainty).

The raw signal timeline is represented in the bottom diagram, after subtracting the mean to highlight the ppm variations. A low-pass filtered version of this timeline is represented on top, averaged using a window of 1 hour. This filtered timeline illustrates the large photometric variation in the middle of the observation, due to the RPE envelope change as discussed previously. The peak-to-peak (Ptp) of this timeline is around 0.7 ppt, while the raw timeline oscillates between around ±4 ppt. The standard deviation of the filtered timeline is around 200 ppm at the timescale of 1 hour, 10 times more than the bright targets

requirement[16]. Photometric variations of this magnitude would severely impact the science data Ariel will provide, if not corrected to within limits. Other features visible in the raw timeline are spikes and the shrinking of the timeline envelope toward the center. The contraction is most significant in correspondence with the “hiccup” feature observed in the jitter timelines and is likely due to this RPE variation occurring suddenly.

4.2.3.2 Decorrelation algorithm

Originally, when investigating this "hiccup" feature, we hypothesized it could be caused by PSF motion. However, we later realized the photometric variations correlated with PSF distortions rather than shifts, thanks to the work led by M.Sc. student Leonardo Altamura, whom I co-advised. In this analysis, I estimate not just image centroids but also the width, skewness, and kurtosis - effectively all low-order moments of the light distribution along x and y - from each exposure as a function of time. Figure 4.11 illustrates the time series of all these moments, showcasing the wealth of spatial information contained in the images and some rather obvious correlations with the signal timeline.

I then cross-correlate these parameters with the photometric signal over time and remove the identified correlations from the data. This processing significantly reduces the photometric variability, demonstrating a nearly one-to-one mapping between the jitter-induced noise and the first four moments of light distribution. By accounting for position, width, skewness, and kurtosis variations, the jitter effect can be almost perfectly removed without having to perform a flat fielding (see later for details).

For sufficiently small jitter, the jitter-induced variability f can be described by a first-order Taylor expansion of those moments:

$$f(t, \vec{X}) = f(\vec{X}_0) + (\vec{X} - \vec{X}_0)^T \nabla f(\vec{X}_0) \tag{4.1}$$

where $\vec{X} = (x, y, w_x, w_y, s_x, s_y, k_x, k_y)$ is the vector containing the position, width, skewness, and kurtosis along x and y[17]. While including higher-order terms[18] can improve the model, it increases the complexity significantly. Moreover, as shown later, in the studied case a first-order approximation yields a decorrelated signal that is already within the specification. Therefore, the jitter model effectively reduces to a linear combination of the moments of the light distribution:

$$f = c_0 + c_1 x + c_2 y + c_3 w_x + c_4 w_y + c_5 s_x + c_6 s_y + c_7 s_x + c_8 k_x + c_9 k_y \tag{4.2}$$

where $c_{0..9}$ are the decorrelation model coefficients, determined by fitting the time-domain photometry data, and we dropped the explicit time dependencies. The obtained model of the jitter-induced variability is then removed from the raw signal timeline.

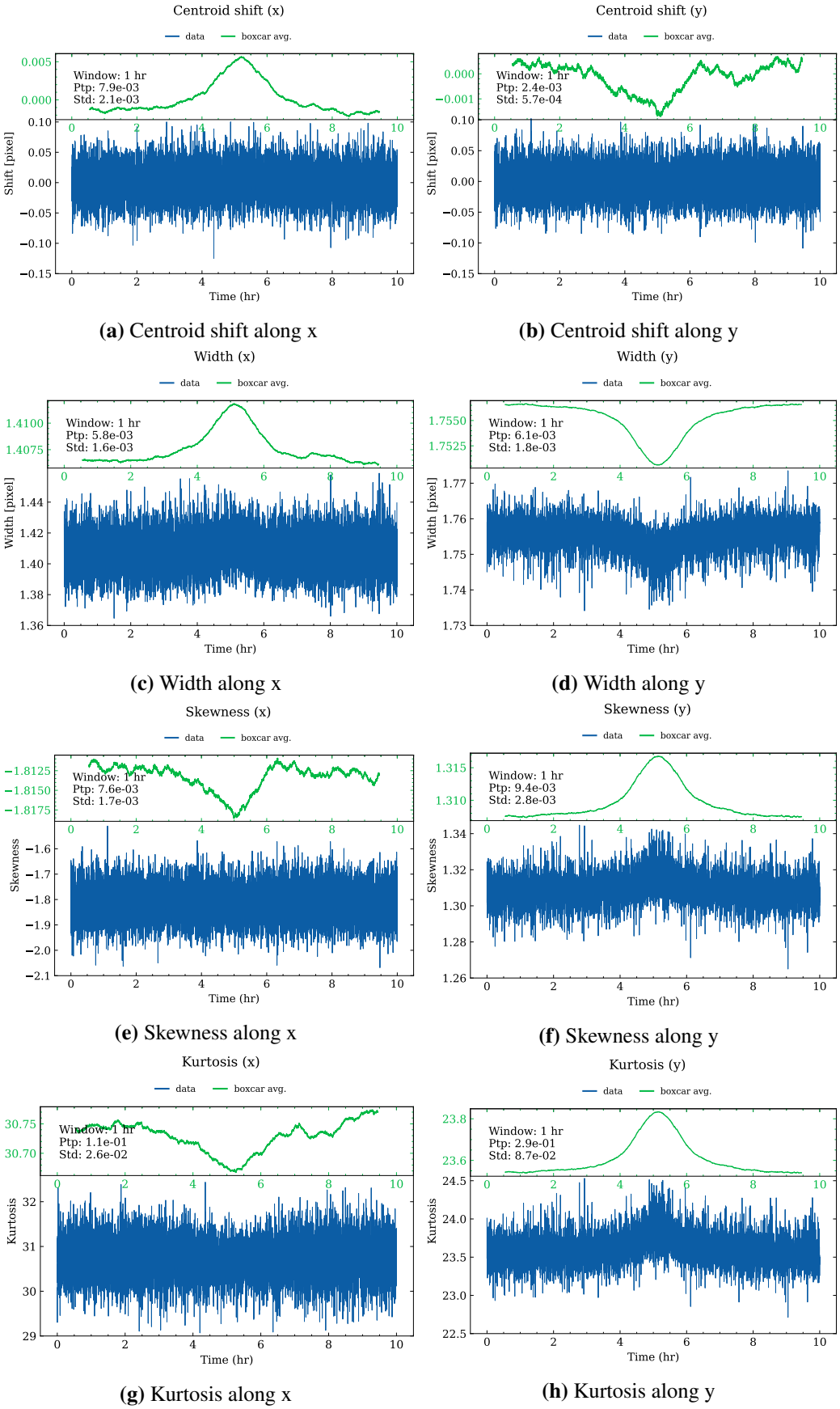
In essence, the decorrelation algorithm implemented is a three-stage process:

1. Estimate position, width, skewness, and kurtosis from the photometric data;
2. Simultaneously correlate the estimated parameters with the time-domain photometry;
3. Remove the identified correlations from the photometric data.

[16] R-PRD-8927

[17] In contrast, for spectroscopic data, for each spectral bin only the spatial dimension is available.

[18] e.g. $\frac{1}{2}(\vec{X} - \vec{X}_0)^T H_f(\vec{X}_0)(\vec{X} - \vec{X}_0)$, where H is the Hessian.



(a) Centroid shift along x

(b) Centroid shift along y

(c) Width along x

(d) Width along y

(e) Skewness along x

(f) Skewness along y

(g) Kurtosis along x

(h) Kurtosis along y

Figure 4.11. Time series of the low-order moments of light distribution for the same FGS-1 timeline.

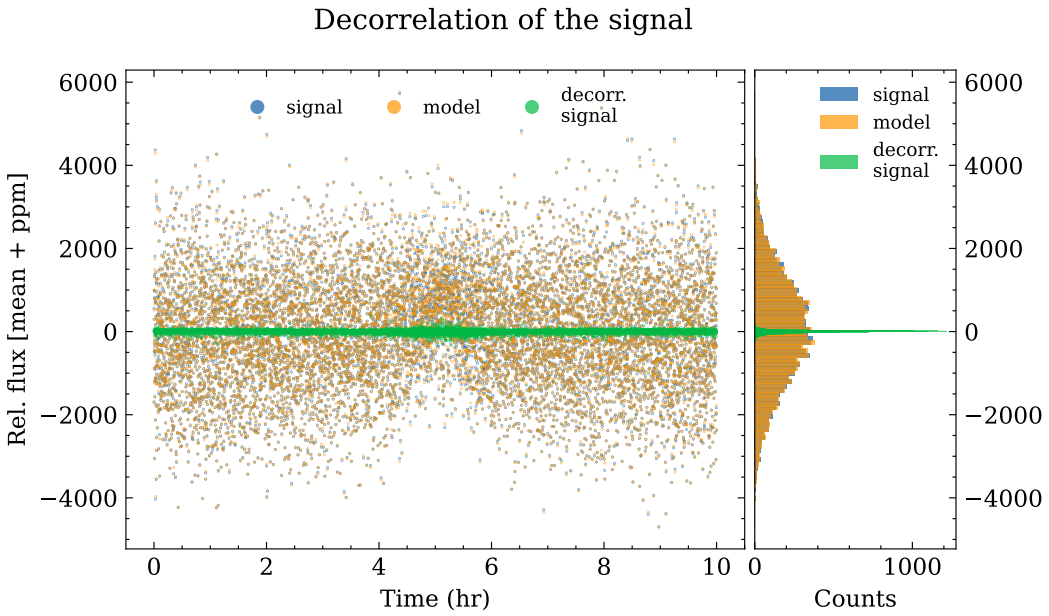
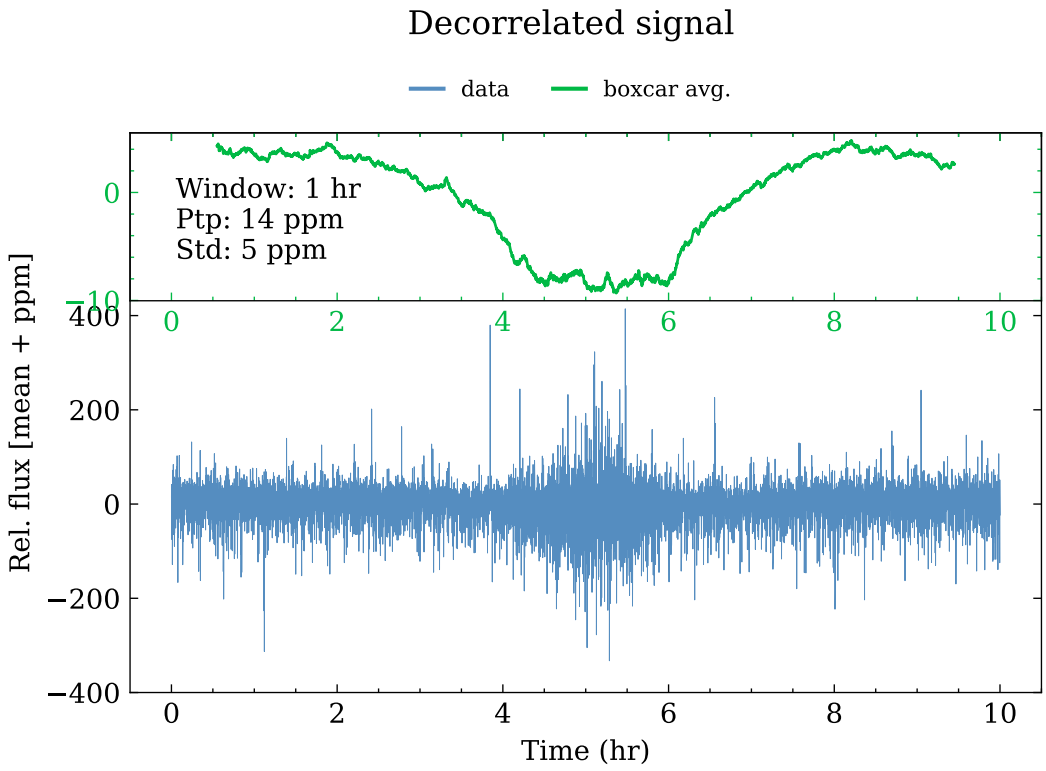


Figure 4.12. An example of jitter decorrelation using the same raw photometric timeline shown in the previous figure. *Left*: the timelines of the raw signal (blue dots), the jitter model (orange), and the decorrelated signal (green), after mean subtraction. Note that the orange dots almost perfectly superimpose the blue dots. *Right*: histogram of the counts obtained from the timelines shown on the left, after collapsing the temporal axis.

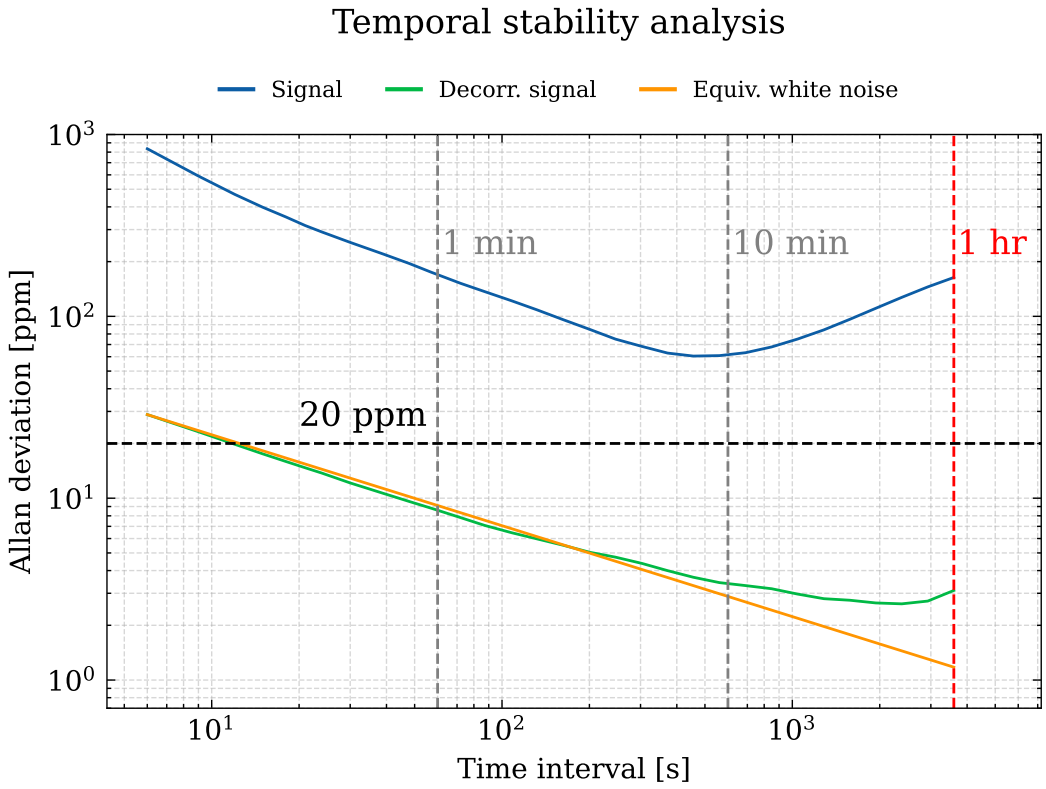
4.2.3.3 Decorrelated signal

As shown in Figure 4.12, the detrending algorithm described above is enough to reduce the variability in the raw signal timeline by more than an order of magnitude, using solely the spatial information contained in the science frames, and no calibration information. There is still a slight bump in the middle of the decorrelated signal timeline, where the model is not perfectly capturing the variability associated with the "hiccup" feature in the pointing. By including higher-order terms, we can achieve a better fit, with the caveats exposed above. Apart from this, the signal after decorrelation is almost perfectly constant, with residual variability that is better illustrated in Figure 4.13a. This figure can be directly compared with Figure 4.10. The residual variability is most significant in the middle of the timeline and has a Ptp of around 14 ppm (decreased by a factor of ~50), and a standard deviation of 5 ppm (decreased by a factor of ~40) when boxcar-filtered using a 1-hour-long window.

To measure the stability of the frequency content in the decorrelated signal, we use the *Allan deviation* (ADEV) (e.g. Riley and Howe, 2008). Also known as *sigma-tau*, the ADEV is a metric designed to overcome the limitations of the standard deviation (Allan, 1987), having the advantage of being convergent for most noise types. This is achieved by using the first differences of the fractional frequency values, rather than deviations from the average as in the standard deviation, which is not stationary for the more divergent noise types. Also, the ADEV is more informative than the standard deviation and can distinguish between different noise types, e.g. white and colored noise. The ADEV is usually employed to present results instead of the *Allan Variance* (AVAR) as it gives the relative amplitude stability and can be directly compared to other sources of errors. For instance, an ADEV of



(a) An example FGS-1 photometric timeline, after jitter decorrelation. Raw timeline (bottom) and low-pass-filtered version (top). To be compared with the same figure for the original timeline.



(b) The stability analysis performed on the same FGS-1 photometric timeline. The overlapping Allan deviation (ADEV) for the original signal (blue) and the decorrelated signal (green). The orange line represents the expected deviation for the decorrelated signal, were it random white noise.

Figure 4.13. Time-domain stability analysis of the decorrelated signal vs. the original.

e.g. 20 ppm at averaging time $\tau = 1$ hr is interpreted as there being a frequency instability between two measurements 1 hour apart with a relative amplitude of 20 ppm RMS. Notably, there is no information difference between the AVAR/ADEV and a power spectral density representation: one can be transformed into the other (and vice versa) by convolving it with the appropriate transfer function (Allan, 2016). Lastly, the original AVAR has been largely superseded by its overlapping version, which improves the confidence of the resulting stability estimate. This version makes maximum use of a data set by forming all possible overlapping samples at each averaging time. The overlapped ADEV is the most common measure of time-domain frequency stability. Therefore, we refer to this form of the metric unless otherwise specified. In addition, throughout this analysis, we use the Python package *Allantools*[19], which provides an extensive library for calculating the ADEV and related time and frequency statistics.

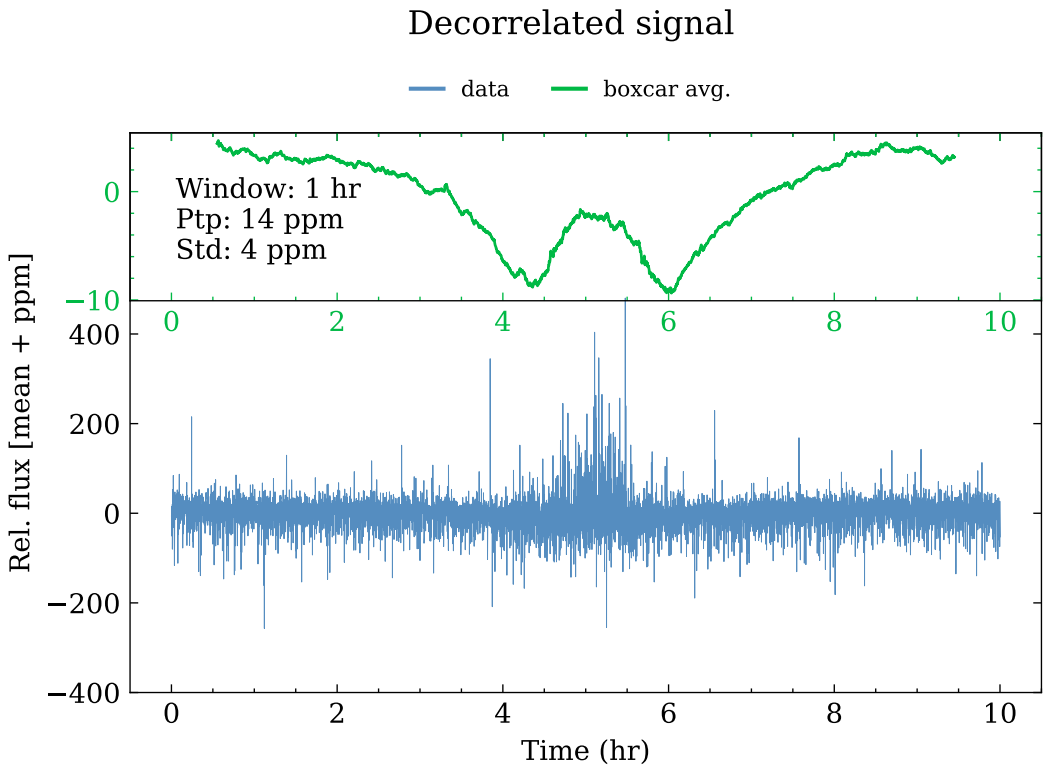
Figure 4.13b illustrates the ADEV of the decorrelated signal vs. the original signal and a reference white noise timeline. The latter is a line with a slope of -0.5, normalized to the value of the decorrelated signal ADEV at the shortest averaging time. Therefore, in the log-log plane, it represents the ADEV curve for the decorrelated signal if it were entirely white noise. This figure allows us to evaluate frequency stability at different averaging times, capped at 1 hour to ensure an adequate number of samples (at least 10, given the length of the simulation) for statistical analysis. The bright targets requirement at 1 hour is represented with a dashed line at 20 ppm. Vertical dashed lines at visualize relevant times (1 min, 10 min, and 1 hour). For both ADEV curves, random white noise processes are dominant at timescales up to ~few minutes and the noise integrates down roughly as a time-uncorrelated process, i.e. as root time. These fluctuations correspond to harmonic fast movements of the line of sight of the telescope. At longer timescales, time-correlated processes become more important, and the ADEV negative slope decreases, i.e. the noise integrates down ever less efficiently with time.

While it is possible to quantify the prevalence of different noise contributions (e.g. Brown noise), it is not the scope of this analysis and is deferred to future work. Here, we focus on demonstrating the level of stability that can be reached given the decorrelation implemented, and we observe a drop in photometric noise by a factor ~20 at short timescales and ~50 at the longest timescales compared to the original signal. At the 1-hour timescale, the decorrelated signal ADEV tapers around 3 ppm, fully compliant with the requirement and consistent with the standard deviation of the low-pass filtered signal discussed previously. As a side note, the fact that at short timescales the noise integrates slightly faster than root time is not yet completely understood and a tentative explanation is that the model is not entirely immune from overfitting. Additionally, with regard to the behavior at longer timescales, we must bear in mind that (i) the reduced number of samples is likely not statistically significant, reducing the level of confidence and (ii) the numerical resolution of the simulation is at a comparable level (~1 ppm).

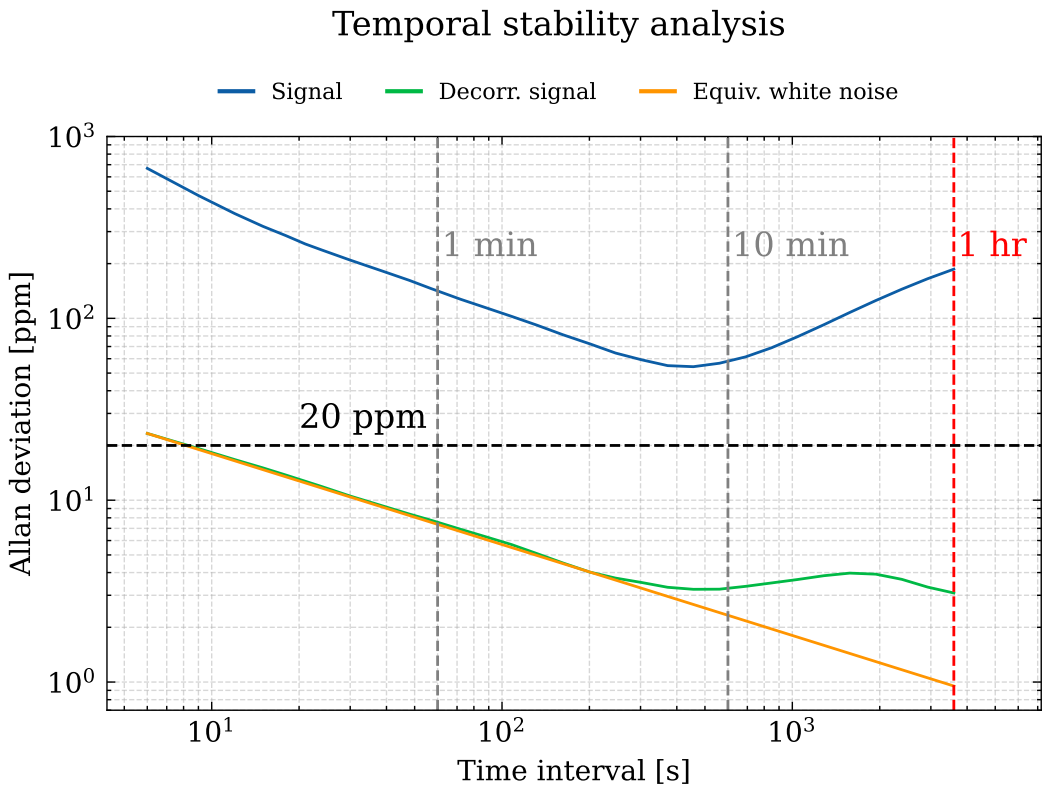
4.2.3.4 Flat fielding

The flat fielding operation is expected to improve our ability to detrend jitter-induced photometric noise compared to the previous results by virtue of removing the impact of inter-pixel responsivity variations. Figure 4.14 illustrates the outcome of the same detrending analysis, preceded by the flat fielding step. The science frames are flat-fielded by dividing

[19] https://allantools.readthedocs.io/en/latest/index.html.



(a) An example FGS-1 photometric timeline, after flat fielding (assuming perfect QE knowledge) and jitter decorrelation. Raw timeline (bottom) and low-pass-filtered version (top).



(b) The stability analysis performed on the same FGS-1 photometric timeline. The overlapping Allan deviation (ADEV) for the original (blue), decorrelated (green), and white noise equivalent signal (orange).

Figure 4.14. Time-domain stability analysis of the decorrelated signal vs. the original one, having flat-fielded the images. To be compared with the same figure for the case of no flat fielding.

with the QE matrix, i.e. a map of the pixel response of the detector to uniform illumination. This QE matrix is a simulation parameter in ExoSim 2 and is provided with the simulation output. The pixel-to-pixel responsivity variations contained in this map are within 5%. The operation of dividing by the "true" QE matrix is the equivalent of perfect flat fielding.

A more realistic scenario can be examined by adding additional uncertainty to this QE map, before dividing the frames. Among the mission baseline calibration requirements is to achieve better than 0.5% uncertainty on the QE knowledge[20]. Although in this analysis we also examined adding an uncertainty of 0.5%, we choose not to include this case for brevity and present solely the results assuming perfect knowledge of QE. This is a limiting scenario that aims to show the full extent of the improvement (if any) that can be obtained by performing the flat field correction.

Compared to Figure 4.13, the stability improvement appears very small even in this idealized condition. The Ptp of the low-pass-filtered timeline at 1 hour is the same (rounded to the nearest ppm), and the standard deviation is ~1 ppm less, possibly due to the model's better fit near the "hiccup" feature in the middle of the timeline. Accordingly, the ADEV is slightly ameliorated in the following: (i) at short timescales, the ADEV perfectly follows the expected trend for white noise, (ii) the knee of the curve, where it deviates from white noise, is slightly lower, and (iii) at 1 hour timescale the noise seems to still integrate down. All this indicates an overall reduced power of time-correlated processes, although the improvement is very minor, in the order of 1 ppm at 1 hour. Under baseline conditions for QE knowledge, the improvement is further reduced, and the flat-field operation could introduce bias. Moreover, this simulation does not contain any other noise source. When including e.g. the photon and read noise contributions, this ~1 ppm improvement is virtually inconsequential.

Overall, this analysis seems to indicate that the decorrelation method implemented fully exploits the spatial information contained in the science frames and is capable of removing the jitter-induced photometric noise efficiently. Flat fielding only marginally reduces noise levels after jitter detrending. The minor gain from flat fielding may not warrant the potential introduction of bias. Since the noise levels achieved through jitter decorrelation alone already satisfy Ariel's stability requirement, flat fielding could potentially be avoided as was done successfully with SST/IRAC observations. This is not to suggest that calibration flats should not be taken meticulously, for (a) this analysis is based on simulated data, and (b) they still contain useful spatial information that can be exploited for sanity checking. However, a flat field calibration is not as critical as previously thought. Using this lesson, the rest of the analysis described here assumes no flat fielding.

4.2.3.5 Noised timeline

In this section, we utilize the same simulation and include photon and read noise alongside jitter to reproduce a more realistic observation than the previous example. Photon noise is by definition a quantized and thus time-uncorrelated process, follows Poisson's law and integrates down as root time. Readout noise is produced in the readout process by the amplifier chain and can be modeled as a Gaussian white noise contribution. The amount of readout noise assumed is 20 electrons per CDS per pixel, as specified in Table 3.2. This analysis aims to demonstrate that, even with the introduction of photon and read noise, the decorrelation procedure can still detrend the variability induced by jitter below the required level.

[20] R-CAL-070

To compare the results obtained post-processing with the jittered signal against what could be achieved in the ideal case where the only noise sources are photon noise from the star and read noise, we generate a reference unjittered timeline. This timeline is obtained by repeating the first frame of the unnoised jittered timeline for the whole duration of the observation. This is equivalent to having a perfectly stable PSF and no jitter. Then, photon and read noise are added to this timeline to obtain the reference unjittered timeline. By comparing the stability of the decorrelated signal with that of the reference unjittered timeline, the best-accomplishable stability, we can evaluate how much of the jitter-induced noise is removed by the decorrelation algorithm.

Compared to the previous case, the introduction of photon and read noise decreases the sensitivity to the tiny photometric variations from exposure to exposure caused by jitter. To counter this, we utilize a weighting algorithm to evaluate the moments of the light distribution. The idea is to implement a Wiener filter[21] that enhances the underlying signal compared to the noise level, decreasing the influence of the noise-dominated tails. As a first approximation, we employ a Gaussian weighting function and compute the moments of the enhanced light distribution, i.e. the light distribution of each frame convolved with the corresponding weighting function. The implementation is analogous to the algorithm described in the centroiding report for the Ariel FGS[22].

Our implementation of the weighting algorithm is as follows. As detailed previously in the photometric aperture discussion, first the median of the stack of images in the timeline is computed to have a reference image. Then, the width of the reference PSF is computed from the reference image, by estimating the width of the light distribution along the corresponding axis from a moment's calculation. Before using this width in the weighting function, it is enlarged by a factor $2\sqrt{2}$ to avoid cutting signal frequencies[23]. The weighting function is then computed as a Gaussian with a standard deviation equal to the enlarged width, centered on the centroid of each frame. Using the same width throughout assures that the gain of the estimator is constant. Moreover, by updating the position of the weighting function for each frame, we ensure that the moments are computed on the same region of the PSF, centered on the centroid. Therefore, they are homogeneous and comparable across frames. The moments of the enhanced light distribution are then computed for each frame and used in the decorrelation algorithm described previously.

Figure 4.15 reports the results obtained for the temporal stability of the decorrelated signal and the reference unjittered signal. 10 different noise realizations were used to estimate the mean and standard deviation of the ADEV curve obtained in both cases. The curve for the signal before decorrelation is also shown. The reference ADEV curve is ∼few ppm lower than the equivalent white noise curve for the mean decorrelated signal ADEV, estimated as discussed above. However, at longer timescales ($\gtrsim$10 min), the decorrelated signal ADEV deviates from the expected behavior for white noise, and reaches 13.7±2.5 ppm at the 1-hour timescale, compared to 9.6±1.8 ppm of the reference. Notwithstanding, this analysis demonstrates compliance of the decorrelated signal with the 20 ppm requirement at 1 hour. The residual correlated noise can be dealt with by including second-order terms in the jitter model; however, this may lead to overfitting.

[21] A type of filter used in signal processing whose goal is to compute a statistical estimate of an unknown signal using a related signal as an input and filtering that known signal to produce the estimate as an output.

[22] ARIEL-UVIE-PL-TN-001

[23] This factor corresponds to the best-performing factor for the FGS-1 photometer from a coarse grid search.

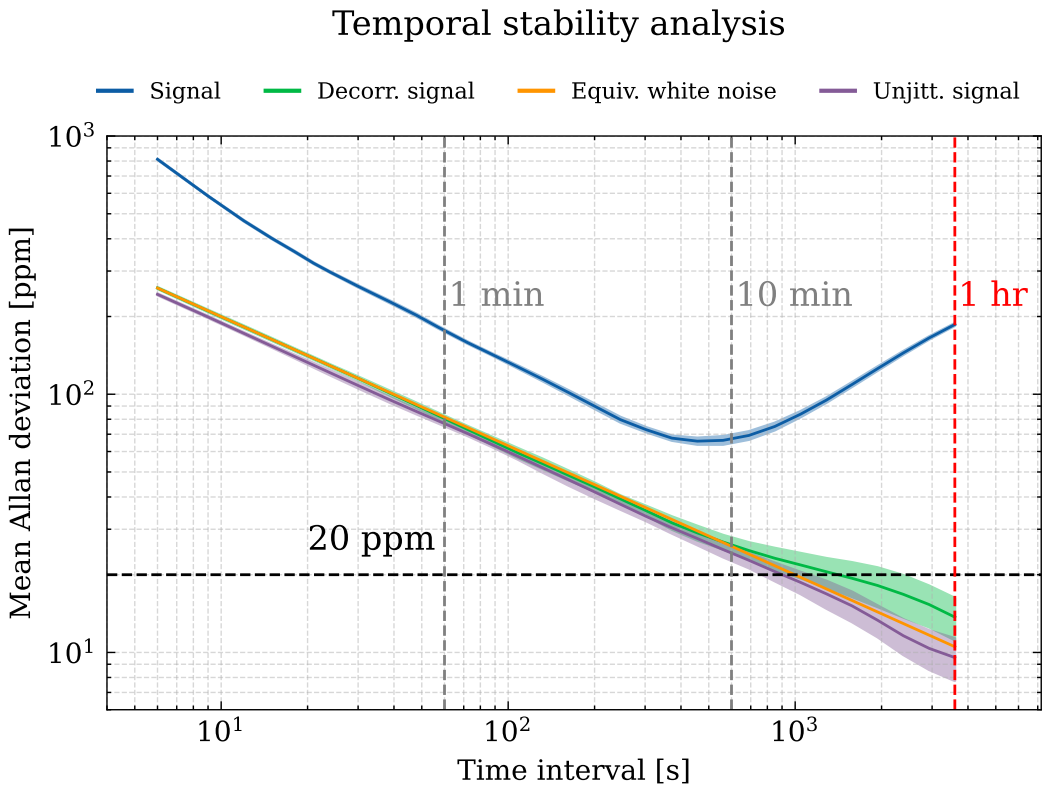


Figure 4.15. Time-domain stability analysis of the same FGS-1 photometric timeline, with the addition of photon and read noise. The overlapping Allan deviation (ADEV) for the original (blue), decorrelated (green), white noise equivalent (orange), and un-jittered signal (purple). The solid lines are the mean of the ADEV from 10 different noise realizations, and the shaded areas represent the corresponding standard deviation.

4.2.3.6 Time averaging

While the decorrelation algorithm described above is capable of removing the jitter-induced photometric noise to within the required level, even in presence of photon and read noise, further improvements are possible, as the decorrelated signal ADEV curve is still above the reference unjittered signal ADEV curve. In this section, we investigate the possibility of stacking frames on time as a way to simulate a longer time exposure. This is conceivable because working at a faster exposure frequency than the astrophysical signal of interest is not mandatory. Stacking frames is equivalent (barring noise contributions) to integrating in time, increasing the SNR and averaging jitter, such that its impact on the photometric stability is reduced.

For transit spectroscopy, the relevant timescale is the transit duration, i.e. up to a few hours. Fully sampling the ingress and egress of the transit event for bright targets[24] requires a cadence of less than about 90 s. In the following, we consider a stacking factor of 10, i.e. simulating a 30 s exposure time by averaging successive groups of 10 exposures together. Figure 4.16 illustrates the resultant ADEV curves for the decorrelated signal and the reference unjittered signal. From the ADEV of the original signal, we can see that compared to the case of no time averaging, the stacking of frames has a very limited effect on improving the stability if no decorrelation is performed. This is expected, as the jitter is a time-correlated process, and averaging does not integrate down the correlated component over longer timescales.

Instead, the decorrelated signal ADEV, shown in green, is significantly improved by averaging the frames. At the 1-hour timescale, the ADEV is 11.6±2.1 ppm compared to 13.7±2.5 ppm for the non-decimated case. This is still above the reference unjittered signal ADEV (9.6±1.8 ppm), however, the difference has now been halved from ~4 ppm to ~2 ppm. This is a significant improvement over the non-decimated case, demonstrating the effectiveness of combining the decorrelation algorithm with time averaging to achieve even greater post-processing stability.

4.2.4 Transit signals

This section presents the jitter-detrending analysis performed using time-varying timelines containing planetary signals. These timelines are produced natively by ExoSim 2, which includes a dedicated module for loading and applying different astronomical signals to the target source (see Section 1.3.3 for details). Here, we simulate the transit light curve of planet HD 209458 b using a simulated atmosphere with a fixed transit depth vs. wavelength and limb darkening coefficients all set to zero as an example of an astronomical signal. The default model for the transit light curve is based on the `batman-package` (Kreidberg, 2015). To investigate what is possibly a worst-case scenario, we implement a mid-transit time at the middle of the timeline, where the RPE amplitude and rate of change are greatest.

4.2.4.1 Decorrelation algorithm

The algorithm to decorrelate the jitter in constant photometric timelines must be updated to account for the presence of the transit signal - the relative flux decrease during planet occultation. Since the transit and jitter effects are orthogonal, the updated model is the

[24] R-PRD-0370

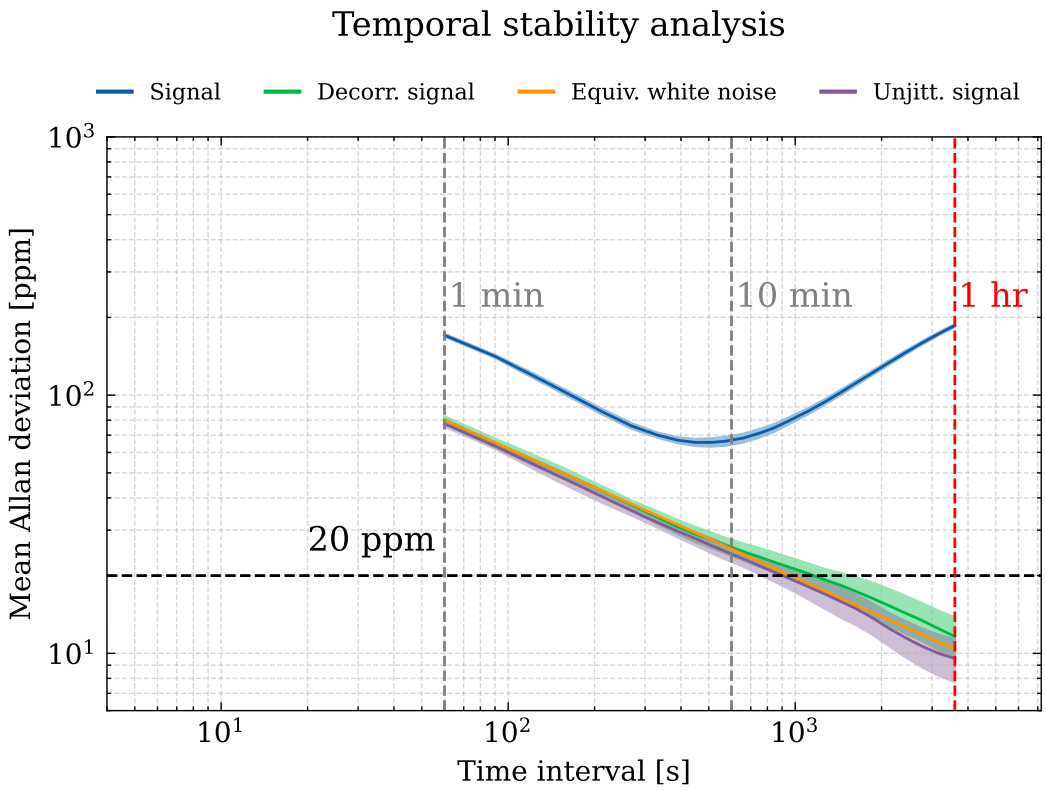


Figure 4.16. Time-domain stability analysis of the same FGS-1 photometric timeline, with the addition of photon and read noise, stacked on time by a factor of 10 (effective exposure time of 30 s). To be compared with the same figure for the case of no time averaging.

product of the jitter model and the transit light curve:

$$\mho = f \times LC \tag{4.3}$$

In the decorrelation step, all free parameters are fitted simultaneously: $c_{0..9}$ for the jitter model and the transit light curve's parameters, e.g. time of mid-transit (t_0), planetary radius over stellar radius (k), orbit semi-major axis (a), inclination (i), and limb darkening coefficients. Here, the free parameters for the light curve are t_0, k, a, and i; all other transit parameters are assumed known.

Therefore, in the presence of a transit signal, the detrending algorithm becomes:

1. Estimate position, width, skewness, and kurtosis from the photometric data;
2. Simultaneously correlate the estimated parameters with the time-domain photometry, together with the science signal model and its parameters;
3. Remove the identified correlations from the photometric data (superfluous).

4.2.4.2 Decorrelated signal

Figure 4.17 illustrates the results obtained by applying the detrending algorithm described above. The residual variability in the decorrelated light curve is largely attributable to the photon and read noise components. However, a small excess noise is visually evident mid-timeline. From inspection of the decorrelated timeline alone, it is difficult to ascertain if any time-correlated behavior remains. To quantify this excess, Figure 4.18a plots the ADEV of the residual signal, $\mathbb{R}$ obtained from the difference between the detrended transit signal and the true transit light curve:

$$\mathbb{R} = \mathbb{D}/f - \text{True LC} \tag{4.4}$$

where $\mathbb{D}$ is the photometric data. As done before, 10 different noise realizations are utilized to estimate the mean and standard deviation of the ADEV curve, shown alongside the equivalent white noise one. The slope for the decorrelated signal follows the expected curve up to time intervals of ~15 min, then increases slightly. At 1 hour, the ADEV is 13.6±2.4 ppm, comparable to what was obtained for the constant noised timeline. Therefore, we demonstrated that including the transit signal does not interfere with the jitter decorrelation, confirming that the astronomical signal and the jitter are fully separable using this algorithm. The jitter noise is corrected to a large extent, and the level of photometric precision that can be achieved is compliant with the 20 ppm requirement at 1 hour.

To further benchmark this result, we analyze a photometric timeline that contains only the transit light curve, photon, and read noise. Figure 4.18b reports the ADEV for the residual obtained in this case, subtracting the true light curve from $\mathbb{D}$, appropriately normalized. Here, we obtain 8.6±1.5 ppm at 1 hour, which effectively represents the lower limit that can be achieved for this simulation and is compatible with the stability achieved for the reference unjittered signal in the case of constant photometry shown previously.

4.2.4.3 Accuracy and bias

An important, yet often overlooked, aspect of detrending is the accuracy of the retrieved parameters. In the case of transit spectroscopy, the planetary radius is the key parameter to

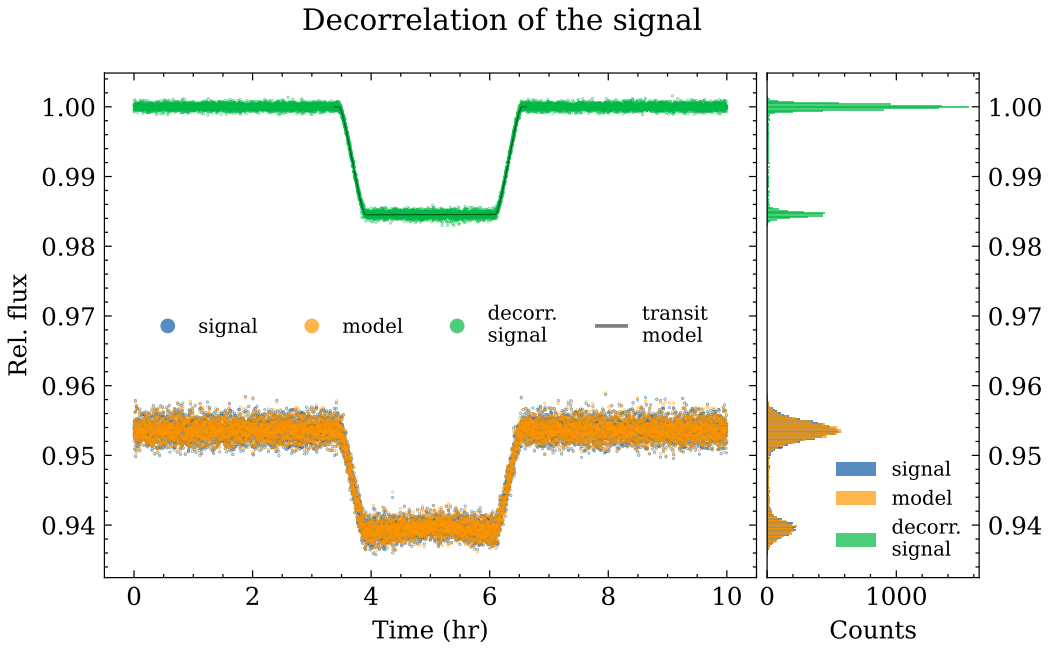
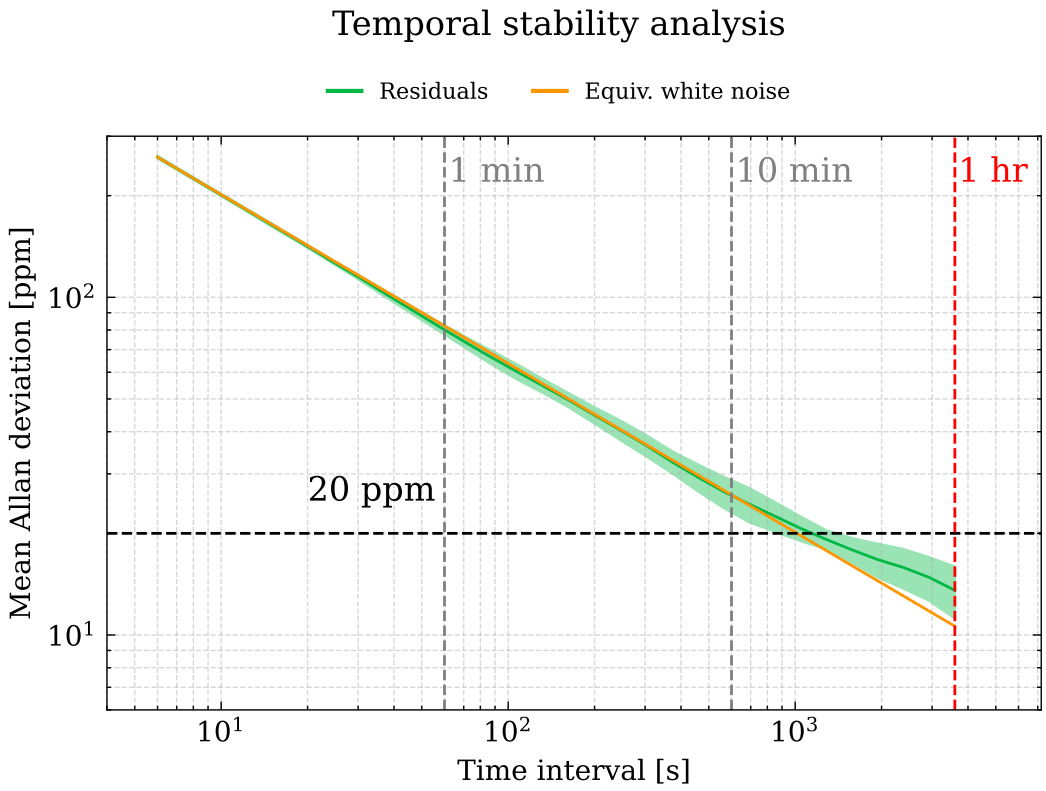
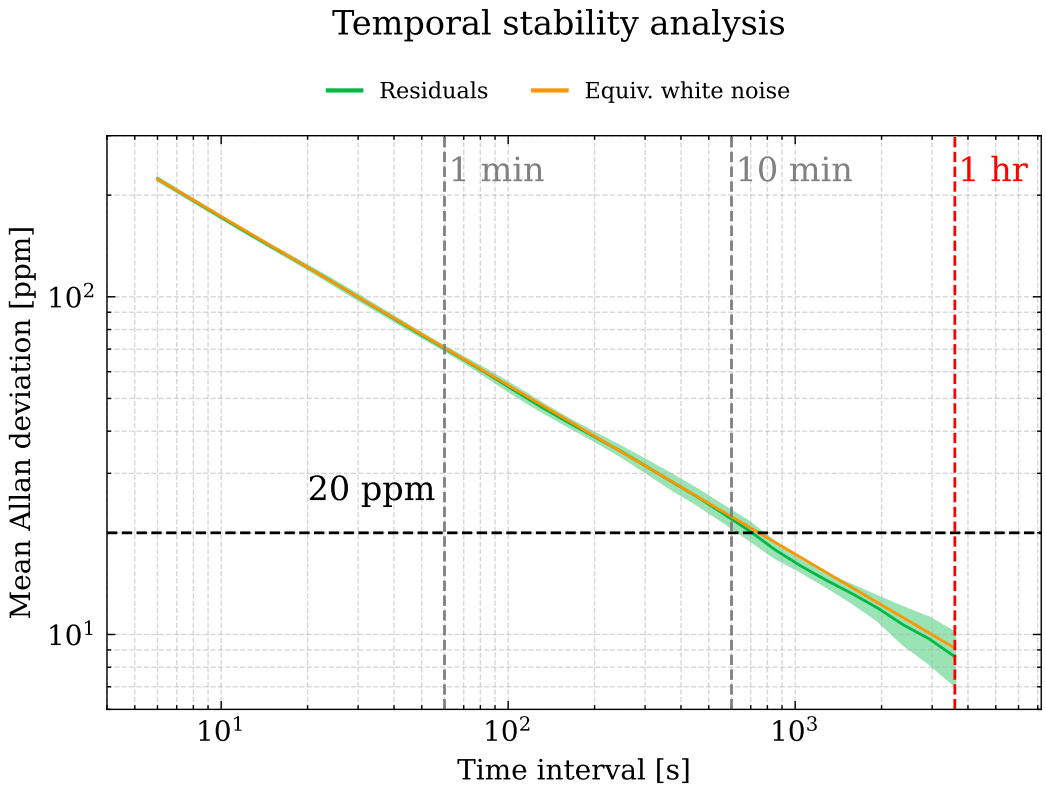


Figure 4.17. An example of jitter decorrelation using the photometric timeline with the transit signal and the addition of photon and read noise. *Left*: the timelines of the raw signal (blue dots), the combined jitter + light curve model (orange), and the decorrelated signal (green). Note that the orange dots almost perfectly superimpose the blue dots. The model transit light curve used in ExoSim 2 is also shown. *Right*: histogram of the counts obtained from the timelines shown on the left, after collapsing the temporal axis.



(a) Decorrelation of the time-varying timeline including the jitter. The overlapping Allan deviation (ADEV) for the decorrelated signal (green) and white noise equivalent (orange). The solid green line is the mean of the ADEV from 10 different noise realizations, and the shaded area represents the corresponding standard deviation.



(b) Residual timeline excluding the jitter: no decorrelation applied. To be compared with the above.

Figure 4.18. Time-domain stability analysis of the residual signal obtained from the FGS-1 photometric timeline modulated by the transit light curve, with the addition of photon and read noise.

reconstruct the atmospheric spectrum. Therefore, it is essential to understand whether it is recovered not only to the required precision but also with no bias. That is, we must ensure (a) that after jitter decorrelation there is no systematics due to pointing jitter still present and biasing the estimate of R_p, and (b) that decorrelating jitter is not introducing unexpected noise or bias to the estimate of R_p.

To investigate this, I have complemented the analysis of the stability of the detrended signal with a study of the distribution of the retrieved planetary radius values. The analysis is performed on the same photometric timeline used in the previous section, containing the transit signal and photon and read noise. The detrending algorithm is applied to the timeline, and the planetary radius is estimated in the decorrelation step described in Equation 4.3. Besides the planetary radius, the other free parameters are the same as in the previous section. Again, all limb darkening coefficients in the simulation are set to zero, which is a simplifying assumption of the light curve shape. The inclusion of non-zero limb darkening coefficients in the simulation is left for future work: while it is not expected to impact the stability of the detrending, it may affect the accuracy of the retrieved parameters.

The analysis is repeated for 10 different noise realizations, and the corner plot of the retrieved parameters is shown in Figure 4.19. Having several realizations is useful to understand not only whether the individual parameters for each realization are unbiased, but also whether the scatter between the different realizations is compatible with the expected experimental uncertainties. We obtain that the mean retrieved uncertainty on the radius estimates is 52 ppm, while the standard deviation of the best-fit radius value across realizations amounts to 48 ppm, i.e. the scatter between realizations is comparable to the retrieved uncertainty.

For better visualization, Figure 4.20 shows the scatter plot of the residuals between the retrieved and true planetary radius values for each realization. While having only 10 realizations is not enough to draw statistically significant conclusions, from this figure it appears that on average, the recovered radius is smaller than the true value by a few tens of ppm. From a χ^2 test with 10 degrees of freedom, the reduced χ^2 is 1.4, indicating that there is a discrepancy (albeit small) between the retrieved and true values. This discrepancy may be due to the residual correlated noise particularly at the center of the timeline, that the decorrelation algorithm is not able to fully remove.

While further work is required to confirm this, here we can investigate whether this discrepancy can be mitigated via time averaging. Figure 4.21 illustrates the same analysis as before, but for the case where the photometric timeline is averaged by a factor of 10 to obtain an effective exposure time of 30 s.

From this figure, we note that the mean retrieved uncertainty on the radius estimates is 69 ppm, while the standard deviation of the best-fit radius value across realizations amounts to 66 ppm. Compared to the previous case (52 ppm), the uncertainties are about 30% larger, and the reduced χ^2 is now 1.2. Therefore, there is a slight improvement in the agreement between the retrieved and true values, although there is still a small discrepancy.

4.2.5 Final remarks

The jitter-detrending procedure developed in this analysis improves upon previous results and has been shown to work well for both constant and time-varying timelines containing a transit light curve. The achievable photometric stability is well within the requirement of 20 ppm at the 1-hour timescale, even when additional noise contributions are included, such as photon and read noise. By implementing a decorrelation based on evaluating the

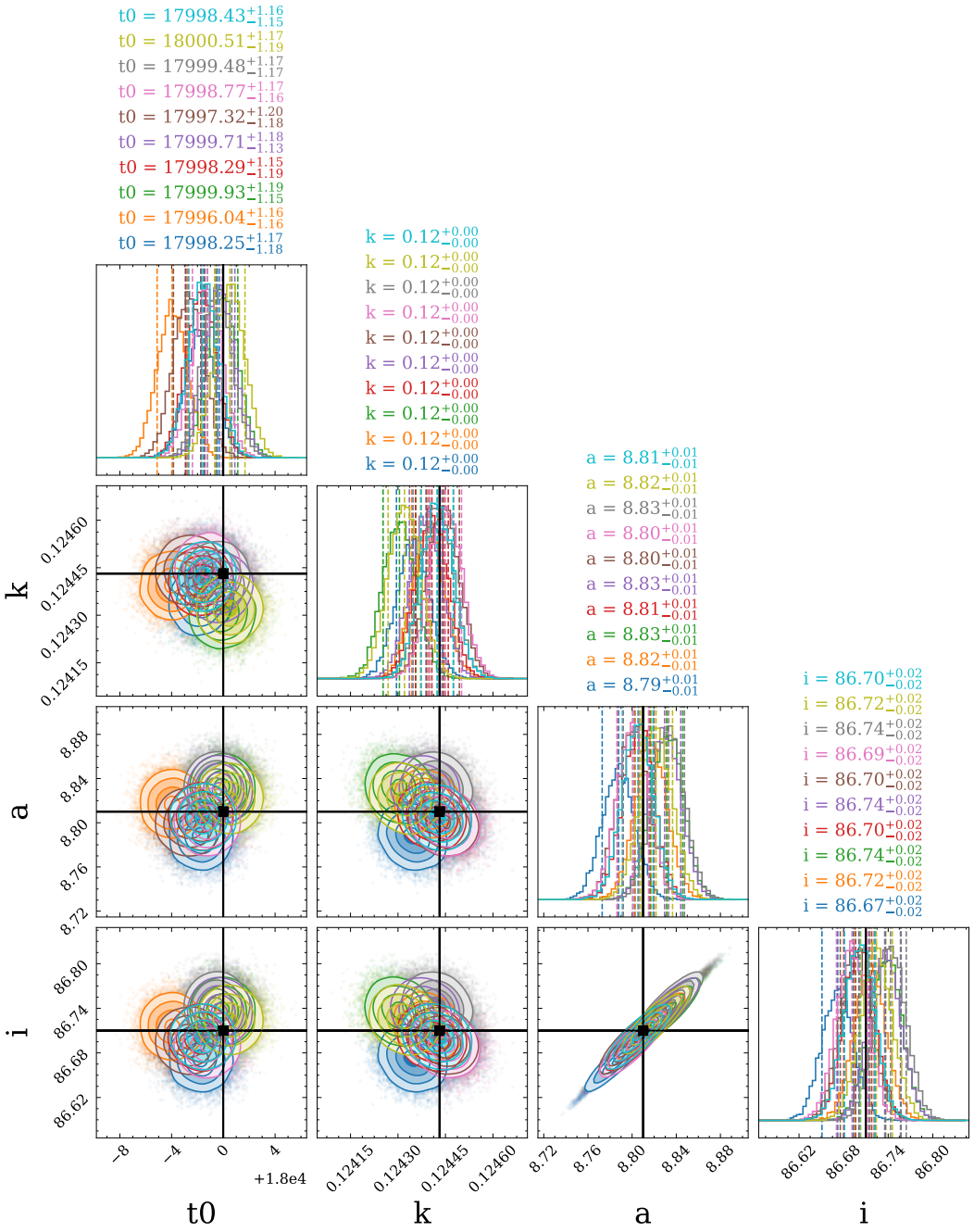


Figure 4.19. Corner plot of the retrieved parameters from the detrending algorithm applied to the FGS-1 photometric timeline modulated by the transit light curve, with the addition of photon and read noise. 10 different noise realizations are used, and the corner plot is obtained by juxtaposing the results from each realization. For each realization, the histograms on the diagonal report the median and associated confidence intervals of the corresponding parameter distribution. The off-diagonal plots show the correlation between the different parameters. The true values of the parameters are shown as black solid lines.

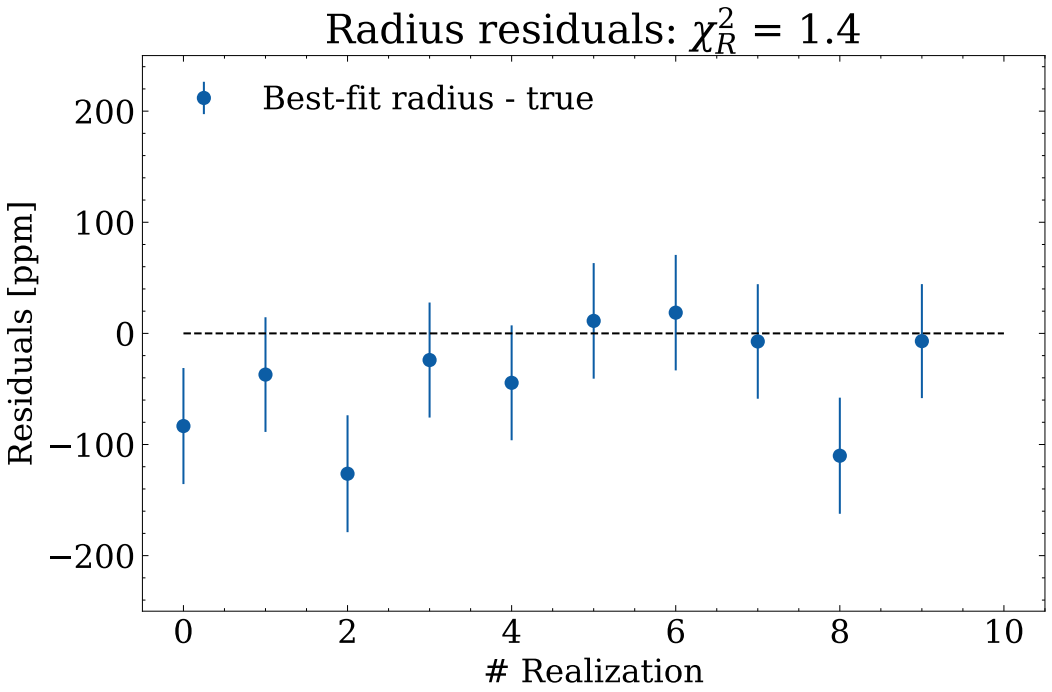


Figure 4.20. Scatter plot of the residuals between the retrieved and true planetary radius values for each realization, with error bars representing the retrieved uncertainty. The horizontal dashed line represents the zero residual. The value of the reduced χ^2 is also reported in the title.

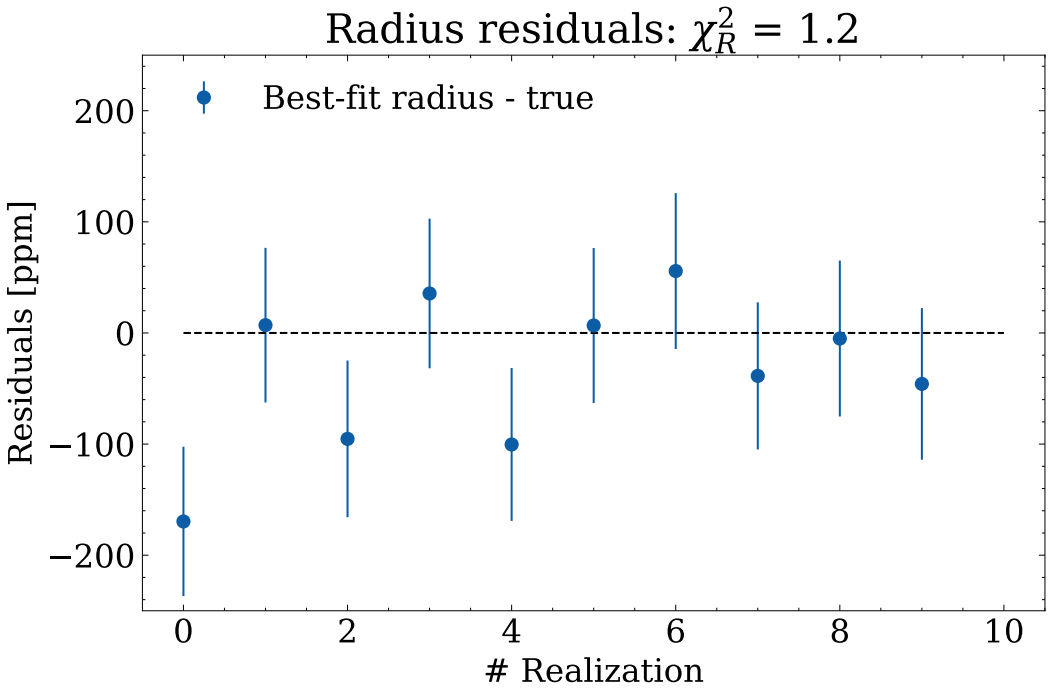


Figure 4.21. Same as Figure 4.20, but for the case where the photometric timeline is averaged via time stacking by a factor of 10.

moments of light distribution, we can fully exploit the spatial information contained in the science images. Calibration information (flat fields) was shown to not improve the achieved precision significantly. In addition, the presented analysis suggests that the decorrelation algorithm may not be fully removing the correlated noise component, and as a result the retrieved planetary radius may be biased by a few tens of ppm. This is a preliminary result, and an investigation is underway to confirm this finding and understand its origin.

The analysis was carried out focusing on FGS-1, but the results apply to all photometric channels. Fully independent verification of the jitter detrending was conducted by Angelos Tsiaras from the S2MD collaboration, and compatible results were obtained. Future work on this topic will include a thorough analysis of the accuracy of the estimated light curve parameters, and the progressive inclusion of more layers of complexity in the simulation, such as nonzero limb darkening coefficients and instrumental effects such as detector 1/f noise, gain drifts, non-linearities, and saturation. The extension of this work to spectroscopic timelines has been initiated by M.Sc. student Angèle Syty, whom I co-supervised. She has adapted the detrending to AIRS channels, significantly improving the results on achieved precision and finding similar small biases in the retrieved planetary radius.

Studying the impact of detrending across multiple wavelength ranges (photometric and spectroscopic) remains a priority, especially given the potential wavelength-dependent biases, however small, that preliminary results suggest may be introduced by the detrending algorithm. A future update of the payload performance analysis will include the results of this work, now carried forward by myself. I am currently preparing a publication that describes the detrending algorithm and presents the results for both photometers and spectrometers.

While only a single Airbus-provided timeline has been used so far, future studies will incorporate diverse representative timelines, to confirm that the achievable photometric precision remains compliant across an ensemble of jitter realizations compatible with the requirements. Such investigations will remain an essential mission preparation activity, with the algorithm presented here to be directly incorporated into the official Ariel data reduction pipeline to ensure optimal exploitation of Ariel's scientific data by effective mitigation of the jitter systematic.

Part III

Pathos
Scientific Impact and Future Prospects

Chapter 5

Alfnoor: a Tool for Quantitative Population Studies

With the Preliminary Design Review (PDR) milestone successfully completed, an important next objective is expanding systematic investigations of Ariel's capability to characterize exoplanetary atmospheres. Our group is leading studies of simulated exoplanet populations representative of the data set observed by Ariel, and is actively contributing to the efforts related to the intermediate milestone of the Dry Run in 2025. This chapter, together with the next, detail our studies on evaluating the expected information content of spectra collected during the mission's reconnaissance survey. This unprecedented survey of hundreds of transit and eclipse spectra aims to address population-level questions, guide further observation, and cement Ariel's legacy. The survey observations will shape our knowledge of the exoplanet population and solar system's place within it for decades.

To simulate this data set, in collaboration with University College London (UCL) we have developed an end-to-end framework for large atmospheric studies, Alfnoor (Changeat et al., 2020a; Mugnai et al., 2021a). Alfnoor combines the atmospheric models of 'forward' spectra produced by TauREx 3 (Al-Refaie et al., 2021) with Ariel's performance estimated by ArielRad (Mugnai et al., 2020). By iterating this process for different planets and compositions, Alfnoor automates the simulation and retrieval of planetary populations and thus a representative data set of Ariel's observations.

We focused on evaluating the reconnaissance survey data set, which poses analysis challenges given that the combination of Signal-to-Noise Ratio (SNR) and spectral resolution might not be sufficient to constrain e.g. abundances of atmospheric species with high confidence. In an initial study (Mugnai et al., 2021a), we developed a metric sensitive to the chemical composition of atmospheres, independent of planet type and fundamental parameters (temperature, mass, and radius), and capable of producing color-color diagrams of planetary populations. Although the classification of Tier-1 observations based on molecular content using this metric is affected by the bias of the estimators used, which depends on instrumental noise, this demonstrated potential for inferring the presence of molecular species and using Tier-1 data to address population-level questions.

* * *

Here we report our work and results, presented in Mugnai et al. (2021a). The analysis

aims are threefold:

1. Show the capability of selecting the planets with featureless spectra, that may not be observed again in successive Tiers, without involving retrieval techniques;

2. Introduce a metric and show its principal applications as a tool to classify Tier 1 observed planets on their molecular content, to aid in the selection of targets to be re-observed in successive Tiers;

3. Show other strategies to exploit Tier 1 data are feasible such as those based on Deep and Machine Learning.

Section 5.1 presents the methodology. Alfnoor is introduced in Section 5.1.1. Section 5.1.2 describes the simulated planets and atmospheric properties. Identification of flat spectra is discussed in Section 5.1.3, with results in Section 5.2.1. The classification metric is detailed in Section 5.1.4, and compared to a K-nearest neighbors algorithm in Section 5.1.4.1. Results are presented in Section 5.2.2, including the relation between the metric and the input molecular abundances, biases, and limitations. Finally, we provide a preliminary assessment of the application of Machine and Deep Learning techniques to the problem of spectra classification in Section 5.1.5, discussing their performance in Section 5.2.3, although thorough investigation is left to future work. The results are discussed and compared in Section 5.3.

5.1 Methodology

5.1.1 The `Alfnoor` software

Ariel will provide a sample of hundreds of planetary spectra. To simulate this data set we developed a new algorithm: Alfnoor, the thousand lights simulator, which was also used for Tier 2 data in Changeat et al. (2020a). Alfnoor is a wrapper of TauREx 3 (Al-Refaie et al., 2021, see Section 1.3.2 for details) and ArielRad (Mugnai et al., 2020, see Section 1.3.1 for details). TauREx 3 is a complete rewrite of the atmospheric retrieval code TauREx (Waldmann et al., 2015b,a). ArielRad is the Ariel Radiometric Model: a software that, given the Ariel payload and mission strategy descriptions, can simulate the signal propagating from a candidate target through the instruments, and return the expected instrument noise. ArielRad, therefore, can compute the number of observations needed to match each of the Ariel Tier requirements (to reach a minimum SNR=7 at the Tier spectral resolution).

By combining the two software, Alfnoor produces the atmospheric high-resolution forward model of a planet with TauREx 3, it bins down the spectrum to the Ariel Tier wavelength grid and adds the expected noise estimated by ArielRad. Consequently, Alfnoor returns a simulation of the planet spectrum as observed in each of the Ariel mission Tiers. Iterating this procedure for different planets or compositions, Alfnoor automates the process of building entire planetary populations and therefore a data set that is representative of the one Ariel will provide.

The Alfnoor and the ArielRad tools are not publicly available, currently. However, both

TauREx 3[1][2] and a generic radiometric simulator called ExoRad 2.0[3][4], are publicly available on GitHub and PyPI. ArielRad is ExoRad 2.0 configured for the Ariel payload.

5.1.2 Planetary populations

To build a diverse sample of planets in terms of masses, radii, and temperatures, we use the Ariel candidates list of Edwards et al. (2019). This list contains 1000 planets, selected from both NASA's Exoplanet Archive and TESS predicted discoveries, and covers a wide range of planetary radii (from $\sim$ 0.4 to $\sim$ 27 $R_\oplus$), masses (from $\sim$ 0.01 to $\sim$ 3000 $M_\oplus$) and equilibrium temperatures (from $\sim$ 200 K to $\sim$ 3900 K). From that list, we extract the parameters listed in Table 5.1. Our goal is not to reproduce accurately the composition of the planets in that list, but to test a diverse sample, and therefore we randomly build an atmosphere for each of the listed targets. We produce three planetary populations that will be of use for this work. We call them POP-I, POP-II and POP-III.

Table 5.1. List of host star and planet information obtained from the Ariel planets candidate list and used to build the planetary populations used in this work.

Star	Planet
mass	mass
radius	radius
effective temperature	equilibrium temperature
distance	distance from the star
	orbital period
	transit duration

5.1.2.1 POP-I

For each planet we randomize the equilibrium temperature, choosing a value between $0.7 \times T_p$ and $1.05 \times T_p$, where T_p is the planet's equilibrium temperature in Edwards et al. (2019). This randomization is biased toward lower temperature values as we probe the terminator region, where the spectral features are affected both by the day side and the night side temperatures (Caldas et al., 2019; Pluriel et al., 2020b; Skaf et al., 2020). The temperature randomization range is consistent with the work presented in Changeat et al. (2020a).

Then, for each planet we consider an isothermal temperature-pressure profile; we add a constant vertical chemical profile (Moses et al., 2011) for every molecule from a list of selected molecules (the abundances are randomized according to defined boundaries). Finally, we add randomly generated gray opaque clouds. We use the plane-parallel approximation, building 100 plane-parallel layers to uniformly sample in log-space the pressure range $10^{-4} \rightarrow 10^{6}$ Pa. Every atmosphere is built with randomised relative abundances of CH_4, H_2O, CO_2 and NH_3 on a uniform logarithmic scale between 10^{-7} and 10^{-2}. Such a

[1] https://github.com/ucl-exoplanets/TauREx3_public
[2] https://pypi.org/project/taurex/
[3] https://github.com/ExObsSim/ExoRad2-public
[4] https://pypi.org/project/exorad/

large range allows us to explore the sensitivity of our developed method to very different abundances. We also randomized the cloud surface pressure varying between 5×10^2 and 10^6 Pa, similarly to what presented in Changeat et al. (2020a), to explore the whole range from overcast to cloud-free atmospheres respectively. Using these boundaries, we obtain that $\sim 40\%$ of the atmospheres in the populations contains clouds to at least 10^4 Pa (surface pressure), as expected from Tsiaras et al. (2018) and Iyer et al. (2016). Every planet is considered filled with a H_2 and He atmosphere with mixed ratio $He/H_2 = 0.17$. A list of the opacities used in this work is reported in Table 5.2.

As already mentioned, we don't focus on the consistency of the atmospheric models used to build the population. The spectra generated will only be used as "transmission spectral shapes" to test our methods against. No information other than the planet transmission spectrum is used in this work.

Table 5.2. List of opacities used in this work and their references.

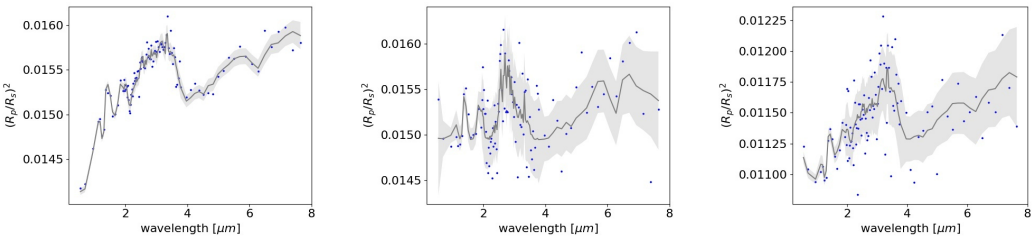
Opacity	Reference
H_2-H_2	Abel et al. (2011); Fletcher et al. (2018)
H_2-He	Abel et al. (2012)
H_2O	Barton et al. (2017); Polyansky et al. (2018)
CH_4	Hill et al. (2013); Yurchenko and Tennyson (2014)
CO_2	Rothman et al. (2010)
NH_3	Yurchenko et al. (2011); Tennyson and Yurchenko (2012)

Each planetary spectrum generated by Alfnoor is binned at Ariel's Tier 3 spectral resolution. These spectra make up the "noiseless spectra" data set. ArielRad then predicts the noise for each spectral bin at the Tier resolution. To reproduce a Tier 1 observation we scatter the data around the true value according to a normal distribution with the mean coinciding with the simulated spectrum, and a standard deviation equal to the noise estimated with ArielRad at each spectral bin. This noise is a re-scaled version of the Tier 3 noise, obtained by combining the number of transit observations needed to match the Tier 1 required SNR. Using these scattered spectra, we build the "observed spectra" data set. Examples of the resulting spectra are shown in Figure 5.1.

We generate POP-I using the full 1000 planets candidate list and we produce one realization for each planet. A similar approach was used by Changeat et al. (2020a) in their investigation of the Ariel Tier-2 observations. We use the POP-I population to test the strategies described later in the text.

5.1.2.2 POP-II

We produce another data set keeping the same 1000 planets from the target list and the randomization rules of POP-I. However, this time we modify the chemical composition to include only H_2O and CH_4. We use POP-II to perform tests against a simpler population, as detailed later in the text.



(a) HD 209458b -like planet. **(b)** GJ 1214b -like planet. **(c)** WASP 79b -like planet.

Figure 5.1. Example of simulated spectra. The gray solid lines are the noiseless spectra simulated and binned at Ariel Tier 3 spectral resolution. The gray bands are the $1-\sigma$ confidence levels centered around the simulated spectra for a number of transit observations needed to match the Tier 1 required SNR. The blue dots are noised data points representing Tier 1 observed spectra. Starting from the left, the first planet is HD 209458 b-like, the second one is GJ 1214 b-like and the third one is WASP-79 b-like. Their atmospheres are built as described in Section 5.1.2.

5.1.2.3 POP-III

To build the last population, we use the same list of 1000 planets, where each planet is repeated 4 times, such that there are 4 randomized atmospheres for each unique set of stellar and planetary properties that define a planet. While the temperature and cloud conditions used are the same as those discussed for POP-I, for each molecule we widen the abundance boundaries to $10^{-9} \rightarrow 10^{-2}$ on a uniform logarithmic scale. We call this population POP-III, and we use it to train our Deep and Machine Learning algorithms.

5.1.3 Flat spectra detection

The first goal of this work, as listed at the beginning of this chapter, is to identify featureless spectra. This will help in the selection of targets to be re-observed in Ariel's higher Tiers. Given the property of the Ariel payload, we divide the spectral wavelength range into four parts or bands:

- from 0.5 to 1.1 μm, sampled by three photometers;
- from 1.1 to 1.95 μm, corresponding to the NIRSpec wavelength range;
- from 1.95 to 3.9 μm, corresponding to the AIRS-CH0 wavelength range;
- from 3.9 to 7.8 μm, corresponding to the AIRS-CH1 wavelength range.

For every planet, and for every band we estimate a χ^2 using all measurements in the band to assess the compatibility with a flat, zero-gradient line: for each planet, there are four χ^2 estimates, one for each band above. We reject the hypothesis of spectral flatness in a given band with a $3-\sigma$ confidence if $\chi^2 > 1 + 3\sqrt{\frac{2}{\nu}}$, where ν are the degrees of freedom. Therefore, if any of the four bands has a χ^2 smaller than this number, we mark the band as flat. If a planetary spectrum has all 4 bands marked as flat, it is classified as a flat spectrum. This strategy is similar to that presented in Zellem et al. (2019), however, while in that work the authors were only focused on the Ariel FGS channels, here we are considering the full Ariel spectral coverage.

5.1.4 An optimized molecular metric

The second goal listed at the beginning of this chapter is to develop a metric, M_{mol}, to assess the presence of a molecule, *mol*, in the planet's atmosphere. We want this metric to work in such a way that by comparing two molecules, the metric produces a diagram similar to that in Figure 5.2. In the diagram, we can distinguish four regions: two regions where the atmospheres are rich in a single molecule and therefore only show its characteristic features; a third region where the atmospheres show features from both molecules; a fourth region where features are absent, either because the planets have flat spectra or because the features from both molecules do not emerge from a thick layer of clouds.

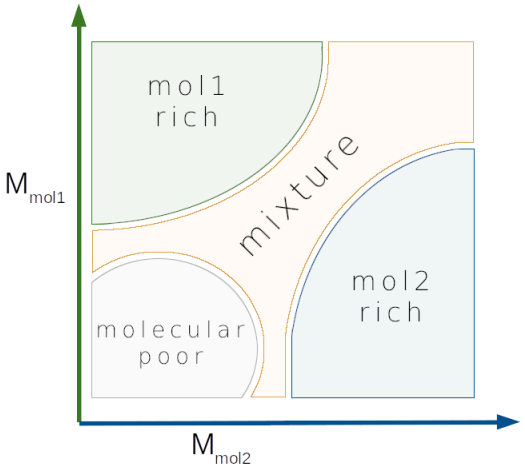


Figure 5.2. Illustration of the diagram we expect to build with our metric. Here, the metric is used to compare two molecules, *mol*1 and *mol*2. By drawing M_{mol1} versus M_{mol2}, we aim to separate four different regions: one rich in the first molecule at the top left (green), where M_{mol1} grows and M_{mol2} is low: a similar region at the bottom right (blue), where the planet atmosphere is rich in the second molecule because M_{mol2} is high and M_{mol1} is low; a region where molecular poor planets are located (gray), or those that have no features in the considered bands, where both M_{mol1} and M_{mol2} are low; a region for mixed atmosphere (yellow) in the central portion of the diagram.

To compare different planets and constrain their atmospheric molecular content, the metric should be (i) sensitive to the spectral signature of molecules, (ii) independent of the planet's size, and (iii) independent of the scale height. Here we present a metric that fulfills these 3 conditions and we show its current limitations.

For each molecule, we select N bands within the Ariel wavelength range, where the molecular features in the transmission spectrum are strong. Then, for each planet, we compute the average in each band, S_{band_i} and its dispersion, σ_{band_i}.

$$S_{band_i} = \frac{1}{M}\sum_{j}^{M} S_j \tag{5.1}$$

$$\sigma_{band_i} = \sqrt{\frac{1}{M}\sum_{j}^{M}(S_j - S_{band_i})^2} \tag{5.2}$$

where M is the number of spectral bins in the band, S_j is the atmospheric transmission spectrum estimated in the j^{th} wavelength bin.

We do the same with a control band where we know there are no major molecular features from the molecule considered, called the "normalization band", obtaining S_{norm} and σ_{norm}. We select a different normalization band for each molecule (Table 5.3).

Thus, for each molecule, *mol*, we define

$$M_{mol} = \frac{1}{N} \sum_{i}^{N} \frac{S_{band_i} - S_{norm}}{\sqrt{\sigma^2_{band_i} + \sigma^2_{norm}}} \tag{5.3}$$

Defined in this way, M_{mol} is similar to a signal-to-noise ratio, where the signals are the molecular features arising above the "normalization band", and the noise is the dispersion in the band. Therefore,

$$\sigma_{M_{mol}} = \frac{1}{\sqrt{N}} \tag{5.4}$$

The metric thus designed, by averaging the contribution of N different bands, corresponding to N different features of the same molecule, reduces the chance of being misled by overlapping features in one of the bands considered. As Ariel's Tier 1 is optimized for low resolution spectroscopy, spectral binning increases the SNR. Also, this metric is (i) sensitive to the presence of molecules, (ii) independent of the planet size, and (iii) independent of the scale height (see Section C.1 for details), at the cost of the introduction of a bias: Equation 5.2 provides an estimate of the spectral dispersion when applied to noiseless spectra, and it is larger for observed spectra because of the presence of measurement noise. Therefore, the absolute value of M_{mol} of Equation 5.3 is always smaller on observed spectra compared to noiseless spectra of the same planet. While the bias effects are further discussed in Section 5.3.1, we note here that a detailed characterization of the instrumental noise would allow to de-bias the metric, but we leave this investigation to future work, and we focus the attention on the performance of the metric in extracting information from Tier 1 observations.

To maximize the metric efficiency, the challenge is to identify the best-performing wavelength range to use: large enough to reduce the uncertainty introduced by the observational noise, but small enough to distinguish the molecular features of interest.

In this work, we consider only H_2O, CH_4, and CO_2, and the bands used are listed in Table 5.3. Even though NH_3 is present in our sample, it is used only to introduce a nuisance and challenge our metric, because NH_3 has features overlapping with those of water. We use 3 feature bands for CH_4 and CO_2 and 5 for H_2O. Examples of the bands used for CH_4 and H_2O are shown in Figure 5.3 where, for the same planetary template, HD 209458 b, we simulate different atmospheres (overcast, CH_4 rich and H_2O rich) to show how the metric captures the relevant spectroscopic features.

In the next section, we show how we intend to use this metric to build a diagram similar to that of Figure 5.2.

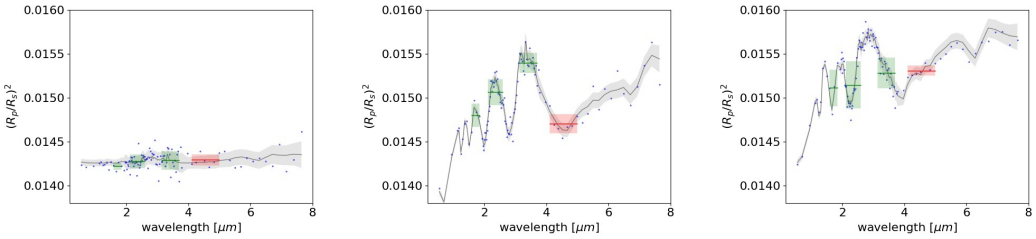
5.1.4.1 Planets classification

The metric requires to be calibrated to assess its capability to estimate the presence of a molecule. The final product is a diagram similar to Figure 5.2, that can be used as a look-up

Table 5.3. Wavelength ranges used to select the molecular features in the spectra (left table) and the normalization bands (right table) for H_2O, CH_4, and CO_2.

H_2O	CH_4	CO_2
1.2 → 1.6 μm	1.5 → 1.8 μm	1.9 → 2.3 μm
1.7 → 2.1 μm	2.1 → 2.6 μm	2.6 → 3.2 μm
2.6 → 3.0 μm	3.1 → 3.7 μm	4.2 → 4.8 μm
5.4 → 6.1 μm		
6.5 → 7 μm		

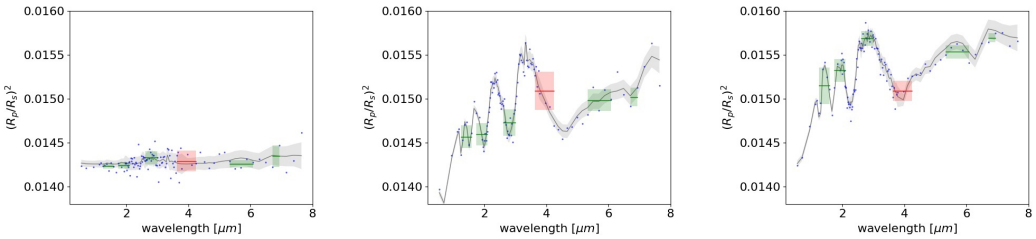
Molecule	Normalization
H_2O	3.6 → 4.2 μm
CH_4	4.0 → 5.0 μm
CO_2	5.0 → 6.0 μm



(a) Overcast HD 209458b - like planet with M_{CH_4} data bands highlighted.

(b) Methane rich HD 209458b -like planet with M_{CH_4} data bands highlighted.

(c) Water rich HD 209458b - like planet with M_{CH_4} data bands highlighted.



(d) Overcast HD 209458b - like planet with M_{H_2O} data bands highlighted.

(e) Methane rich HD 209458b -like planet with M_{H_2O} data bands highlighted.

(f) Water rich HD 209458b - like planet with M_{H_2O} data bands highlighted.

Figure 5.3. Here are shown three examples of randomized spectra. For the same planet, HD 209458 b, we present three different realizations: a flat atmosphere (first column), a methane-rich atmosphere (second column), and a water-rich atmosphere (third column). Each column shows the same planetary spectra. gray solid lines are the original binned spectral data (Tier 3 spectral resolution), the filled gray areas are the $1-\sigma$ uncertainties (Tier 1), and blue dots are the simulated observation data used in this work. The top row highlights the M_{CH_4} feature bands from Table 5.3, while the bottom row shows the M_{H_2O} bands. In green are reported the molecular feature bands values, with their dispersion, while in red are reported the normalization bands'. Comparing the rows we see how the bands selected match the relevant molecular spectral features.

table, such that, given an observed spectrum, its corresponding M_{mol} can be located on the diagram, and its possible composition inferred.

To assess the ability of the metric to separate the atmospheres in the sample, we use the k-Nearest Neighbors (kNN) algorithm, a non-parametric pattern recognition algorithm (Hastie et al., 2009). Note that the k-Nearest Neighbors (kNN) algorithm is a very simple algorithm that is known to return noisy class-division boundaries. More expressive and flexible clustering algorithm such as DBSCAN (Ester et al., 1996) would have likely been more appropriate for this task. However, the kNN algorithm is what we chose to use in this work because it is simple and easy to implement, and we deemed it sufficient to demonstrate the potential of the metric.

This algorithm, after a training process, assigns a class to an element given the properties of its neighbors. The goal is to classify observed spectra by their molecular content, according to their M_{mol}. Considering two molecules at a time, we first define four classes of planets: molecular poor, *mol*1 rich, *mol*2 rich, and mixture, as defined in Table 5.4.

Table 5.4. Diagram classes and conditions.

Class	Condition
molecular poor	$Ab_{mol1} < 10^{-5}$ and $Ab_{mol2} < 10^{-5}$
*mol*1 rich	$Ab_{mol1} > 10^{-4}$ and $Ab_{mol1} > 10 \times Ab_{mol2}$
*mol*2 rich	$Ab_{mol2} > 10^{-4}$ and $Ab_{mol2} > 10 \times Ab_{mol1}$
mixture	everything else

The kNN algorithm used classifies each planet according to the 20 (k = 20) nearest planets, in the M_{mol1} vs M_{mol2} space, in the same data set. We choose to use 20 neighbors (2% of the full data set) to minimize the number of misclassified planets. The closest neighbors are uniformly weighted, and we verified that weighting the neighbors with their Euclidean distance in the metric space does not affect the results significantly. All other parameters are set to their default values of the scikit-learn[5] implementation of the algorithm (Pedregosa et al., 2011) which we use in this work.

The analysis involves three separate steps, summarised in Figure 5.4, applied to POP-I.

Step 1. We estimate the (M_{mol1}, M_{mol2}) on the POP-I observed spectra. We assign a class to each POP-I planet using its input molecular abundance values, Ab_{mol}, that are stored during the population production. This process is described in the top branch of Figure 5.4

Step 2. To calibrate the metric, we map the metric space grid by training the kNN algorithm on the (M_{mol1}, M_{mol2}) estimated from the noiseless POP-I planetary spectra. We assign again a class to each planet using its input molecular abundance, Ab_{mol}, and the training is performed on a randomly chosen selection accounting for 70% of the data set, while we use the remaining 30% to test the success of the training. Finally, we classify each point (M_{mol1}, M_{mol2}) of the M_{mol} space grid M_{mol} sampled at a step width of 0.2 M_{mol}, obtaining a map comparable to Figure 5.2. This part of the procedure corresponds to the central branch of Figure 5.4.

Step 3. Since the noiseless planetary spectra are not expected to sample the parameter space uniformly, we build a mask to select a region of the (M_{mol1}, M_{mol2}) space that is sufficiently well sampled to achieve a reliable classification. To do so, we replace each

[5] https://scikit-learn.org/0.22/

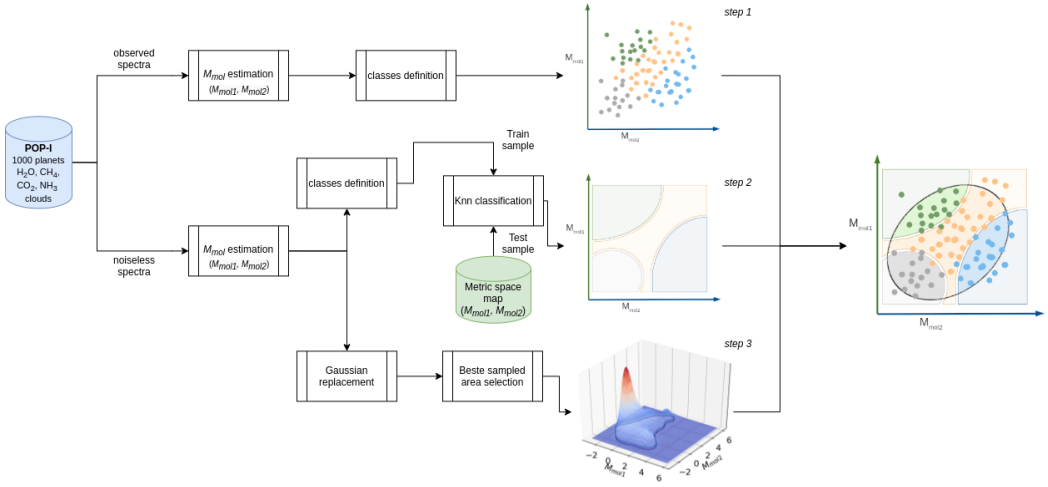


Figure 5.4. Planets classification summary. The figure reports the steps implemented to build the diagram in Figure 5.2. Starting from POP-I, for each planet we compute (M_{mol1}, M_{mol2}) for the considered molecules and for both observed and noiseless data. Following the top branch, classes are assigned to the observed spectra (*step 1* in the text). Following the middle branch, a kNN classification is performed on noiseless spectra to calibrate the metric space (*step 2* in the text). Following the bottom branch, the distribution of noiseless metric data points is convolved with a 2D Gaussian with varying widths to generate a unit-normalized volume. The intersection between this volume and the calibration of *step 2* selects the best sampled (i.e. calibrated) region in the metric space (*step 3* in the text). The combination of these three steps is shown in the rightmost diagram to be compared with Figure 5.2.

(M_{mol1}, M_{mol2}) point representing a noiseless planetary spectrum with a two-dimensional Gaussian distribution using the metric dispersion in the two directions as σ. We sum the Gaussian volumes on the parameter space, ending up, after volume normalization, with a statistical distribution of our data points on the parameter space grid. Then, we select a region in the metric space that results in a total volume of 95%, therefore removing all under-sampled areas from the grid. This last step is represented in the bottom branch of Figure 5.4.

The combination of the three steps is shown in the rightmost panel of Figure 5.4 and it is the equivalent of Figure 5.2 calibrated for the metric on the investigated population.

5.1.5 Deep and Machine Learning

The metric presented in Section 5.1.4 is based on binning the spectra, and therefore is equivalent to using Ariel as a multi-band photometer. This strategy is in line with the Tier 1 definition of Tinetti et al. (2018). However, we are also investigating different strategies to classify spectra by their molecular content (the third goal listed at the beginning of the chapter). Deep and Machine Learning (ML) techniques are promising because these algorithms can learn to classify planets from their spectral shape over the whole wavelength range sampled by Ariel. Another advantage over the metric is that ML techniques are not supposed to be biased by the instrumental noise, or at least they can be made to learn how to deal with the bias provided that a sufficiently large and representative set of examples is provided in training. To train the algorithms we use the POP-III observed spectra and

their known abundances as a training sample. Each example spectrum is normalized to zero mean and unit dispersion. The normalization facilitates the training process but might introduce a bias that may be very similar to that affecting the metric. A detailed investigation of these aspects concerning ML is left to future work. Knowing the input abundance of each planet, Ab_{mol}, we can define a threshold and flag a planet as bearing a certain molecule if Ab_{mol} is larger than the threshold. This means that, for each molecule, the algorithm learns to flag the planets as bearing that molecule by looking at characteristic spectral shapes. Then we measure the algorithm's ability to "learn" by how much they can generalize their predictions to unknown shapes, testing it on POP-I observed spectra, used as a test data set. The comparison of the ML classification with the known input abundance of each POP-I planet provides an estimate of the success rate.

A detailed investigation of the use of these algorithms and their limitations will be discussed in future work: here we report only an example of how these tools might be used and we compare some preliminary results with the outcomes of the metric of Section 5.1.4. We implemented all algorithms in Python using the scikit-learn package.

The first ML algorithm we use is the kNN algorithm described above. This time we want to simply classify the planets and not produce a map as in Section 5.1.4.1. For this exercise, we use all the scikit-learn default kNN settings, including $k = 5$ and uniform weight for the neighbors. Other ML algorithms can be used to classify planets. Here we also present our preliminary results using a Multi-layer Perceptron (MLP) classifier, a Random Forest Classifier (RFC), and a Support Vector Classifier (SVC) (e.g. Goodfellow et al., 2016; Sturrock et al., 2019). More sophisticated algorithms such as Convolutional Neural Networks (CNNs) (Lecun and Bengio, 1995) or Long Short-Term Memory Networks (LSTMs) (Hochreiter and Schmidhuber, 1997) can also be used, but we leave their investigation to future work.

The MLP is a feed-forward neural network composed of multiple layers of perceptrons largely used in classification problems. To produce the results shown later in the text we use an MLP network keeping the scikit-learn default settings (a single hidden layer made of 100 units) and we classify the spectra with the same procedure used for the kNN. The RFC is an ensemble of decision trees used for classification, where each decision tree is a directed graph and each vertex is a binary test. In this work, we use an RFC set-up commonly used in binary decision problems, which has a number of features equal to the square root of the number of input data points, again, as per scikit-learn is the default configuration. The SVC is a Support Vector Machine method, a family of non-probabilistic linear classifiers that construct hyper-planes to separate the data points. Here, we implemented a simple SVC shaping the decision function in "one-vs-one" mode, as it is the default configuration in scikit-learn at the moment of writing.

The rationale behind using the default scikit-learn settings for all these algorithms is to show how they perform "out of the box", without any optimization. All the information about the algorithms and their settings can be found in the scikit-learn documentation[6], e.g. convergence criteria, minimizers, training duration, pruning regimes (for RFC), non-linearities and depth (for MLP), etc.

[6] https://scikit-learn.org/0.22/

5.2 Results

5.2.1 Flat spectra detection

Shown in Figure 5.5 is the frequency of observed planets in the POP-I population that have a certain number of flat bands. In this population, 16% planets are to be considered "flat" as all of the four spectral bands considered are flat. From the figure, we notice that around 46% of the planets in the population have three or more flat bands, which is consistent with POP-I's known properties and with the ground truth (Tsiaras et al., 2018; Iyer et al., 2016), as mentioned in Section 5.1.2. In the same figure, it is shown the same statistic for the 100 planets of POP-I most covered in clouds (corresponding to a cloud surface pressure of roughly $< 10^3$ Pa), and for the 100 planets of POP-I with fewer clouds (corresponding to a cloud surface pressure of roughly $> 10^{5.5}$ Pa). This comparison shows how overcast planets on average present more flat bands than clean planets, demonstrating how this approach is sensitive to the presence of clouds.

This result clearly shows that Tier 1 observations are effective in the identification of atmospheres with no detectable molecular absorption features.

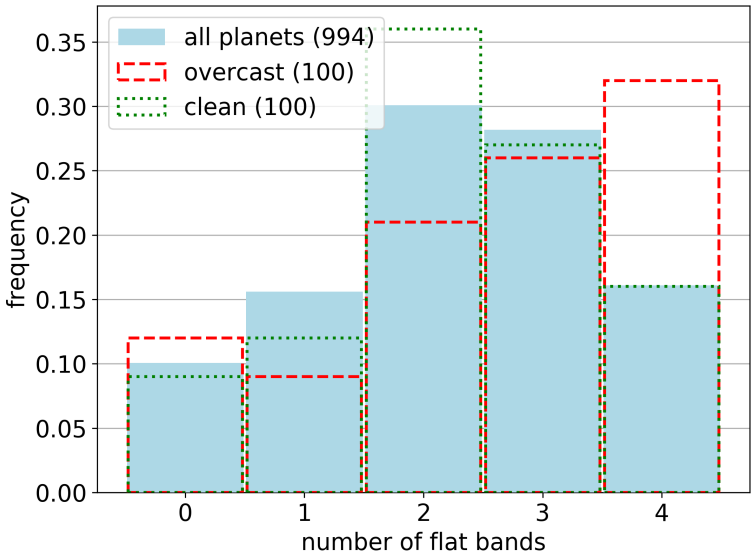


Figure 5.5. The histograms show the frequency of planets in the population vs. the number of flat bands. We consider four bands: one for the photometers (VisPhot, FGS-1, FGS-2) and one for each spectrometer (NIRSpec, AIRS CH0 and AIRS CH1). Each band is compared with a constant value using a χ^2 test to determine its compatibility with flatness. The light blue histogram shows the frequency of planets in the POP-I population with flat bands. The red dashed histogram shows the same statistic but for a selection of the 100 planets of POP-I that are more overcast. The green dotted histogram shows the opposite situation, for a selection of the 100 planets in POP-I where the cloud pressure is the highest (i.e. it is deeper down in the atmosphere; therefore, no clouds appear in the spectrum). We notice that the overcast planets show more flat bands than planets with fewer clouds.

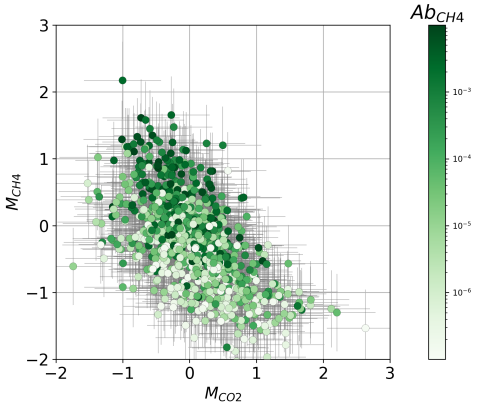
5.2.2 Spectra classification

The M_{mol} (Section 5.1.4) estimated for the observed POP-I planets are shown in Figure 5.6 for different pairs of molecules: CH_4 - CO_2 and CH_4 - H_2O. Comparing the top left and right panels in Figure 5.6, we notice from the color scale that our metric can separate between planets bearing more or less methane (dark and light green colored dots respectively) or carbon dioxide (dark and light orange colored dots respectively). The bottom panels, and the bottom-right panel in particular, show that it is harder to separate planets bearing more or less water (dark and light blue colored dots respectively). Water data appear more clustered around the axes' origin than the top row, and the water-colored data points are not as clearly separated according to their color gradient as the methane or the carbon dioxide data points are. A possible explanation is that CH_4 and CO_2 have strong spectral features, with isolated transmission features in the range $3 \rightarrow 4$ μm and $4 \rightarrow 5$ μm respectively, while H_2O features are less obvious and frequently overlap with the ones of NH_3, that is present in the population (Tinetti et al., 2013). An alternative explanation is that involving a bias in the metric that affects more strongly the water bands.

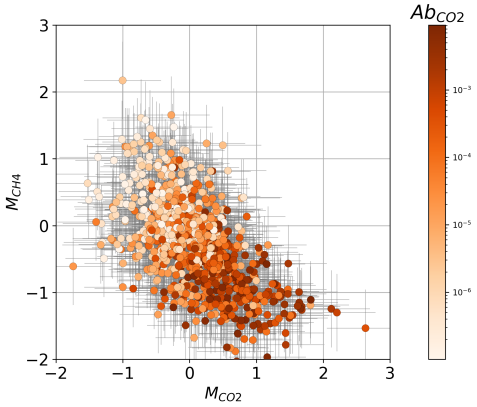
The diagrams of Figure 5.6 are reproduced in Figure 5.7, where the data points are now color-coded following the assigned classes (*step 1*, Section 5.1.4.1) and the background colors, constructed by training the kNN on noiseless spectra (*step 2* and *3*, Section 5.1.4.1), serve as reference and calibrated regions in the metric space. It can be noticed that the metric has the desired response from the similarities between the reference regions in Figure 5.7 with those of Figure 5.2, with a good separation in the metric space. However, it can be seen that the classification boundaries are not as sharp as in Figure 5.2. In some instances, the data belonging to a class do not adhere to their shaded regions. Moreover, the data points tend to cluster towards the origin of the grid more strongly than the reference regions. This is the effect of the bias, further discussed in Section 5.3.1.

While better clustering results are expected in higher dimensions, for instance with the inclusion of bulk parameters (radius, mass, stellar radius, semi-major axis, etc.), we stress that the aim of these plots is to show the potential of the molecular metric to separate the atmospheres in the sample and construct a diagram similar to Figure 5.2. This is a first attempt at producing one of the Holy Grails of exoplanet science, that will be enabled by Ariel observations: color-color diagrams for the exoplanetary population based on the molecular content of their atmospheres. Future work will focus on the exploration of the higher dimensional space to optimize the clustering and classification of the atmospheres.

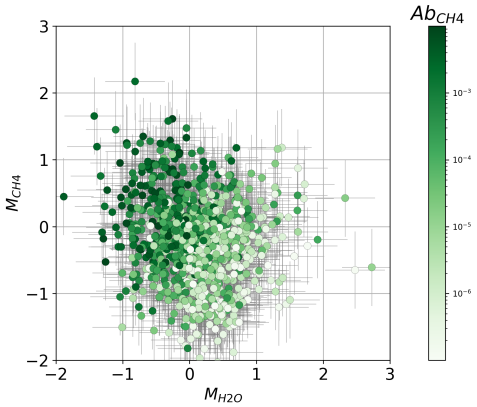
Figure 5.8 shows the relation between the metric, M_{mol}, estimated on POP-I observed spectra, and the input abundances, Ab_{mol}. The coefficients of the linear trends of M_{mol} vs the logarithm of Ab_{mol} are listed in Table 5.5. An appreciable trend is detected with log abundances of CO_2 and CH_4, while the H_2O metric shows only a weak trend with input log abundance. Anti-correlations between e.g., M_{CH_4} - $\log(Ab_{CO_2})$, or M_{H_2O}-$\log(Ab_{CH_4})$ are present as we are considering juxtaposed bands to size these molecules, as listed in Table 5.3. The logarithmic abundances of H_2O and NH_3 show similar correlations with M_{H_2O}. While this is expected, as the two molecules manifest similar spectral shapes, the water sensitivity of the metric to the abundance may also be limited by the noise, by a bias squeezing the metric to small values, or both, and further investigation is required in future work. However, the metric is an estimator for the classification of atmospheres on the basis of their molecular content, and it would be misleading to expect the metric to provide robust estimates of abundances, for which spectral retrieval techniques are more appropriate. These



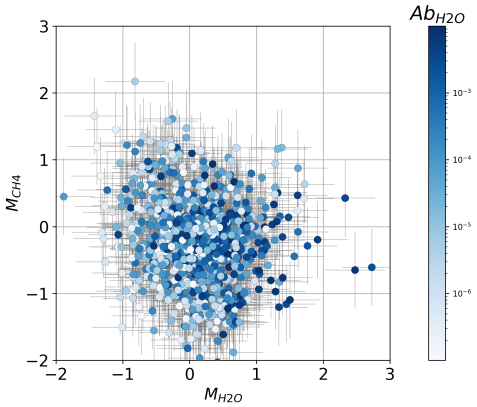
(a) POP-I estimated $M_{CO_2} - M_{CH_4}$ colored by CH_4.



(b) POP-I estimated $M_{CO_2} - M_{CH_4}$ colored by CO_2.

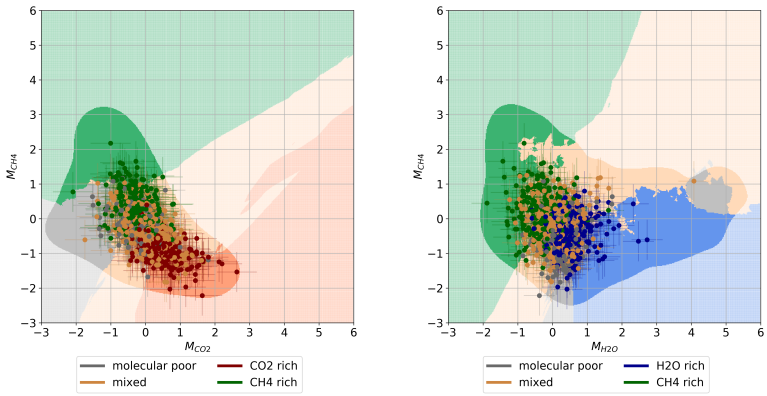


(c) POP-I estimated $M_{H_2O} - M_{CH_4}$ colored by CH_4.



(d) POP-I estimated $M_{H_2O} - M_{CH_4}$ colored by H_2O.

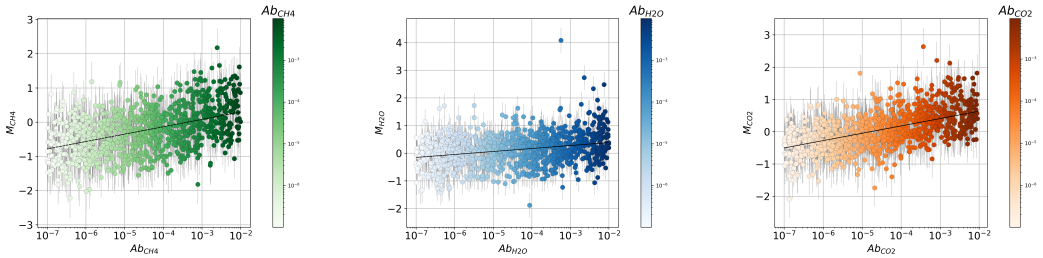
Figure 5.6. Diagrams for comparison between M_{CO_2} - M_{CH_4} and M_{H_2O} - M_{CH_4}. In these figures, each point represents an observed POP-I planet, and the color scale reflects the input abundances. gray horizontal and vertical lines are the metric estimated dispersion. By comparing the diagrams on the left with those on the right we can see that planets bearing more CH_4 are located on the top left, while the ones bearing more CO_2 and H_2O are on the bottom right.



(a) $M_{CO_2} - M_{CH_4}$ - observed spectra. **(b)** $M_{H_2O} - M_{CH_4}$ - observed spectra.

Figure 5.7. k-Nearest Neighbors analysis results with $k = 20$ for CO_2-CH_4 (left) and H_2O-CH_4 (right) cases. The superimposed dots are from the POP-I observed spectra and the error bars represent the metric dispersion. Colors correspond to classes described in Table 5.4. gray dots: planets that contain less than 10^{-5} in mixing ratio for the considered molecules; green points: planets that contain 10 times more CH_4 than the other molecule and $Ab_{CH_4} > 10^{-4}$; red points: planets that hold 10 times more CO_2 than CH_4 and $Ab_{CO_2} > 10^{-4}$; blue points: planets with 10 times more H_2O than CH_4 and $Ab_{H_2O} > 10^{-4}$; yellow dots: all the other possible configurations. The same color scheme applies to the painted region of the diagram, built from the noiseless spectral data. gray area: planets with low quantities of water and methane; green area: where we expect to have methane-rich planets, blue: for water-rich planets; yellow: for mixed atmospheres. The regions best sampled by the noiseless data, as described in Section 5.1.4.1, are fully colored, while other regions are transparent.

aspects are further discussed in Section 5.3.2 as well as in Section 5.3.4, where we show with an example how a retrieval exercise is effective in constraining the input abundances of the molecules considered, water included.



(a) M_{CH_4} versus CH_4 true abundance value.

(b) M_{H_2O} versus H_2O true abundance value.

(c) M_{CO_2} versus CO_2 true abundance value.

Figure 5.8. Comparison between the M_{mol} estimates for each planet and the true molecular abundance value, Ab_{mol}, in the atmospheres. CH_4, H_2O, and CO_2 cases are shown in the leftmost, middle, and rightmost panel, respectively. Data points with error bars represent POP-I planets. The color scale gives a visual representation of the molecular abundance in the atmosphere. A linear fit is shown by the solid black lines in each panel, with coefficients listed in Table 5.5. The fitted lines superimposed to the data highlight a positive correlation between the true molecular abundance values and the values estimated by the metric, M_{mol}.

Table 5.5. Here we report the fitted C_0 (top table) and C_1 (bottom table) coefficients for $M_{mol} = C_0 \cdot \log(Ab_{mol}) + C_1$ for all the possible combination of considered molecules. The bands used for M_{mol} are reported in Table 5.3.

(a) C_0 coefficients.

	$\log(Ab_{H_2O})$	$\log(Ab_{CH_4})$	$\log(Ab_{CO_2})$	$\log(Ab_{NH_3})$
M_{H_2O}	0.108(±0.022)	−0.193(±0.022)	−0.033(±0.022)	0.104(±0.022)
M_{CH_4}	0.003(±0.022)	0.215(±0.022)	−0.258(±0.022)	0.104(±0.022)
M_{CO_2}	−0.094(±0.022)	−0.057(±0.022)	0.228(±0.022)	−0.104(±0.022)

(b) C_1 coefficients.

	$\log(Ab_{H_2O})$	$\log(Ab_{CH_4})$	$\log(Ab_{CO_2})$	$\log(Ab_{NH_3})$
M_{H_2O}	0.599(±0.102)	−0.756(±0.106)	−0.028(±0.105)	0.590(±0.105)
M_{CH_4}	−0.241(±0.102)	0.725(±0.106)	−1.419(±0.105)	0.215(±0.105)
M_{CO_2}	−0.357(±0.102)	−0.202(±0.106)	1.088(±0.105)	−0.411(±0.105)

We can use Figure 5.8 to obtain an estimate of the probability that a molecule *mol* has abundance in excess of 10^{-4}, conditioned to the metric being larger than some value $M_{mol,*}$, i.e. $P(Ab_{mol} > 10^{-4} | M_{mol} > M_{mol,*})$. For this, we can use the well-known chain rule for the conditional probability that states that $P(A|B) = P(A \cap B)/P(B)$, where A and B are two separate events. We estimate the number of data points found in a region of the diagrams of Figure 5.8 where both conditions are satisfied (favorable outcomes) divided by the number of data points for which only the condition $M_{mol} > M_{mol,*}$ is satisfied (total outcomes). From

POP-I observed spectra, we can obtain a single realization of P. Therefore we simulate 1000 realizations of POP-I observed spectra, using the same input noiseless POP-I population spectra, and randomizing the noise realizations. In this way, we simulate 1000 realizations of P from which medians and 1-σ confidence levels are computed.

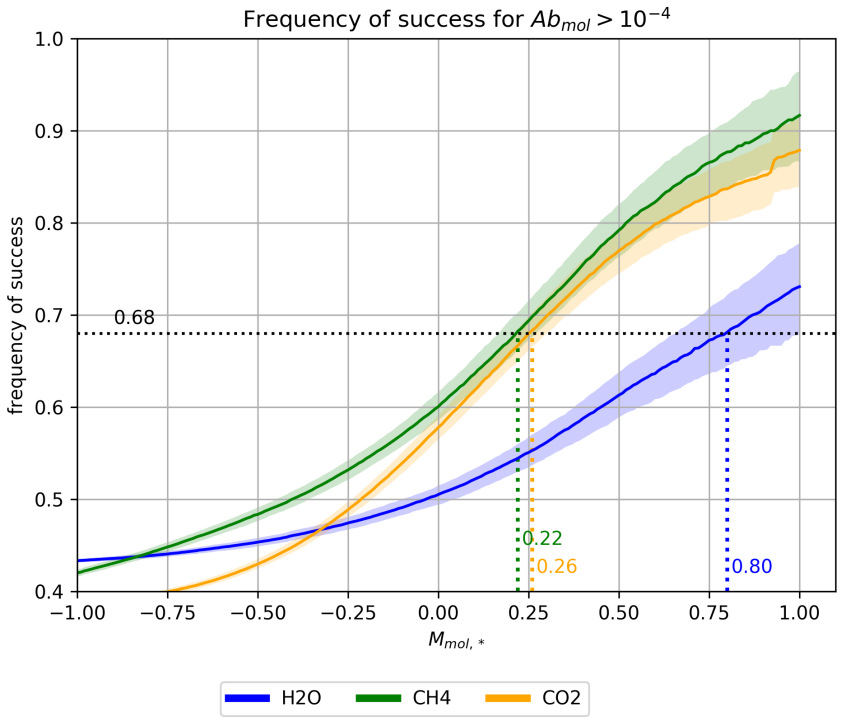


Figure 5.9. Probability that a molecule *mol* has abundance in excess of 10^{-4}, conditioned to the metric being larger than some value $M_{mol,*}$, i.e. $P(Ab_{mol} > 10^{-4} | M_{mol} > M_{mol,*})$. CH_4, H_2O, and CO_2 cases are shown by the green, blue, and orange lines, respectively. The lines are computed as the median of the probability estimates from 1000 different realizations of the POP-I observed population. The shaded regions are the $1-\sigma$ confidence levels associated with the median probability. Vertical dotted lines mark metric values, $M_{mol,*}$, corresponding to a probability of 68%.

Figure 5.9 suggests that the metric can be used to classify planetary primary atmospheres for the presence of CH_4 and CO_2, and to a lesser extent H_2O, and atmospheres that are likely missing these molecular contributions. With reference to Figure 5.9, it can be seen that when $M_{CH_4} \geq 0.5$, the number of planets wrongly classified to have $Ab_{CH_4} > 10^{-4}$ is only 20%, or 1 out of 5 are false positives. However, as expected, the case of water is different, and our metric is not as effective in detecting the presence of water as it is for the other molecules. Even for large values of M_{H_2O}, the rate of false positives is close to 40%.

5.2.3 Deep and Machine Learning

The percentages of correct classifications for all considered molecules and for different minimum input abundances are reported in Table 5.6a for kNN, in Table 5.6b for MLP, in

Table 5.6c for RFC and in Table 5.6d for SVC.

Table 5.6. Percentages of correct identifications for the considered molecules and with different thresholds. In each column, we report a different minimum Ab_{mol} and in each row a different molecule. The percentages represent how many of the atmospheres have been correctly identified by the algorithm to have at least the specified minimum amount of that molecule, and therefore they represent the algorithm's accuracy. Each ML algorithm has been trained on POP-III and tested on POP-I.

(a) KNN percentages of success to identify spectra bearing different minimum amounts of molecules.

Molecule	$Ab_{mol} > 10^{-5}$ [%]	$Ab_{mol} > 10^{-4}$ [%]	$Ab_{mol} > 10^{-3}$ [%]
CH_4	79	83	85
CO_2	77	79	82
H_2O	64	71	82
NH_3	75	82	84

(b) MLP percentages of success to identify spectra bearing different minimum amounts of molecules.

Molecule	$Ab_{mol} > 10^{-5}$ [%]	$Ab_{mol} > 10^{-4}$ [%]	$Ab_{mol} > 10^{-3}$ [%]
CH_4	78	85	87
CO_2	77	81	83
H_2O	70	76	84
NH_3	80	86	87

(c) RFC percentages of success to identify spectra bearing different minimum amounts of molecules.

Molecule	$Ab_{mol} > 10^{-5}$ [%]	$Ab_{mol} > 10^{-4}$ [%]	$Ab_{mol} > 10^{-3}$ [%]
CH_4	77	82	87
CO_2	76	79	83
H_2O	69	74	82
NH_3	78	85	87

(d) SVC percentages of success to identify spectra bearing different minimum amounts of molecules.

Molecule	$Ab_{mol} > 10^{-5}$ [%]	$Ab_{mol} > 10^{-4}$ [%]	$Ab_{mol} > 10^{-3}$ [%]
CH_4	79	86	89
CO_2	79	83	84
H_2O	69	78	84
NH_3	81	87	87

Table 5.6 shows that for all Deep and Machine Learning algorithms, the percentages of success in identifying the presence of molecules inside the atmosphere grow with the minimum molecular abundances that we set as a threshold for the classification. While this is expected, it may come as a surprise that in general, these algorithms appear to be effective in detecting the presence of all individual molecules with a relatively small fraction of false positives (about 30% or smaller) even at low abundances. This is perhaps because ML algorithms learn to classify atmospheres by recognizing spectral shapes. These algorithms' performances can be to a certain level independent of the molecules considered, as long as the training set contains sufficiently diverse spectra to allow a secure identification,

including water in the presence of ammonia or biases, that is where our metric shows its more severe weaknesses. We also notice from Table 5.6 that kNN, MLP, RFC, and SVC show comparable overall performance, and that CH_4 and CO_2 are the most straightforward molecules to identify in Tier 1 planetary spectra.

A comparison between these results and our metric is presented in Section 5.3.5.

5.3 Discussion

In this section, we discuss the metric results shown in Section 5.2.2. We first discuss the bias (Section 5.3.1), then we focus on the metric characteristics, such as the relation between the metric estimates and the input molecular abundances (Section 5.3.2) and the detection limits (Section 5.3.3). Then we compare the metric performance with a spectral retrieval (Section 5.3.4), and with Deep and Machine Learning algorithms (Section 5.3.5).

5.3.1 Metric bias

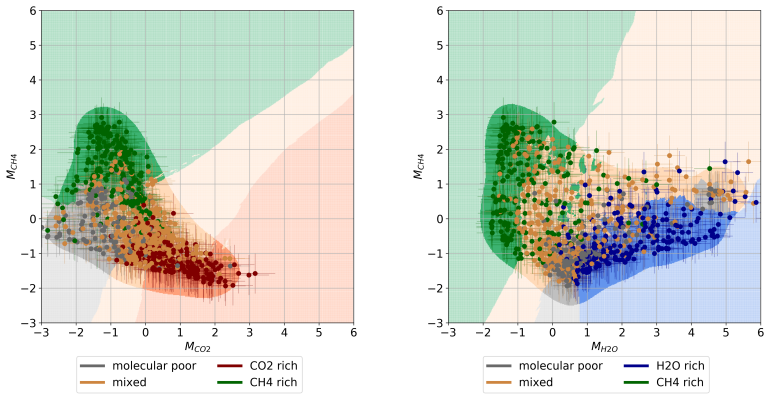
The kNN analysis discussed earlier and shown in Figure 5.7 is trained on POP-I noiseless spectra, and the data points shown in that figure are obtained estimating the metric on POP-I observed spectra, as described in Section 5.1.4.1. To verify if the metric is biased, the kNN analysis is repeated with data points obtained estimating the metric on POP-I noiseless spectra. This is shown in Figure 5.10 that should be compared with Figure 5.7. The background colors are very similar in either case, with small variations due to the training process that selects randomly 70% POP-I noiseless examples. In the absence of biases, we expect the distribution of observed data points to be that of noise-less data points, convolved with the distribution of the noise. However, it can be noticed from the comparison of the two figures, that the distribution of the observations is more clustered towards the origin of the coordinate axes, compared to noiseless data points. This is a consequence of the bias introduced by the metric normalization discussed in Section 5.1.4: normalization is required such that the metric response is insensitive to the atmospheric scale height, and sensitive only to the presence of molecular signatures, at the cost of biasing the estimator. We should additionally point out that Figure 5.9 results are also affected by the bias. The observing noise reduces the M_{mol} average estimates, and therefore for smaller observing noise, the three colored lines in the figure are shifted to the right, and the 68% of success corresponds to higher M_{mol} values.

The work presented here demonstrates that the metric we have designed is a powerful tool capable of revealing the presence of a molecule in an atmosphere and that the prediction is independent of the type of the planet and its basic parameters (such as temperature, radius, and pressure) within the limits explored here. However, this comes at the cost of biasing the estimator by a quantity that depends on the instrumental noise as discussed in Section 5.1.4. Provided that the metric can be de-biased, it can be used in a predictive way where an observation (along with its dispersion estimate) can be compared to the calibrated (trained) metric space to infer the possible molecular content of the target. Because instrumental noise can be well characterized, it would be possible to de-bias the metric estimator. This requires a detailed noise analysis, taking into account the uncertainties in the noise estimates, which is beyond the scope of this thesis. In the rest of this section, we focus on what we can learn from this kind of analysis provided that the metric can be de-biased, and we leave to future work a detailed study on how this de-biasing can be secured.

5.3.2 Relation with the input abundances

We see in Figure 5.8 that the correlation between M_{mol} and $\log(Ab_{mol})$ is in general not strong enough to quantify the input molecular abundances. This is because atmospheric spectra are made of complex non-linear contributions from all the molecules. Therefore, a method based only on spectral shapes (i.e., this metric), is inadequate to quantify molecular abundances. However, the goal of this metric, provided that the bias can be removed, is not to assess the abundance of a certain species in the planet's atmosphere, but only its possible presence, avoiding the use of spectral retrieval techniques, that may not be indicated for Tier 1 data.

Focusing on Table 5.5 and looking at the coefficients fitted for M_{H_2O} over $\log(Ab_{H_2O})$ and over $\log(Ab_{NH_3})$ we may infer that the metric may not be effective to distinguish between water and ammonia. However, the degeneracy can be broken by performing a spectral retrieval if the target was observed at Ariel Tier 2 SNR, as shown in an example in Section 5.3.4. This population analysis is based on the study of spectral shapes only, and it does not make use of parameters such as planetary mass, radius, and temperature. Although it has proven difficult to distinguish between water and ammonia with this metric, using some knowledge of planetary properties may help us to disentangle the two molecules in future work; for example, while a Neptune can hold ammonia, a Hot Jupiter planet is not expected to. One of the goals of Tier 1 is to identify targets with interesting spectra to be re-observed in higher SNR Tiers. From this point of view, even if the metric cannot clearly separate between water and ammonia, it can suggest the presence of interesting molecules in the spectrum. This can in turn be used to make informed decisions about targets to be selected for further studies.



(a) $M_{CO_2} - M_{CH_4}$ - noiseless spectra. **(b)** $M_{H_2O} - M_{CH_4}$ - noiseless spectra.

Figure 5.10. This figure is the equivalent of Figure 5.7, but the superimposed dots are now from the POP-I noiseless spectra, and the errorbars represent the metric dispersion on the spectra before the application of Ariel's observing noise. The parameter space area best sampled by the noiseless data is now well-filled with the dots.

5.3.3 Metric detection limit

To explore the detection limit of molecules by the metric, we examine the molecular poor/spectral flat region of Figure 5.2. A planet spectrum would be found in that region

because of i) clouds, ii) a low temperature (i.e. small scale height), iii) low molecular abundances, or a combination of the three. In all cases, the spectrum is expected to be featureless, i.e. flat. Point iii) is defined from input abundances smaller than 10^{-5} (Table 5.4). The metric detection limit can then be investigated by removing flat spectra before training the kNN, by rising before training the molecular poor spectra threshold to above 10^{-5}, and by monitoring the kNN classification results. As the threshold increases, we expect the kNN to begin failing the molecular poor/flat classification when spectra can no longer be considered flat.

We perform the kNN training on the noiseless spectra of both POP-I and POP-II, the latter containing only CH_4 and H_2O, the former containing all molecules considered in this work. Each noiseless spectrum has its associated observed spectrum. Flat spectra are identified on observed spectra, and the corresponding noiseless spectra are ignored in the KNN training.

The motivation behind using POP-II is as follows. If we have a population containing only CH_4 and H_2O and we properly remove all planets with a flat spectrum, there should be no targets left with non-detectable molecular features. In the case of POP-I, however, we do not expect all the planets with Ab_{CH_4} and $Ab_{H_2O} < 10^{-5}$ to be flat, because other molecules (CO_2 and NH_3) can show features. Therefore, the flat spectra removal procedure will not empty the molecular poor planets class in this population. Using POP-II instead, we expect that, after removing all flat planets, there will not be molecular poor atmospheres anymore. The procedure is summarized in Figure 5.11.

The outcome of this analysis is shown for POP-I and POP-II in respectively Figure 5.12 and Figure 5.13. Only the calibrated regions are shown and data points have been omitted for clarity. Figure 5.12a shows the POP-II kNN analysis with all planets and planetary classes of Table 5.4, in Figure 5.12b the kNN is trained removing flat spectra from the training set, and in Figure 5.12c the training is done removing flat spectra first, and raising the threshold of molecular poor spectra from $Ab_{mol} < 10^{-5}$ to $Ab_{mol} < 10^{-4}$. We notice that Figure 5.12b shows no molecular poor atmosphere after excluding spectrally flat cases. This confirms that our metric is able to separate the more complex atmospheres from the flat ones in the simple case of only two molecules. By contrast, Figure 5.12c still shows a gray area, signifying that atmospheres with $10^{-5} < Ab_{mol} < 10^{-4}$ cannot be considered flat. This can be interpreted as a molecular detection limit. We also notice from the figure that these spectra populate the bottom left corner of the best-sampled area of the diagram, meaning that they are classified as having the smallest spectral features of the samples. This confirms the relation between the metric and the molecule abundance. The detection limit is expected to improve in Tier 2 observations, and Changeat et al. (2020a) find that the detection limit using spectral retrieval techniques on Tier 2 is about two orders of magnitude smaller compared to that of the metric.

In Figure 5.13 we remove all flat spectra from the planetary population POP-I and we report the results of kNN analysis. Here we see that, as expected, while removing all flat spectra from POP-II also removes all molecular-poor instances, the same does not occur in POP-I. In this case, molecular-poor spectra in any two molecules, such as CH_4-CO_2 or CH_4-H_2O, may appear non-flat because of the presence of the other two molecules, i.e. NH_3-H_2O or NH_3-CO_2, respectively.

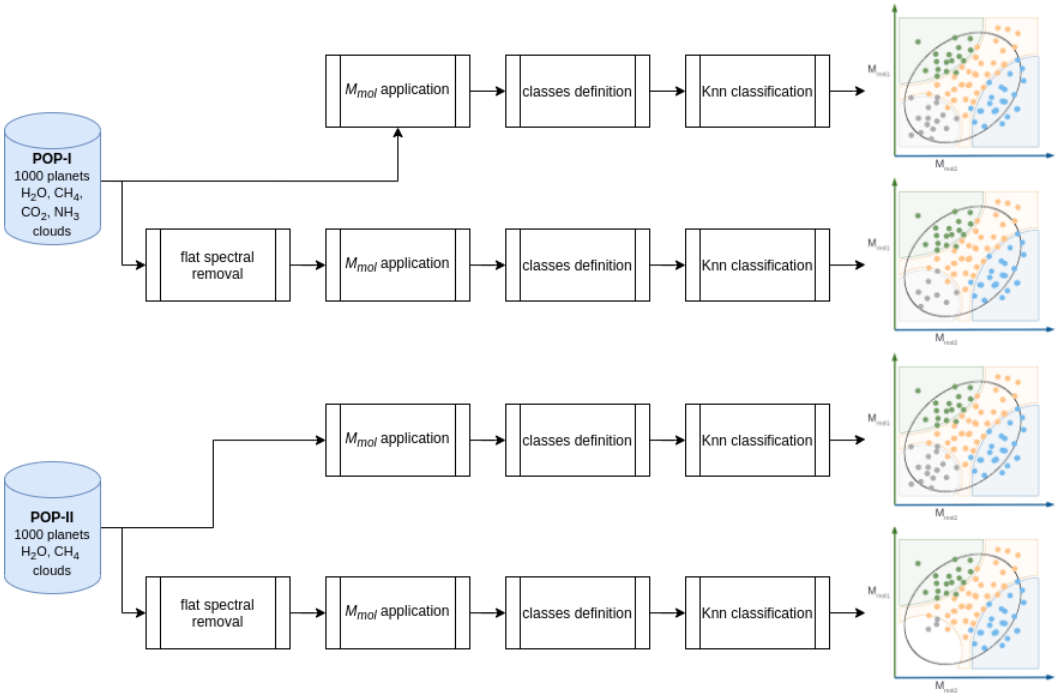
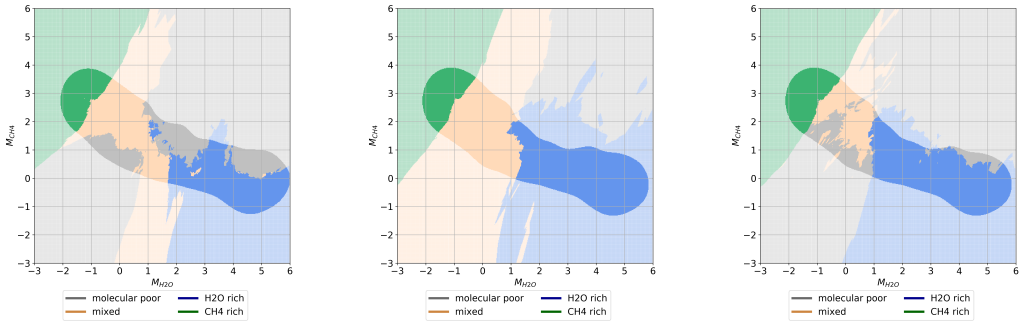


Figure 5.11. The figure shows the strategy adopted to identify the molecular detection limit for the developed metric. Starting from *POP-I*, we classify the planets as described in Section 5.1.4.1. Without removing the flat spectra from the population, we would end up with the same results described in Figure 5.4; by contrast, if we remove flat spectra, we end up with similar results but with fewer molecular poor planets, because even without flat spectra atmospheres, there will be planets bearing molecules different from the couple investigated by the plot. Different is the case of *POP-II*: here we have only two molecules in the population, and therefore if we remove the flat spectra planets, we will end with no molecular poor atmospheres.

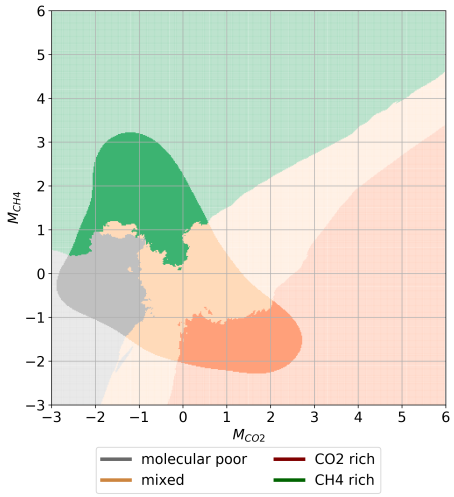


(a) KNN classification map for CH_4 and H_2O, including flat spectra. Molecular poor planets defined as Ab_{CH_4} and $Ab_{H_2O} < 10^{-5}$.

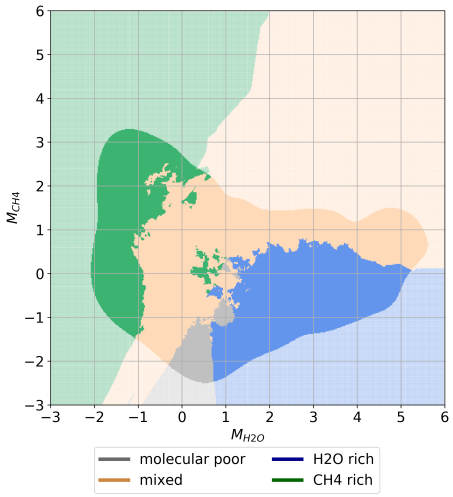
(b) KNN classification map for CH_4 and H_2O without flat spectra. Molecular poor planets defined as Ab_{CH_4} and $Ab_{H_2O} < 10^{-5}$.

(c) KNN classification map for CH_4 and H_2O without flat spectra. Molecular poor planets defined as Ab_{CH_4} and $Ab_{H_2O} < 10^{-4}$.

Figure 5.12. KNN analysis for the POP-II population, considering the full data set (left) and excluding flat spectra (center and right). The diagrams are obtained following the bottom branches of Figure 5.4: we used the noiseless planetary spectra to classify the metric space and to select the best-sampled regions.



(a) k-Nearest Neighbors for POP-I population without flat spectra. CH_4-CO_2 case.



(b) k-Nearest Neighbors for POP-I population without flat spectra. CH_4-H_2O case.

Figure 5.13. The figure shows the population POP-I where the flat spectra have been masked. On the left is reported the case of CH_4 and CO_2 and on the right is the case of CH_4 and H_2O. The diagrams are obtained following the bottom branches of Figure 5.4: we used the noiseless planetary spectra to classify the parameter space and to select the best-sampled areas. Comparing this figure with Figure 5.7, we notice that the "molecular poor" area is still present because even if there are no CO_2 and CH_4 in the planet atmosphere, there could be NH_3 and H_2O having features (left case) or if there are no H_2O and CO_2 there could be NH_3 and CO_2 (right case).

5.3.4 Input abundances retrieval

We compare here two atmospheric retrievals of the same planet observed both in Tier 1 and in Tier 2. This exercise has two goals:

1. to confirm that a spectral retrieval is capable of disentangling water and ammonia, and to constrain the atmospheric composition of POP-I targets observed in Tier 2 with Ariel;

2. to show that even though it is possible to perform a spectral retrieval on Tier 1 data for some selected planets, its performance is comparable with that of the metric.

From the POP-I planets, we select one that has water and ammonia in high abundance, low cloud presence, high temperature, and a diameter larger than Jupiter's. Such selection will help us to investigate the capability of Tier 2 observed data (simulated as described in Section 5.1.2) to break the water-ammonia degeneracy, as well as to estimate the uncertainties from a retrieval using Tier 1 observed data only.

To perform the retrieval, we use TauREx 3 (Al-Refaie et al., 2021). The parameters fitted with fit boundaries, true, and retrieved values are listed in Table 5.7, while the retrieved solutions and posteriors are shown in Figure 5.14.

For the selected planet, we notice that in Tier 2 the abundances of the molecules considered are well constrained, and, as expected, low-level (high-pressure) clouds are undetected in both cases.

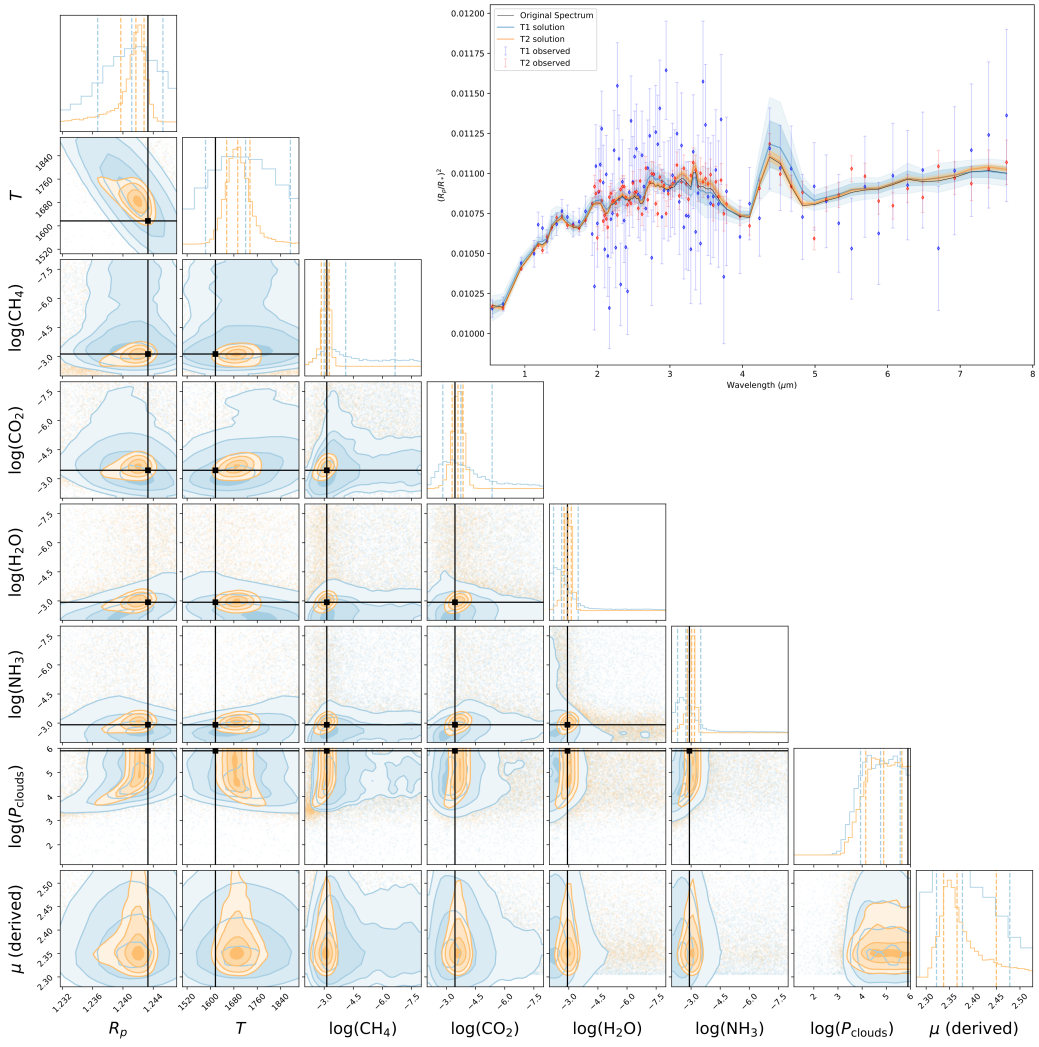


Figure 5.14. Retrieved spectra and posteriors. The corner plot shows the posteriors for each retrieved parameter using Tier 1 (blue) and Tier 2 (orange) observed data. Input values are shown by the black lines. The panel in the top right corner shows the retrieved spectra from Tier 1 (blue) and Tier 2 data using colored shaded bands for 1 and 2 $-$ σ uncertainties, and the input (black solid line). The notation log(X), where X is one of CH_4, CO_2, H_2O or NH_3, represents the logarithm of the molecular abundance of the given species and should be compared to log(Ab_X).

Table 5.7. Retrieval parameter table showing fit boundaries, true inputs, and retrieved parameters with uncertainties for Tier 1 and Tier 2 observations. As in Figure 5.14, the notation $\log(X)$, where X is one of CH_4, CO_2, H_2O or NH_3, represents the retrieved logarithm of the molecular abundance of the given species and should be compared to the input $\log(Ab_X)$.

Name	Boundaries	True value	Tier 1 retrieved	Tier 2 retrieved
R_p $[R_{Jup}]$	$[0.5 \rightarrow 2]$	1.24	$1.241^{+0.004}_{-0.004}$	$1.2412^{+0.0020}_{-0.0011}$
T_p $[K]$	$[800 \rightarrow 2400]$	1617	1720^{+153}_{-137}	1693^{+42}_{-38}
$\log(CH_4)$	$[-8 \rightarrow -2]$	-3.13	$-4.11^{+1.11}_{-2.55}$	$-3.08^{+0.23}_{-0.18}$
$\log(CO_2)$	$[-8 \rightarrow -2]$	-3.44	$-3.74^{+0.93}_{-1.61}$	$-3.59^{+0.30}_{-0.26}$
$\log(H_2O)$	$[-8 \rightarrow -2]$	-2.93	$-2.63^{+0.42}_{-0.85}$	$-2.96^{+0.20}_{-0.17}$
$\log(NH_3)$	$[-8 \rightarrow -2]$	-2.91	$-2.73^{+0.43}_{-0.77}$	$-3.03^{+0.22}_{-0.15}$
$\log(P_{clouds})$	$[-3 \rightarrow 6]$	5.90	$4.76^{+0.81}_{-0.83}$	$4.89^{+0.75}_{-0.74}$

The Tier 1 results can be linked to our previous analysis on molecular input abundance detection (Section 5.3.2). We compute the probability of having molecular abundances greater than 10^{-4} from the retrieval posteriors and compare these with the probability obtained with our metric (Figure 5.9). In this case, the measured M_{mol} are: $M_{CH_4} = -0.47$, $M_{CO_2} = 0.54$ and $M_{H_2O} = 0.29$. The results are listed in Table 5.8. Tier 2 observations provide a confident detection of methane, carbon dioxide, and water, while Tier 1 retrievals are broadly comparable to our metric approach in detecting the presence of these molecules.

These results appear to confirm that spectral retrievals may not be best suited or at the very least necessary to analyze Tier 1 data. Retrievals are model-dependent, and one needs to define planet parameters, as well as cross-sections, pressure-temperature profiles, etc. Priors might need to be imposed to ensure convergence. Retrievals are also computationally expensive, making it not trivial to conduct the analysis on hundreds of targets. A photometric metric instead, is model-independent, which may be an advantage when assessing a planet observation for the first time. The full analysis takes only minutes on a desktop computer to reduce 1000 observations.

Table 5.8. Probability to have $Ab_{mol} > 10^{-4}$ for each molecule computed from M_{mol} and from Tier 1 and Tier 2 retrieval posteriors. The numbers refer to the planet case discussed in Section 5.3.4.

Molecule	M_{mol} [%]	Tier 1 [%]	Tier 2 [%]
CH_4	49	48	100
CO_2	78	58	94
H_2O	56	89	100

5.3.5 Comparison with Deep and Machine Learning

ML techniques are difficult to interpret, and so a comparison between their performance and that of our metric can help us in gaining confidence in the outcomes from ML classifiers. For this purpose, we consider a planet as bearing a molecule if $Ab_{mol} > 10^{-4}$. Then with our metric, we select all planets that have $M_{CH_4} \geq 0.22$ that according to Figure 5.9 corresponds to a probability of $\sim 68.3\%$ to have a $Ab_{mol} > 10^{-4}$ for CH_4.We repeat the same procedure,

letting $M_{CO_2} \geq 0.26$ for CO_2 and $M_{H_2O} \geq 0.80$ for H_2O. In each sample, we check how many of the selected planets have molecular abundances in excess of 10^{-4}, obtaining a percentage of success for our metric (or metric precision). In the same way, we check how many of the planets flagged by each of the ML algorithms in the full sample actually bear the molecules, such that we can compare their precision performance in Table 5.9.

We notice a marginally better success rate for ML algorithms in the cases of kNN and MLP, while RFC and SVC algorithms suggest a better performance when compared to that of the metric. Better performances are expected because, while our metric considers only specific bins in the spectrum, the classification algorithms gather information from all the spectral data points. The comparable performance of the metric with the kNN and MLP suggests that the molecular bands chosen for the metric are not far from ideal, but the comparatively better performances of RFC and SVC provide an indication that margins for improvement may exist.

While more work is required along this path, which is beyond the scope of this work, Deep and Machine Learning appear to be very promising for this classification problem, and we shall leave to dedicated works, as the one presented in Hou Yip et al. (2020), a more exhaustive investigation of these techniques, their comparison with more physically motivated strategies similar to the metric, and a thorough investigation of biases that may affect all these techniques.

Table 5.9. Percentages of positive detection for our metric, compared to Deep Learning algorithms precision. To assess the presence of a molecule we flag a planet if $Ab_{mol} > 10^{-4}$. We investigate CH_4 in the first row, CO_2 in the second and H_2O in the third, selecting the planets with $M_{CH_4} \geq 0.22$ (first row), $M_{CO_2} \geq 0.26$ (second row) and $M_{H_2O} \geq 0.80$ (third row).

Molecule	M_{mol} [%]	**KNN** [%]	**MLP**[%]	**RFC**[%]	**SVC**[%]
CH_4	69*	75.4	84.2	92.5	90.1
CO_2	68.3	71.4	75.8	83.1	83.5
H_2O	68.3	74.5	79.0	96.7	99.4

* These percentages arise from a discrete distribution of data and therefore we cannot exactly identify the 68.3% quantity. In this case 69% is the closest possible value.

Chapter 6

Detection of Molecules and Promotion to Higher Tiers

This chapter details the follow-up study on evaluating the information content of Ariel's reconnaissance survey spectra, initiated in the study presented in Chapter 5. To overcome the bias of the molecular metric discussed there, we developed a different data analysis technique that provides quantitative predictions for the molecular classification of planets. The goal is a reliable and unbiased method to address: *Can we use Tier 1 transmission spectra to identify the presence of a molecule, with an associated calibrated probability?*. Hence, we aim to find calibrated probabilities that can also guide the decision-making process to select Tier 1 targets for re-observation in Ariel's higher Tiers.

In Bocchieri et al. (2023), we developed a technique based on a P-statistic that measures the probability that a molecule is present in an atmosphere observed in Tier 1 in excess of a given threshold. The input data is the marginalized posterior distribution for a molecular abundance from spectral retrieval, and the probability is estimated as the integral of the distribution on all abundance values above the threshold. We focused our investigation on (i) assessing the reliability of the classification, (ii) evaluating its robustness, i.e. whether it provides valid results across a variety of conditions, and (iii) its predictive power. We found that, at least in the parameter space explored, the P-statistic can be used to implement a reliable and powerful predictor of the presence of molecules, as long as the retrieval model matches the complexity of the data.

* * *

Here we detail the work and results from this second study of Tier 1 spectra. Section 6.1 outlines the methodology used in this analysis. Section 6.1.1 describes our data analysis strategy for detecting a molecule in these spectra. Section 6.1.2 details our experimental data set, including the planetary population, forward model parameters, atmosphere randomization, and noise estimation. Section 6.1.3 summarizes the spectral retrievals performed, discussing the optimization algorithm and the priors used. Section 6.1.5 describes the data analysis tools used to evaluate the probability forecasts of the method. Section 6.2 details the results obtained in terms of forecast reliability (Section 6.2.1), predictive power (Section 6.2.2), and bias of the abundance estimator utilized (Section 6.2.3). Finally, Section 6.3 provides an overall discussion of the results.

6.1 Methods

Tier 1 transmission spectra contain sufficient information to infer the presence of several atmospheric molecules (Mugnai et al., 2021b), but Tier 1 observations are in general non-ideal for quantitative spectral retrievals in terms of molecular abundances, as they are required to achieve a SNR $\geq$ 7 when binned in only seven effective photometric data points in the 0.5–7.8 μm wavelength range (Edwards et al., 2019). Abundance posterior probabilities from retrievals can however still be informative and here we develop a new method to identify the presence of molecules in Tier 1 transmission spectra starting from these posteriors.

6.1.1 Analysis strategy

Given a marginalized posterior distribution of a molecular abundance, we compute an empirical probability, P, that the molecule is present in the atmosphere of a planet, with an abundance above some threshold, $\mathbb{T}_{Ab}$, as:

$$P \simeq \int_{\mathbb{T}_{Ab}}^{\infty} \mathcal{P}(x)dx \tag{6.1}$$

where $\mathcal{P}$ is the marginalized posterior distribution and x represents the abundance values. Thus, the predicted P depends on the assumed atmospheric model and the selected abundance threshold $\mathbb{T}_{Ab}$. If the assumed atmospheric model is representative of the observed atmosphere, then a clear correlation (above noise) between P and the true abundance in Tier 1 data implies that P can be used to identify the most likely spectra that contain a molecule, providing a preliminary classification of planets by their molecular content. Thus, this P-statistic can be considered robust (Wall and Jenkins, 2012), even when $\mathcal{P}(x)$ is too broad to constrain the abundance.

To test whether this method is sensitive enough, we need to simulate transmission spectra as observed in Tier 1, using an atmospheric model that includes a certain number of molecules. Then, we need to perform a spectral retrieval with the same atmospheric model and compare each input molecular abundance with the predicted P corresponding to that molecule. The test is successful if, for an agreed $\mathbb{T}_{Ab}$, we recover a high P for each large input abundance and a low P for each small input abundance. To understand how well the method behaves under conditions similar to the Ariel reconnaissance survey, we repeat this test on a large and diverse planetary population.

In this study, we employ a simulated population of approximately 300 transmission spectra of H_2-He gaseous planets, which contain CH_4, H_2O, and CO_2 trace gases with randomized input abundances. Additionally, we introduce NH_3 with randomized abundances as a nuisance parameter since its spectral features overlap with those of water and other molecules. We utilize NH_3 to test the P-statistic's efficacy and investigate the robustness of its predictions under various assumptions, such as the exclusion of NH_3 from retrievals or the inclusion of additional molecules not present in the population.

Therefore, we can study whether this method provides reliable predictions under less favorable conditions when the assumed model is not fully representative of the observed atmosphere. This might provide some insight into how robustly the method can reveal the presence of a molecule in a real observation when the atmosphere is unknown. For this, we add or remove molecules from the retrieval model (hereafter, "fit-composition") with

respect to the simulated composition. Then, we perform different spectral retrievals, that use different fit-compositions, and compare the predictions obtained from the *P*-statistic with the input abundances.

6.1.1.1 Model exploration

We consider three cases in our analysis. In the first case (referred to as R_0), we use an atmospheric model that includes CH_4, H_2O, CO_2, and NH_3 as trace gases, which matches the composition used in the forward model generation of the population.

In the second case (referred to as R_1), we consider a fit-composition that includes only CH4, CO2, and H2O, omitting NH3. In this case, there is a possibility of inadequate representation of the data because NH_3's molecular features could overlap with the observed features of other molecules (hence its adoption as a nuisance), particularly H_2O (Encrenaz et al., 2015). As a result, the retrieved values of P may not accurately reflect the input abundances of H_2O, leading to decreased reliability of the predictions.

In the third case (referred to as R_2), we expand the fit-composition beyond the input composition by including also CO, HCN, and H_2S. It should be noted that the spectral features of these additional molecules could also overlap with the observed features of the other molecules. For instance, CO and CO_2 exhibit a spectral overlap around 4.5 μm. Hence, even in this case, obtaining reliable predictions of the input composition may not be obvious.

Table 6.1 provides a summary of the molecules included in the fit-composition for each retrieval. For more detailed information on the retrievals performed, please refer to Section 6.1.3.

Table 6.1. Molecules included in the fit-composition for each retrieval.

Retrieval	CH_4	CO_2	H_2O	NH_3	CO	HCN	H_2S
R_0	✓	✓	✓	✓			
R_1	✓	✓	✓				
R_2	✓	✓	✓	✓	✓	✓	✓

6.1.2 Experimental data set

As a simulated population, we use a planetary population generated using the Alfnoor software (Changeat et al., 2020a; Mugnai et al., 2021b). Alfnoor is a wrapper of TauREx 3 (Al-Refaie et al., 2021) and ArielRad (Mugnai et al., 2020). Given a list of candidate targets and a model of the Ariel payload, it automatically computes simulated exoplanet spectra as observed in each Ariel Tier.

Specifically, we use a subset of the POP-I planetary population of Mugnai et al. (2021b). POP-I consists of 1000 planets from a possible realization of the Ariel Mission Reference Sample (MRS) of Edwards et al. (2019). That MRS (hereafter, MRS19) comprises known planets in 2019 from NASA's Exoplanet Archive and TESS forecast discoveries. Here we ignore the TESS forecasts, thus obtaining a sub-population of around 300 planets, that we label POP-Is. Using POP-Is planets ensures that, in principle, we can compare our results with those of Mugnai et al. (2021b).

Figure 6.1 shows that POP-Is comprises a diverse sample of planets mostly with large radii ($\gtrsim 5$ $R_\oplus$), short orbital periods ($\leq 4/5$ days), warm to hot equilibrium temperatures ($500 - 2500\,^\circ K$) and stellar hosts with different magnitudes in the K band of the infrared spectrum ($8 - 12$ m_K). Compared to the parameter space sampled by the entire POP-I, this data set has more occasional statistics on smaller and longer-period planets around brighter stars.

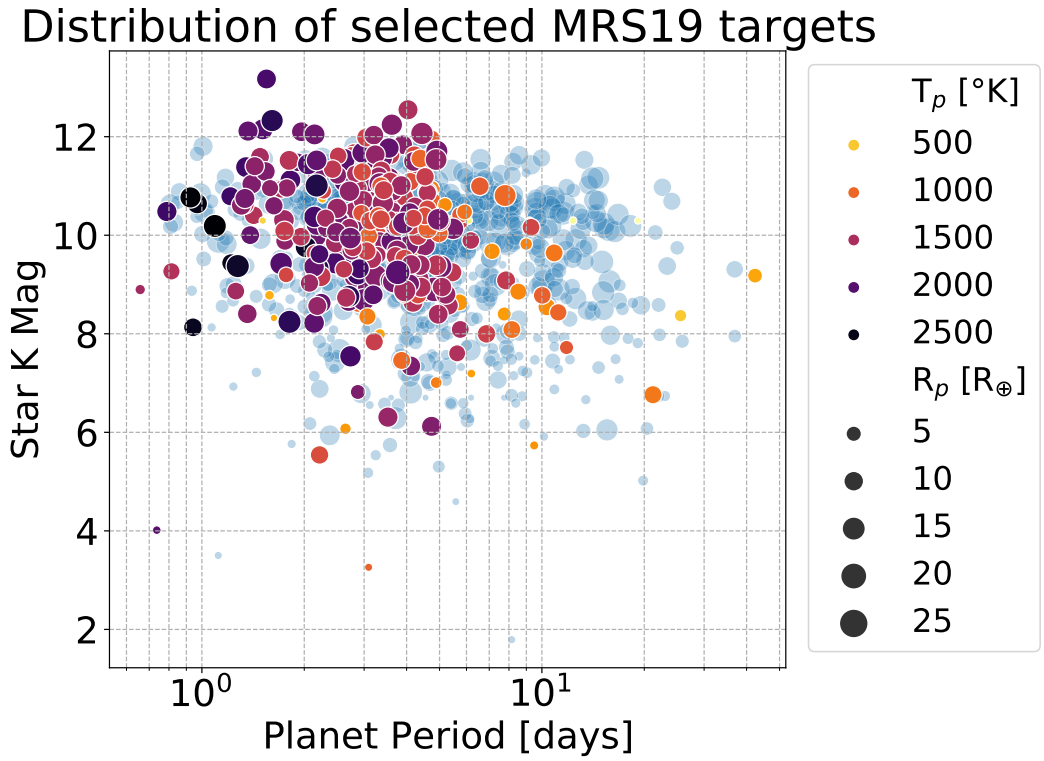


Figure 6.1. Parameter space distribution of the POP-Is planetary population used in this work, which comprises about 300 selected planets from MRS19. The horizontal axis reports the planetary orbital period in days; the vertical axis reports the stellar magnitude in the K band. Each data point represents a planet; the symbol size is proportional to the planetary radius in Earth's radii; the symbol color shows the expected planetary equilibrium temperature. Light blue data points in the background show the entire MRS19/POP-I parameter space for reference.

The detailed properties of POP-I (and therefore POP-Is) are discussed in Mugnai et al. (2021b) and briefly summarized here. The forward model parameters are randomized to test diverse planetary atmospheres. The baseline atmosphere is a primordial atmosphere filled with H_2 and He with a solar mixing ratio of $He/H_2 = 0.17$. The vertical structure of the atmosphere comprises 100 pressure layers, uniformly distributed in log space from 10^{-4} to 10^6 Pa, using the plane-parallel approximation. The equilibrium temperature of each planet is randomized between $0.7 \times T_p$ and $1.05 \times T_p$, where T_p is the equilibrium temperature of the planet listed in MRS19; the atmospheric temperature-pressure profile is isothermal. Constant vertical chemical profiles are added for H_2O, CO_2, CH_4, and NH_3, with abundances randomized according to a logarithmic uniform distribution spanning 10^{-7} to 10^{-2} in Vertical Mixing Ratios (VMR). Randomly generated opaque gray clouds are also added with a surface pressure varying from $5{\times}10^2$ to 10^6 Pa to simulate cloudless

to overcast atmospheres. Table 6.2 summarizes the randomized parameters of the POP-I forward models. For each planet, POP-I contains the raw spectrum binned at each Ariel Tier resolution ("noiseless spectra"), the associated noise predicted by the Ariel radiometric simulator, ArielRad, for each spectral bin, and the number of transit observations expected to reach the Tier-required SNR. To simulate an observation, we scatter the noiseless spectra according to a normal distribution with a standard deviation equal to the noise at each spectral bin. The "observed spectra" data set is built by repeating this process for each planet in POP-Is. As in Mugnai et al. (2021b), the Tier 1 data used in this work are binned on the higher resolution Tier 3 spectral grid: R = 20, 100, and 30, in NIRSpec, AIRS-CH0, and AIRS-CH1, respectively. The noise is that of Tier 1, which yields a SNR > 7 if data were binned on the Tier 1 spectral grid. This is to prevent the loss of spectral information that may occur in binning.

Table 6.2. Forward model randomized parameters in POP-I.

Parameter	Unit	Range	Scale
T_P / $T_{P;MRS19}$	$^\circ K$	0.7; 1.05	linear
CH_4	VMR	10^{-7}; 10^{-2}	log
CO_2	VMR	10^{-7}; 10^{-2}	log
H_2O	VMR	10^{-7}; 10^{-2}	log
NH_3	VMR	10^{-7}; 10^{-2}	log
P_{clouds}	Pa	5×10^2; 10^6	log

6.1.3 Retrievals summary

To perform the retrievals, we use the TauREx 3 retrieval framework (Al-Refaie et al., 2021), the same used to generate the raw POP-Is spectra. In the retrieval model, we include opaque gray clouds, pressure-dependent molecular opacities of various trace gases, Rayleigh scattering, and Collision-Induced Absorption (CIA) of H_2-H_2 and H_2-He. Table 6.3 reports a referenced list of CIA and all molecular opacities used in this study.

Table 6.3. List of opacities used in this work and their references.

Opacity	Reference(s)
H_2-H_2	(Abel et al., 2011; Fletcher et al., 2018)
H_2-He	(Abel et al., 2012)
H_2O	(Barton et al., 2017; Polyansky et al., 2018)
CH_4	(Hill et al., 2013; Yurchenko and Tennyson, 2014)
CO_2	(Rothman et al., 2010)
NH_3	(Yurchenko et al., 2011; Tennyson and Yurchenko, 2012)
CO	(Li et al., 2015)
H_2S	(Azzam et al., 2016)
HCN	(Barber et al., 2014)

The free parameters of the retrievals are the radius and mass of the planet, as well as the molecular mixing ratios, as listed in Table 6.4. We use broad logarithmic uniform priors for

the molecular abundances, ranging from 10^{-12} to 10^{-1} in VMR. For the mass and radius of the planet, we select uniform priors of 20% and 10% around the respective values listed in MRS19. The gray cloud pressure levels are not included as free parameters in the retrieval because of their degeneracy with other parameters such as the radius Changeat et al. (2020b).

Table 6.4. Fit parameters and their priors for the retrievals.

Parameters	Units	Priors	Scale
M_P	M_J	±20%	linear
R_P	R_J	±10%	linear
CH_4	VMR	10^{-12}; 10^{-1}	log
CO_2	VMR	10^{-12}; 10^{-1}	log
H_2O	VMR	10^{-12}; 10^{-1}	log
NH_3	VMR	10^{-12}; 10^{-1}	log
CO	VMR	10^{-12}; 10^{-1}	log
HCN	VMR	10^{-12}; 10^{-1}	log
H_2S	VMR	10^{-12}; 10^{-1}	log

We take a conservative approach by choosing larger bounds for the priors than those used for the random forward spectra generation, reported in Table 6.2.

We set the evidence tolerance to 0.5 and sample the parameter space through 1500 live points using the MultiNest algorithm[1] (Feroz et al., 2009; Buchner, 2021a). We disable the search for multiple modes to obtain a single marginalized posterior distribution of each molecular abundance to insert in Equation 6.1.

We then perform the three different retrievals (respectively R_0, R_1, and R_2) described in Section 6.1.1 on each POP-Is planet. We use the Atmospheric Detectability Index (ADI) (Tsiaras et al., 2018) to assign statistical significance to the results of these retrievals. Given the Bayesian evidence of a nominal retrieval model, E_N, and of a pure-cloud/no-atmosphere model, E_F, the ADI is:

$$\text{ADI} = \begin{cases} log(E_N) - log(E_F), \text{if } log(E_N) > log(E_F) \\ 0, \text{otherwise} \end{cases} \tag{6.2}$$

ADI is a positively defined metric, equivalent to the log-Bayesian factor (Kass and Raftery, 1995; Jenkins and Peacock, 2011) where $\log(E_N) > \log(E_F)$. To compute E_F, we perform an additional retrieval for each planet with a flat-line model with the planet radius being the only free parameter.

6.1.4 Abundance threshold

We utilized the marginalized posteriors to estimate the *P*-statistic using an abundance threshold of $\mathbb{T}_{Ab} = 10^{-5}$, which is considered "molecular-poor" according to the definition by Mugnai et al. (2021b). This threshold is higher by 1–2 orders of magnitude compared

[1] v3.11, Release April 2018

to the Tier-2 detection limits reported by (Changeat et al., 2020a). The "molecular-poor" condition is met for approximately 40% of the atmospheres due to the randomization boundaries set for each molecule (see Table 6.2). The ability to detect a molecule depends on factors such as opacities, correlations among molecules, and noise in the measured spectrum. Therefore, $\mathbb{T}_{Ab}$ can be optimized for each molecule in future work, although we applied the same abundance threshold for all in this pilot study.

6.1.5 Data analysis tools

The *P*-statistic can be used to reliably classify planets for the presence of a molecule with an abundance above $\mathbb{T}_{Ab}$ when *P* correlates with the *Ab* true value. The stronger the correlation above noise fluctuations, the larger the predictive power. Because this classification is binary and *P* is defined in the range $0 \rightarrow 1$, we can use standard statistical tools such as calibration curves and ROC curves (Sanders, 1963; Wilks, 2019) to evaluate the performance of this method in revealing the presence of molecules and in selecting Tier 1 targets for higher Tiers. These curves are routinely utilized by the Machine Learning community[2], as they present the forecast quality of a binary classifier in a well-designed graphical format.

6.1.5.1 Calibration curves

A calibration curve (e.g. Wilks, 2019) plots the forecast probability averaged in different bins on the horizontal axis and the fraction of positives, in each bin, on the vertical axis (see Figure 6.2 for a generic example). In this work, the fraction of positives is the fraction of POP-Is planets with true abundance larger than $\mathbb{T}_{Ab}$, and the forecast probability is the corresponding *P*-statistic. Calibration curves provide an immediate visual diagnosis of the quality of binary classifier forecasts and the biases that the forecasts may exhibit.

For well-calibrated predictions, the forecast probability is equal to the fraction of positives, except for deviations consistent with sampling variability. Therefore, the ideal calibration curve follows the 1:1 line. Miscalibrated forecasts can be biased differently depending on whether the calibration curve lies on the left or on the right of the 1:1 line. A curve entirely to the right of the 1:1 line indicates an over-forecasting bias, as the forecasts are consistently too large relative to the fraction of positives, as seen in the calibration curve of Classifier 1 in Figure 6.2. On the contrary, the calibration curve of Classifier 2 shows the characteristic signature of under-forecasting, being entirely on the left of the 1:1 line, indicating that the forecasts are consistently too small relative to the fraction of positives. There may also be more subtle deficiencies in forecast performance, such as an under-confident forecast, with over-forecasting biases associated with lower probabilities and under-forecasting biases associated with higher probabilities, as seen in the calibration curve of Classifier 3.

Calibration curves paint a detailed picture of forecast performance, often summarized in a scalar metric known as the Brier Score (B-S, Brier, 1950), which is defined as the mean square difference between probability forecasts and true class labels (positive or negative); the lower the B-S, the better the predictions are calibrated. From Figure 6.2, we see that Classifier 3 achieves the best B-S, although the forecasts are not well calibrated. In general, uncalibrated forecasts can be calibrated using calibration methods such as Platt scaling

[2]In Python, the package scikit-learn (Pedregosa et al., 2011) (v1.0) provides the method `calibration_curve` in `sklearn.calibration` and the method `roc_curve` in `sklearn.metrics`.

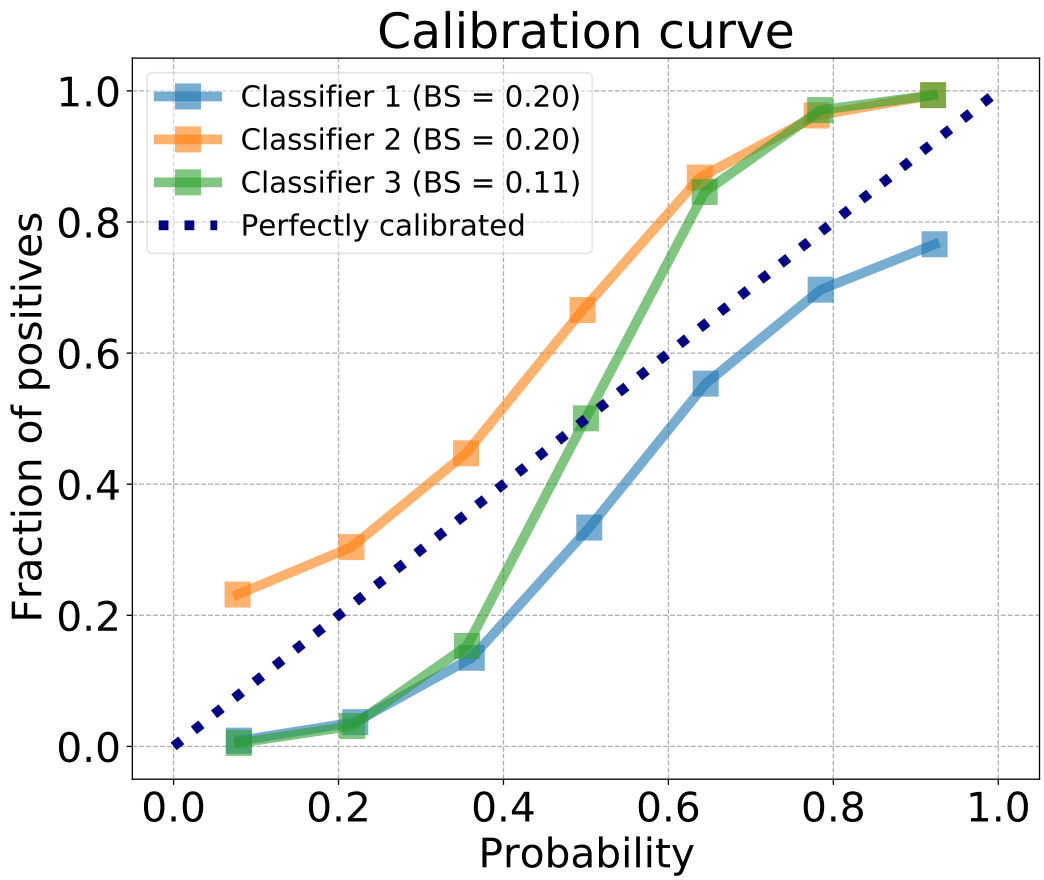


Figure 6.2. Calibration curves of three mock classifiers, exhibiting different forecast quality and biases. The legend reports the B-S of the forecasts of each classifier. The calibration curve for perfectly calibrated forecasts is reported for reference.

and Isotonic regression (Platt et al., 1999; Zadrozny and Elkan, 2002; Niculescu-Mizil and Caruana, 2005).

6.1.5.2 ROC curves

Given the predicted probabilities of a classifier, and a selected probability threshold $\mathbb{P}$, the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), are defined in Table 6.5.

Table 6.5. Contingency table formulating all four possible outcomes of a binary classification problem.

		True label	
Forecast	Forecast label	Yes	No
$P \geq \mathbb{P}$	Yes	TP	FP
$P < \mathbb{P}$	No	FN	TN

A binary classifier with high predictive power assigns larger P to positive observations (true label "Yes") and smaller P to negative (true label "No"). This maximizes TP and TN, and minimizes FP and FN.

A ROC curve (e.g. Wilks, 2019) is a square diagram that illustrates the predictive power at different values of the probability threshold $\mathbb{P}$. It plots the False Positive Rate (FPR) on the horizontal axis and the True Positive Rate (TPR) on the vertical axis (see Figure 6.3 for a generic example), defined as:

$$\text{FPR} = \frac{\text{FP}}{\text{Negatives}} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{6.3a}$$

$$\text{TPR} = \frac{\text{TP}}{\text{Positives}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6.3b}$$

FPR and TPR are commonly known as "false alarm" and "hit" rates. ROC curves are constructed by calculating the TPR and FPR from the number of TP, TN, FP, and FN as $\mathbb{P}$ decreases from 1 to 0. The ideal classifier minimizes the FPR while maximizing the TPR; thus, its ROC curve is the unit step function. On the other hand, the worst possible classifier is a random classifier with a ROC curve along the 1:1 line. Real-world classifiers have intermediate ROC curves ranked by how close they are to the unit step function. As seen in Figure 6.3, Classifier 3 exhibits the highest predictive power, as the corresponding ROC curve arcs everywhere above the ROC curves for Classifiers 1 and 2.

ROC curves portray a detailed picture of predictive power, often summarized in a scalar metric known as the Area Under the Curve (AUC), the fraction of the unit square area subtended by a ROC curve. The higher the AUC, the higher the predictive power. The ideal classifier has AUC = 1.0; the random one has AUC = 0.5. From Figure 6.3, we see that, as expected, Classifier 3 also achieves the largest AUC.

ROC curves can also be used to select the optimal classification threshold $\mathbb{P}$, which roughly corresponds to the position on the curve where the TPR cannot be raised without

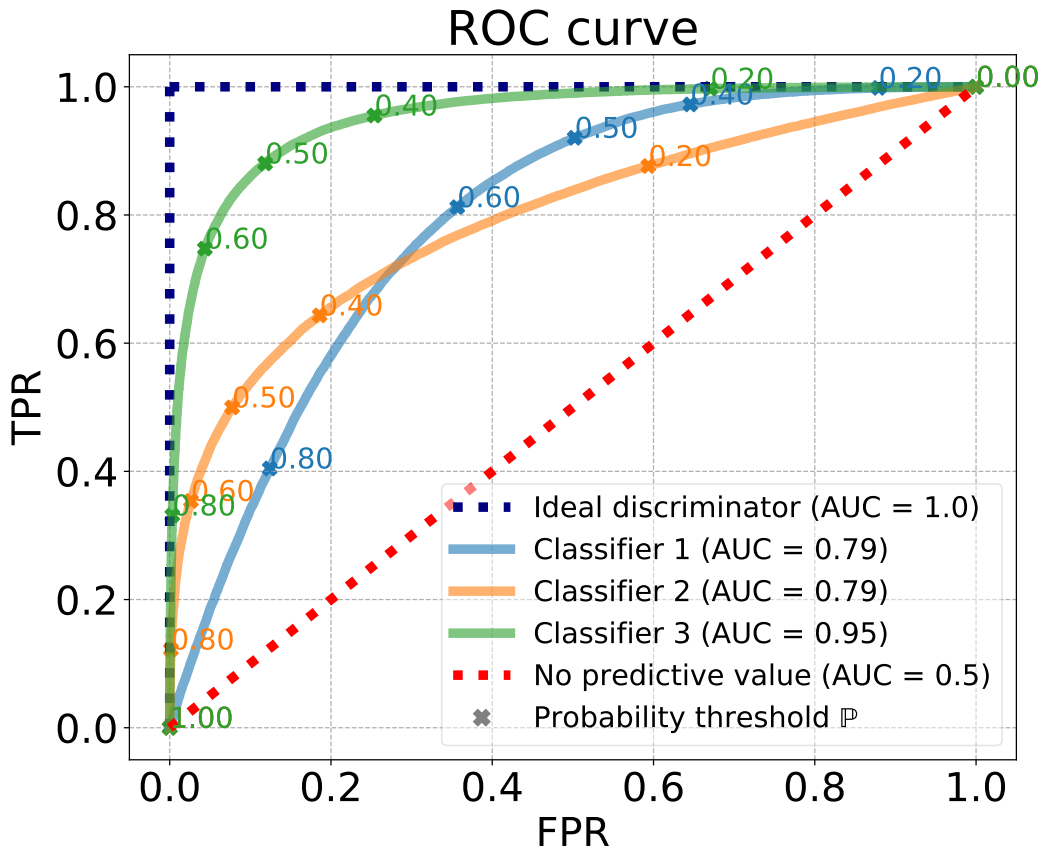


Figure 6.3. ROC curves of the same mock classifiers shown in Figure 6.2, exhibiting different predictive powers. The legend reports the AUC associated with each ROC curve. The ideal and worst possible classifier ROC curves are reported for reference. Several probability thresholds $\mathbb{P}$ at regularly spaced intervals are also displayed on each curve.

significantly increasing the FPR. For example, as seen in Figure 6.3, the optimal $\mathbb{P}$ for Classifier 3 is around 0.5, where it achieves a TPR of nearly 0.9 at a low FPR of approximately 0.1. Reducing $\mathbb{P}$ to 0.4 is not advantageous, as it only increases the TPR to approximately 0.95, at the expense of increasing the FPR to almost 0.3.

6.1.6 Using calibration and ROC curves

Using calibration curves and the B-S metric, we can immediately diagnose the forecast quality of the *P*-statistic and its potential biases. Suppose that the forecast probability P matches the fraction of planets with input abundances greater than $\mathbb{T}_{Ab}$ (fraction of positives) in each probability bin. In that case, the prediction of the method is well-calibrated. Moreover, we can compare the forecast quality achieved for different molecules using the B-S metric. If the forecasts are not well calibrated, we can infer which kind of bias affects the predictions of the method by inspecting the shape of the calibration curve. If the forecasts show an over-forecasting bias (as in the example of Classifier 1, Fig. 6.2) and therefore incorrectly classify a fraction of planets as bearing a molecule, too many Tier 1 planets may be selected for re-observation in higher Tiers, resulting in less optimal scheduling of observations. On the contrary, an under-forecasting bias (as in the example of Classifier 2, Fig. 6.2) may imply that fewer Tier 1 planets than possible would be scheduled for re-observing in higher Tiers.

Using ROC curves and the AUC metric, the power of the *P*-statistic to predict the presence of molecules can be assessed. The closer the ROC curve approaches the unit step function (AUC $\simeq$ 1, Fig. 6.3), the higher the predictive power. Moreover, we can directly compare the predictive power achieved for different molecules by analyzing the shape of the corresponding ROC curves and the AUC values.

The shape of the ROC curve provides a way to select the optimal classification threshold, $\mathbb{P}_*$, for the problem under study. For instance, $\mathbb{P}_*$ can be chosen in a trade-off process that maximizes the TPR while keeping the FPR at an acceptable low value.

This choice can aid the selection of Tier 1 targets for re-observation in a higher Tier: a large FPR would result in a poor allocation of observing time while a low TPR would result in a reduction of observational opportunities. It can also benefit population studies where one might need to track the presence of certain molecules across families of planets and extrasolar systems. These types of studies are outside the scope of this work, but can profit from the methodology developed here.

6.2 Results

As detailed in Section 6.1.1, we designed a method based on the *P*-statistic to reveal the presence of a molecule in Tier 1 spectra. In the following sections, we use the statistical tools described in Section 6.1.5 to show the performance of the *P*-statistic in predicting the presence of several molecules in our simulated planetary population. In particular, in Section 6.2.1, we use calibration curves to assess the reliability of the predictions of the method and related biases, while in Section 6.2.2, we use ROC curves to assess the predictive power of the method and discuss the optimal classification threshold, $\mathbb{P}_*$. In Section 6.2.3, we use the median abundance as an estimator of the true abundance and investigate its biases in the low SNR regime to explain the biases observed in the calibration curves.

6.2.1 Detection reliability

6.2.1.1 Retrieval R_0

Figure 6.4 shows the analysis performed to evaluate the reliability of the method when using the abundance posteriors of the retrieval R_0, which uses the same atmospheric composition as the one used in the generation of the simulated atmospheres (see Table 6.1). The subplots in each column share the same horizontal axis with the predicted probability P that a molecule is present with an input abundance, Ab_{mol}, above the selected abundance threshold $\mathbb{T}_{Ab} = 10^{-5}$ (see Section 6.1.4). The figure reports the results for CH_4, H_2O, and CO_2, shown from left to right, respectively.

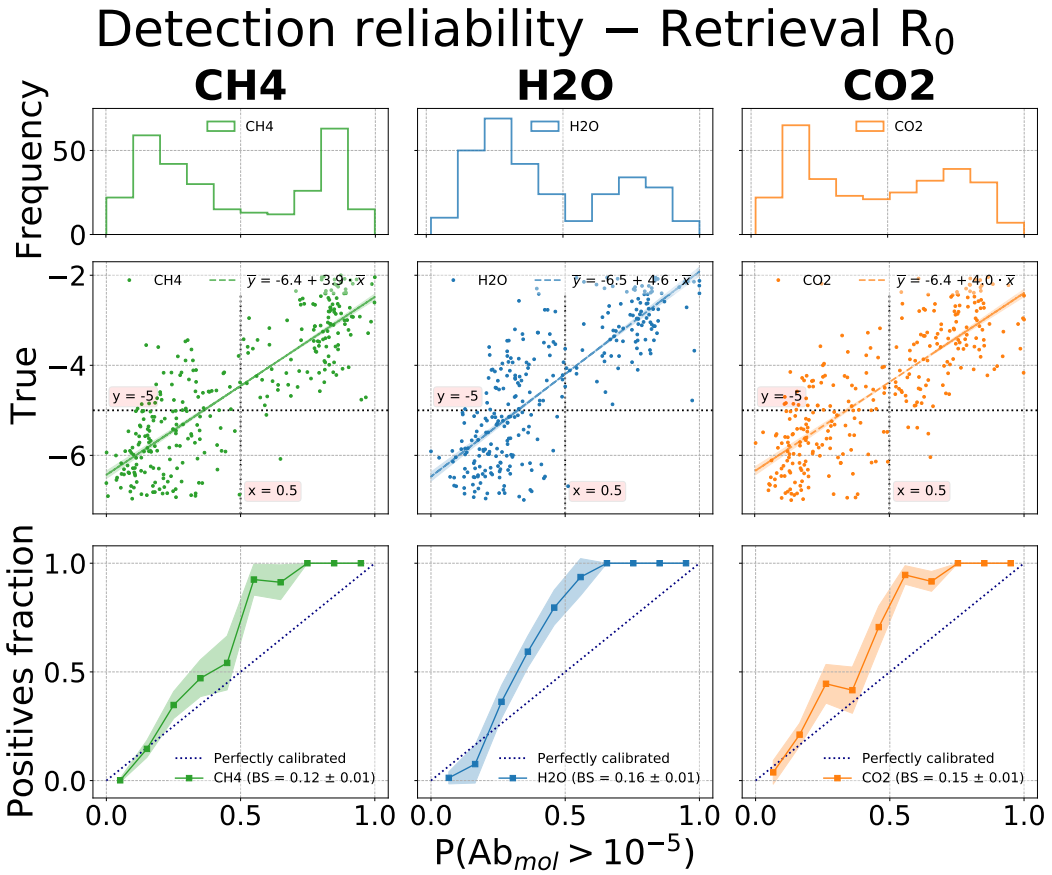


Figure 6.4. Detection reliability analysis for CH_4, H_2O, and CO_2 from the R_0 retrievals, that implement a model that is fully representative of the simulated atmospheres. All plots in the same column share the same horizontal axis with the predicted probabilities, $P(Ab_{mol} > 10^{-5})$, that a molecule is present in the atmosphere of a planet, with an abundance above the selected abundance threshold, $\mathbb{T}_{Ab} = 10^{-5}$. Top row: histogram with the frequency of the P forecasts. Middle row: diagrams showing the correlation between P values on the horizontal axis and input abundances on the vertical axis. The linear fit parameters of the data points are reported on each legend. For visual reference, the dotted horizontal lines show the position of $\mathbb{T}_{Ab}$ and the dotted vertical lines the value 0.5 on the x-axis. Bottom row: calibration curves with associated bootstrap confidence intervals; each legend shows the B-S of the forecasts.

The top row displays histograms of the P-statistic realizations, which exhibit a bimodal

distribution. Two peaks are observed in the distribution, with one located at $P \approx 0.2$ and the other at $P \approx 0.8$, with the former being more prominent. Additionally, a valley is observed at intermediate values, with $P \approx 0.5$.

The middle row shows the correlation between the predicted probabilities on the horizontal axis and the input abundances of each molecule on the vertical axis. We take a rough measure of the correlation by calculating the angular coefficient of the data points from a linear fit. These coefficients are listed in Table 6.6. The lower right quadrant of these diagrams ($P \gtrsim 0.5$ and $Ab_{mol} < 10^{-5}$) is almost empty of data points, indicating that whenever the method predicts a high P, the corresponding input abundance is likely higher than $\mathbb{T}_{Ab}$. However, not all planets with an input abundance greater than $\mathbb{T}_{Ab}$ are associated with a high P, as the upper left quadrants of these diagrams ($P \lesssim 0.5$ and $Ab_{mol} > 10^{-5}$) are not empty of data points.

The bottom row shows the calibration curves computed for each molecule; each curve is shown with a bootstrap confidence interval calculated using 1000 bootstrap samples. That is, following Press et al. (1992), we randomly remove $\sim 1/e \approx 36\%$ of the data from each of these samples and replace them by repeating some randomly chosen instances of the ones kept. It should be stresses that the bootstrap confidence intervals are addressing the uncertainty of the sample itself (i.e. the epistemic uncertainty) but not the uncertainty of each sample point (i.e. the aleatoric uncertainty). The latter is not addressed in this work and would necessitate estimating the uncertainties of the abundance posteriors, to be propagated through the P-statistic.

For each molecule, we calculate the B-S using the `brier_score_loss` method of `sklearn.metrics` (Pedregosa et al., 2011), with the associated uncertainty estimated from the same bootstrap samples. Table 6.6 lists the B-S values obtained.

Table 6.6. Best-fit value for the angular coefficient m from the linear fit $\log(Ab_{mol}) \propto m\, P(Ab_{mol} > \mathbb{T}_{Ab})$, with $\mathbb{T}_{Ab} = 10^{-5}$, and Brier Score for the calibration curves for all possible combinations of retrievals and molecules.

Retrieval	molecule	m	B-S [%]
R_0	CH_4	3.9	12 ± 1
	H_2O	4.6	16 ± 1
	CO_2	4.0	15 ± 1
R_1	CH_4	3.2	15 ± 1
	H_2O	3.8	17 ± 1
	CO_2	3.7	14 ± 1
R_2	CH_4	3.9	13 ± 1
	H_2O	4.4	16 ± 1
	CO_2	3.9	16 ± 1

The calibration curves show an under-forecasting bias (curve to the left of the 1:1 line; see Section 6.1.5.1) especially associated with larger forecast probabilities, giving a fraction of positives ≈ 1.0 for $P \gtrsim 0.6$. On the contrary, the probabilities are better calibrated for $P \lesssim 0.4$. From the B-S values (less accurate forecasts receive higher B-S), we see that CH_4 is the best-scoring molecule, probably due to its strong absorption spectral features.

It is possible that the observed under-forecasting of the calibration curves and the

bimodality of the *P*-statistic distribution are both related to the sampling of the parameter space. This is briefly discussed further in Section 6.3.2.

6.2.1.2 Retrieval R_1

Figure 6.5 shows the same analysis for the retrieval R_1, which includes only CH_4, CO_2, and H_2O in the fit-composition and excludes NH_3, although this molecule is present in the data set (see Table 6.1). Comparing the histograms from the top row of this figure with those

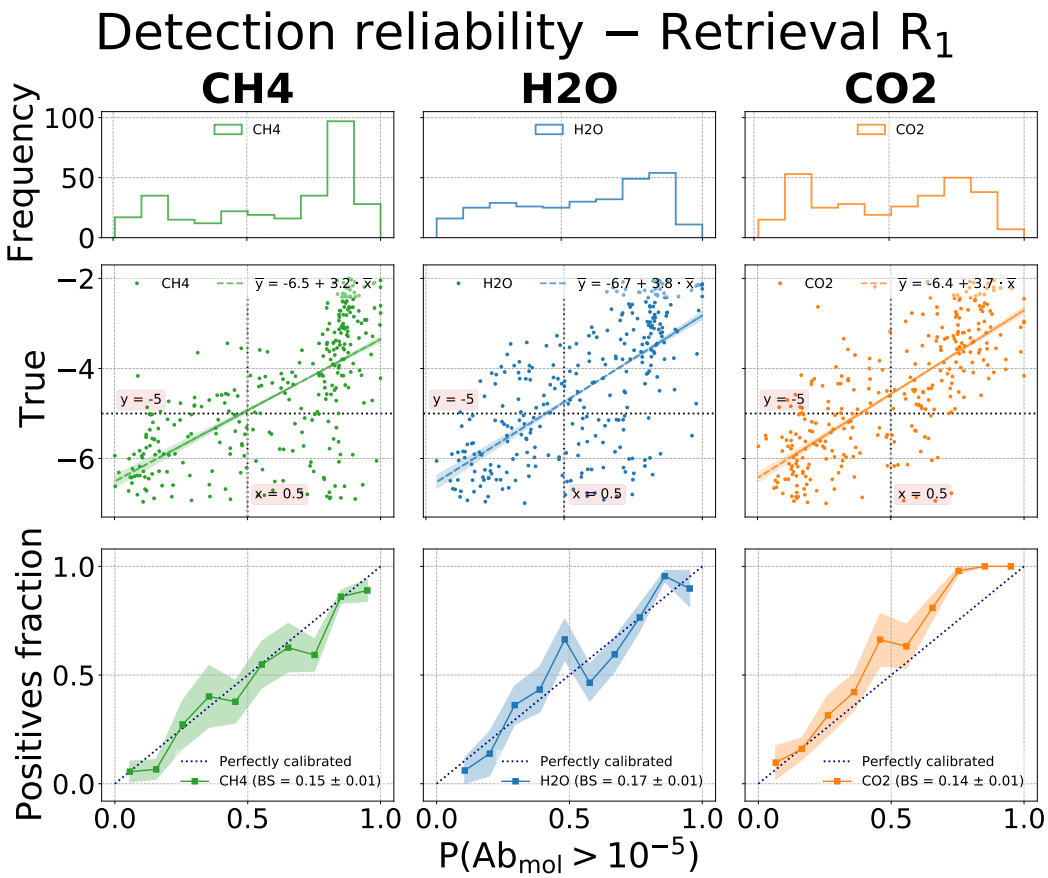


Figure 6.5. Same as Figure 6.4. Detection reliability for the R_1 retrievals, implementing a model that excludes NH_3 from the fit-composition.

obtained for the retrieval R_0 (Figure 6.4), we notice a decrease in the forecast frequency at low *P*, especially for CH_4 and H_2O, with a reduced peak at *P* around 0.2. On the contrary, high values of *P* are more frequent, enhancing the peak at *P* around 0.8: for CH_4, more than 30% of the data set receives *P* between 0.8 and 0.9. These are samples with high input abundance.

The plots in the middle row show an increase in the scatter in the data points compared to R_0. In this case, we find a decrease in the correlation between *P* and the input abundances, and the angular coefficients of the linear fit are reported in Table 6.6. Planets that receive $P \gtrsim 0.8$ have high input abundance, $Ab_{mol} > 10^{-5}$.

The calibration curves for H_2O and CH_4 in the bottom row are, within the uncertainties, closer to the 1:1 line than for R_0, both for high and low forecast probabilities. Although this

might appear closer to the ideal behavior, it could be misleading. The B-S is higher than for R_0, because the mean squared difference between the forecasts and true class labels is larger. This is visualized in the middle plots: for $Ab_{mol} < 10^{-5}$ (negative true class label), there are many forecast values with $P > 0.5$. In other words, the correlation between the P-statistic and the true input abundances is weaker. In contrast, the entire CO_2 calibration curve shows the signature of under-forecasting. The curve for CO_2 is almost the same as for R_0, likely because the missing NH_3 affects less the CO_2 abundance posteriors. On the other hand, the overlap of NH_3 with H_2O but also CH_4 makes the model used in the retrieval less suitable to describe the data.

The reduced correlation between probability forecasts and input abundances, as well as the higher B-S values, suggest that excluding NH_3, despite its presence in the data set, leads to less representative abundance posteriors. However, predictions for CO_2 are less affected, possibly because this trace gas has less spectral overlap with NH_3 compared to H_2O or CH_4.

6.2.1.3 Retrieval R_2

The results of the same analysis for the retrieval R_2, which includes CO, HCN, and H_2S as additional molecules to the fit-composition (see Table 6.1) are very similar to those of R_0 (see Section 6.2.1.1). Therefore, we refer the reader to Table 6.6 that summarizes the results for the correlation between predicted probabilities and input abundances, along with the B-S values, and to Figure C.1 in Section C.2 of the Appendix.

6.2.2 Predictor assessment

6.2.2.1 Retrieval R_0

Figure 6.6 shows the analysis performed to assess the predictive power of the P-statistic (ability to maximize TP and TN while minimizing FP and FN) when using the abundance posteriors from the retrieval R_0. The figure reports the results for CH_4, H_2O, and CO_2, shown in different columns from left to right, respectively.

The upper row shows the calculated ROC curves for each molecule. Each curve is reported with a bootstrap confidence interval calculated using 1000 bootstrap samples, with the same random removal and replacement of the data as discussed in Section 6.2.1, involving $1/e \approx 36\%$ of the data. For each molecule, we calculate the AUC using the `roc_auc_score` method of `sklearn.metrics` (Pedregosa et al., 2011), with the associated uncertainty estimated from the same bootstrap samples. The AUC values thus obtained are collected in Table 6.7. For all molecules, the ROC curves are close to ideal behavior (curve near the unit step function, see Section 6.1.5.2), showcasing that the P-statistic has significant predictive power. Consequently, the corresponding AUC values are > 0.9, with no considerable variation between molecules, implying similar predictive power.

For each molecule, the bottom row shows the number of TP, TN, FP, and FN (see Table 6.5), used to construct the ROC, versus the probability threshold $\mathbb{P}$. Also shown are the associated confidence intervals estimated from the same bootstrap samples. These diagrams provide information on how the predictive power of the method changes as $\mathbb{P}$ varies from 1 to 0 and aid in the selection of the optimal classification threshold $\mathbb{P}_*$ (see Section 6.1.6).

Given the randomization of trace gas abundances in the forward model (10^{-7} to 10^{-2} on a uniform logarithmic scale, see Table 6.2), and the selected abundance threshold ($\mathbb{T}_{Ab} = 10^{-5}$), the data set contains $\sim 60\%$ positive observations and $\sim 40\%$ negative observations. By

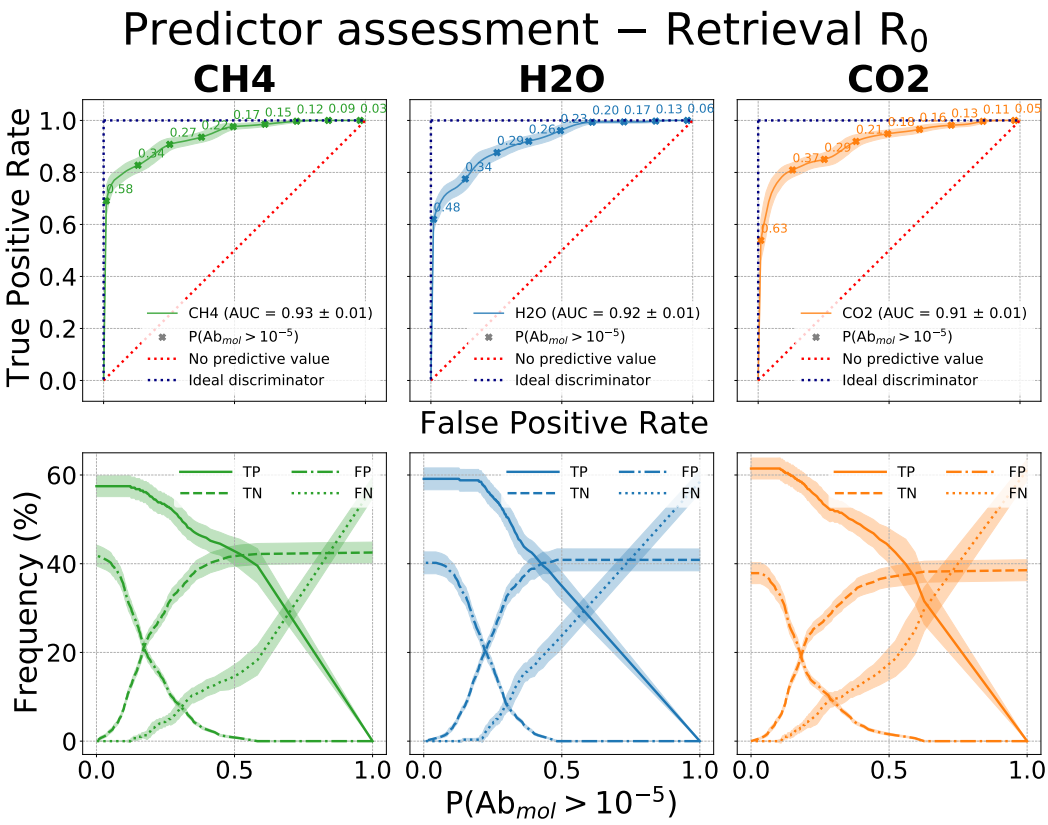


Figure 6.6. Predictor assessment analysis for CH_4, H_2O, and CO_2 from the R_0 retrievals, that implement a model that is fully representative of the simulated atmospheres. Top row: ROC curves with associated bootstrap confidence intervals. The ideal and worst possible classifier ROC curves are reported for reference. The legends report the AUC associated with each ROC curve. Several probability thresholds $\mathbb{P}$ at regularly spaced intervals are also displayed on each curve. Bottom row: TP, TN, FP, and FN curves plotted as a function of the probability threshold $\mathbb{P}$, with confidence intervals from the same bootstrap estimation.

Table 6.7. AUC of the ROC curves and probability odds at the probability threshold $P = 0.5$ for all possible combinations of retrievals and molecules.

Retrieval	molecule	AUC [%]	TP [%]: FP [%]	TN [%]: FN [%]
R_0	CH_4	93 ± 1	43 ± 3 : < 1	42 ± 2 : 15 ± 3
	H_2O	92 ± 1	37 ± 3 : < 1	41 ± 3 : 23 ± 3
	CO_2	91 ± 1	45 ± 4 : 1.7 ± 0.3	37 ± 2 : 17 ± 3
R_1	CH_4	86 ± 2	51 ± 3 : 16 ± 1	27 ± 2 : 7 ± 2
	H_2O	82 ± 2	47 ± 3 : 15 ± 1	26 ± 2 : 13 ± 3
	CO_2	90 ± 1	48 ± 3 : 5.6 ± 0.5	33 ± 2 : 14 ± 2
R_2	CH_4	93 ± 1	41 ± 3 : < 1	42 ± 2 : 17 ± 3
	H_2O	92 ± 1	37 ± 4 : < 1	41 ± 2 : 23 ± 3
	CO_2	91 ± 1	45 ± 3 : 1.7 ± 0.3	37 ± 2 : 17 ± 3

definition, for $\mathbb{P} = 1$, the number of positive forecasts, $N_P = TP + FP$, is zero, and the number of negative forecasts, $N_N = TN + FN$, is equal to the size of the data set. Therefore, at this probability threshold, $TN \simeq 40\%$ and $FN \simeq 60\%$. As $\mathbb{P}$ decreases, N_P increases (TP and FP increase), while N_N decreases (TN and FN decrease). For $\mathbb{P} = 0$, N_N is zero and N_P is equal to the data set size; at this classification threshold, $TP \simeq 60\%$ and $FP \simeq 40\%$.

In those cases where there are no external constraints on which misclassification is more bearable (FP or FN), the intersection of their curves gives an optimized classification threshold $\mathbb{P}_*$.

From this intersection, we obtain $\mathbb{P}_* \approx 0.3$ for all molecules. For confirmation, we can trace this $\mathbb{P}_*$ on the ROC curves. As expected, it roughly corresponds to the point where we cannot significantly increase TPR without increasing FPR, which is at $TPR \approx 0.8$. If, instead, we need a more conservative number of FP, we can choose a higher $\mathbb{P}_*$, for example $\mathbb{P}_* = 0.5$, the default classification threshold for a binary classifier.

A concise way to demonstrate the effectiveness of the *P*-statistic in rejecting misclassifications is by computing the odds TP:FP and TN:FN, estimated from the curves in the bottom row of Figure 6.6. Odds relate to the probability that a molecule is correctly identified at the selected $\mathbb{P}$, with an example shown in Table 6.7, estimated at $\mathbb{P}_* = 0.5$. The table shows that the *P*-statistic is quite effective in rejecting FP, as they are negligible for all molecules at this threshold. Moreover, TPR at $\mathbb{P}_* = 0.5$ indicates that more than 60% of the positives in the dataset is correctly identified, with TP values of approximately 45%, 35%, and 45% for CH_4, H_2O, and CO_2, respectively (rounded to the nearest 5% from the odds values listed in the table). However, at this $\mathbb{P}$, FN increases to approximately 15–25% of the dataset (as seen in the bottom row of Figure 6.6 at $\mathbb{P}_* = 0.5$), resulting in TN:FN odds of less than 3:1.

6.2.2.2 Retrieval R_1

Figure 6.7 shows the same analysis for the retrieval R_1.

Comparing the ROC curves in the top row with those obtained for the retrieval R_0 (see Section 6.2.2.1), we notice a decrease in the predictive power of the method, measured by a reduction in AUC for CH_4 and H_2O, as reported in Table 6.7. On the contrary, the CO_2 ROC achieves the highest AUC, similar to that of R_0, possibly caused by the limited overlap between NH_3 and CO_2, when compared to the case of CH_4 and H_2O.

The plots in the bottom row show a significant reduction in the performance of the FP curve compared to that achieved for R_0: for CH_4 and H_2O, it is above 10% up to $\mathbb{P} \simeq 0.6$, instead of $< 1\%$ at $\mathbb{P} \simeq 0.5$. The TN curve also shows a decrease in performance: it remains below 30% to $\mathbb{P} \simeq 0.6$, instead of reaching 40% at $\mathbb{P} \simeq 0.4$ in R_0. Although the TP and FN curves demonstrate relatively better performance, the optimal classification threshold denoted as $\mathbb{P}_*$, determined at the intersection of the FP and FN curves, increases to approximately $\mathbb{P}_* \sim 0.65, 0.5, 0.4$ for CH_4, H_2O, and CO_2, respectively. Tracing these $\mathbb{P}*$ values on the ROC curves reveals that they correspond to a TPR of approximately 0.8 for all molecules, similar to R_0, but with a significantly worse FPR, as a consequence of the reduced predictive power.

Table 6.7 reflects this, showing the odds of TP:FP and TN:FN at the same probability threshold $\mathbb{P}_* = 0.5$, which was used for R_0. In this case, the method is less efficient in rejecting FP, despite having TP of approximately 50% and 45% for CH_4 and H_2O,

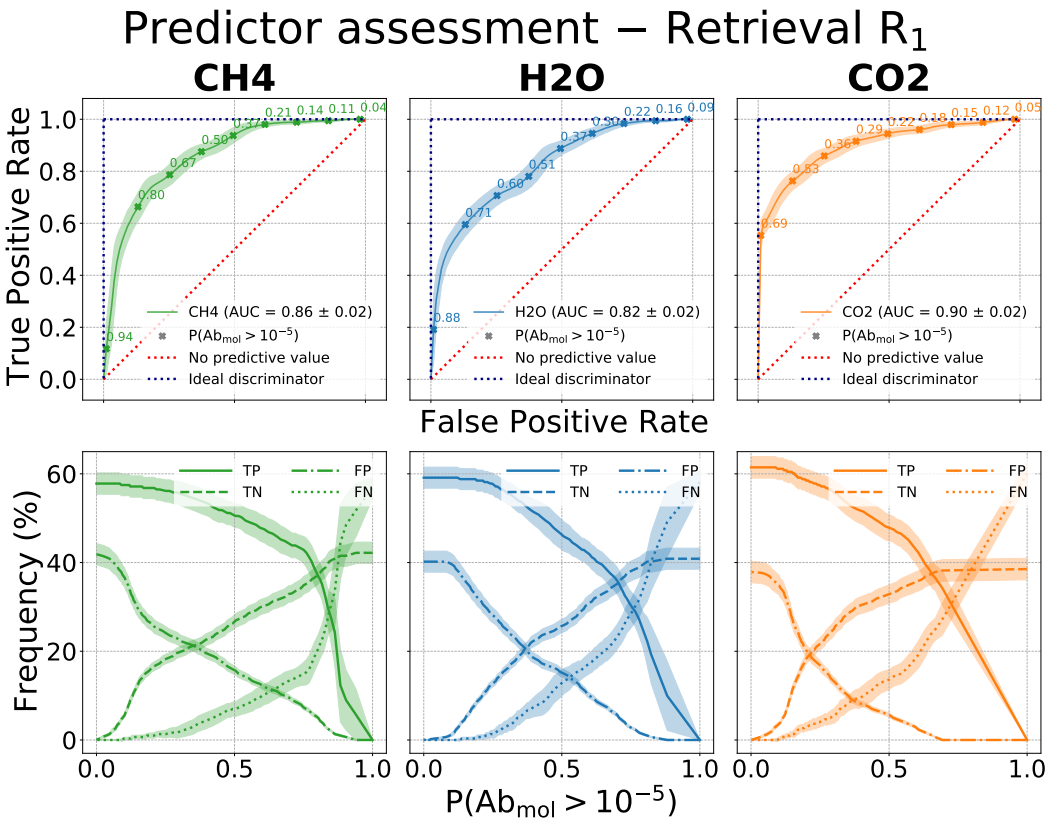


Figure 6.7. Same as Figure 6.6. Predictor assessment for the R_1 retrievals, implementing a model that excludes NH_3 from the fit-composition.

respectively, resulting in only about 3:1 odds for TP:FP. However, the method is still effective in correctly identifying planets with CO_2, with TP:FP odds of about 9:1. As for TN:FN, the results are similar to R_0, with a slightly better rejection of FN in the case of CH_4 (4:1 instead of 3:1).

6.2.2.3 Retrieval R_2

The results from the same analysis for the retrieval R_2 are very similar to R_0's (see Section 6.2.2.1). Therefore, we refer the reader to Table 6.7 that summarizes the AUC values obtained and the odds TP:FP and TN:FN at the probability threshold $\mathbb{P}_* = 0.5$, and to Figure C.2 in Section C.2 of the Appendix.

6.2.3 Abundance estimates

Tier 1 might not be adequate for reliable abundance retrieval, for which higher Ariel Tiers are better suited. Therefore, we study the retrieved Tier 1 abundances to investigate trends in their distribution that may clarify some of the behavior observed in the calibration and ROC curves seen in the previous sections. The abundance estimator used is obtained from the median of the marginalized posterior distribution of the $\log Ab_{mol}$ with asymmetric error bars estimated from the 68.3% confidence level around the median. In particular, we are interested in investigating the regime of input abundances under which this median-based estimator is unbiased.

6.2.3.1 Retrieval R_0

Figure 6.8 reports the analysis performed to investigate potential biases affecting the median of the marginalized posteriors when used as an estimator of the log-abundances. The figure reports the results for CH_4, H_2O, and CO_2, shown in different columns from left to right, respectively. NH_3 exhibits similar behavior to the other three molecules, but it is not included in the figure in line with the decision to treat it as a nuisance in this study.

Panels in the top row show the molecular log-abundance input vs. the retrieved with the error bar. A solid black line serves as the ideal trend (1:1 line) for visual reference. The color bar indicates the distances between the input and retrieved log-abundance, expressed in units of the uncertainty σ on $\log Ab_{mol}$, estimated by averaging the asymmetric error bars. Blue colors denote distances up to 1σ; red colors represent distances in the range of $1 \rightarrow 2\sigma$. Larger distances are marked with black circles, which serve to diagnose potential trends and biases that may affect the retrieval results. In addition, the symbol size reflects the Signal-to-Noise Ratio (SNR) of each observation as estimated in the AIRS-CH0 spectroscopic channel, providing insight into possible trends between the distance to the input abundance and the SNR condition.

The retrieved abundances exhibit good agreement with the input abundances in the large abundance regime, characterized by limited scatter around the ideal trend and by low retrieved uncertainties. This regime is generally observed for $Ab_{mol} \gtrsim 10^{-4}$, but starts to break down at $10^{-5} \lesssim Ab_{mol} \lesssim 10^{-4}$. For $Ab_{mol} \lesssim 10^{-5}$, the input abundances are rarely retrieved accurately. This analysis can provide insights into the detection limits of CH_4, H_2O, and CO_2 in Ariel Tier 1, which are estimated to be around 10^{-4}. These values can be compared with the expected detection limits of the same molecules in Ariel Tier 2, which are

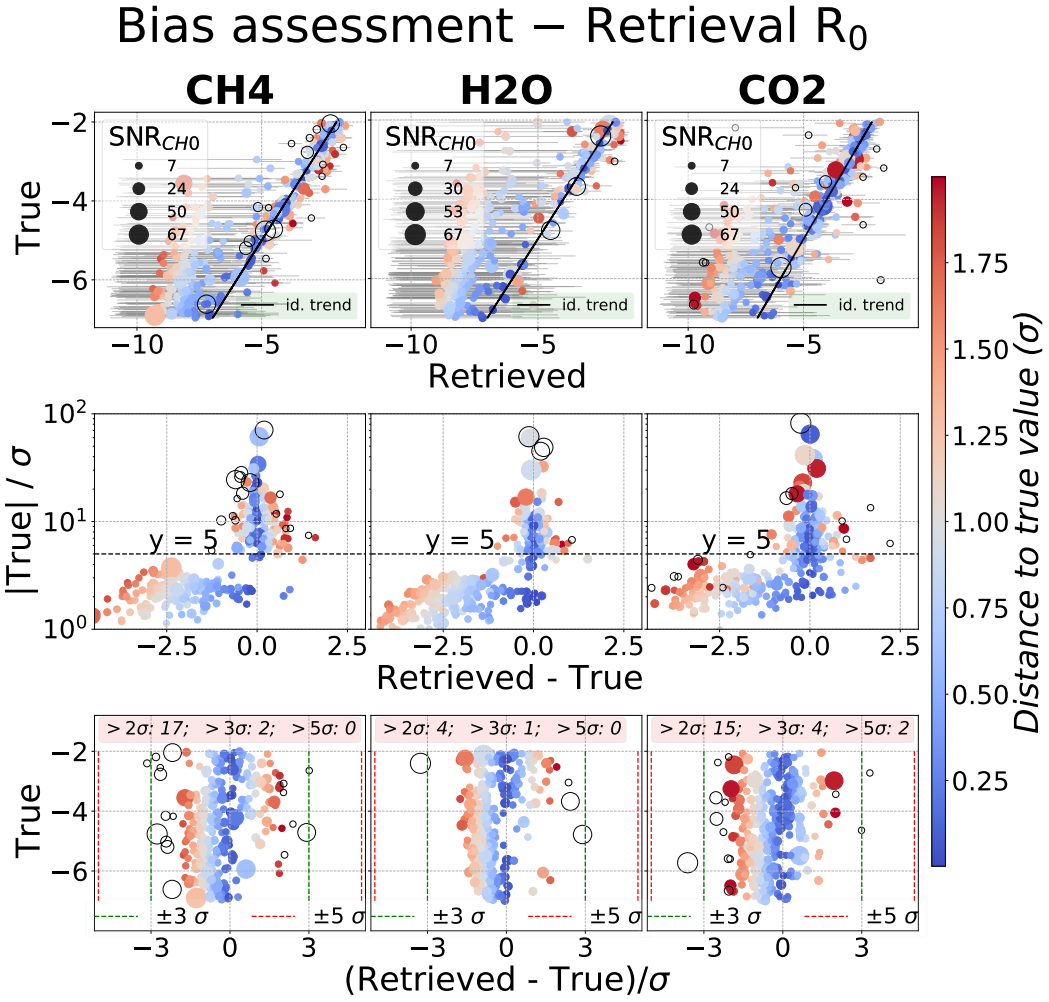


Figure 6.8. Comparison between the retrieved molecular abundances and their true values is shown from the R_0 retrievals. The estimator for the retrieved log-abundances is the median of the posterior distributions from the retrievals. Top row: retrieved vs. input molecular abundances. The solid black line represents the ideal trend, and the color bar visualizes the distance between input and retrieved abundances in units of uncertainty σ. The symbol size is proportional to the SNR in the AIRS-CH0 spectroscopic channel. Middle row: log-abundance SNR vs. the difference between the retrieved and input log-abundances. A black dashed line is drawn at a value of 5 on the vertical axis for visual reference. Bottom row: true abundances vs. the difference between the retrieved and true log-abundances, in units of σ. Dashed vertical lines are drawn at 3 and 5-σ. Text boxes show the number of 2-, 3-, and 5-σ outliers.

anticipated to be significantly lower, with previous studies (Changeat et al., 2020a) reporting limits between 10^{-7} and $10^{-6.5}$.

Let the log-abundance SNR be defined as $\frac{1}{\sigma} \mid \log Ab_{mol} \mid$, where Ab_{mol} is the true value of the molecular abundance. The middle row panels in Figure 6.8 show the plot of log-abundance SNR vs. the difference between the retrieved and input log abundances. It can be observed that the distribution of data points is broadly separated into two sub-populations at a SNR of about 5. Data points with high SNR correspond to cases where the input is confidently retrieved and aligned along the 1:1 line in the upper row diagrams, indicating unbiased estimation. On the other hand, data points with low SNR cluster in the bottom left portion of the diagram. In these cases, the median is no longer an unbiased estimator of the true value, as the corresponding data points lie to the left of the 1:1 line in the upper row diagrams. As discussed further in Section 6.3.2, these cases have posteriors dominated by the prior imposed in the retrieval and are best treated as upper limits.

In the bottom row of Figure 6.8, the true abundances are shown vs. the difference between the retrieved and true abundances, in units of σ. The diagrams provide a visualization of how many samples are 2-, 3-, and 5-σ outliers, allowing verification that the distribution is compatible with the tail of the abundance posteriors. The number of outliers is shown in the text box inserted in the diagrams and (converted into percentages) in Table 6.8. Assuming that the abundance posteriors are representative of the data, the fraction of expected outliers outside is 5%, 0.3%, and $\ll 1\%$, respectively at 2-, 3-, and 5-σ. We find good agreement between the percentages reported in Table 6.8 and these values, with minor deviations compatible with the statistical fluctuations of a random variable.

Table 6.8. Percentage of data points counted outside three confidence intervals for all possible combinations of retrievals and molecules.

Retrieval	molecule	$> 2\sigma$ [%]	$> 3\sigma$ [%]	$> 5\sigma$ [%]
R_0	CH_4	5.6	0.7	$\ll 1$
	H_2O	1.3	0.3	$\ll 1$
	CO_2	5.0	1.3	0.7
R_1	CH_4	32.9	19.6	11.6
	H_2O	17.9	13.6	9.6
	CO_2	16.6	10.3	6.6
R_2	CH_4	6.0	0.7	$\ll 1$
	H_2O	1.3	0.3	$\ll 1$
	CO_2	5.3	1.7	1.3

6.2.3.2 Retrieval R_1

Figure 6.9 shows the same analysis for the retrieval R_1. The top row shows that, although there is still a correlation between the retrieved and input abundances, it is less significant than for R_0. Furthermore, comparing the retrieved and input abundances yields different regimes for each molecule. However, the main difference from R_0 is the significant number of data points at distances greater than 2σ (marked by black circles), corresponding to 2-σ outliers. In particular, for all molecules, most of these points are located to the right of the

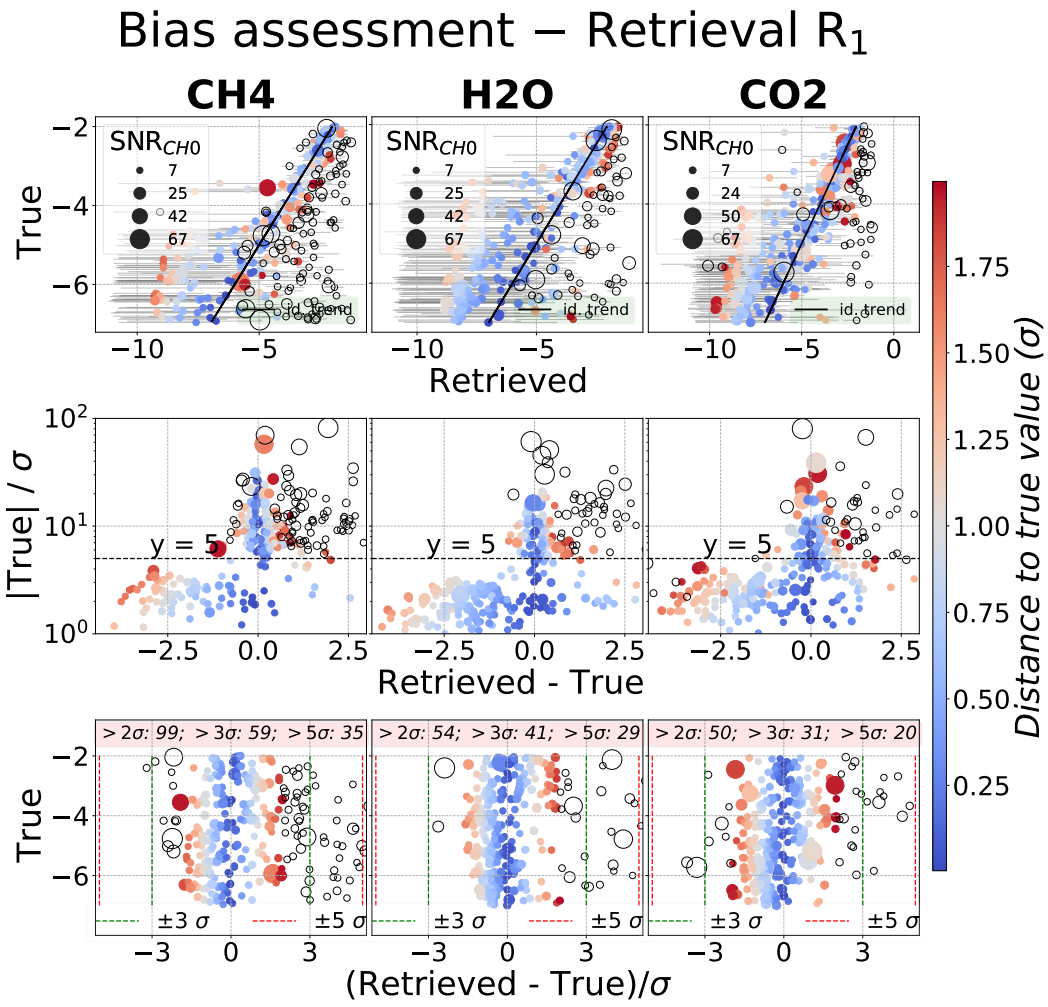


Figure 6.9. Same as Figure 6.8 for the R_1 retrievals, implementing a model that excludes NH_3 from the fit-composition.

ideal trend, indicating the presence of an overestimation bias for the retrieved abundances. These data points are located in the region $y \gtrsim 5$ and $x > 0$ in the plots in the middle row. Therefore, in addition to the overestimation bias for the abundances, their retrieved uncertainties are underestimated. Furthermore, the bottom-row diagrams show a larger number of outliers compared to the R_0 case: too many for the posterior to be considered representative. This is a consequence of an atmospheric model which is not representative of the data, biasing the likelihood, the abundance posteriors, and the median estimator of the abundances.

6.2.3.3 Retrieval R_2

The results of the same analysis for the retrieval R_2 are very similar to those of R_0, including the number of outliers that are compatible with the expectations for a model that is representative of the data. Therefore, we refer the reader to Table 6.8, and to Figure C.3 in Section C.2 of the Appendix. Here, we only stress that adding molecules to the fit-composition that are not present in the data set does not appear to significantly bias the abundance posteriors, compared to R_0. This is further discussed in Section 6.3.2.

6.3 Discussion

In this section, we first discuss the similarities between the results from the retrievals R_0 and R_2, shown in Sections 6.2.1 and 6.2.2. Then we apply the ADI metric to compare all retrievals from the point of view of the Bayesian evidence (Section 6.3.1). Finally, we expand the discussion to the role of the priors in the retrieved abundance posteriors (Section 6.3.2).

The results of Sections 6.2.1 and 6.2.2 show that the predictions of the *P*-statistic for the retrievals R_0 and R_2 are comparable, despite the quite different fit-compositions, while the reliability of the *P*-statistic is lower in the R_1 case. The R_0 model and its parameters are identical to those used to generate the POP-Is population, and the R_2 extends the parameter space with new molecules. In R_2, the abundance posteriors for CH_4, H_2O, and CO_2 do not appear to be significantly affected by the addition of CO, HCN, and H_2S in R_2, despite that the latter three spectral signatures partially overlap with those of CH_4, H_2O, and CO_2 (Encrenaz et al., 2015). It should be noted that the absence of the three molecules from the simulated atmospheres is correctly revealed in R_2 by their low *P*-statistic, shown in Figure 6.10, that take values smaller than 40% for CO, HCN, and H_2S, respectively. The extension of the analysis to include the calibration and ROC curves to these molecules is left to future work.

The analysis, therefore, suggests that the *P*-statistic is robust (that means, provides reliable results) against retrieval models that are over-representative of the observed atmosphere. However, the *P*-statistic can no longer be considered robust when the retrieval models are under-representative of the observed atmosphere.

In the current study, the threshold abundance used to estimate the *P*-statistic remains constant for all molecules. While it is possible to optimize this threshold for individual molecules, we leave this aspect for future research as discussed in Section 6.1.4. Lowering the threshold reduces the information provided by the ROC curves. To achieve the optimal point of operation, one must balance the True and False Positive Rates, which is necessary to promote a Tier-1 target to higher Tiers. It is important to note that ROC curves calculated

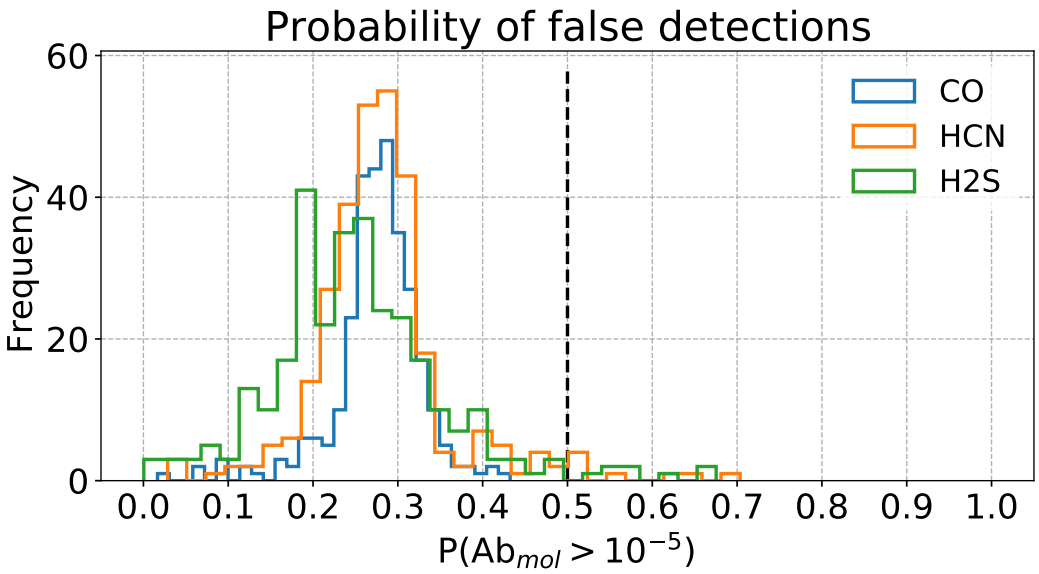


Figure 6.10. Histogram of the frequency of use of each possible P forecast for CO, HCN, and H_2S, using the abundance posteriors from the retrieval R_2. The dotted vertical line marks the default binary classification threshold $P = 0.5$ for reference.

at different threshold levels provide a statistical estimation of the sample's completeness, enabling the inference of population-wide properties such as the fraction of planets containing certain molecules. While this aspect requires further investigation in future research, it should be noted that the fraction of positive, Σ (planets with true abundance in excess of $\mathbb{T}_{Ab}$) is related to the fraction of Tier-1 targets, $\tilde{\Sigma}$, selected with $P(> \mathbb{T}_{Ab}) > \mathbb{P}$ by

$$\Sigma = \frac{\tilde{\Sigma} - FPR}{TPR - FPR}.$$

The similarities between the R_0 and R_2 models are further discussed in the next section.

6.3.1 ADI comparison

The ADI metric, described in Section 6.1.3, is used to assess the statistical significance of a model atmosphere with respect to a featureless spectrum using the log-Bayesian factor. A large ADI suggests that a featureless spectrum is less favored by the data. From the ADI definition, the log-Bayesian factor of two competing models is the difference between their respective ADI. The ADI is used in this section only to compare the Bayesian evidence of the retrievals, and not for anything else in this study.

Figure 6.11 shows the ADI differences between the R_0 model and the two competing models, R_1 and R_2, plotted against NH_3 abundances. A large, positive difference indicates that the competing models are less representative of the data compared to R_0. The median ADI values for all retrievals are approximately 91, 86, and 92 for R_0, R_1, and R_2, respectively, as shown in the text box within Figure 6.11. This suggests that a featureless atmospheric model is not favored by the data, and R_1 is the least representative, as expected. This is further supported by the fact that the ADI difference between R_0 and R_1 increases with increasing NH_3 abundance, indicating that higher NH_3 abundances make R_1 less

representative compared to R_0, in agreement with the analysis of Section 6.2. In contrast, the ADI difference between R_0 and R_2 is close to zero, with a scatter described by a standard deviation of approximately 0.5, which is independent of NH_3 abundance. This confirms that R_2 is similarly representative of the data compared to R_0, despite describing a wider parameter space.

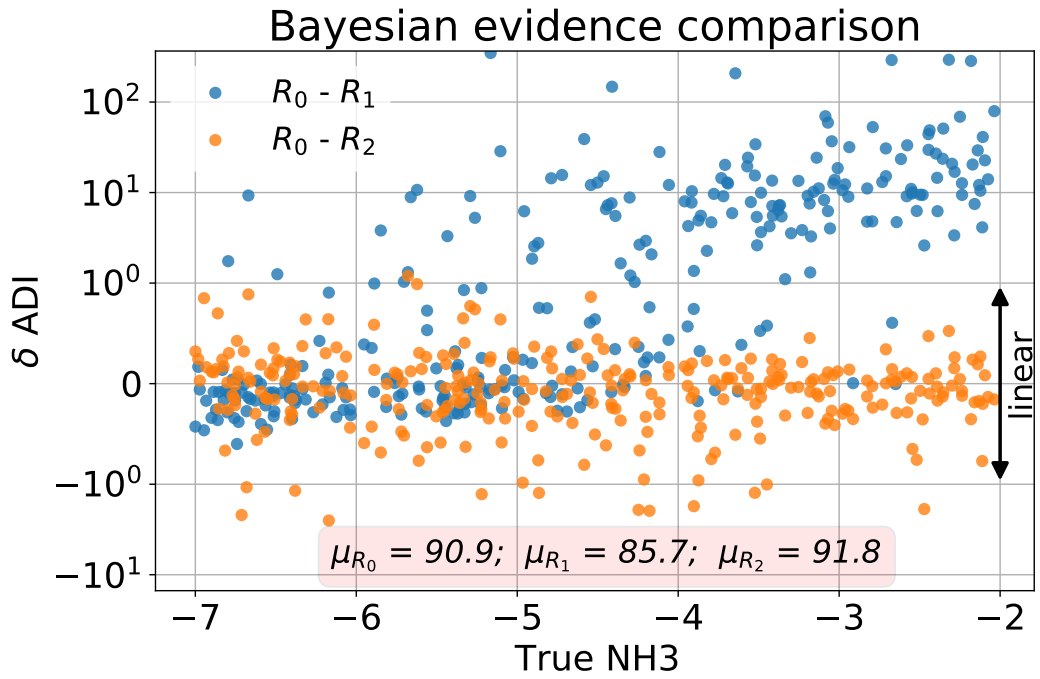


Figure 6.11. Bayesian evidence comparison of the retrievals R_0, R_1, and R_2, measured in ADI. The horizontal axis plots the input abundances of NH_3; the vertical axis reports the ADI difference between R_0 and the other two retrievals, R_1 and R_2. The y-axis uses a `matplotlib` "symlog" scale with the linear threshold set at 1 for better visualization. The text box on the bottom shows the median ADI reported by each retrieval.

6.3.2 Priors

In this section, we discuss the impact of the log-uniform priors adopted in the analysis on the results presented. The consequence is a non-Gaussian posterior distribution, and the mean, mode, and median are not equivalent moments of the distribution. In particular, the median is not an unbiased estimator of the true abundance as shown in Figure 6.8 for low log-abundance SNR (hereafter, "abundance SNR"). This can be explained in terms of the Bayesian formulation of the posterior, $\mathcal{P}$, which is proportional to the product of the likelihood, $\mathcal{L}$, and the prior, Π.

$$\mathcal{P} \propto \mathcal{L} \times \Pi \tag{6.4}$$

Because $\Pi(\log x)$ is uniform, $\Pi(x) \sim 1/x$, for large abundance SNR, the likelihood dominates, the posterior is Gaussian (because of the central limit theorem), and the median estimator is unbiased. For low abundances, the prior dominates, $\mathcal{P}(x) \propto 1/x$, and the median is an estimator of the molecular abundance that is biased towards low abundances. This is shown in Figure 6.12. Each panel shows the Probability Density Function (pdf) of

the likelihood, prior and posterior normalized to 1 at the peak, for three cases where the abundance SNR is 4.0, 5.5, and 7.0, respectively, from the top to the bottom panel, assuming an input abundance of 10^{-5}. The posterior is likelihood-dominated when the abundance SNR is 7 and is prior-dominated when the abundance SNR is 4.

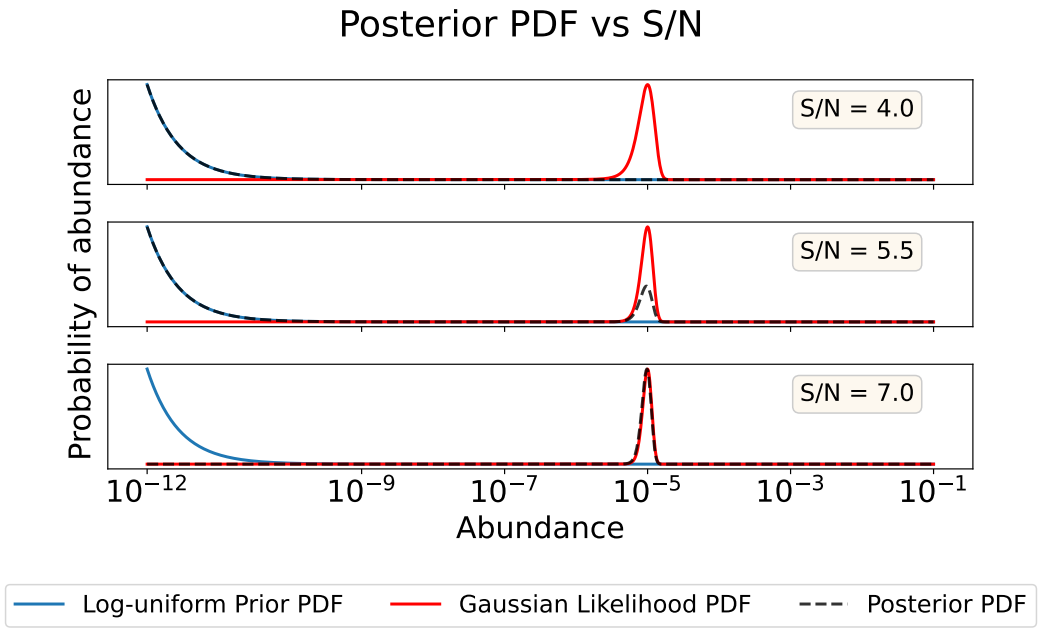


Figure 6.12. The pdfs of the likelihood, prior and posterior are shown by the red, blue, and black lines, respectively. The pdfs are normalized to 1 at their peak. The assumed abundance SNR is 4.0, 5.5, and 7.0, respectively, from the top to the bottom panel. An input abundance of 10^{-5} is assumed.

Although logarithmic uniform priors are often assumed in spectral retrieval studies, they are certainly not "uninformative priors" (Trotta, 2008; Oreshenko et al., 2017). Clearly, using these priors biases the median estimator of the molecular abundance in the low SNR regime, explaining the trends seen in Figure 6.8. As a side note, log-priors on molecular abundances could as well introduce biases on the derived elemental abundances, therefore the issue has to be investigated carefully in future studies.

The low abundance SNR targets are those that contribute to the leftmost peak in the bimodal distribution of the *P*-statistic (Figure 6.4). Further investigation is however needed to fully understand the origin of the *P*-statistic bimodality and its under-forecasting properties.

Chapter 7

ARES VI: Interpreting 3D Biases in Transmission Spectra

This chapter follows in the steps of the work done to address questions on the interpretability of transmission spectra given the 3D nature of exoplanetary atmospheres. Biases can occur when 1D retrieval models are used to interpret high-quality observations from JWST and Ariel, which can be affected by 3D thermal and chemical effects. While previous works have begun to investigate and address this type of bias, here we focus on untangling these effects via examining biases in the retrieval results when using simulated spectra. This study was initiated at the second edition of the Ariel Retrieval of Exoplanets School (ARES)[1], held in 2021 in Biarritz. The proper investigation was carried out in the subsequent years and involved a tight collaboration between diverse expertise in the field: Global Circulation Models (GCMs), atmospheric chemistry, and simulators to estimate the expected experimental uncertainties with JWST and Ariel, to name a few.

* * *

A major scientific question in the exoplanet field is: "How does the 3D atmospheric structure affect the transmission spectra of exoplanets, from a cold planet to an ultra-hot Jupiter?". Recent works (Pluriel et al., 2020b; Lacy and Burrows, 2020; Wardenier et al., 2022; Pluriel et al., 2022) concluded that, for ultra-hot Jupiters, the 3D structure plays a major role in shaping the transmission spectra. If the temperature of the atmosphere is not high enough to dissociate a molecule on both the day-side and the night-side, the amplitude of its spectral features will be larger than predicted by a 1D plane parallel approach. In addition, Falco et al. (2022) showed how the changes in planet orientation during the transit allow us to probe the horizontal variations in the atmosphere. Pluriel et al. (2022) investigated 3D effects in the transmission spectra of hot Jupiters, dividing them into three main groups: vertical effects, horizontal effects along the limb, and horizontal effects through the limb.

Another key open question is: "Can 1D retrievals find consistent parameters (T-P profile, abundances, C/O ratio, metallicity, and clouds)?". This question is closely related to the first one: if the 3D atmospheric structure strongly affects transmission spectra – can 1D retrieval models find correct atmospheric parameters? In this regard, MacDonald

[1] ARES II: Ariel School, Biarritz, 2021.

et al. (2020) investigated the following conundrum: "Why are most inferred temperatures from transmission spectra far colder than expected from the equilibrium temperature?". They concluded that a 1D model can fit the transmission spectra of planets with asymmetric terminators, but retrieved atmospheric parameters may not represent true terminator averages. Also, the retrieved temperatures of planetary terminators may be biased by hundred degrees below their real value, and this bias is most extreme in the case of ultra-hot Jupiters. This study also mentions the biases in chemical abundances derived from 1D retrievals. Pluriel et al. (2020b) concluded the same – if the temperature and the chemical composition vary across the limb, which is the case for 3D structures, 1D retrievals cannot find the correct molecular abundances. This also affects the inferred C/O ratio, which is an indirect estimate based on the abundances of all C- and O-bearing molecules. Recently Zingales et al. (2022) demonstrated that the choice of the retrieval model is critical for correctly retrieving the thermal structure of the atmosphere. Lastly, Pluriel et al. (2022) provided a "cheat sheet" of the minimum model assumptions needed to avoid biases in interpreting atmospheric properties. When optical absorbers are present, 1D models are adequate to describe transmission spectra for atmospheric equilibrium temperatures lower than 1400 K; in their absence, the 1D assumption can extend up to 2000 K. Above these temperatures, retrievals with 1D models return biased estimates of the parameters in the forward model.

To address the above-mentioned questions and related issues, we need two distinct branches of action: one that can produce simulated transmission spectra for different planets, accounting for 3D structures, and one that can perform atmospheric retrievals and compare the inferred parameters with the forward models. Our goal is to assess the extent to which our retrievals can reconstruct the true atmospheric composition. This end-to-end system allows us to investigate the interpretation of our atmospheric retrievals consistently. The first branch starts from precomputed GCMs for three planets (see Section 7.1.1):

- GJ1214 b, a Neptune-like planet orbiting an M-type star;
- HD189733 b, a hot Jupiter around a K-type star;
- WASP-121 b, an ultra-hot Jupiter orbiting an F-type star.

Then, it uses the *Pytmosph3R* code to produce 3D transmission spectra (see Section 7.1.2). In this step, we use two different configurations for each planet: an equilibrium chemistry model and a model with chemical profiles constant with altitude. The final products of the first branch are spectra with attached errorbars, to reproduce spectra "as observed" by JWST and Ariel. The expected errorbars are estimated using PandExo and ArielRad, two simulators of the noise performance of JWST and Ariel, respectively (see Section 7.1.3). The second branch starts from these simulated observations and performs a Bayesian retrieval to estimate the best-fitting parameters of the model. As a retrieval tool, we use the retrieval framework Tau Retrieval for Exoplanets (TauREx 3), briefly described in Section 7.2.1. Section 7.2.3 details the retrieval procedure, the chemical configurations, and the other atmospheric parameters. The retrieval procedure is the same for all spectra – the same set of atmospheric models is applied, to remove our a priori knowledge and compare the obtained results correctly. We discuss the results from this comparison and their implications for interpreting transmission spectra with the 1D assumption in Section 7.3.

7.1 Transmission spectra simulations

The following three planets (GJ1214 b, HD189733 b, WASP-121 b) have been chosen from warm Neptune to ultra-hot Jupiter in order to study the transmission spectra and retrieval biases depending on the temperature of the planets. Following Al-Refaie et al. (2022a) study, we go a step further by focusing on biases arising from 1D vertical thermal variation as well as full 3D thermal structure. We simulate JWST (NIRSpec + MIRI) and Ariel observations.

7.1.1 Global Climate Models

GJ1214 b has been simulated using the generic Planetary Climate Model. This model has been specifically developed for exoplanets and paleoclimate studies (Charnay et al., 2015; Leconte, J. et al., 2013). The dynamical core solves the primitive hydrostatic equations of meteorology on an Arakawa C grid, using a finite difference scheme. Radiative transfer is solved using the correlated-k model. Radiative effects of H_2, He, H_2O, CH_4, NH_3, CO, and CO_2 are taken into account, assuming a 100x solar metallicity. The horizontal resolution is 64×48 and we use 45 vertical layers between 80 bar and 3 Pa, equally spaced in log pressure. The star is taken as a blackbody at 3026 K, and we assume an internal temperature of 60 K. The dynamical time-step is 60s and the physical/radiative time-step is 300s. The model was integrated for 1600 days. For a more complete description of the model and the simulation, we refer the reader to Charnay et al. (2015).

For HD189733 b, we make use of the Met Office Unified Model (Drummond et al., 2018). The model solves the deep atmosphere, non-hydrostatic Navier–Stokes equations on an Arakawa C grid. Radiative transfer is handled through the SOCRATES[2] code adapted for hot Jupiters (Amundsen et al., 2016). A chemical relaxation scheme is used and the radiative transfer is computed in 32 wavelength bins. The simulation is integrated for 1000 Earth days. For a more complete description of the model, see Drummond et al. (2018).

WASP-121 b has been simulated using the SPARC/MIT global circulation model (Showman et al., 2009). The model solves the same primitive equations of meteorology as the generic Planetary Climate Model on a cubic-sphere grid. It has been widely used for various hot Jupiters (Showman et al., 2009; Kataria et al., 2015; Parmentier et al., 2016, 2018, 2021) and has also been applied to ultra-hot Jupiters (Kreidberg, 2018; Arcangeli et al., 2019). For this study, we use the model published in Parmentier et al. (2018). The horizontal resolution is $C32$, equivalent to 128 cells in longitude and 64 in latitude and 53 vertical levels with pressure ranging from 200 bar to 2 µbar. Radiative transfer is handled using the two-stream approximation with 11 wavelength bins, as done in Kataria et al. (2013). The model assumes chemical equilibrium, taking into account the thermal dissociation of water and hydrogen. However, H_2 recombination is neglected despite its non-negligible impact on the thermal and dynamical structure (Tan and Komacek, 2019). For a more complete description of the model and the simulation, we refer the reader to Parmentier et al. (2018).

For each planet, we build a pseudo-1D version of the 3D model. The temperature profiles of the whole grid are replaced by the same 1D profile. This profile is the average temperature profile over all latitudes of the eastern terminator for each model. Thus, the pseudo-1D models are representative of each temperature condition from warm Neptune to ultra-hot Jupiter. The purpose of these pseudo-1D models is to control the correct behavior

[2] https://code.metoffice.gov.uk/trac/socrates

of the retrieval code. If we consider the 1D retrieval code, it should correctly retrieve the pseudo-1D models. Thus, we can confidently untangle the 3D effects.

7.1.2 Pytmosph3R

Based on the planetary atmospheres described in Section 7.1.1, we used the latest version of *Pytmosph3R* (Falco et al., 2022) to generate the transmission spectra. It takes into account the 3D structure of the atmosphere and uses the monochromatic cross sections calculated by ExoMol (Yurchenko et al., 2011; Tennyson and Yurchenko, 2012; Barton et al., 2013; Yurchenko et al., 2014; Barton et al., 2014). Species abundances were established in two different ways: (i) a constant chemistry model, e.g., where abundances are independent of temperature and pressure, thus constant everywhere, (ii) an equilibrium chemical model.

The abundances of the constant chemistry model have been chosen close to the equilibrium ratio for a given temperature representative of each planet. We have taken into account only the main species: H_2O, CO, CH_4, CO_2, HCN, C_2H_2, NH_3, C_2H_4, and in addition TiO, VO, K, Na, SiO, FeH for HD189733 b and WASP-121 b. The values are given in Table D.2, D.3, and D.4. This theoretical construction, not representative of the real atmosphere, removes one degree of freedom (chemistry) to check how 1D retrieval models handle 3D atmospheric thermal structures and thus what can be the biases.

Equilibrium chemistry leads to variable chemical profiles. In the modeled atmospheres, equilibrium chemistry can be expected in the hottest and densest regions. However, we know that non-equilibrium chemistry must be accounted for, especially in the upper atmosphere (Cooper and Showman, 2006; Moses et al., 2011, 2013; Venot et al., 2012, 2020b; Molaverdikhani et al., 2019; Tsai et al., 2021, 2022). This has been very recently implemented in the retrieval code (FRECKLL code developed by Al-Refaie et al. (2022b)) but this is computationally expensive. We used ACE chemistry (see Section 7.2.2 for details) to model the input equilibrium chemistry of GJ1214 b and the chemical abundances used in Parmentier et al. (2016) to model the input equilibrium chemistry of HD189733 b and WASP-121 b. This theoretical construct, relative to the constant case, focuses on the research biases that can arise from the chemical retrieval models.

7.1.3 Uncertainty model

7.1.3.1 PandExo - JWST

To simulate JWST observables (spectra Figure 7.1), we utilize the *PandExo* package (Batalha et al., 2017). This program is a noise simulator designed for JWST transiting observations of exoplanets. We make use of the model to simulate one transit of each planet using NIRSpec-PRISM and one transit using MIRI-LRS. For each planet, we consider a saturation limit of 80% of the full well and a fraction of time out-of-transit to in-transit of 1. For each instrument and planet, we use the *optimize* option for the number of groups per integration, which automatically defines the best settings to carry the observations. The planet and star-specific parameters used for producing the observables are summarized in Table D.1. *PandExo* neither includes the varying stellar noise nor takes into account the transit ingress and egress. However, we estimate that the produced observables are a good enough approximation to what one will observe with the JWST facility and should not change the conclusions of this work. It is worth noting that the three systems studied here have a J-band stellar magnitude below 11.4, which is the estimated saturation limit of NIRSpec-PRISM. Thus,

these systems are not observable with this instrument configuration. Yet, we still performed the computations and studied these systems, as the goal of this analysis is not to prepare JWST observations *per se*, but to highlight possible biases introduced by retrievals on JWST-like data sets.

7.1.3.2 ArielRad - Ariel

This section partially duplicates Section 1.3.1, which is more exhaustive.

7.1.3.2.1 The Ariel Radiometric model ArielRad (Mugnai et al., 2020) is the radiometric simulator of the Ariel payload, developed and maintained by the Ariel Consortium. We will briefly describe ArielRad here; for a technical description, including the detailed noise model, the reader is encouraged to read the original article. Given the description of the payload and a list of candidate exoplanets, ArielRad outputs the expected experimental uncertainty on their measured atmospheric transmission or emission spectra.

The simulation propagates the stellar light through the payload, accounting for each transmission or dispersion by interposing optical components until reaching the focal planes. Then, ArielRad evaluates the noise contributions (with margins) from stationary processes, i.e., stellar photon noise, detector noise, dark current, zodiacal background, and instrument emission. Jitter noise is computed externally by ExoSim (Sarkar et al., 2021), the end-to-end time-domain simulator of Ariel observations, and included in the final noise budget. Then, ArielRad returns the uncertainty estimates on a single transit or eclipse observation. Ariel defines an observation to last 2.5 times the time between the first and the last contact between the planetary and the stellar disks to obtain a sufficiently long baseline integration for the light curve fit and the transit depth estimation (Mugnai et al., 2020). Because the astronomical measurement is the contrast ratio with the signal from the stellar host, ArielRad uses the contrast ratio to compute the observed spectrum's expected Signal-to-Noise Ratio (SNR).

The Ariel mission adopts a four-tier observation strategy in which, after each observation, the resulting spectrum from each spectrometer is binned in data analysis according to specific requirements to optimize the SNR and the mission scientific return (Edwards et al., 2019). ArielRad, knowing the binning and spectral resolution implemented in the different tiers, computes the SNR in the spectral bin according to the tier of interest. Then, it estimates the number of observations required for each planet to reach the tier's required SNR. From the number of observations, ArielRad obtains the final noise estimate for a planet by rescaling the noise on a single observation. These uncertainties can be attached to simulated forward models of transmission or emission spectra, binned down to the tier spectral resolution to obtain the simulated observed spectra.

7.1.3.2.2 Computing the errorbars We utilize the general procedure described above to calculate the Ariel observation uncertainties for the three planetary targets of interest. We use an updated version of the code, which is a wrapper of ExoRad 2.0, the instrument-independent version of the radiometric simulator publicly available on GitHub[3],

[3] https://github.com/ExObsSim/ExoRad2-public

and ArielRad-Payloads, the repository of configuration files for the payload maintained by the Ariel Consortium. For reproducibility, we report the code versions in Table 7.1.

Table 7.1. Versions of the codes used to generate the Ariel spectra.

Code	Version
ArielRad	2.4.25
ExoRad	2.1.111
ArielRad-Payloads	0.0.16

For each planet, we assume a mean molecular weight of 2.3 a.m.u. to simulate H_2-He dominated atmospheres. We use this parameter to calculate the atmospheric scale height and, consequently, the contrast ratio of the transit. We utilize the Ariel strategy for collecting data during an observation; therefore, the ratio of observing time in and out of transit is 1/1.5. Then, for each planet, we estimate the noise and the SNR for a transit observation in Tier 3. We find that 1, 3, and 4 observations are needed to achieve the Ariel Tier 3 required SNR for HD189733 b, GJ1214 b, and WASP-121 b, respectively. Then, we rescale the noise by the square root of the corresponding number of observations, assuming each observation has a Gaussian noise distribution. Finally, we attach the rescaled noise estimates to the respective transmission spectra, binned down at the Tier 3 wavelength grid.

Alternatively, given the noise for a single transit observation and the atmospheric spectrum, we could calculate a more realistic SNR that does not rely on the assumed atmospheric scale height (Mugnai et al., 2020). However, the SNR would depend on the assumed atmospheric spectrum, changing the number of observations on a single target. The resulting analysis would be more akin to an observability study which, again, is not the scope of this chapter.

All the spectra calculated with the methodology described in this section are shown in Figure 7.1. This includes 6 input configurations for each planet (listed Table 7.2).

Table 7.2. Spectra input configurations.

Instrument	JWST				Ariel	
Dimension	1D		3D			
Chemistry	constant	equilibrium	constant	equilibrium	constant	equilibrium

7.2 Transmission spectra retrievals

7.2.1 TauREx

This section partially duplicates Section 1.3.2, which is more exhaustive.

As a retrieval tool, we used Tau Retrieval for Exoplanets (TauREx 3) [4] a fully Bayesian

[4] https://github.com/ucl-exoplanets/TauREx3_public

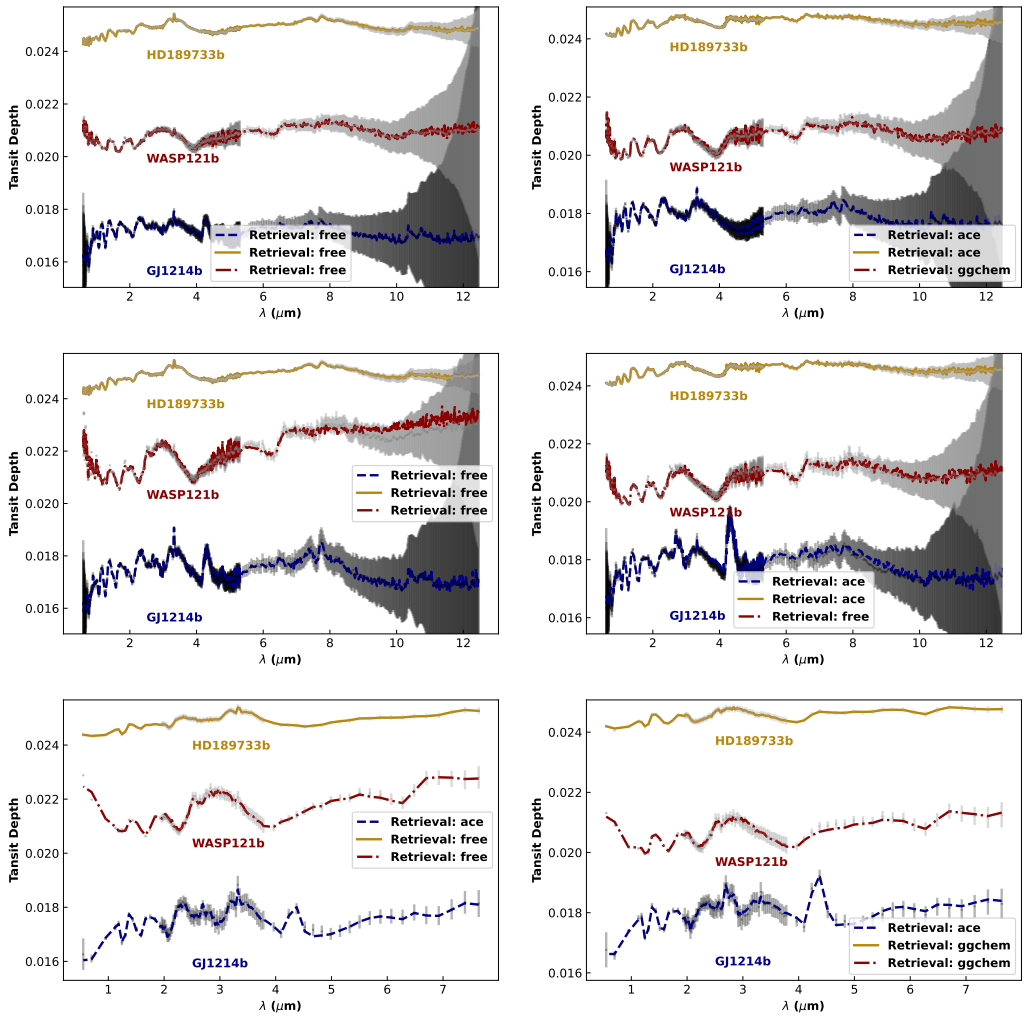


Figure 7.1. Simulation of observations for HD189733 b, WASP-121 b, and GJ1214 b (gray) with the corresponding best retrieval configuration in solid yellow, red dashed, and dash-dot blue lines respectively. **Left panels:** constant input chemistry. **Right:** equilibrium input chemistry. **Top and middle:** JWST simulated observation for atmospheric with 1D and 3D assumption respectively. **Bottom:** Ariel simulated observation for a 3D atmospheric assumption.

inverse atmospheric retrieval framework (Al-Refaie et al., 2021). TauREx 3 consists of two main frameworks: the Forward Model framework and the Retrieval framework. The goal of the Retrieval framework is to fit a Forward Model to an observation. A Forward Model framework is necessary to provide information about the planet, the host star, temperature-pressure profiles as well as chemistry and contributions (e.g. CIA, Rayleigh, gray clouds). TauREx 3 adopts the layer-by-layer approach for the temperature profile, which can be parameterized in different ways, such as isothermal, a radiative two-stream approximation, a custom profile loaded from a file, or a multi-point temperature profile. The vertical pressure profile is equally spaced in log-pressure, between a P_{max} and a P_{min} value specified by the user along with a number of layers N_l. The cloud model provided by TauREx 3 is discretized along $P(z)$, allowing the user to define an opacity value in square meters for layers between the pressure at the top and the pressure at the bottom of the cloud deck.

TauREx 3 supports equilibrium chemistry using the ACE chemical code (Agúndez et al., 2012, 2020), FastChem (Stock et al., 2018), GGchem (Woitke et al., 2018), and the Free chemistry model (Al-Refaie et al., 2022b). In our study, we explore different combinations of chemistry and contribution parameters. We performed retrievals with and without clouds, using ACE, FastChem, GGchem, or Free chemistry. To perform retrievals, TauREx 3 can use several sampling techniques PyMultiNest and MultiNest, PolyChord, or dyPolyChord. For our study, we used the nested sampling retrieval algorithm Multinest with its Python version PyMultiNest (Feroz et al., 2009). TauREx 3 is a full Bayesian Retrieval framework, which returns the final exoplanetary spectrum along with all parameter posterior distributions and the Bayesian Evidence. For model comparisons, we use the Bayesian Evidence as defined by Waldmann et al. (2015b) to compute the logarithmic Bayes factor,

$$\Delta logE = \log \frac{E_{model_A}}{E_{model_B}} = \log E_{model_A} - \log E_{model_B}, \tag{7.1}$$

where E_{model_A} and E_{model_B} are the evidences of two competing models. According to Benneke and Seager (2013), by translating these Bayes factors into a statistical significance (Kass and Raftery, 1995), $\Delta\log E \leq 2$ does not favor one of the two models, while we can start to favor one compared to the other with $\Delta\log E \geq 3$.

7.2.2 Chemical model

The assumption of chemical equilibrium is a classic hypothesis when considering exoplanetary atmospheres (Seager and Sasselov, 2000; Burrows et al., 2007, 2008; Fortney et al., 2008; Madhusudhan et al., 2011; Kataria et al., 2014; Al-Refaie et al., 2021). A system is at a thermodynamic equilibrium state when there is thermal, mechanical, and chemical equilibrium at the same time. This equilibrium is characterized by the minimum of a thermodynamic potential, such as the Gibbs free energy. It happens in exoplanets' atmospheres when the dynamical timescales can be considered longer than the chemical reaction timescales and when we suppose negligible the irradiation by a dissociating or ionizing source (photochemistry or cosmic rays induced processes). For very hot planets this approximation is close to reality, on the other hand, for cooler planets, vertical mixing and photodissociation have an effect on the chemical composition and the atmospheres are no longer at a thermodynamic equilibrium state. This disequilibrium chemical composition must then be taken into account with a more complex kinetic model (Cooper and Showman,

2006; Moses et al., 2011, 2013; Venot et al., 2012, 2020b; Molaverdikhani et al., 2019; Tsai et al., 2021, 2022; Al-Refaie et al., 2022b).

Although several different algorithms have been made to calculate the chemical composition at equilibrium, we focused on three algorithms for this study: ACE (Agúndez et al., 2012, 2020), GGchem (Woitke et al., 2018) and FastChem (Stock et al., 2018). The basic principle of these three models is the same, they start from an initial composition made up of initial abundances for each molecule taken into account then they iterate until a convergent state. However, each model has a slightly different procedure, whether in the molecules taken into account as input, the iteration method, or the network of chemical reactions. We will study in more detail these differences in this part.

ACE minimizes the total Gibbs free energy by applying the algorithm first introduced by White et al. (1958). ACE is based on an algorithm implemented in the NASA/CEA program and presented in detail in Gordon and McBride (1994). For a closed system of N chemical compounds at a certain temperature and pressure, in the absence of disturbance (transport, UV radiation, etc.), the equilibrium chemical composition can be calculated theoretically, thanks to standard-state chemical potential expressed as a function of the standard-state enthalpy and entropy of the species. These thermodynamic quantities can be calculated using NASA polynomial coefficients (see e.g. McBride et al. (2002)) in databases such as NASA/CEA (McBride et al., 2002) or the Third Millennium Thermochemical Database (Goos et al., 2016). The chemical species used include 105 neutral species composed of C, H, O, and N, more specifically species up to 2 carbon atoms and the main nitrogen species (NH_3, HCN, N_2, NO_x). It has been validated for temperatures as low as 300K. The reader is encouraged to consult Venot et al. (2012) for more details on the ACE code thermodynamic coefficients and calculation of thermochemical equilibrium.

Both FastChem and GGchem use a second type of method for determining the chemical composition at the equilibrium state. These two programs use the law of mass action and equilibrium constants, with some subtleties (for FastChem equilibrium constants are based on Gibbs free energy while for GGchem they are based on partition functions). This amounts to solving a system of N algebraic equations with N unknowns, which correspond to the conservation equations for N elements. The partial pressure of each molecule is defined as the partial pressure of the constituent atoms by the atomization equilibrium constant. This system of equations can then be solved by any root-finding algorithms like the Newton–Raphson method for example (Russell, 1934; Brinkley, 1947; Tsuji, 1973).

The thermodynamical data used in FastChem are mainly from the NIST-JANAF database detailed in Chase (1998). The list of species used has been modified to take into account molecules that may be of interest in astrophysics with data from Tsuji (1973); Barin and PLatzki (1995); Burcat et al. (2005); Goos et al. (2016). The total of species used amounts to 396 neutral and 114 charge species and the code has been validated for parameters down to 100 K and up to 1000 bar. The reader is directed to Stock et al. (2018) for more details on the list of species used. We note that the FastChem code has recently been updated with FastChem 2 (Stock et al., 2022) which is more efficient but not yet available in combination with *TauREx*.

Compared to the other two codes we are using, GGchem takes condensation into account. In fact, the formation of liquids and solids in the atmosphere will have an effect on the composition at thermodynamic equilibrium. Condensed species can consume certain elements leaving a significant difference in composition between before condensation and after condensation (Woitke and Helling, 2004; Juncher et al., 2017). Condensation will

have an effect, especially at temperatures below 2000 K so this mechanism will mainly affect GJ1214 b and HD189733 b in our work. The data included in this code includes 552 molecules and 257 condensates, including 38 liquids, and GGchem has been proven robust down to 100 K. All elements from hydrogen to zirconium are included, as well as the option to add tungsten and charges. We refer the reader to Woitke et al. (2018) for more details on the list of species.

7.2.3 Retrieval procedure

All spectra configurations are retrieved with the same set of retrieval models. These models are representative of all input configurations. We expect each input configuration to be best retrieved by the corresponding retrieval models.

Each retrieval model assumes a four points temperature-pressure (TP) profile (T_{top}, $T_{surface}$, T_1, and T_2) with the corresponding pressure level P_1 and P_2 free to converge between $P_{surface}$ and P_{top}. It has already been shown (Rocchetto et al., 2016; Pluriel et al., 2022) that retrieving an isothermal temperature profile generates biases for hot planets. We also retrieve the radius of the planet. We duplicate our retrievals by adding a gray cloud level as a parameter. Finally, we took into account two chemical configurations in the retrievals, which we call Free and equilibrium chemistry.

Free chemistry is based on vertical constant abundances, taking into account the following species: H_2O, CO, CH_4, CO_2, HCN, NH_3, FeH, SiO, Na, K, TiO, and VO. Note that for GJ1214 b we did not retrieve Na, K, TiO, and VO because these species cannot be in gaseous form under these temperature and pressure conditions. This configuration gives the model a certain degree of freedom, as it imposes no physical or chemical constraints on what will be retrieved. In this way, it is possible to retrieve abundances corresponding to non-equilibrium chemistry, or various other species distributions since the species are not correlated with each other. However, it could also retrieve non-realistic chemical abundances. On the other hand, it should be remembered that this is a 1D model and we are therefore limited when faced with strong vertical variation (which could be compensated for by the two-layer method of Changeat et al. 2019).

Chemical equilibrium is based on temperature and pressure conditions. Thus, abundances vary with altitude according to the retrieved TP profile, as we assume they are at chemical equilibrium. As the chemical profiles are not forced to be vertically constant, this approach should be more accurate for real atmospheres. However, if the observed planet's atmosphere exhibits non-equilibrium chemistry (due to quenching and/or photochemistry), or longitudinal/latitudinal heterogeneities, the retrieved parameters will be erroneous. In such cases, it may not be possible to find a consistent fit, or the retrieval may find an adequate fit that corresponds to erroneous parameters. Equilibrium chemistry is calculated using three different models included in *TauREx*: ACE, FastChem, and GGchem (details Section 7.2.2). Metallicity (Z) and C/O ratio, from which abundance profiles are derived, are retrieved. ACE, FastChem, and GGchem have already been shown in Al-Refaie et al. (2022b) to be equivalent using the same molecules and without condensation for GGchem. Thus, we consider all the molecules of each model as well as condensation for GGchem.

Therefore, we end up with 4 retrieval models (Free, ACE, FastChem, GGchem) for each of the 3 planets (GJ1214 b, HD189733 b, WASP-121 b) considering each of the 6 input configurations (listed Table 7.2). This makes 4×3×6 = 72 retrievals (×2 with clouds). For better readability of the large number of results, we do not show GGchem results

for GJ1214 b, where the temperature is not high enough to be affected by the additional condensation considered by GGchem, and we do not show ACE results for WASP-121 b, where the missing species in ACE mean that the model is not representative of this type of planet (see Section 7.2.2 for more details on these chemical models). Retrievals are compared to each other considering their relative Bayes Factor as described in Sec. 7.2.1, equation 7.1, following the same idea developed in Tsiaras et al. (2018) with a baseline. However, we define here the Bayes factor (ΔlogE) as the difference with the logarithmic evidence of the worse model (lower one). This allows us to compare the different models between each other as it is done in Panek et al. (2023).

Table 7.3 shows free parameters and priors for the retrievals. We used a uniform sampling in log space for the species, the metallicity, the pressure P_1 and P_2, and the pressure of the clouds, and a uniform sampling in linear space for the temperatures, the radius, and the C/O ratio.

Parameters	**Bounds**		
	GJ1214 b	**HD189733 b**	**WASP-121 b**
T_{top} [K]	100 to 2000	500 to 2500	450 to 3750
T_{bot} [K]	100 to 2000	500 to 2500	450 to 3750
T_1 [K]	100 to 2000	500 to 2500	450 to 3750
T_2 [K]	100 to 2000	500 to 2500	450 to 3750
$\log_{10}(P_1)$ [Pa]	2 to 6	2 to 6	2 to 6
$\log_{10}(P_2)$ [Pa]	-1 to 6	-1 to 6	-1 to 6
$\log_{10}(P_{clouds})$ [Pa]	-2 to 6	-2 to 6	-2 to 6
radius [R_{jup}]	0.1 to 0.5	0.5 to 2.0	1.0 to 2.5
Chemistry			
Equilibrium			
log(Z)	-1 to 3	-1 to 3	-1 to 3
C/O	0.01 to 2	0.01 to 2	0.01 to 2
Constant			
$\log_{10}(H_2O)$	-12 to -1	-12 to -1	-12 to -1
$\log_{10}(CO)$	-12 to -1	-12 to -1	-12 to -1
$\log_{10}(CH_4)$	-12 to -1	-12 to -1	-12 to -1
$\log_{10}(CO_2)$	-12 to -1	-12 to -1	-12 to -1
$\log_{10}(HCN)$	-12 to -1	-12 to -1	-12 to -1
$\log_{10}(NH_3)$	-12 to -1	-12 to -1	-12 to -1
$\log_{10}(FeH)$	-12 to -1	-12 to -1	-12 to -1
$\log_{10}(SiO)$	-12 to -1	-12 to -1	-12 to -1
$\log_{10}(Na)$	-12 to -1	-12 to -1	-12 to -1
$\log_{10}(K)$	-12 to -1	-12 to -1	-12 to -1
$\log_{10}(TiO)$	-12 to -1	-12 to -1	-12 to -1
$\log_{10}(VO)$	-12 to -1	-12 to -1	-12 to -1

Table 7.3. Free parameters and priors for the retrievals.

7.3 Results and discussion

7.3.1 From 1D to 3D

We present how the transmission spectra are affected considering 1D or 3D GCM models, with or without equilibrium chemistry.

7.3.1.1 Constant chemistry

We will first focus on the simulated spectra assuming constant chemistry simulations which correspond to the left panels of Figure 7.2.

We consider in this case that the abundances are constant everywhere in the atmosphere no matter the temperature and pressure conditions. Thanks to this assumption, the effects on the spectra are only due to the thermal structure of the atmosphere. For our coldest case GJ1214 b, we see strong differences comparing 1D and 3D transmission spectra. Indeed, the 3D GCM model of GJ1214 b (Charnay et al., 2015) shows large day-night asymmetries with an extended hot day side combined with a 30° eastward shifted hot spot. Due to this shift, the light coming from the star thus probes through the hot day side with a larger scale height compared to the limb (which is used to compute the 1D temperature profile of the pseudo-1D model). This implies hundreds of ppm differences in the transit depth in particular in the major absorbers, such as carbon dioxide and methane. However, on the water-dominated bands in the far infrared and in the visible, we see very small differences between 1D and 3D transmission spectra, in the order of 50 ppm. The altitude where the atmosphere is opaque on these bands is indeed deep in the troposphere (around 100 mbar). Thus, even when we take into account the inflated day side in 3D, the effective radius observed is very similar to that in 1D, because the region where the atmosphere is opaque remains at the same place at the limb. Therefore, the impact of the 3D structure of the atmosphere will depend on the wavelength and the composition.

Similar results are shown concerning the hottest study case WASP-121 b. 3D GCM models show that for highly irradiated atmospheres in tidal locking, the radiative timescale becomes substantially smaller than the dynamic timescale implying that almost no heat is transported from the day to the night side. It results in an extremely large day-side scale height compared to the night side because of the large day-night temperature contrast. Indeed, Keating et al. (2019) showed that regardless of the day-side temperature, the night-side temperature of short-period gas giants is relatively uniform, around $\sim$ 1100 K. This very inflated day side is thus mainly probed during the transit which explains why 3D transmission spectra are by thousands of ppm larger than in 1D. Unlike the previous case GJ1214 b, all absorption bands are shifted. The atmosphere is so inflated for this ultra-hot Jupiter that the altitude at which the atmosphere is opaque is much higher due to the much greater scale height.

Interestingly, the results are very different in our intermediate case HD189733 b. Here, despite a hotter day side compared to the night side, we see less than a 50 ppm difference between 1D and 3D transmission spectra. This means that the limb represents well the observable and that such atmospheres are more homogeneous than colder or hotter atmospheres. Here, despite a strong day-night temperature contrast, we observe less than 50 ppm differences between the 1D and 3D transmission spectra. This is due to the high surface gravity of the planet, which mitigates the scale height differences between the different sides

of the planet. Thus, using only the limb is a fair representation of the observable.

7.3.1.2 Equilibrium chemistry

We now look at the equilibrium chemistry simulations which are shown in the right panels of Figure 7.2. Now, the simulated spectra are impacted by both the effects of the chemical and the thermal heterogeneities which is more realistic.

For GJ1214 b, the global picture has completely changed compared to constant chemistry. The largest difference (1000 to 2000 ppm) concerns CO_2 bands at 2.5 and 4.5 μm. The temperature profile of the 1D model is not hot enough to obtain abundant CO_2 at the equilibrium state, whereas the day side of the planet in the 3D model reaches a temperature where CO_2 is abundant enough to show strong signatures. Furthermore, as explained above, we probe a non-negligible part of the planet's day-side, hence the presence of broad bands of CO_2. We also see differences in the water bands (between 1 and 2 μm) which weren't present in the constant chemistry model. The reason for these differences is the longitudinal variation in water abundance, which is lower on the day side of the planet. Light from the star then probes deeper regions corresponding to lower transit depth. Between 5 and 9 μm, as well as around 3.5 μm, in the region of the methane bands, the 1D and 3D spectra show fewer discrepancies than the rest of the spectra. Indeed, looking at the methane abundances in Figure D.7, we see that its presence is drastically reduced in the day side above 100 mbar which is deeper than where we probe. It results that in these bands, we are not affected by the hot day side and we are mainly probing at the limb which is equivalent to the 1D spectrum.

For WASP-121 b, the 1D and 3D spectra show few differences in the whole wavelength range using equilibrium chemistry compared to constant chemistry. As shown in Figure D.9, the abundances of almost every species drastically diminished on the day side, mainly due to thermal dissociation (Parmentier et al., 2018), with the exception of carbon monoxide which is only divided by two due to its dilution in an H-dominated day side instead of a H_2-dominated atmosphere. This implies that, on the water bands, the spectrum is not affected by the day side of the atmosphere because water is almost not present there. That's why a 1D model manages to fit the spectrum as shown in Pluriel et al. (2020a). However, we can see in the residuals of Figure 7.2 that in some bands, the fit is clearly less good than in the other bands. This is particularly true for the CO bands around 2.5 and 4.5 μm, as well as for the TiO and VO bands in the visible. As we explained, CO is present on the day side of the atmosphere, where extreme temperatures induce a scale height far greater than the scale height of the limbs. Consequently, the 1D model does not represent this behavior well, resulting in a large difference (around 300-400 ppm) at these regions of the spectrum.

The differences between 1D and 3D for HD189733 b with equilibrium chemistry are very low, as with constant chemistry. Indeed, we see in Figure D.8 that the main species such as H_2O, CH_4, CO_2, and NH_3 do not display strong longitudinal variations. In addition, we have shown before that due to similar scale height, there is no significant impact comparing 1D and 3D spectra with constant chemistry. As a consequence, a 1D model at the limb is equivalent to the 3D model meaning that for such warm atmospheres, we are probing near the limb.

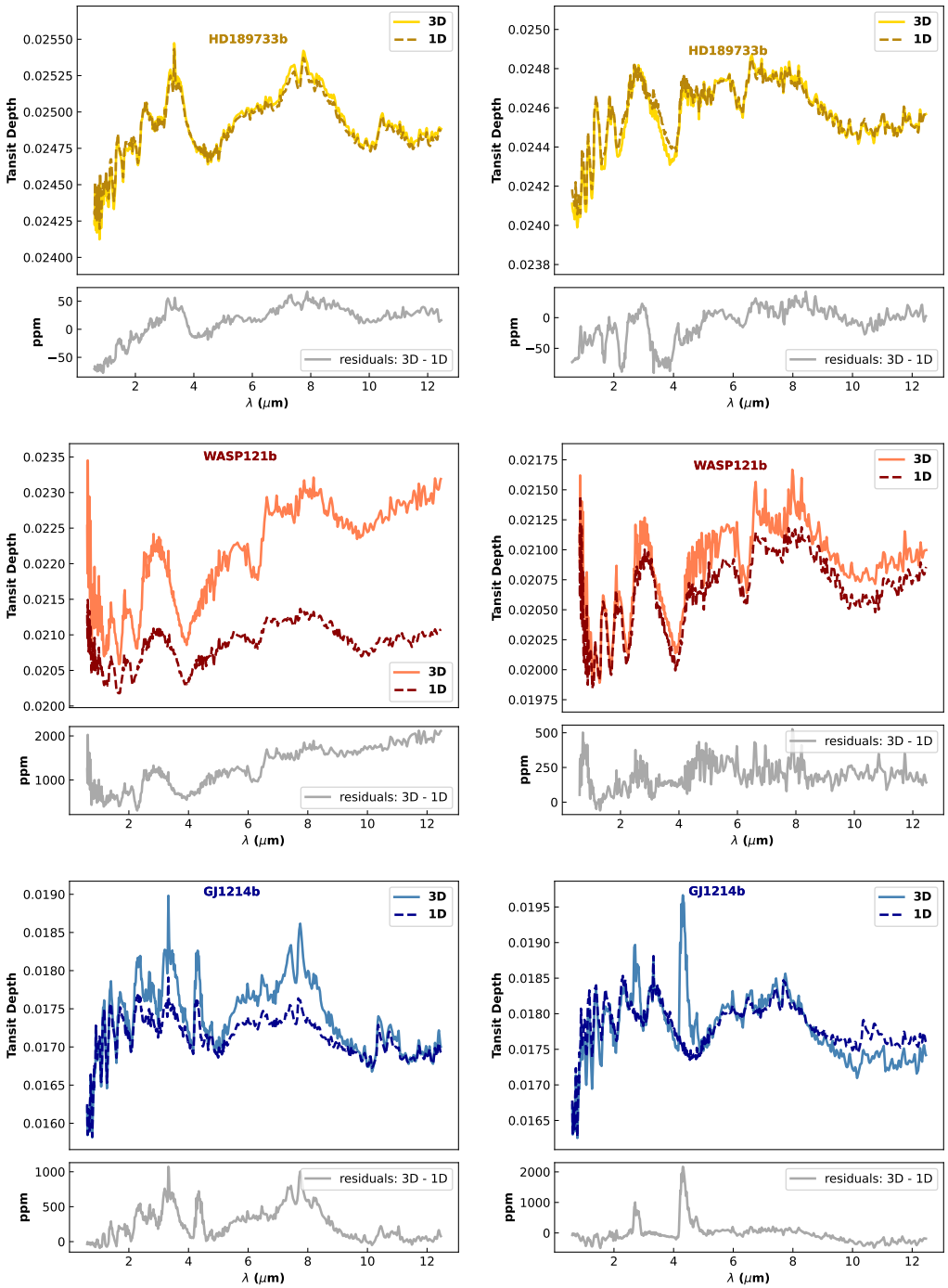


Figure 7.2. Transmission spectra simulated with *Pytmosph3R* (Falco et al., 2022) for HD189733 b (**top**), WASP-121 b (**middle**) and GJ1214 b (**bottom**). Each panel compares two transmission spectra based on a 1D and a 3D atmosphere, respectively in dashed and solid lines. **Left panels:** constant input chemistry. **Right:** equilibrium input chemistry. The differences between the 3D and the 1D spectra are plotted below each panel in grey.

7.3.2 Cloud effect

Even if we have good reasons to think that cloud decks are present in many exoplanets (Parmentier et al., 2016; Tsiaras et al., 2018), each atmospheric model used in this study, 1D or 3D, for the three planets, is cloudless. It would be interesting to add clouds in the simulations, in particular, because they affect the short wavelengths observed by JWST and Ariel. However, the aim of this study is to see the impact of chemical and thermal 3D effects on the transmission spectra and how to deal with these 3D effects in the context of atmospheric characterization using 1D retrieval models. For this reason, we chose to not over-complicate our models. Clouds would require a dedicated chapter. Clouds are nevertheless part of the retrieval parameters in the *TauREx* model. This can break up possible degeneracies (Pluriel et al., 2022; Changeat et al., 2020b), and we verified that the model works correctly by not retrieving a cloud layer when we knew that none was implemented. For each retrieval performed, the Bayes factor has always privileged a cloudless model, and in the retrievals assuming clouds, the cloud deck was always pushed near to the surface pressure thus without impact on the spectra. We thus decided not to present retrievals including clouds in this chapter as they bring no more information compared to cloudless retrievals.

7.3.3 What if atmospheres are 1D?

We used the theoretical 1D atmosphere (see Section 7.1.1 for construction) to check the correct behavior of the 1D retrieval code. We remind that the 3 planets have been chosen to study the retrieval biases depending on the effective temperature of the planets, from warm Neptune to ultra-hot Jupiter. Also, the chemical construction with a constant profile or equilibrium chemistry has been considered to unravel the retrieval biases from temperature (constant chemistry) and chemistry (equilibrium chemistry). To do so, Free retrievals (constant chemistry) and equilibrium retrievals are both performed for each configuration (more details Section 7.2.3).

Figure 7.3 and Table D.2, D.3, D.4 shows that for all configurations the best retrieval is consistent with the input model configuration. This means that Free retrievals fit better input constant chemistry and equilibrium retrievals fit better input equilibrium chemistry. However, the best retrieval is not always more significant than the others and the equilibrium retrieval models are not all adapted for the different configurations. Figures D.1, D.4, D.5 and D.6 show that the temperature and species profiles are mostly not well retrieved below the probed altitudes (deeper than $\sim 10^2$ Pa). This part of the atmosphere does not contribute to the features of the spectra, which explains why they are not well retrieved. However, the retrieved values could be well constrained while the input values are often outside the uncertainty. Thus, we cannot trust the retrieved profiles of the lower atmosphere.

7.3.3.1 Temperate-Warm planet: GJ1214 b

The best retrieval is consistent with the input model configuration and significantly better than the other models with $\Delta\log E \geq 19$ for constant input chemistry and $\Delta\log E \geq 7$ for equilibrium input chemistry (Figure 7.3 and Table D.2).

- Constant input chemistry: For lower pressure than $\sim 10^2$ Pa (high altitudes), Free retrieval shows a significantly better retrieved temperature profile within 60 K compared to the input profile (Figure D.1). The main absorber CH_4 is retrieved by the Free

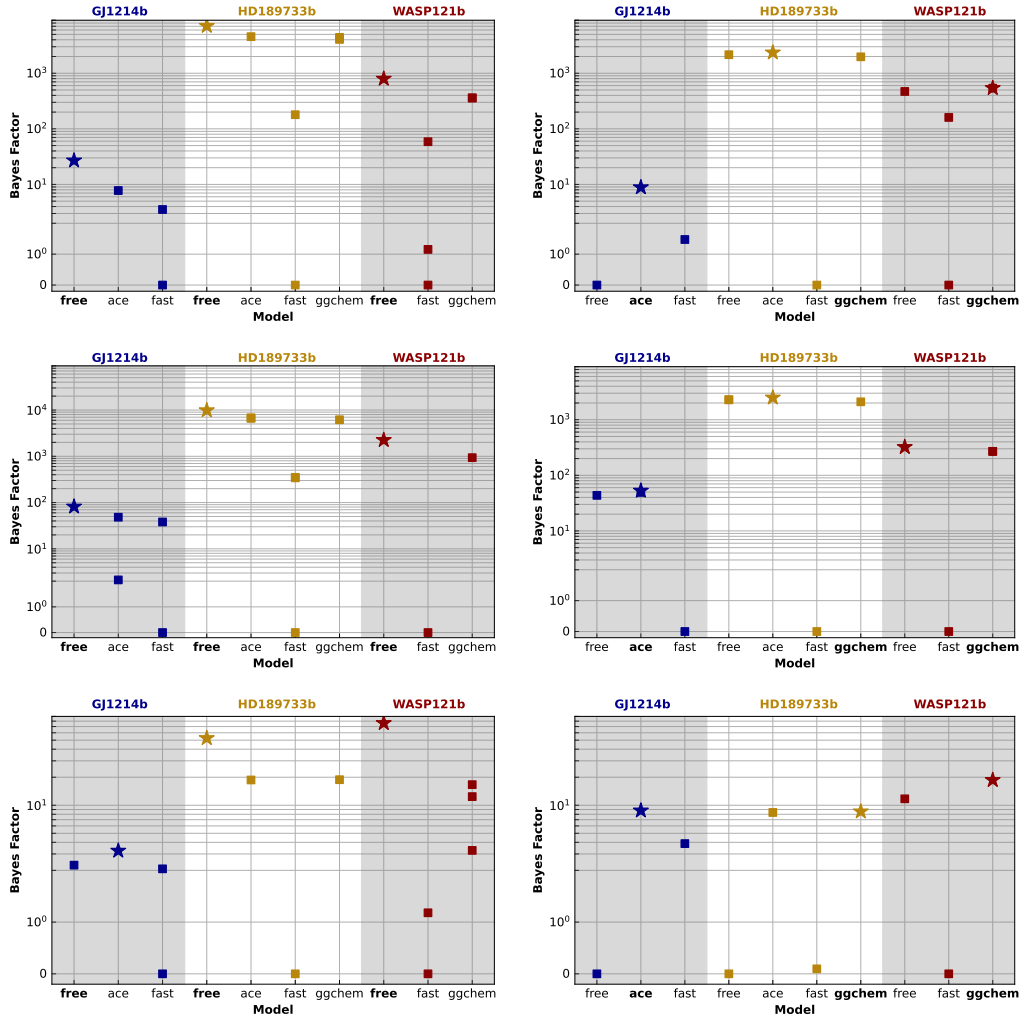


Figure 7.3. Bayes factors for each retrieval (Free, ACE, FastChem and GGchem). By definition, we put at 0 the retrieval with the lowest Bayesian evidence as it is the reference of the comparison. **Left panels:** constant input chemistry. **Right:** equilibrium input chemistry. **Top and middle:** JWST simulated observation for atmospheric 1D and 3D assumption respectively. **Bottom:** Ariel simulated observation for a 3D atmospheric assumption. The star represents the highest Bayes factor. The expected best retrieval is highlighted in bold.

retrieval within 22% deviation while H_2O, CO_2 and NH_3 giving secondary features are retrieved within 29% deviation (Figure D.4). This allows the best model to be significantly in agreement with the input model even if not all input profiles are included in the uncertainty.

- Equilibrium input chemistry: For lower pressure than $\sim 10^2$ Pa, all retrievals give the same temperature profile within 90 K of the input model (Figure D.1). All retrievals also give the same main absorber profiles (CH_4 and H_2O) below $\sim 10^2$ Pa, where the input equilibrium chemistry is constant at these temperature and pressure conditions (Figure D.4). However, ACE is better at retrieving NH_3 profile, which varies by 4 orders of magnitude and contributes at 3 μm, which makes ACE the best model in agreement with input ACE chemical composition. We can highlight that Free and ACE retrievals prefer to retrieve at least 104% more CH_4 and down to 32% less H_2O than input values to better fit the spectrum, with input values out of the uncertainty. FastChem displays some discrepancies with ACE, as pointed out in Al-Refaie et al. (2022a) when using an isothermal profile. ACE fits better than FastChem which is consistent with input ACE composition. However, none of the equilibrium models can reproduce the input composition. The metallicity is retrieved by ACE within 19% and the C/O ratio only at 96% deviation to the input value. While FastChem, which is not considered the best retrieval, has a closer value to input for C/O ratio (76%) and metallicity (9%) (see Table D.2 and Figure D.10). It shows that even at these "low" temperatures we already have difficulties to perfectly retrieve the input model.

To summarize for GJ1214 b: the retrievals work well on the temperature, but not without some biases on the chemistry. Constant chemistry and 1D temperature profiles are retrieved by Free retrieval within 29% deviation. However, even at such "low" temperatures, equilibrium input chemistry is best retrieved by equilibrium models because of species with strong vertical variations (here NH_3). In addition, *TauREx* not only has difficulty in accurately retrieving the C/O ratio and the metallicity (respectively 96% and 19% deviation for the best model ACE), but also the values closest to the input are not retrieved by the more significant model ACE, despite the fact that the best retrieval ACE is consistent with the input ACE equilibrium chemistry.

7.3.3.2 Warm-hot planet: HD189733 b

The best retrieval is consistent with the input model configuration and significantly better than the other models with $\Delta\log E \geq 2536$ for constant input chemistry and $\Delta\log E \geq 200$ for equilibrium input chemistry (Figure 7.3 and Table D.3).

- Constant input chemistry: For lower pressure than $\sim 10^2$ Pa (high altitudes), Free retrieval shows a significantly better fit of the temperature profile within 150 K compared to the input profile (Figure D.1). However, the abundance of the main absorbers CH_4, H_2O, NH_3, TiO, and VO, are overestimated by 70 to 130%. This remains a better fit compared to equilibrium chemistry retrievals that show strong vertical variation in these temperature-pressure conditions (Figure D.5).

- Equilibrium input chemistry: For lower pressure than $\sim 10^2$ Pa, Free and FastChem retrievals overestimate the temperature profile by more than 250 K while ACE and GGchem stay mostly within 250 K deviation from the input values. However, these

two chemical models underestimate the temperature in the upper atmosphere by more than 500 K (Figure D.1). In this peculiar configuration, with warm-hot temperature, molecules such as TiO, VO, and K are close to their condensation temperature. Compare to the isotherm configuration of Al-Refaie et al. (2022a), condensation occurs at mid and high altitudes where the temperature is lower than in the deep atmosphere. Thus, a strong bias occurs with FastChem where it considers these molecules but not their condensates. Therefore, FastChem cannot retrieve properly this configuration where TiO, VO, and K have strong features in the visible while they should not be because of their condensation, Figure D.1 shows a wrong temperature profile, Table D.3 shows wrong retrieved parameters and Figure D.5 shows wrong species profiles. Thus, GGchem, which considers condensation of these species, should in theory solve this issue but it surprisingly does not. It shows a better fit for the temperature profile and species abundances with regards to FastChem. But still, it does not manage to give a consistent abundance of K, overestimated by a factor of 2, and shows a strong feature in the visible which bias all the spectra, since all chemical abundances are correlated. This is explained by the discrepancy between the K abundance calculated by GGchem and the input abundance calculated by Parmentier et al. (2016). Al-Refaie et al. (2022a) found a good agreement between ACE, FastChem, and GGchem because the same molecules are considered in the three models. However, we show here chemical discrepancies between the two models, illustrating that using imperfect chemical models with regards to what is actually observed leads to biased interpretations. ACE, which considers only C, H, O, and N atoms, ends up giving the best-fit model by getting rid of TiO, VO, and K condensation issues. The retrieved thermal profile is similar to the one retrieved by GGchem but the main absorber H_2O is now perfectly constrained under 1% deviation (see Figure D.5). All these biases can be seen in Table D.3 and Figure D.11 on the C/O ratio and the metallicity. FastChem model is far from input values while ACE has a 24% deviation. Furthermore, even if the Free is not the more significant model, it is closest to input values.

To summarize for HD189733 b: the retrievals work well on the temperature for constant input chemistry (within 150 K) but for equilibrium input chemistry they begin to show difficulties in retrieving the top of the atmosphere (underestimated by 500 K), as well as the bottom of the atmosphere. Only the pressures corresponding to the highest atmospheric contribution (between $\sim 10^2$ and $\sim 10^0$ Pa) are well retrieved (within 250 K). In addition, we encounter even more chemical abundance bias compared to GJ1214 b. Constant chemistry profiles are best retrieved by the Free retrieval model (which is consistent) with an overestimation of 70 to 130%. Due to the strong vertical variation, the equilibrium configuration is better retrieved by an equilibrium model (ACE), which is not the expected one (GGchem), but which fits very well the main absorber H_2O under 1% deviation. However, the study performed on this planet shows all the limitations of the 3 chemical equilibrium models for retrievals. At this temperature, the role of condensates (such as TiO, VO, and K here) is essential with their feature in the visible. The discrepancies between the K chemical modeling of GGchem and the K chemical modeling of Parmentier et al. (2016), bias GGchem retrieval. This is a good example of what can be encountered when fitting observations with a model that does not perfectly reproduce the true chemistry. The C/O ratio and the metallicity are well retrieved by ACE (best model) but not as well as Free. Our results show that we can

retrieve values closest to input with a model statistically not favored.

7.3.3.3 Ultra hot planet: WASP-121 b

The best retrieval is consistent with the input model configuration and significantly better than the other models with $\Delta\log E \geq 434$ for constant input chemistry and $\Delta\log E \geq 71$ for equilibrium input chemistry (see Figure 7.3 and Table D.4). We observe a higher overall uncertainty in the retrieved temperature and chemical species profiles compared to the other two planets. The very hot temperature coupled with a sharp day-to-night gradient brings a complexity that is more difficult to retrieve with a simpler model.

- Constant input chemistry: in the upper atmosphere, the retrieved thermal profiles are either overestimated or underestimated by $\sim$ 1000 K (Figure D.1). This is far from the input temperature profile ($\sim$ 50% deviation). However, close to a pressure corresponding to the highest atmospheric contribution, between $\sim 10^2$ and $\sim 10^0$ Pa, retrieved profiles remain within 500 K (less than 25% deviation), while the best model (Free) is not significantly better. The abundance of the main absorbers H_2O, CH_4 and TiO, FeH in the visible, are overestimated at 66 to 95% for the Free retrieval (best model). Despite these high values, this is still a better fit than equilibrium chemistry. The strong vertical variation in equilibrium chemistry under these temperature and pressure conditions cannot match the constant input chemistry (see Figure D.6).
- Equilibrium input chemistry: we observe here the same behavior as constant input chemistry for the retrieved temperature profiles (see Figure D.1). There is less than a 10% deviation between $\sim 10^2$ and $\sim 10^0$ Pa for the best retrieval GGchem, but the retrieved profiles are mostly erroneous outside this range (reaching more than 50% deviation). Considering the strong temperature gradient between $\sim 10^2$ and $\sim 10^0$ Pa, we observe the same behavior as HD189733 b regarding the condensation of species such as TiO, VO, and K. This makes GGchem a better model than FastChem in this case, as shown in Figures 7.3, D.6. Although GGchem gives the best fit, none of the retrieval models fit all the main absorbers H_2O, TiO, CO_2, CO, and VO better than the others. The Free retrieval is better for H_2O and CO_2, while GGchem is better for the other absorbers (despite a strong deviation for both models, reaching 200%). In addition, the best model, GGchem, retrieves an erroneous C/O ratio with a deviation of 80% and an erroneous metallicity with a deviation of 530%. In contrast, the Free model ends up giving values close to the input within 13% (see Table D.4 and Figure D.12). Thus, all the models retrieved different parts of the input, but none of them obtained the entire structure.

To summarize for WASP-121 b: the retrievals perform poorly outside the pressures corresponding to the highest atmospheric contribution (between $\sim 10^2$ and $\sim 10^0$ Pa). Due to strong large-scale vertical variation in temperature and species, 1D temperature profiles for constant and equilibrium chemistry are best retrieved at only 25% between $\sim 10^2$ and $\sim 10^0$ Pa. Constant input chemistry is best retrieved by Free retrieval, overestimated by 66–95%, and equilibrium input chemistry is best retrieved by GGchem retrieval, but with a deviation reaching ~200%. The retrieval models are consistent with the input configurations, but not without bias on absolute values. GGchem retrieves the wrong C/O ratio and metallicity, while the Free model is within 13%. The strong vertical gradient on a large scale brings a

complexity that is difficult to retrieve correctly with a simpler model, which also translates into greater uncertainty in the retrievals.

7.3.3.4 Summary

The theoretical 1D analysis validates the consistency of the 1D retrievals, but not without a few biases. This shows that we can trust neither the retrieved parameters below the probed altitudes (as well as at the top of the atmosphere when moving towards hotter planets), nor the uncertainties given, which are largely underestimated.

Furthermore, equilibrium chemical models are not equivalent and give significantly different results:

- ACE cannot retrieve ultra-hot planets whereas it might be a better approximation for cooler planets.
- FastChem is biased towards warm, even hot planets, where species are close to or below condensation temperature. This is never the best retrieval model.
- GGchem, which should be a complete model, will still be in competition with a simplified model for cold and warm planets. Apparently the best option for hot planets, but Free retrieval can give better retrieved values.

We need to be careful with equilibrium models. Our study shows that if one part of the chemistry modeling is wrong, all the chemical abundances will be biased since everything is correlated. This agrees with the conclusion of Al-Refaie et al. (2022a) using an isotherm configuration.

7.3.4 But atmospheres are 3D!

7.3.4.1 Constant chemistry

Given the 3D thermal structure, we used theoretical atmospheric models with constant chemistry (see Section 7.1.2) to disentangle the 3D effects of temperature without being biased by chemistry. Section 7.3.3 confirms the overall correct behavior of the retrieval code or highlights any biases we may encounter. So, we can be confident in this approach to focus on thermal 3D effects. To this end, Free retrievals (constant chemistry) and equilibrium retrievals are performed for each configuration (more details in Section 7.2.3). The aim is also to compare the biases of the JWST and Ariel instruments.

Figure 7.3 shows that the Free retrieval finds the best solution compared to equilibrium chemistry (except for GJ1214 b in the Ariel configuration). This is consistent with the input constant chemistry. For GJ1214 b in the Ariel configuration, ACE retrieval has a better Bayes factor but within 1.5 deviations compared to the others. Therefore, all models are statistically equivalent and none is preferred. We observed a stronger deviation of the Bayes factor for the JWST configuration compared to the Ariel one. This could simply be due to the higher resolution on a larger wavelength range for the JWST configuration or this may be linked to some particular wavelength band such as the lack of data points in the visible for the Ariel configuration where strong features for hot planets (such as TiO, VO, and K) are located.

7.3.4.1.1 Temperate-Warm planet: GJ1214 b All retrieval models for the Ariel spectrum are within a Bayes factor deviation of 1.5 which makes all models equivalent (see Table D.2). However, the JWST spectrum is better retrieved by the Free retrieval with ΔlogE= 33 compared to the second-best model. The higher resolution of the JWST spectrum, compared to the Ariel spectrum, gives more constraints on H_2O and CH_4 features at low wavelength (between 1 and 2 μm), but probably also at higher wavelength (above 4 μm). Yet, this does not translate to closer-to-truth retrieved profiles, although the retrieved uncertainties are smaller. Figure D.2 and D.3 show consistent inputs and retrieved temperature profiles for GJ1214 b. At low altitudes (deeper than $\sim 10^2$ Pa) there is high uncertainty because these altitudes are not probed. Higher in the atmosphere (probed altitudes), the temperature day-night variation is within 300K and all models retrieve the limb profiles. H_2O and CH_4 are responsible for the main features of the spectra. Both retrieved values of the Free retrieval are around 20% lower than the input. For the JWST spectrum, the uncertainty is lower but the input value is not within the uncertainty. While, for the Ariel spectrum, the CH_4 input value is within the uncertainty. Thus, even with more data point, the retrieved chemical abundances is not closer to the input and the uncertainties cannot be trusted.

7.3.4.1.2 Warm-hot planet: HD189733 b The best retrieval (Free) is consistent with the input model configuration and significantly better than the other models with ΔlogE≥ 3048 for the JWST spectrum and ΔlogE≥ 34 for the Ariel spectrum (see Figure 7.3 and Table D.3). Discrepancies between retrieval models are the same as those explained in Section 7.3.3. Figure D.2 and D.3 show that the biases on the temperature profiles are the same as in Section 7.3.3. Temperatures retrieved below $\sim 10^2$ Pa correspond mainly to those of the limb. For the Ariel spectrum, the temperature is slightly warmer, but the solution for the JWST spectrum remains within the Ariel uncertainties, which are significantly higher. Equilibrium chemistry cannot reproduce the constant input chemistry, while Free retrieval gives consistent results but not without significant deviation (between 26% to 134%). Even with a variation of less than 500 K between day and night side, only temperature variation can bias the retrieval of species abundances. While VO is detected in the JWST spectrum, it is not in the Ariel spectrum. Retrieved uncertainties on the Ariel spectrum are larger than on the JWST spectrum. However (and to the exception of VO) the same molecules are detected with both observatories. The lack of detection of VO with Ariel is due to the coarser spectral resolution in visible light which is still sufficient to detect TiO.

7.3.4.1.3 Ultra hot planet: WASP-121 b The best retrieval (Free) is consistent with the input model configuration and significantly better than the other models with ΔlogE≥ 1310 for the JWST spectrum and ΔlogE≥ 60 for the Ariel spectrum (see Figure 7.3 and Table D.4). Discrepancies between retrieval models are the same as those explained in Section 7.3.3. Figure D.2 and D.3 show that the Free retrieval (best model) finds a temperature at the top of the atmosphere higher than the input model, and a temperature at the bottom of the atmosphere lower than the input model. The temperature transition occurs in the atmosphere where species absorption contributes the most, around $\sim 10^2$ Pa. The temperature gradient is steep, crossing all possible temperatures from day to night side. The chemical abundances of the main absorbers H_2O, TiO, FeH, VO, CO are retrieved between 44% and 100% for the JWST spectrum and between 48% and 100% for the Ariel spectrum, with the exception

of TiO with a deviation of 7% from the input value. However, TiO feature does not match the input spectrum for both JWST and Ariel configurations. The retrieved spectra are outside the uncertainties, by several sigma. The much higher JWST resolution in the visible compared with Ariel surprisingly does not provide better constraints on TiO. While we already encounter difficulties in retrieving input values with a 1D atmosphere, the huge temperature gradient between day and night brings even more biases. The end result is a temperature from both the day and night sides that does not allow the retrieval models to find input spectra and species profiles. There is not even thermal inversion retrieved by the best model.

7.3.4.1.4 Summary The conclusion in Section 7.3.3 remains the same in 3D. Secondly, using for retrievals a 4-point temperature profile gives good results for the cooler planets but not for hotter ones, which need at least 2D retrievals, as has already been pointed out in more detail by Pluriel et al. (2022). The higher resolution of the JWST spectrum, particularly in the visible, reduces uncertainties but does not provide a better fit. In addition, the input values will be out of uncertainty, making them unreliable. The lower atmosphere is still poorly retrieved, especially as we move towards hotter planets. Nevertheless, this is still a good detection of input species. Finally, the low resolution in the visible wavelength range of the Ariel spectrum has missed the detection of the visible absorber VO but never TiO with a retrieval similar to that of the JWST spectrum.

7.3.4.2 Equilibrium chemistry

Here the 3D thermal structure, as well as equilibrium chemistry, are considered as input. Considering the conclusion of Section 7.3.3 and 7.3.4.1, this will highlight biases due to the variability of chemical abundances in the atmospheres of warm to ultra-hot planets. Section 7.3.4.1 has already shown the temperature biases from the 3D structure. As previously done, Free retrievals (constant chemistry) and equilibrium retrievals are both performed for each configuration (more details in Section 3.3).

Figure 7.3 shows that the equilibrium retrievals always find the best solution compared to Free chemistry (except for the WASP-121 b JWST configuration). This is consistent with the input equilibrium chemistry. The retrieval models of the Ariel spectrum have less deviation from each other than the JWST spectrum. This is due to the lower spectral resolution across all wavelength bands, but particularly in the visible bands. Between visible and infrared the chemical species contributing to the spectral signatures are different, TiO, VO, and K against H_2O, CH_4, CO_2, CO, and NH_3. If we try to retrieve both parts at the same time, both will be biased, as the signature may come from different parts of the atmosphere, at different temperatures.

7.3.4.2.1 Temperate-Warm planet: GJ1214 b The ACE model provides the best fit to both JWST and Ariel spectra with $\Delta\log E \geq 10$ and $\Delta\log E \geq 5$ respectively (see Table D.2), which is consistent with the ACE input chemistry modeling. Figure 7.3 also shows that the JWST spectrum is secondly best retrieved by the Free model, while the Ariel spectrum is secondly best retrieved by the FastChem model. FastChem's retrievals poorly fit the CO and the visible bands, because of the chemical modeling differences with the input ACE chemistry modeling. Retrievals on the Ariel spectrum circumvent this issue thanks to its low spectral resolution in visible light. The main difference from the previous input

configuration is that the equilibrium chemistry results in a strong dichotomy of CO_2 between the day and night side (see Figure D.7). As a result, the retrieved temperature profiles correspond to the day side (see Figure D.2 and D.3). Table D.2 and Figure D.13 shows that the temperature bias on the day side still keeps a good agreement on the C/O ratio and the metallicity retrieved, within 20%.

7.3.4.2.2 Warm-hot planet: HD189733 b The ACE model retrieves the JWST spectrum much better than the other models with $\Delta\log E \geq 208$, as already explained in Section 7.3.3. However, the Ariel spectrum is equivalently retrieved by ACE and GGchem (Figure 7.3 and Table D.3), again due to the lack of constraint in the visible wavelength bands where discrepancies between chemical models appear. ACE is as significant as GGchem, but Table D.3 and Figure D.14 show that GGchem better retrieves the C/O ratio and metallicity. Figure D.3 shows that the temperature at the top of the atmosphere is unconstrained by the huge uncertainty. This part of the atmosphere, therefore, makes no significant contribution to the features of the spectra. In contrast, Figure D.2 shows that increasing resolution adds an erroneous constraint on the temperature of the top of the atmosphere. Only the temperature around pressures corresponding to the highest atmospheric contribution (around $\sim 10^2$ Pa), is consistent between equilibrium models and the input temperature profiles. The temperature retrieved at these pressures is that of the limb. Figure D.8 shows a good agreement between the retrieved species profiles.

7.3.4.2.3 Ultra hot planet: WASP-121 b The GGchem model retrieves the Ariel spectrum better than the other models with $\Delta\log E \geq 7$, as already explained in Section 7.3.3. However, the JWST spectrum is better retrieved by the Free model with $\Delta\log E \geq 52$ (see Figure 7.3 and Table D.4). Table D.4 and Figure D.15 show that GGchem retrieves for Ariel configuration the C/O ratio very well at 5% but not the metallicity (75% deviation), while for JWST configuration it is higher than 44% considering all retrieval models. The high spectral resolution of JWST imposes too tight constraints on the thermal contrast and thus, on the chemical distribution. This shows that such a contrasted atmosphere cannot be retrieved by a 1D model with correlated chemistry. However, the higher degree of freedom of the Free retrieval allows a better match. The temperature profiles retrieved between the Free model and the GGchem model are similar (within 500 K below 10^4 Pa, see Figure D.2 and D.3). The conclusions on temperature biases are the same as for HD189733 b. But Figure D.9 shows that species abundances are more difficult to retrieve.

7.3.4.2.4 Summary In addition to the previous biases from Section 7.3.3 and 7.3.4.1, the biases coming from the chemistry show that, even on a warm planet, it would make sense to fit the different molecular features separately to disentangle the temperature variation that brings chemical variability. Otherwise, using a 1D retrieval model will bias all different spectral contributions. Furthermore, only the pressure where the contribution is highest should be considered as a significantly good retrieval of the observation. The rest should be treated with caution. All models remain good at detecting input molecules.

	JWST												Ariel					
	1D						3D											
	constant			equilibrium			constant			equilibrium			constant			equilibrium		
	GJ	HD	WASP	GJ	HD	WASP	GJ	HD	WASP	GJ	HD	WASP	GJ	HD	WASP	GJ	HD	WASP
Species detection	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
C/O				×	~	×				✓	~	×				✓	✓	✓
Metallicity (Z)				✓	✓	×				✓	~	×				✓	✓	×
Top of the atmosphere [~ 10^0 Pa - ~ 10^{-4} Pa]																		
Temperature profile	✓	✓	×	✓	×	×	Limb	Night	×	Day	×	Night	Night	Day	×	Day	Day	Day
Chemical profile	✓	~	~	✓	~	×	✓	~	~	Day	×	×	✓	~	~	Day	Day	×
Middle of the atmosphere [~ 10^2 Pa - ~ 10^0 Pa]																		
Temperature profile	✓	✓	~	✓	✓	✓	Limb	Day	Limb	Limb	Limb	Limb	Limb	Day	Limb	Limb	Limb	Limb
Chemical profile	✓	~	~	✓	✓	~	✓	~	~	Limb	Limb	×	✓	~	~	Limb	Limb	×
Bottom of the atmosphere [~ 10^6 Pa - ~ 10^2 Pa]																		
Temperature profile	~	×	×	×	✓	×	✓	×	×	✓	Night	Night	✓	×	×	✓	×	×
Chemical profile	✓	~	~	×	✓	~	✓	~	~	×	~	×	✓	~	~	✓	×	×

Table 7.4. Overview of the ability of the best retrievals to recover the input values. For species detection, C/O ratio, metallicity (Z), and chemical profiles, it focuses only on the main absorbers. The temperature and chemical profiles are split depending on the region of the atmosphere, where around the highest atmospheric contribution (between ~ 10^2 Pa and ~ 10^0 Pa) the atmosphere is globally well retrieved, contrary to the bottom of the atmosphere (between ~ 10^6 Pa and ~ 10^2 Pa). See Section 7.3.3 and 7.3.4 for more details on the specific biases.

Chapter 8

Proposed Observations

A defining moment in the field has been the recent launch and start of the scientific operations of the James Webb Space Telescope (JWST), coinciding with my "coming of age" as a Ph.D. student in my second year. Soon after, at a specialization course on exoplanet atmospheres[1], we were challenged to write a proposal to observe exoplanets using the JWST, inspired by the observing time awarded in Cycle 1 General Observers (GO). Teamwork and a compelling target were of the essence. I teamed up with fellow student Jiří Žák to take on this challenge, identifying the young multi-planet system TOI-942 as an enticing target. Section 8.1 reports the scientific and technical justification given in Cycle 2 GO. The idea being to leverage the exceptional infrared sensitivity and photometric stability of the JWST to carry out a comparative study of disequilibrium chemistry in TOI-942's two Neptune-like planets early in the era of stellar erosion.

While waiting for the outcome of the referee process, we decided to complement this proposal with ground-based, high-resolution observations with the Very Large Telescope (VLT). Targeting the same system, we submitted two more proposals in European Southern Observatory (ESO) Cycle P112. Section 8.2 details the proposed observation with the Echelle SPectrograph for Rocky Exoplanets and Stable Spectroscopic Observations (ESPRESSO) to constrain the dynamical evolution of the system. The state-of-the-art ESPRESSO is essential to improving current measurements taken with the High Accuracy Radial velocity Planet Searcher for the Northern Hemisphere (HARPS-N). On the other hand, Section 8.3 reports the proposal to use the CRyogenic high-resolution InfraRed Echelle Spectrograph (CRIRES) to study the atmospheric escape of the highly irradiated Neptunes.

* * *

This chapter aims to show that, by strategically interconnecting these cutting-edge facilities from ground and space, we can comprehensively probe the atmospheres and evolution of the TOI-942 planets. Only the CRIRES proposal was scheduled for observation at the time of writing. The JWST proposal was graded in the second quintile, with only the first quintile being awarded observing time. The referee pointed to minor weaknesses related to the current knowledge of the masses and eccentricities while praising the overall scientific merit of the observations and instrument choice. These issues are currently being

[1] Exo-Atmospheres Summer School in Les Houches

addressed using several years' worth of HARPS-N now public data, and an updated proposal is being prepared for Cycle 3 GO at the time of writing. The ESPRESSO proposal was in the second quartile, which we will resubmit targeting a late 2024 transit of TOI-942 c.

During my doctoral work, I have contributed to several additional observing proposals as Co-Investigator: 1 for HST in Cycle 31, 1 for ESPRESSO in Cycle P113, and 4 for JWST in Cycle 3 GO. These proposals are not discussed in this chapter for brevity. However, the experience I gained from these collaborations has been invaluable in shaping my scientific interests and future research directions.

8.1 JWST Cycle 2 GO (PI) A comparative study of disequilibrium chemistry in the atmospheres of very young Neptunes

8.1.1 Scientific Justification

8.1.1.1 Background information

The formation of planetary systems and the origins of life are among the early science motivations (McElwain et al., 2023) of the James Webb Space Telescope (JWST). With the discovery of more than 5000 exoplanets, a deeper understanding of the true nature of exoplanets and their formation-evolution histories is finally within reach.

Planetary atmospheres are unique windows into the interior composition of planets, which holds the record of how and where they formed. At the same time, they connect the planet to its host star and are shaped by the stellar environments. Transiting exoplanets provide detailed access to their atmospheres, as the signals from the planet and the host star can be effectively separated (Sing, 2018). In this respect, multiband photometry and spectroscopy of transiting exoplanets are currently the most promising techniques for characterizing exoplanetary atmospheres (Madhusudhan, 2019). However, studies of atmospheric composition and thermodynamics have been limited in sensitivity and spectral range (Bean et al., 2018; Birkmann et al., 2022). JWST, with its broad spectral coverage and unprecedented sensitivity and stability, is already revolutionizing exoplanetary science (Ahrer et al., 2023; Alderson et al., 2023; Greene et al., 2016; Fu et al., 2022; Rustamkulov et al., 2023; Tsai et al., 2023a). JWST allows detailed studies of the composition and thermodynamics of exo-atmospheres and of fundamental processes shaping the atmospheres such as disequilibrium chemistry, mass loss, and thermal heating (Komacek et al., 2022; Roudier et al., 2021; Showman et al., 2020). Understanding these processes is the key to disentangling the present interaction between the planet and its star, and the formation-evolution history of the system (Barnes, 2010; Owen, 2020; Rogers et al., 2011; Welbanks et al., 2019).

In the early stages of planetary evolution (first ≈100 Myr), processes such as mass loss and tidal heating are most influential (David et al., 2016; Ginzburg and Sari, 2017; Owen, 2020; Valsecchi et al., 2015). They shape the subsequent atmospheric properties and the trends in the observed planetary population (bulk planetary parameters and orbits). Transit spectroscopy of young planets can characterize these processes (Cridland et al., 2016; David et al., 2016), thus providing evolutionary links to more evolved systems (Berezutsky et al., 2022; Kubyshkina et al., 2022a,b). Recently, it has been found that the ratio of C-, O-, and N-bearing species can be used to probe the interiors of planets and their evolutionary

history (Cridland et al., 2019; Moses et al., 2011; Pacetti et al., 2022; Schneider and Bitsch, 2021; Turrini et al., 2021). The ratio of CH_4/CO_2 in particular, because these molecules are strong absorbers in warm gas giants (Fortney et al., 2020). Their presence is mutually linked through the temperature as they are the main atmospheric carbon carriers. For instance, the more abundant CH_4, the more depleted CO_2 is. As the abundance of methane above the clouds is independent of metallicity and depends solely on the thermal structure of the atmosphere, this molecule can be used as an effective "thermometer" (Fortney et al., 2016). Therefore, methane can probe disequilibrium effects that modify the composition and thermal structure of the atmosphere. Moreover, multi-planet systems offer us opportunities to test planetary evolution and migration theories, e.g., the link between migration and composition, or the formation of compact architectures (Choksi and Chiang, 2020; Krissansen-Totton and Fortney, 2022; Martioli et al., 2021; Newton et al., 2021; Owen and Campos Estrada, 2020).

Multi-planet transiting systems around young stars (< 100 Myr) are thus ideal laboratories for obtaining a comprehensive characterization of extrasolar systems (Weiss et al., 2022), and can be used to address the following questions: "What is the chemical abundance of trace gases in the atmospheres of young planets?", "How can the inferred abundances be compared to the predictions of theoretical models?", "What can we learn about their thermal structure?", "What are the properties and origin of clouds/hazes, if present?", and "What can the chemical abundances in multi-planet systems tell us about the differences in the formation pathways of the individual planets?".

8.1.1.2 Program's goals

We propose to perform transmission spectroscopy of two planets in the TOI-942 system to probe the molecular composition and the thermodynamical state of their atmospheres. Moreover, we will perform comparative planetology studies on their characteristics, including their compositional differences, caused solely by the different formation histories.

8.1.1.3 The multi-planet transiting system TOI-942

TOI-942 is a ≈50 Myr old star of spectral type K2.5V that hosts two transiting exo-Neptunes at close-in orbits. The planets were first discovered by TESS and later confirmed by ground-based photometric and spectroscopic observations (Carleo et al., 2021; Zhou et al., 2021; Wirth et al., 2021). Both planets have a mass comparable to that of Neptune and a slightly larger radius. The inner and outer planets have orbital periods of 4.3 and 10.2 days and equilibrium temperatures of about 1010 and 750 K, respectively. Detailed information on the system is presented in Table 8.1.

	TOI-942		TOI-942b	TOI-942c
V mag	11.982 ± 0.026	M_p (M_{Jup})	0.050 ± 0.020	0.084 ± 0.060
J mag	10.231 ± 0.022	R_p (R_{Jup})	0.409 ± 0.011	0.505 ± 0.011
K mag	9.639 ± 0.023	Period (d)	$4.324241 \pm 9.2 \times 10^{-6}$	$10.15622 \pm 1.8 \times 10^{-5}$
M_s ($M_\odot$)	$0.788^{+0.037}_{-0.031}$	a (AU)	0.0472 ± 0.0007	0.0860 ± 0.0015
R_s ($R_\odot$)	$1.022^{+0.018}_{-0.020}$	T_{eq} (K)	1010^{+26}_{-19}	750^{+19}_{-14}
T_{eff} (K)	4928^{+125}_{-85}	Incl. (deg)	$89.994^{+0.050}_{-0.053}$	$89.696^{+1.24}_{-0.65}$
Age (Myr)	50^{+30}_{-20}	Ecc. (deg)	$0.34^{+0.10}_{-0.14}$	$0.32^{+0.23}_{-0.16}$

Table 8.1. Properties of the host star TOI-942 and the exoplanets TOI-942 b and c.

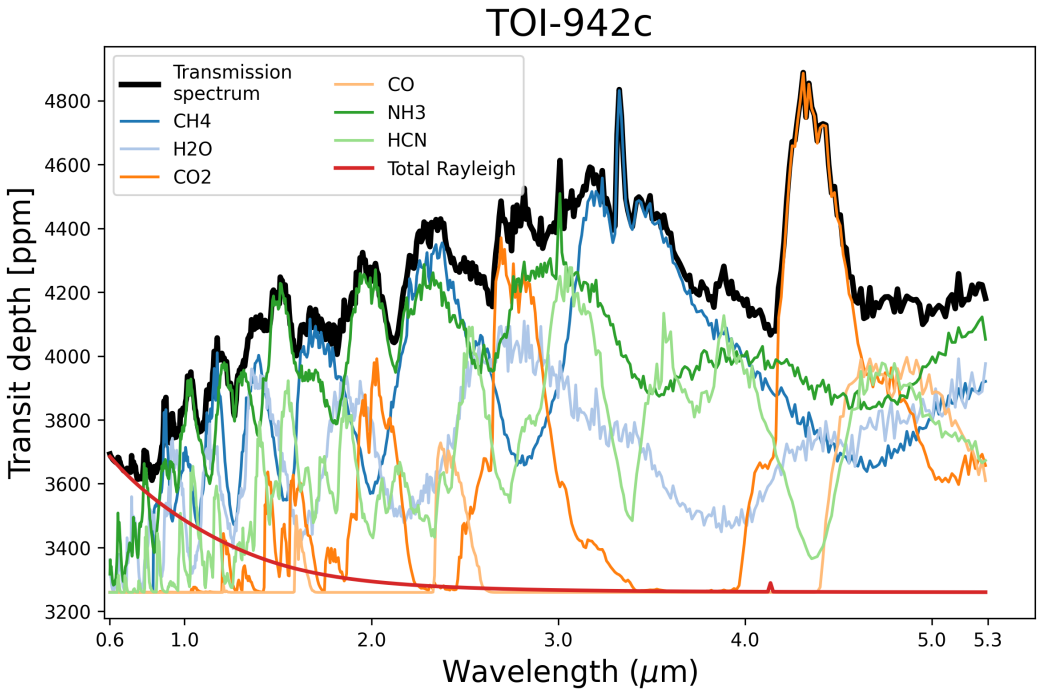


Figure 8.1. Transmission spectrum of TOI-942c with an atmosphere containing constant vertical abundances of the trace gases CH_4, H_2O, CO_2, HCN, CO, and NH_3, each with a mixing ratio of 10^{-4}, with line lists from ExoMol (Tennyson et al., 2016). Absorption contributions from the trace gases are shown individually, along with the total contribution from Rayleigh scattering.

Both planets are excellent targets for transmission spectroscopy observations needed to characterize their atmospheres (Kempton et al., 2018). The molecular signatures of expected major atmospheric gases, such as H_2O, CH_4, NH_3, CO_2, CO, and HCN, are simultaneously observable only with JWST instrumentation (Figure 8.1), which provides the necessary wavelength coverage and sensitivity to constrain their abundances with high confidence (Rigby et al., 2022).

8.1.1.4 Chemical disequilibrium

Often, theoretical models of exo-atmospheres assume a molecular composition given by chemical equilibrium where the detailed molecular abundances depend solely on the temperature, pressure, and elemental composition. However, this is not always the case (Baxter et al., 2021; Miller-Ricci Kempton et al., 2012; Roudier et al., 2021; Tsai et al., 2023a). Chemical disequilibrium is caused by different physical processes such as photolysis from stellar irradiation, vertical mixing, the presence of clouds/hazes, atmospheric escape, tidal heating, and circulation mechanisms. These processes affect pressure-temperature profiles and atmospheric abundances, leading to deviations from chemical equilibrium and a wide range of inferred atmospheric metallicities (Fortney et al., 2020; Moses, 2014). Warm exo-Neptunes can be used to probe disequilibrium chemistry: planets with equilibrium temperatures between 650 and 750 K are expected to have CH_4-dominated atmospheres (Moses et al., 2022). However, observations showed increased abundances of CO and CO_2 and a lack of CH_4 (Benneke et al., 2019), difficult to explain with current atmospheric models.

Missing methane was also reported in the atmosphere of a temperate planet (Tsiaras et al., 2019). Reports of CH_4 depletion have become known as the "missing methane problem". A possible explanation is an increased internal temperature from tidal heating, enhancing the CO and CO_2 abundances at the expense of CH_4 (Fortney et al., 2020).

As listed in Table 8.1, the eccentricities for TOI-942 b and c are 0.34 and 0.32 (Wirth et al., 2021), quite high for such close-in planets. Thus, we can expect tidal heating due to eccentricity damping (Agúndez et al., 2014; Leconte et al., 2010; Millholland et al., 2020) will occur in both planets. For TOI-942b, the ratio between the tidal heat flux and the energy received from the star is about $L_{\rm tide}/L_{\rm irr} \approx 10^{-2}$ while for TOI-942c about $\approx 10^{-3}$. In addition to changes in atmospheric chemistry, tides can significantly inflate the radii of exo-Neptunes when $L_{\rm tide}/L_{\rm irr} \geq 10^{-5}$ (Millholland, 2019). The total positive energy budget of the planet is composed mainly of two contributions: the stellar flux and the internal heat of the planet, which originates from planetary formation and tidal damping (Baraffe et al., 2010; Millholland et al., 2020). Therefore, placing observational constraints on pressure-temperature profiles can link the thermal budget of the highly irradiated outer atmospheric layers and the very interior of the planet.

But how can we measure the effects of disequilibrium chemistry in TOI-942 b and c? Theoretical models (Fortney et al., 2020) that consider disequilibrium chemistry effects in planets with comparable equilibrium temperatures show that the abundances of C-, N-, and O-bearing molecules will have a diverse and complex behavior. For instance, we can use the CH_4/CO_2 and CH_4/CO ratios from detailed transmission spectroscopy of TOI-942 b and c as probes to quantify processes responsible for disequilibrium chemistry (Fortney et al., 2020; Stevenson et al., 2010).

Above the lower, denser layers of the atmospheres of gas giants, generally governed by chemical equilibrium reactions, different disequilibrium processes become important at different altitudes. Vertical transport of material can create quenching, where chemical reactions progressively slow down with altitude until the quenching point is reached and the chemical profiles become constant (Madhusudhan et al., 2016). Using state-of-the-art retrieval codes, e.g. (Al-Refaie et al., 2021; Mukherjee et al., 2023), we will measure the abundances of CH_4, CO, CO_2, H_2O, HCN, NH_3 at varying altitudes and by computing their ratios, we can infer where the quenching point is for different reactions. Another disequilibrium mechanism in still higher layers is photochemistry, caused by high-energy photons from the star that photodissociate various chemical compounds. The incoming stellar flux and spectral distribution strongly depend on the stellar type and age. Hence, for TOI-942, we expect a considerable flux of extreme ultraviolet radiation to impact the upper layers of the atmospheres, driving chemical disequilibrium via photodissociation.

SO_2 from photochemistry was recently inferred in the upper atmosphere of WASP-39b using JWST data (Alderson et al., 2023; Rustamkulov et al., 2023; Tsai et al., 2023a). Given the inverse relationship between planetary mass and bulk metallicity (Thorngren et al., 2016), it is plausible that SO_2 could be present in higher abundances in TOI-942's planets. For exo-Neptunes, theoretical predictions (Guzmán-Mesa et al., 2022) of observable photochemistry products include increased CO and CO_2 abundances at the expense of CH_4. Our results from synthetic retrievals (Figure 8.2) show that we will be able to constrain the abundances of CH_4 and CO_2 to 2 orders of magnitude below the expected equilibrium values. Furthermore, Moses et al. (2011) predicts that HCN is formed via photodissociation of NH_3 and recombination with C atoms from CH_4. In cool to moderately warm gas giants, HCN has extended vertical distribution and high abundances (MacDonald and Madhusudhan,

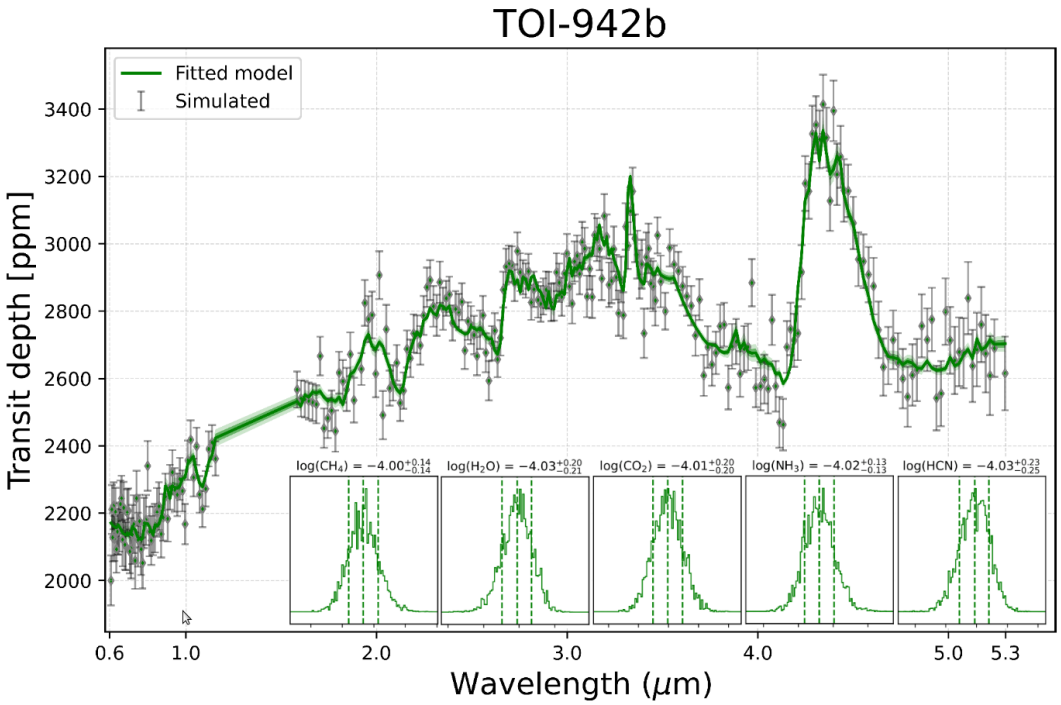


Figure 8.2. Synthetic spectrum of TOI-942b for one transit, with an atmosphere containing the trace gases CH_4, H_2O, CO_2, NH_3, and HCN, each with a vertical mixing ratio of 10^{-4}. The spectrum was binned to R=100, and scattered with errorbars from PandExo (Batalha et al., 2017). The fully saturated region is masked and does not affect the retrieval results significantly. The marginalized posteriors for each molecule show that we can retrieve the input abundances with high confidence (the retrieved values are well within 1-σ from the true values).

2017) and thus could be a candidate for detection on planet c.

8.1.1.5 Aerosols

The presence of aerosols has been inferred in the atmospheres of exoplanets (Chachan et al., 2019; Kreidberg et al., 2014; Wakeford et al., 2017); however, their true nature is still not fully understood (He et al., 2023). Possible candidates include mineral condensate clouds (Helling et al., 2020) and hazes based on sulphur or hydrocarbon precursors (Adams et al., 2019). Both clouds and hazes mute the strength of spectral features. Clouds damp the features more uniformly (Wakeford and Sing, 2015) while hazes produce a detectable slope in the visible to near-infrared part of the transmission spectrum (Helling, 2019). Haze formation is strongly linked to the UV radiation received from the host star (Kawashima and Ikoma, 2019). Kawashima et al. (2019) investigated transmission spectra of warm exo-Neptunes with JWST in the presence of strong hydrocarbon hazes and concluded that several strong molecular features would still be accessible in the near-infrared beyond ≈2 μm. Constantinou et al. (2023) recently emphasized the need for retrievals that include the contribution of aerosols. By obtaining transmission spectra with JWST for both planets, we will characterize hazes and clouds (if present) with high confidence from the spectral slope at short wavelengths (below 1.1 μm) and the strength of the features and compare their properties with theoretical calculations and previous observations in the literature. We

will be able to consistently test models that predict aerosol formation in different stellar environments for exo-Neptunes by directly comparing their predictions to the observations.

8.1.1.6 Distinct evolutionary paths of TOI-942b and TOI-942c

We will perform comparative planetology of the two exo-Neptunes and search for similarities or differences in the formation histories through their composition. For instance, we can address the following questions: "Did both planets form in the same compositional region of the proto-planetary disk?", and "Did TOI-942c form instead farther away, and the current compact architecture is the result of convergent migration of the two planets?". The high eccentricities of TOI-942 b and c (see Table 8.1) indicate that the dynamical history of the system has been significantly more violent than that of the solar system (Carleo et al., 2021). Combined with the fact that their orbital configuration is currently stable, this suggests that the system underwent a phase of dynamical instability in its past. The young age of the system offers the unique opportunity of studying the product of dynamical instability before the interactions with the star alter the atmospheric composition of one of the planets.

Both planets are found within the Neptunian desert (Mazeh et al., 2016; Szabó M et al., 2023) (see Figure 8.3). They are expected to have a detectable mass loss, which is one of the possible shaping mechanisms of the desert (Demangeon et al., 2018). However, due to their different orbital separations, they are predicted to have distinct evolutionary paths. The mass loss of TOI-942b should be strong enough to strip the atmosphere of the planet, which will bring the planet out of the desert, with only the stripped core remaining. In contrast, it should not be sufficient to remove the atmosphere completely on TOI-942c (Kubyshkina et al., 2022b). This scenario can be readily tested as part of this program.

8.1.1.7 Stellar activity

Stellar activity introduces a systematic variation in the measured transit depth of the planets, biasing the inferred atmospheric properties and their physical interpretation. Stellar activity correlates with spectral type and age (France et al., 2016; Soderblom et al., 1991). Our target is an early K-type dwarf; thus, stellar activity is expected (Carleo et al., 2021). An attempt will be made to perform multi-epoch photometric measurements of TOI-942 from ground for each transit. These observations will be used to characterize the stellar activity (e.g., spots and faculae) and to reduce the uncertainty on the stellar SED needed to compute the transmission model. Even without these measurements, it is possible to reliably recover the transmission spectrum within the noise level (McCullough et al., 2014) using an approach such as in (Cracchiolo et al., 2021a,b; Zellem et al., 2017), with the extended wavelength coverage of JWST that allows for a detailed characterization of the host star.

8.1.1.8 Evolutionary links to older systems

Comparison between TOI-942 and other exo-Neptunian systems holds the key to answering important questions, such as: "To what extent can stellar irradiation impact the atmospheres of Neptune-class planets?" and "How does internal heat shape the thermal structure of their atmospheres and the resulting molecular spectral features?". The obtained molecular abundances in the atmosphere of TOI-942b will be directly compared to those of another Neptune, HAT-P-26b, observed in JWST Cycle 1 GO. While TOI-942b and HAT-P-26b have a similar type of host star, planetary mass, and equilibrium temperature, HAT-P-26b has

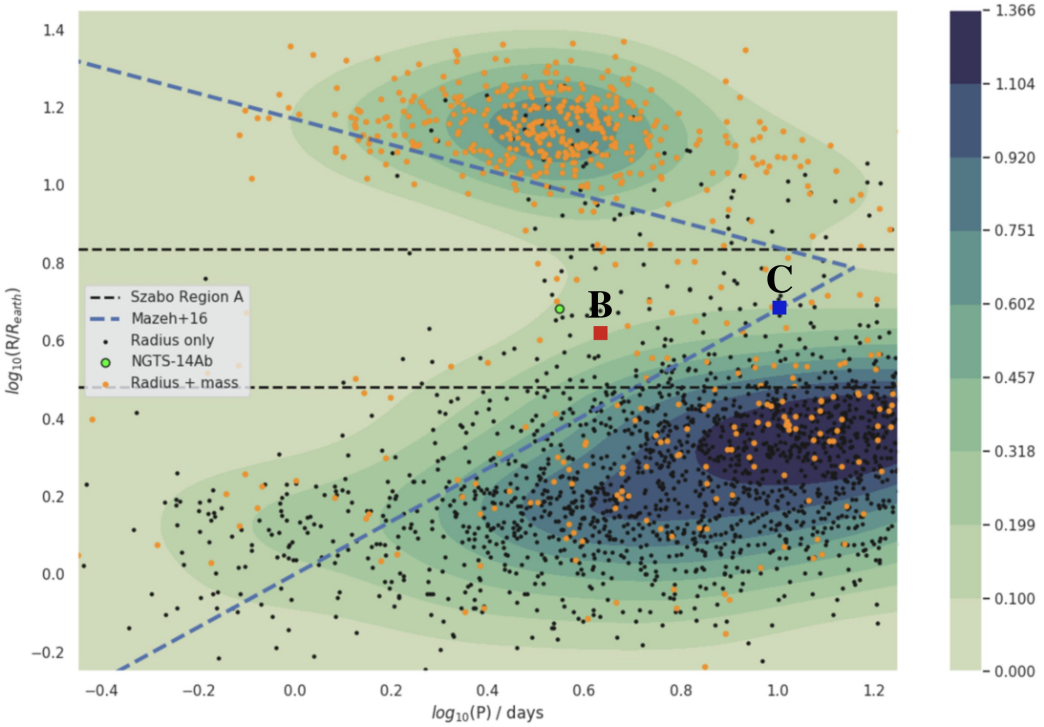


Figure 8.3. Positions of TOI-942 b and c on the diagram showing the logarithm of the period (in days) on the x-axis and the logarithm of the planetary radius (in Earth radii) on the y-axis (figure adapted from (Smith et al., 2021)). Both planets are within the Neptunian desert (Mazeh et al., 2016), the region to the left of the blue dashed lines. The color scale indicates the number of planets per grid element.

a less eccentric orbit and a considerably larger radius. Furthermore, HAT-P-26 is significantly older, at ≈ 9 Gyr. TOI-942b is expected to have a hotter interior, given its younger age (shorter internal cooling) and significant eccentricity (tidal heating). We will test this by comparing the abundance ratios CH_4/CO_2 and CH_4/H_2O between the two systems (Fortney et al., 2020).

8.1.1.9 Program's significance to astronomy in general

"Understanding the connections between stars and the worlds that orbit them" (National Academies of Sciences, 2021) and "[...] the conditions for planet formation and the emergence of life" (ESA, 2015) are science priorities identified by scientists in the USA, EU and worldwide. While thousands of exoplanets have been discovered to date, spanning a huge range of the parameter space in terms of masses, sizes and orbits, there is still no certain link between the presence, size, or orbital parameters of a planet and the stellar environment in which it has formed and evolved. Planets are the end product of star formation, and the understanding of their true nature has become a focal point of the astronomical community. Spectroscopic observations of transiting exoplanets provide us with unique opportunities to link the chemistry of a planet's surface and atmosphere to its formation environment. Moreover, they are essential to understanding whether the stellar type drives the physics and chemistry of the planet's birth and evolution. Characterizing planetary atmospheres will inform us of the planetary composition which, in turn, will tell us about the stellar and disk environment within which it has formed.

8.1.1.10 Program's importance to the specific sub-field of astronomy it addresses

By observing two warm exo-Neptunes in the very young (≈50 Myr) system TOI-942 and obtaining their transmission spectra with JWST, we will characterize their atmospheres in detail, including their chemical abundances and thermal structure. We will focus on disequilibrium effects linked to the stellar environment at an unparalleled age. This information will enable us to test models of the formation and evolution of exoplanetary systems and address open questions on underlying processes shaping the observed population of exo-Neptunes. These observations will lead to a more comprehensive understanding of the Neptunian desert. The retrieved abundances will shed new insight into atmospheric chemistry and the "missing methane problem" by revealing the interplay between CH_4 and CO_2 at a very young age.

8.1.2 Technical Justification

8.1.2.1 Overall experiment design

We will perform transmission spectroscopy of two exoplanets using JWST time series observations with the NIRSpec instrument using the PRISM configuration. For each planet, we will observe one transit and a baseline equal to the transit duration evenly split before and after the transit event to ensure proper analysis and detrending of the light curve.

8.1.2.2 Selection of instrument, modes, exposure times

Given the magnitude of the star (J_{mag} = 10.23) and the need for broad wavelength coverage, we select the NIRSpec PRISM. This setup allows observing the wavelength range from 0.6

to 5.3 μm continuously. This spectral range covers prominent molecular bands (Figure 8.1), as well as strong atomic features such as the Balmer alpha, the potassium doublet and the helium triplet. Our scientific case cannot be satisfied by other facilities, either ground-based (due to telluric contamination) or from space (given the limited wavelength range and sensitivity of other telescopes). The resolution of the NIRSpec PRISM is R≈100. This allows sampling molecular bands with several data points and also detecting strong atomic features (Rustamkulov et al., 2023).

We used the JWST Exposure Time Calculator and the ExoCTK website to finalize our modes and exposure times selection. For NIRSpec, we select the 1.6 x 1.6 fixed slit aperture with the Bright Object Time Series mode optimized for transiting exoplanets, enabling high photometric precision time-series spectroscopy. We have chosen the SUB512 subarray as it provides faster read-out time, saturates fewer pixels and also contains un-illuminated pixels in the trace that may be used as reference pixels to correct for 1/f noise (Rustamkulov et al., 2023; Schlawin et al., 2020).

For both planets, we select five groups per integration to maximize the duty cycle while avoiding unnecessary saturation. We reach an observing efficiency of 79 %. Over the course of the transit plus the time-base out-of-transit, a total of 18370 and 23162 integrations will be taken for TOI-942 b and c, respectively. We acknowledge that a portion of the spectrum will be saturated. We can retrieve the signal from the partially saturated pixels as shown in Rustamkulov et al. (2023); however, our science case is feasible even without it. Our synthetic retrievals (Figure 8.2) show that the NIRSpec PRISM is superior (based on the retrieved properties) to any other setup with the equivalent required time.

8.1.2.3 Justify time-critical observations

Our observations are time-critical as they are exoplanetary transits. The APT is updated with our refined ephemerides from ground-based photometry with 1-m class telescopes. Thus, our transit opportunities have an uncertainty on the time of observations of less than 10 minutes. There are 36 available transits for TOI-942b and 15 for TOI-942c during JWST cycle 2. We have discarded transit opportunities when both planets are transiting.

8.2 VTL/ESPRESSO Cycle P112 (PI)

Unruly Neptunes: constraining the evolution of the very young transiting system TOI-942

8.2.1 Scientific Rationale

The formation and evolution of planetary systems are among the most outstanding and challenging topics in exoplanet research (Winn and Fabrycky, 2015). Many exoplanets have similar observable properties despite having different formation and migration histories (Dawson and Johnson, 2018). By measuring the **spin-orbit alignment** we can infer fundamental aspects related to their **evolutionary history** (Campante et al., 2016). In particular, this parameter reveals valuable clues about the dynamical evolution of multi-planet systems and their interactions with protoplanetary disks or other bodies (Matsakos and Königl, 2017).

The Rossiter-McLaughlin (R-M) effect is a spectroscopic phenomenon that occurs during

exoplanet transits and results in a distortion in the stellar spectral lines. This distortion allows us to measure the sky-projected spin-orbit angle λ between a planet's orbital plane and its host star's equatorial plane (Queloz et al., 2000). There are several techniques to characterize the R-M effect using spectroscopic data. The classical approach is to fit a parametric model to the radial velocity anomaly. This method is widely used, but it suffers from some limitations, such as degeneracies between parameters, and is sensitive to stellar activity. Recent work updates this method to allow to measure the spin-orbit alignment even in the presence of stellar spots or high rotational broadening, the so-called "re-loaded approach" (Cegla et al., 2016, 2018). Another approach to characterize the R-M effect is Doppler tomography (Collier Cameron et al., 2010), which takes advantage of temporal variations of a large number of stellar lines to reconstruct the "Doppler shadow". Measurements of λ reveal a large diversity in orbital orientations, varying from well-aligned to slightly misaligned planets to planets on polar and even retrograde orbits (Albrecht et al., 2012). Furthermore, λ can be used to investigate the exchange of angular momentum between the planet and its host star (Winn et al., 2005). Using λ and an independent measurement of i_s, which is the inclination angle of the stellar spin axis with respect to the line-of-sight, we can derive ψ, which is the true obliquity angle between the planetary orbit and the stellar equator. ψ gives us more information about the dynamical history and evolution of exoplanet systems than λ alone, because it reflects the actual orientation of both axes without projection effects.

Planets emerge from their native circumstellar discs on circular orbits coplanar with the equatorial plane of their stars. Violent dynamical events like planet-planet scattering, resonance breaking, and global phases of chaotic evolution will excite these initially circular and planar orbits, increasing their eccentricities and inclinations. The Normalized Angular Momentum Deficit (NAMD) is a metric that measures the dynamical excitation of a planetary system produced by such violent events during its dynamical history (Turrini et al., 2020), encoded in the planetary eccentricities and inclinations. The NAMD can be used to compare the violence of the dynamical histories of planetary systems with different orbital architectures. One way to constrain the absolute orbital inclinations and thus the NAMD of exoplanet systems is to measure their spin-orbit alignment. However, currently there are very few R-M effect measurements for planets younger than 100 Myr, and even fewer for young multi-planet transiting systems, with observations indicating that all of these planets have orbits with projected obliquity $\lambda < 15°$ (Albrecht et al., 2022).

TOI-942 is a ≈**60 Myr old** star of spectral type K2.5V that hosts **two transiting exo-Neptunes** at close-in orbits. TOI-942b and c were first discovered by TESS and later confirmed by ground-based photometric and spectroscopic observations (Carleo et al., 2021; Zhou et al., 2021). Both planets have a mass comparable to that of Neptune and a slightly larger radius. The two planets have orbital periods of 4.3 and 10.2 days and equilibrium temperatures of about 1010 and 750 K, respectively. Both have significantly eccentric orbits (0.34 and 0.32 for b and c (Wirth et al., 2021)). The inner planet resides within the Neptunian desert and the outer one at its border (Figure 8.4); both planets are predicted to be undergoing significant mass loss with different amplitudes (Kubyshkina et al., 2022a).

TOI-942 offers a unique opportunity to study the dynamical evolution and migration history of close-in planets. The range of possible NAMD values derived for the system using the relative orbital inclinations (Figure 8.5) provides a lower limit to its excitation, but nevertheless reveals that TOI-942 experienced a violent and chaotic past and requires planet-planet interactions to account for its observed state as the planets are not in resonance

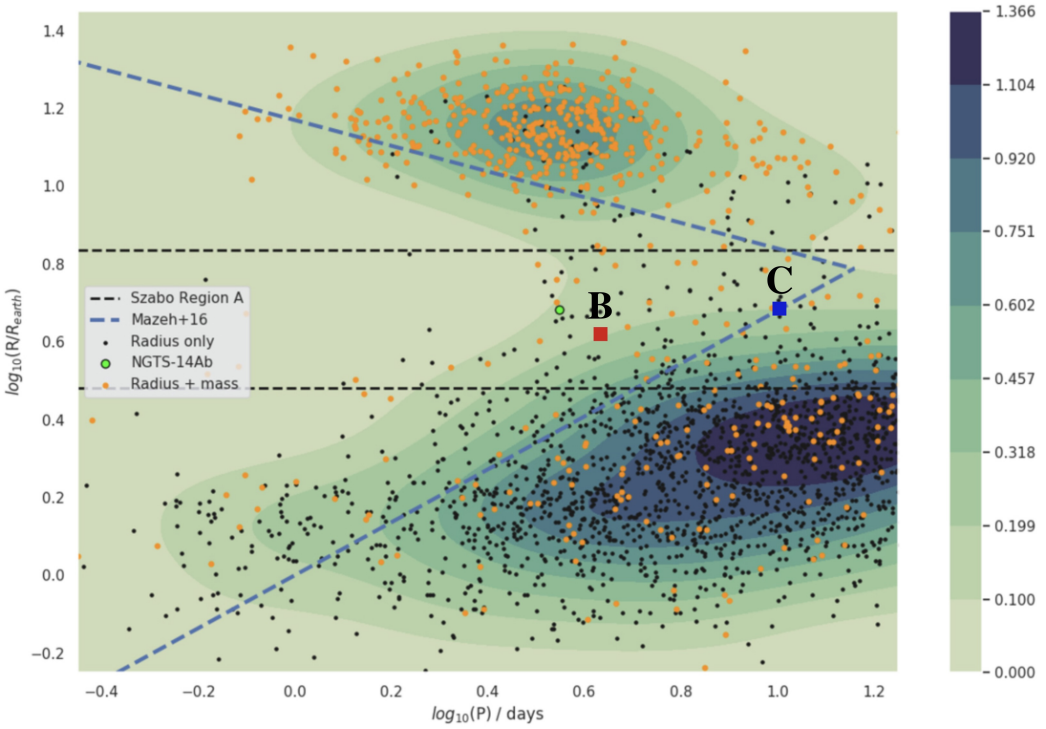


Figure 8.4. Positions of TOI-942b and c on the diagram showing the logarithm of the period (in days) on the x-axis and the logarithm of the planetary radius (in Earth radii) on the y-axis (adapted from (Smith et al., 2021)). The Neptunian desert (Mazeh et al., 2016) is the region inside the blue dashed lines.

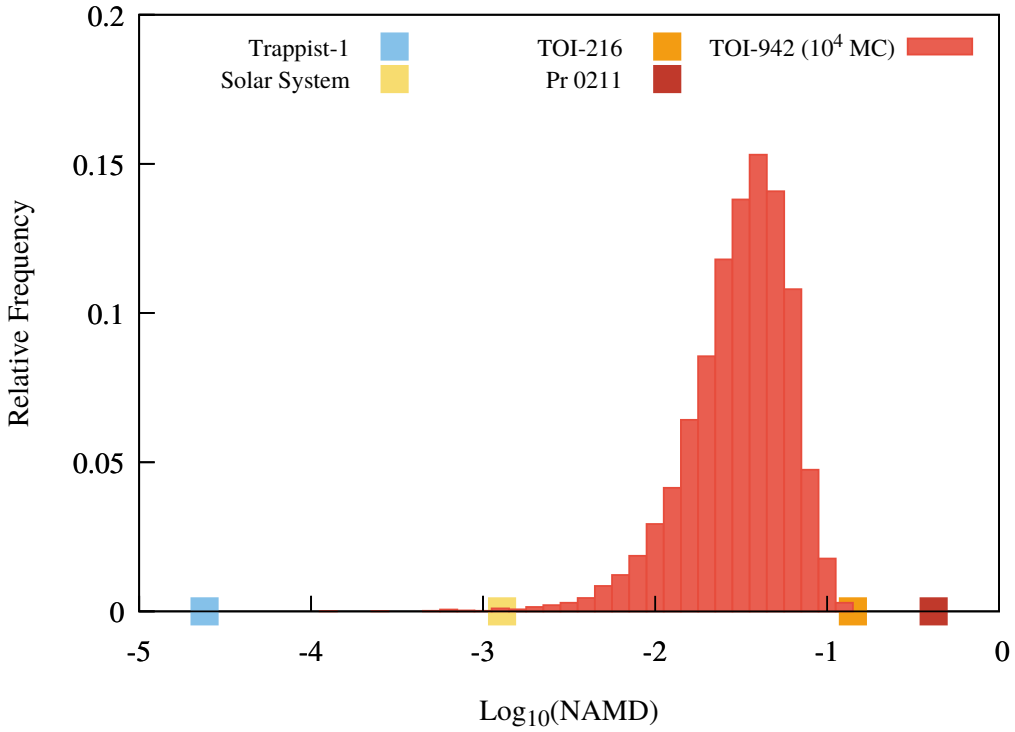


Figure 8.5. NAMD lognormal distribution of TOI-942 computed varying the orbital and physical parameters of the two planets within their confidence intervals (Carleo et al., 2021). Also shown are the NAMD values of the solar system, Trappist-1, TOI-216 and Pr 0211 for comparison. TOI-216 has the highest NAMD among the compact systems, Pr 0211 has the highest NAMD among all analyzed multi-planet systems. The larger the NAMD, the more violent has been the dynamical history of the system. Increasing the obliquity of TOI-942c would shift the NAMD of the system toward that of TOI-216.

and have eccentric orbits (Carleo et al., 2021). However, the spin-orbit alignment of both planets is crucial for quantifying how violent this history was and for understanding how these events occurred. (Wirth et al., 2021) measured the obliquity of the inner planet using two transits and found a projected obliquity $\lambda = 1^{+41}_{-33}$ deg and $\psi = 2^{+27}_{-23}$ deg. This implies that the events that shaped the orbits of the two planets were more effective in the planar direction than in the vertical one, which is consistent with the low mutual inclinations of the two planets (Carleo et al., 2021). By measuring the obliquity of the outer planet, we can unravel the history of the system and the migration of the planets, and determine whether they formed in-situ or farther from the star and whether their compact configuration is due to convergent high-eccentricity migration. Moreover, a misaligned orbit of TOI-942c could indicate a hidden perturber on an outer orbit. These migration mechanisms are still poorly understood, especially given the scarcity of R-M effect measurements of young systems: our observations would add to the precious few R-M effect measurements for a multi-planet transiting system of 100 Myr or younger. In addition, this would enable us to constrain the ongoing migration and test the predictions of migration theories for multi-planet systems at such a young age. Finally, given the constraints on their mutual inclinations and the

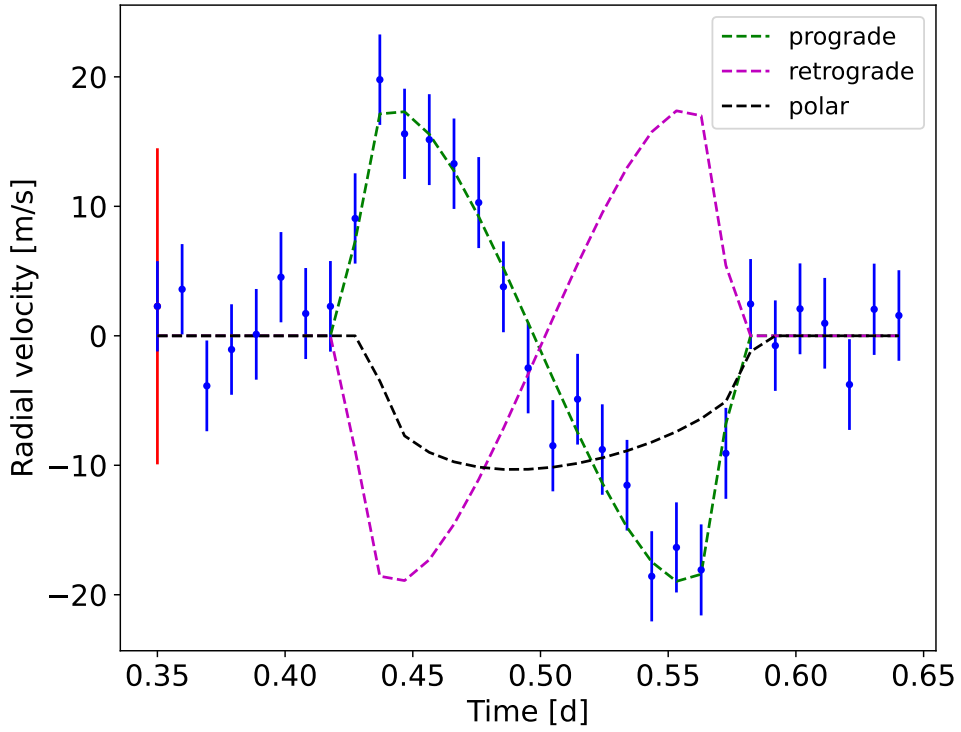


Figure 8.6. Simulated R-M effect for TOI-942c using the formulation by (Hirano et al., 2011). Simulated ESPRESSO data points for prograde orbit are shown in blue. For comparison, the much larger green errorbar is the mean uncertainty taken from 1800s HARPS-N exposures (Carleo et al., 2021). Distinct possible alignments of the orbit are shown in colored dashed lines.

obliquity of the outer planet, we will be able to reduce the uncertainty on the obliquity of the inner planet as well. This will allow to **constrain the three-dimensional geometry of the system and its implications for planet formation and evolution**.

8.2.2 Immediate Objective

We propose to use the state-of-the-art ESPRESSO instrument to perform high-resolution spectroscopic measurements of TOI-942c during a transit event. Our immediate objectives are:

1. To measure the Rossiter-McLaughlin effect and constrain the spin-orbit alignment of TOI-942c (Figure 8.6).

2. To investigate the migration history and dynamical excitation of the very young (≈60 Myr) multi-planet system and test various scenarios for its formation and architecture.

We will employ both the classical and the "re-loaded" approach to characterize the R-M effect, as well as Doppler tomography to rule out any biases in the analysis. We will use indicators from the spectra that correlate directly with the red noise (increased

by the stellar activity) to model it and detrend the RV time series (Huerta et al., 2008). Furthermore, our team has access to 1-m class telescopes able to capture a simultaneous transit to independently assess the level of stellar activity. The high stellar activity and the presence of spots limited previous R-M effect observations, which constrained the obliquity of the inner planet to within 50 deg of its host star's plane of rotation (Wirth et al., 2021). ESPRESSO provides the stability and precision that are essential for this science case, **allowing us to detect even a slight misalignment** in the orbit of the outer planet (our estimated uncertainty on the spin-orbit alignment is less than 10 deg). This will shed light on whether TOI-942c has experienced past interactions with the inner planet (low-obliquity scenario) or whether it interacted with a larger, currently undiscovered body (high-obliquity scenario). As a by-product of these observations, we will also search for strong optical absorbers (sodium, hydrogen, potassium) to characterize the atmosphere of TOI-942c and compare it with older Neptune-like planets.

8.3 VTL/CRIRES Cycle P112 (Co-PI) Vanishing Worlds: Comparative Study of Atmospheric Mass Loss of Two Very Young Neptunes

8.3.1 Scientific Rationale

Under the action of X-ray and ultraviolet irradiation from their host stars, exoplanets can lose part of their atmosphere. Such atmospheric evaporation significantly influences planetary evolution, including the early evolution on Mars and other planets in the solar system (Lammer et al., 2008). Photo-evaporation may also explain the population of observed exoplanets. In particular, evaporation is suspected to be the cause of the "Neptune desert", a region close to the host stars which presents a deficit in Neptunian exoplanets (Mazeh et al., 2016). The radii of these planets lie between the terrestrial planets that are believed to form in-situ (Matsumoto and Kokubo, 2017) and the giant planets that are thought to form outside the snow line before migrating inwards (Armitage and Rice, 2005). **Neptunes in close orbits are proposed to form helium-dominated atmospheres through hydrodynamic escape of hydrogen** (Hu et al., 2015), however current observations suggest that at least the "escaping atmosphere" has a higher than expected H/He ratio. The atmospheric characterization of planets undergoing hydrodynamic escape provides information about the planetary properties and the escaping mechanism, and hints about their formation history and evolutionary pathways (Lampón et al., 2023).

Escaping atmospheres of transiting planets were previously studied using the Lyman-alpha line (Vidal-Madjar et al., 2003). However, this line is heavily affected by interstellar absorption and geocoronal emission. Furthermore, due to its location in the UV region, there is a very limited number of instruments capable of investigating this line.

The presence of helium in an evaporating atmosphere was predicted by theory (Seager and Sasselov, 2000), and the metastable helium line at 1083 nm was soon identified as a suitable indicator of atmospheric escape (Oklopčić and Hirata, 2018; Spake et al., 2018). Not all escaping atmospheres have a helium signature amenable for detection, as the excitation of helium atoms is more favorable at small orbital separations and around stars with higher extreme-ultraviolet flux (which ionizes the helium ground state), and lower mid-ultraviolet flux (which ionizes the helium metastable state (Oklopčić, 2019)). As can be

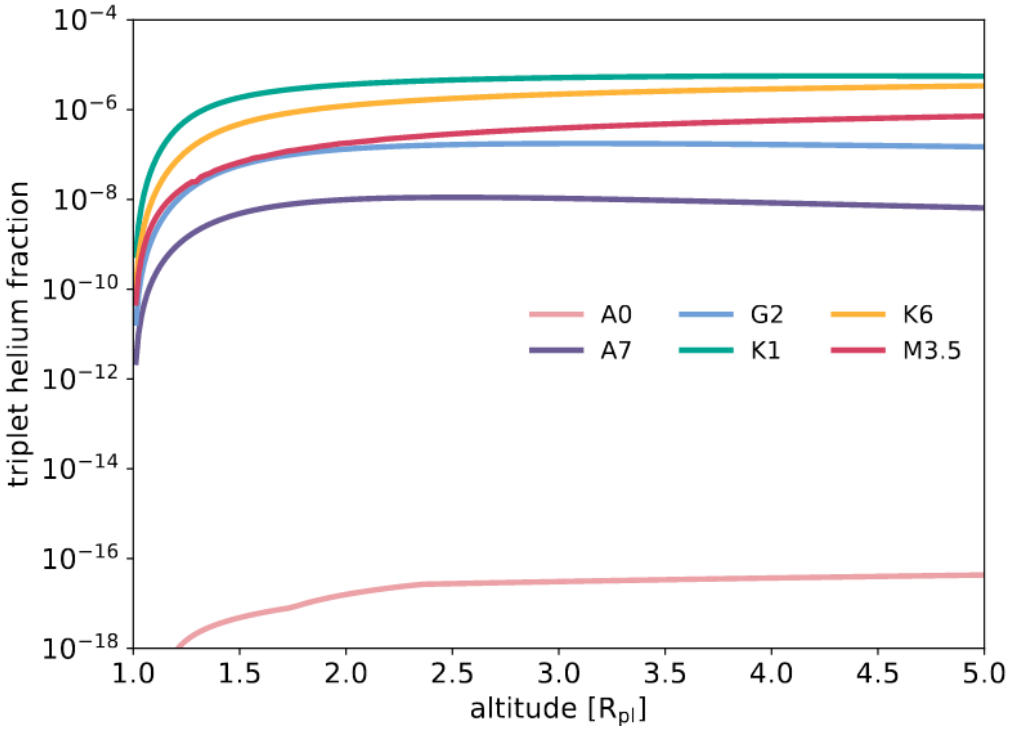


Figure 8.7. Properties of the host star play a significant role in populating the metastable triplet state. The fraction of helium (relative to all He atoms/ions) in a planetary upper atmosphere that populates the excited, metastable triplet state, is shown as a function of altitude (in planetary radii). Different curves represent exoatmospheres irradiated with stellar SEDs of different spectral types. Higher fraction of triplet helium in a planetary atmosphere produces stronger transit absorption in the 1083 nm line (Oklopčić, 2019).

seen in Figure 8.7, planets orbiting K-dwarfs have well-populated metastable triplet states and are thus most suitable for detection studies.

Large photometric surveys such as TESS, WASP, and KELT have detected excellent candidates for atmospheric studies of evaporating exoplanets. Such transiting planets are amenable to high-resolution transmission spectroscopy, which analyses the imprint of the planetary atmosphere as the planet transits in front of the star. Spectra are obtained before and after the transit to create a high SNR stellar template. A series of high-cadence spectra is obtained during the transit to resolve the moving signal originating in the planetary atmosphere. Subsequently, the in-transit data are compared to the template to extract the planetary signal. This method was used by (Wyttenbach et al., 2015) and (Allart et al., 2018), among others, to detect the presence of strong atomic absorbers. This technique was also successful recently in detecting helium ((Nortmann et al., 2018) – see Figure 8.8, (Kirk et al., 2020)).

TOI-942 is a K2.5V star with an age of ≈**60 Myr** that hosts **two transiting Neptunes** with close-in orbits, as discovered by TESS and later confirmed by ground-based radial velocity and transit follow-up observations (Carleo et al., 2021; Zhou et al., 2021). The inner and outer planets have masses of 0.05 and 0.08 M_{Jup}, radii of 0.41 and 0.51 R_{Jup}, orbital periods of 4.3 and 10.2 days and equilibrium temperatures of about 1010 and 750 K,

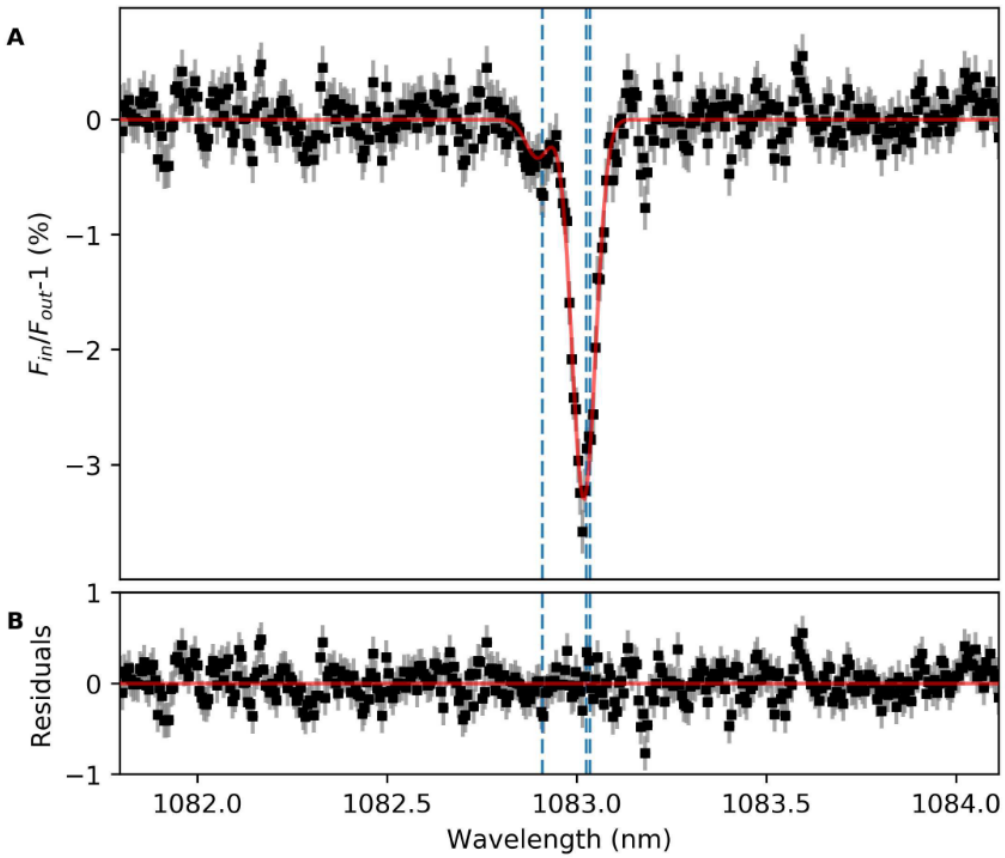


Figure 8.8. The transmission spectrum of WASP-69b observed by CARMENES. The predicted positions of the helium triplet lines are marked by blue vertical dashed lines. The best-fit model (red) shows a net blue shift, which indicates an escaping atmosphere (Nortmann et al., 2018).

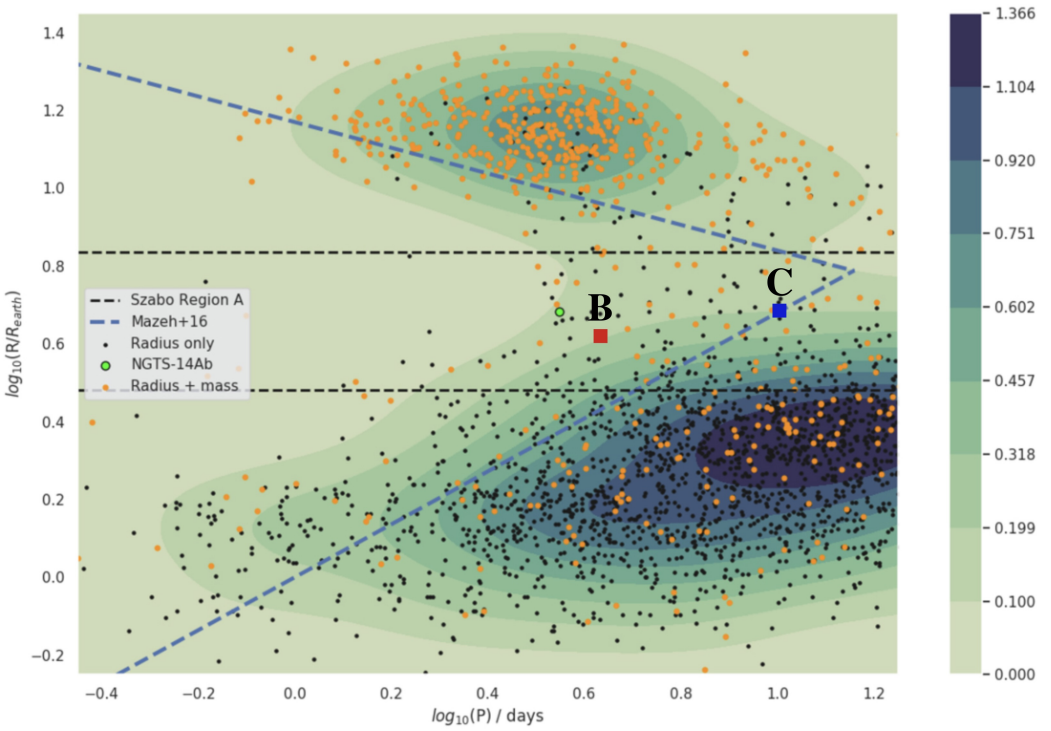


Figure 8.9. Positions of TOI-942b and c with the logarithm of the period on the x-axis and the logarithm of the planetary radius (in Earth radii) on the y-axis (figure adapted from (Smith et al., 2021)). The Neptunian desert (Mazeh et al., 2016) is the region inside the blue dashed lines. The color scale indicates the number of planets per grid element.

respectively. The physical parameters place TOI-942b inside the Neptunian desert while TOI-942c is located just on the border of the desert (Figure 8.9). This suggests that TOI-942b is undergoing **significant atmospheric escape** due to the high XUV flux from its host star. The detection of atmospheric escape signatures around TOI-942b would support this scenario. Moreover, TOI-942c is also expected to exhibit a detectable mass loss rate due to its proximity to the star (Kubyshkina et al., 2022b). As TOI-942 is one of the youngest known multi-planet systems, this study would provide the first evidence of atmospheric evaporation at such an early stage for both planets. Furthermore, this would also mark **the first detection of atmospheric escape on two planets in the same system** and allow us to test theoretical models of mass loss as a function of orbital distance and planetary properties within the same stellar environment.

The mass loss rate of TOI-942b (Figure 8.10) is predicted to be $\sim 1.2 \times 10^{12}$ g s^{-1} (Kubyshkina et al., 2022b) – about 10 times higher than the similarly sized, but significantly older planet GJ 3470b, for which a strong helium detection was reported (Palle et al., 2020). Our detection will be able to challenge hydrodynamical models at an unprecedented age. TOI-942b will likely lose most of its atmosphere and turn into a super-Earth-like planet of half its current size and will reside on the lower boundary of the Neptunian desert.

The mass loss of TOI-942c is predicted to be $\sim 3.5 \times 10^{11}$ g s^{-1} (Kubyshkina et al., 2022b) – about 3 times higher than the mass loss of GJ 3470b. However, the scale of the mass loss will not be enough to completely strip the atmosphere in the future as for the

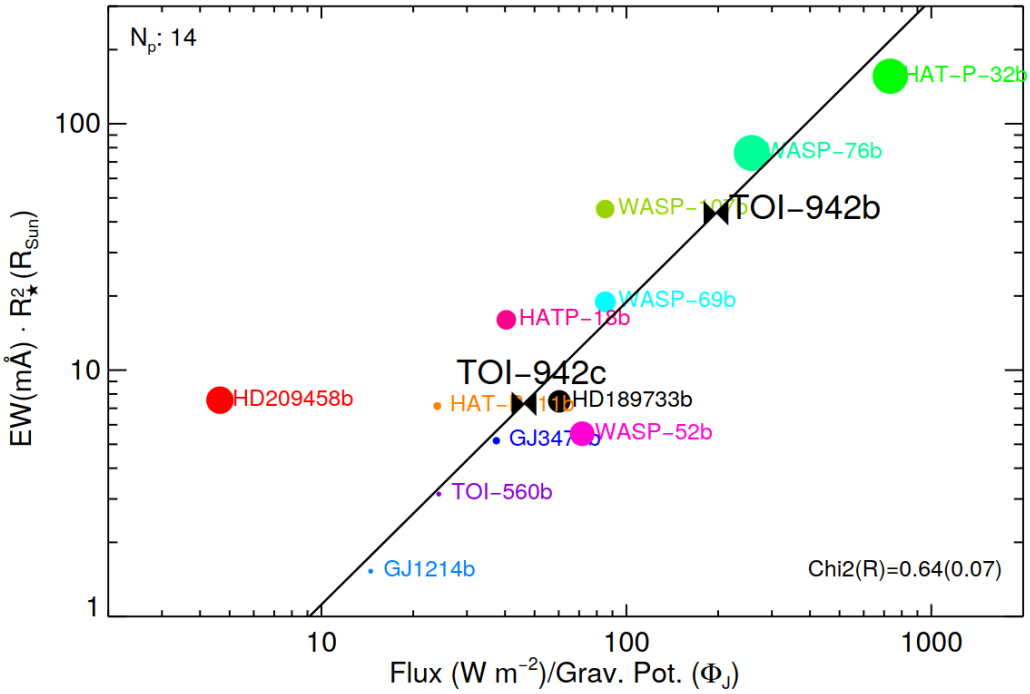


Figure 8.10. Estimated equivalent width of the helium triplet scaled with a stellar radius for TOI-942b and c based on empirical relationship we found for the planets with the observed helium triplet absorption.

inner planet; hence, the two planets are expected to end up on two distinct evolutionary paths due to the atmospheric escape. Detection of significant mass loss could be used to test whether **photoevaporation is the driving mechanism** in shaping atmospheric evolution for these planets.

8.3.2 Immediate Objective

We will observe single transits of TOI-942b and TOI-942c with CRIRES ($\mathcal{R} \sim 40,000$), covering the spectral range from 950 to 1100 nm (spectral orders 51 to 59). This region contains the metastable **helium line** which is present in evaporating atmospheres and **which we aim to detect**.

CRIRES's high-spectral resolution and the large collecting capability of the VLT will allow both spectral and temporal resolution of the helium signature via transmission spectroscopy. Its detection will allow us to constrain **mass loss**, planetary winds, and possible pre- or post-transit absorption hinting at a leading/trailing atmosphere. We will use 1D hydrodynamic spherically symmetric models coupled with a non-local thermodynamic model to analyze the He I triplet state (Lampón et al., 2020). In case of non-detection, we will still be able to set strong limits on the planetary thermospheric temperature and mass loss rate which will put tight constraints on the future evolution of the planets. We will be able to place observations of other exo-Neptunes with escaping atmospheres in a broader context and test theoretical models at a **yet unexplored age regime**. Mass loss detection would also be in synergy with a submitted JWST cycle 2 proposal to investigate the atmospheric composition of TOI-942b and c as mass loss is known to alter the composition of planetary

atmospheres (Hu et al., 2015).

Thanks to CRIRES's high spectral resolution and stability, we will be able to clearly disentangle the faster varying planetary signal from the stellar activity signal. This has been a major concern for previous studies that employed low-resolution spectroscopy. Moreover, CRIRES will enable us to study atmospheric details inaccessible to the low-resolution method and retrieve multiple parameters from the resolved helium signature, such as atmospheric winds, rotation, thermospheric temperature, H/He ratio, heating efficiency of the upper atmosphere, and the hydrodynamic escape regime(Lampón et al., 2020, 2021).

Chapter 9

Conclusions and Future Perspectives

The tumultuous growth of the relatively new field of exoplanets - planets outside our solar system - set the stage for this thesis (Section 1.1). The confirmed existence of thousands of such celestial bodies opened the door to numerous questions on their true nature and to make sense of the peculiarities of our own planetary system in a suddenly huge parameter space. Whilst some underlying trends behind the observed diversity of exoplanets have found convincing explanations, observations are still in the early phase and significant selection biases exist. The observation of a statistically significant sample of exoplanets has become a major goal in the community. Various observing techniques are available and can give complementary cues to address the hottest questions, such as "What are the main physical processes that shape planetary atmospheres?", "What are exoplanets made of?", "How do planets and planetary systems form and evolve?", and even "What are the conditions for the emergence of life?". Among the various techniques, transit spectroscopy is among the most promising, and it has been successfully utilized to obtain the first atmospheric characterization of exoplanets (Section 1.1.5). With transit spectroscopy, we can measure the wavelength-dependent modulation of the stellar flux as the planet transits in front of the star, and we can use this information to infer the composition and structure of the planet's atmosphere. So far, a small sample of exoplanets has measured spectra, with limited spectral coverage and a strong bias toward hot Jupiters. New ground and space-based observatories have recently come online, most notably the JWST. While the JWST is undoubtedly the most sensitive space telescope of the next decade, it is not intended to conduct a survey of the exoplanet population as a whole. Instead, the Atmospheric Remote-Sensing Infrared Exoplanet Large-survey (Ariel) - to be launched in 2029 - is the first space telescope entirely dedicated to conducting the first spectroscopic survey of a large and diverse sample of hundreds of exoplanets (Section 1.2).

Ariel is uniquely equipped to perform simultaneous observations over the full 0.5 to 7.8 μm spectral band, and an optimized observing strategy will ensure maximum target coverage while addressing layered questions from the population to individual planets. To extract the information content of the spectroscopic data, retrieval tools such as TauREx 3 have been developed that given an observation return the most probable set of atmospheric parameters with an associated confidence interval. Ariel integrates these tools with simulators of the expected experimental uncertainties and systematics to produce representative spectra and learn about their information content ahead of the mission launch. Among

these simulators are ArielRad (Section 1.3.1) and ExoSim 2 (Section 1.3.3). ArielRad is the Ariel Radiometric Model that, given the description of the payload and the target, returns the estimates for the measured signal and the noise. ArielRad is the cornerstone to perform detailed investigations of the payload and the target selection observable by Ariel. ExoSim 2, on the other hand, is the time-domain simulator of a target observation, and given the detailed payload model, can assess time-varying noise and systematics with the highest fidelity. It constitutes a formidable tool to produce realistic observations and prepare the data reduction pipeline.

In the three years of my Ph.D. research, I have navigated deep into this field with a focus on the link between instrumentation and observation. This leitmotiv has led me to develop a new simulator, and a fresh mind about how to interpret the data, ultimately bridging the two communities. The following sections draw the curtain and offer a review of my technical and scientific contributions and a glimpse of future work in this ever-fascinating field.

9.1 Technical Contributions

A major contribution to the software side is the development of the Physical Optics Simulator (PAOS). Accurate assessment of the optical performance of advanced telescopes and imaging systems is essential to achieve an optimal balance between optical quality, system complexity, costs, and risks. PAOS is an open-source code that implements Physical Optics Propagation (POP) in Fresnel approximation and paraxial ray tracing to analyze complex waveform propagation through both generic and off-axes optical systems, enabling the generation of realistic Point Spread Functions (PSFs) across various wavelengths and focal planes. Developed using a Python 3 stack, PAOS includes an installer, documented examples, and a comprehensive guide. It addresses limitations in other POP codes, offering extensive customization options and the liberty to access, utilize, and adapt the software library to the user's application. With a generic input system and a built-in Graphical User Interface (GUI), PAOS ensures seamless user interaction and facilitates simulations. The versatility of PAOS enables its application to a wide array of optical systems, extending beyond its initial use case. PAOS presents a fast, modern, and reliable POP simulation tool for the scientific community, enhancing the assessment of optical performance in various optical systems and making advanced simulations more accessible and user-friendly.

In this thesis, I have included reports on different analyses using PAOS to investigate several aspects related to Ariel's optical performance. Specifically, to ensure that Ariel achieves (i) sufficiently compact PSFs to operate as light-bucket and (ii) sufficiently sampled PSFs to mitigate Line of Sight (LoS) pointing jitter, as per design. Concerning the former, Section 2.3 reports on the analysis of the maximum amplitude of wavefront aberrations that are compatible with the top-level performance requirement on optical quality. The results of this analysis, done from first principles in optics, were validated with PAOS, and are currently being updated to further support the activities related to the manufacturing of the Primary Mirror (M1). With regard to the PSF sampling analysis, in the study summarized in Section 4.1, I investigate the PSF sampling across the Ariel focal planes. Critically sampled focal planes are key to mitigating systematic uncertainties, such as photometric noise from pointing jitter. The PSFs produced with PAOS were imported in ExoSim 2 and used to produce realistic observations of the Ariel Fine Guidance System (FGS) focal planes.

These simulations were used in the payload performance analysis for the Preliminary

Design Review (PDR). I contributed to this analysis, and the results are reported in Chapter 3, detailing the current Ariel performance parameters and margins, optical efficiency, and noise budget. Using ArielRad, we demonstrated that the mission can perform the observations of the Mission Reference Sample 2019 (MRS19), a list of 1000 candidate targets presented in Edwards et al. (2019), during its 4-year nominal lifetime. Moreover, observations of 557 targets in the *deep survey* tier can be completed, 57 more than the required 500. Finally, in this chapter, I included a detail-rich report on the photometric stability that can be achieved by the mission, presented at the payload PDR. Photometric stability is a critical aspect of the mission, and it is a key driver for the mission's scientific performance. A result of this analysis was a small non-compliance with the jitter noise requirement for bright targets from 1.95-2.1 μm in Ch0 of the Ariel Infra-Red Spectrometer (AIRS). This non-compliance does not impact overall science capabilities and was expected to be resolved with improved detrending techniques.

This has been confirmed by the analysis presented in Section 4.2, where I led the effort by the Simulators Software, Management, and Documentation (S2MD) working group to develop a more effective detrending algorithm. This algorithm was tested on simulated photometric data with ExoSim 2 for constant and time-varying signals, using a representative jitter timeline provided by Airbus. The decorrelation is based on the assumption that the first few moments of light distribution along both spatial axes are sufficient to capture the spatial information of the PSF modulated by the jitter. The results show that, even absent calibration information, the algorithm can effectively decorrelate the jitter signal, with a residual jitter noise compliant with the requirement. In addition, the main transit parameters such as the planetary radius can be retrieved with good accuracy, although small discrepancies are observed between the measured planetary radius and the input value that need to be further investigated. The algorithm is now being applied across the entire Ariel wavelength range, with results to be included in a future update of the payload performance analysis as well as in a dedicated publication.

Other technical work I conducted during the course of my Ph.D. includes a vignetting analysis of the AIRS PSFs in the presence of pointing jitter, which was instrumental in re-designing the size of the AIRS field stops to avoid cutting signal frequencies. However, this work was not included here for length constraints. Other notable analyses not covered here are a study to support the Radius of Encircled Energy (rEE) requirement for the Ariel Optical Ground Support Equipment (OGSE), a validation with PAOS of Ariel PSFs using Structural, Thermal, and Optical Performance (STOP) analysis inputs, and an investigation of astrophysical sources for flat fielding including zodiacal light and solar system planets.

9.2 Scientific Contributions

On the science side, I have been involved in a series of studies of simulated transmission spectra. An important focus of investigation in preparation for the Ariel mission is the expected information content of the spectra collected during the mission's *reconnaissance survey*. This unprecedented survey of hundreds of transit and eclipse spectra aims to address population-level questions and guide further observation in higher tiers for detailed characterization. While rich in information content that can be exploited, spectra collected in the first tier of the mission are not designed to enable high-confidence spectral retrieval of atmospheric parameters. Instead, alternative techniques to exploit the spectra are being

developed, and Chapter 5 presents a systematic study of the full MRS19 population simulated with gaseous atmospheres observed in transit. Several techniques are successfully introduced that enable the extraction of information from the spectra, including (i) a χ^2 metric based on band averages to identify planets with no above-noise spectral features, (ii) a model-independent molecular metric comparing different selected spectral bands to a normalization band to predict the atmospheric composition, and (iii) a suite of four different machine learning methods that can successfully predict whether a molecule is present in a majority of cases. These techniques are not without bias, and I led further investigations focused on determining whether an atmospheric molecule is present with an associated calibrated probability.

This study is reported in Chapter 6, where we introduce a *P*-statistic that utilizes abundance posteriors from spectral retrieval to infer the probability of a molecule's presence in a given planet's atmosphere in Tier 1. We employ calibration and Receiver operating characteristic (ROC) curves to assess the reliability of the P-statistic on our target population under different atmospheric models assumed in the retrievals. This analysis demonstrates that the P-statistic predicts probabilities that correlate well with input abundances when retrieval models have comparable or higher complexity than the data. However, the P-statistic's representativity declines when the retrieval model's complexity is lower, expressed as the inclusion of fewer than the expected molecules. We study forecasting biases and find them not to adversely affect the classification of the survey. The P-statistic can effectively identify interesting atmospheres, detect molecular presence, and guide the promotion of targets from Tier 1 to higher Tiers. In addition, we discuss how to quantitatively assess the completeness of the Tier 1 sample, an aspect that requires further investigation. This information can be used to infer population-wide properties such as the fraction of planets that bear a molecule, which can inform theoretical models of exoplanetary formation and evolution, capitalizing on the reconnaissance survey observations. More work has been initiated to further develop the P-statistic and extend it to rocky planets and secondary atmospheres, and we plan to present the results in a future publication.

Another study I co-authored during my doctoral work is reported in Chapter 7. In this chapter, we investigate how the 3D atmospheric structure biases the interpretation of exoplanet transmission spectra using 1D retrieval models. Simulated observations with JWST and Ariel are generated for three exoplanets (GJ1214b, HD189733b, WASP-121b) spanning a range of temperatures. Retrievals are performed on the simulated data to assess biases in retrieved parameters compared to the known input models. We find several important biases arise from using 1D retrieval models on 3D atmospheres. First, temperatures are only accurately retrieved at pressures corresponding to the highest atmospheric contribution. Outside this region, temperatures can be biased by hundreds of K. This indicates a limitation in how well 1D models can characterize the full thermal structure. Second, with equilibrium chemistry, the abundances of species with strong vertical variations are inaccurate. This goes on to bias all correlated parameters derived from these abundances. One way to mitigate this issue is by retrieving different spectral features separately, which helps disentangle the effects of temperature variation. Third, different equilibrium chemistry models can give significantly different results when applied to the same data. In some cases, simpler models actually outperform more complex ones, especially when the input assumptions of the complex model are imperfect. This demonstrates the challenges in selecting appropriate chemistry models for retrieval. Fourth, higher spectral resolution (e.g. with JWST) does not necessarily provide better constraints compared to lower resolution (e.g. Ariel). This is

due to the biases mentioned above that affect the interpretation of extremely high-quality observations with JWST, indicating that more complex retrieval models are needed. Finally, uncertainties from 1D retrievals are underestimated and should not be trusted, especially for hotter exoplanets where biases are more severe. In summary, through this work, we demonstrate that 1D retrieval assumptions can lead to significantly biased interpretations of 3D atmospheres. While improved 3D modeling is needed, caution must be exercised when deriving physical parameters from transmission spectra, especially with high-quality observations.

9.3 Future Directions

During my Ph.D., I have been involved in several studies that have led to a series of publications. I have been conducting research in a field that is still in its early days, and I have been fortunate to be part of a team and a mission that will shape the future of exoplanetary science. However, there is still much work to be done, and I have identified several areas that I believe are worth pursuing. In the following, I will briefly discuss some of these areas and how my work can inform future studies in this field.

First, I feel that there are still under-explored synergies between ground and space-based observatories. The two communities have been working primarily in parallel for a long time, and there is a lot of potential for further collaboration. I have been submitting observing proposals to ground-based facilities to complement the observation of an intriguing system from JWST (see Chapter 8). While I have been successful in obtaining observing time for one of the ground-based instruments I have been requesting, I still have to obtain the JWST time. This is a lengthy process, but I believe that once the space and ground-based observations are completed, the insights from combining the two datasets will be worth it.

Second, synergies between space missions (existing and proposed) also have yet untapped potential. I have been working on a paper to explore synergies between Ariel and Twinkle, a small satellite mission with expected launch in 2025. Twinkle will be able to observe a large number of targets in the near-infrared, and it is expected to provide a first glimpse of the atmospheric composition of some of the Ariel targets. Therefore, accurate simulations of spectra obtained with Twinkle will be essential to guide the selection of Twinkle targets in a synergistic way with Ariel.

Finally, and most importantly, some results obtained so far by the community using brute-force approaches in atmospheric studies would likely benefit from analytical approaches to improve interpretability. I have identified a formalism that I believe can be used to this end. The Fisher information formalism is a well-known mathematical-statistical framework that can be used to assess the information content of observations before the data is ever taken. It was first employed in the field of cosmology and some recent studies have begun using it for exoplanets, however not to characterize the information content of atmospheric spectra. During the last part of my Ph.D., I have been working on a publication to leverage this powerful approach in atmospheric studies. I hope it will become standard practice to use this analytical formalism alongside spectral retrieval tools and machine learning techniques to interpret the data and prepare for future missions and observations.

Appendices

Appendix A

Bayesian Inference

In this appendix, we provide a brief introduction to Bayesian inference, and we describe the Bayesian inference methods used in this thesis. Bayesian inference provides a framework to solve the inverse problem of inferring the model parameters given some data. Figure A.1 illustrates the difference between the direct and the inverse problems. The direct problem is usually straightforward to solve. However, the inverse problem is usually ill-posed, and it may have multiple solutions, requiring a statistical approach to select the best model that describes the data.

Bayesian inference has grown in popularity in astrophysics in the last decades, as the inverse problem is rather ubiquitous. In our case, the inverse problem is to infer the properties of a planet given its observed transmission spectrum. Immediately, we can see a number of issues arising from this problem. First, the observed spectrum is noisy, and the noise distribution is unknown. Second, the forward model that describes the atmospheric processes that produce the spectrum is only approximate. Third, the atmospheric processes are non-linear, and the model parameters are correlated. Therefore, we need Bayesian inference and sophisticated optimizing strategies to map the parameter space and find the best model that describes the data.

A.1 Bayes' theorem

Bayesian inference is based on Bayes' theorem, named after the English mathematician Thomas Bayes. Few ingredients are required to write Bayes' theorem. The data obtained from an experiment or an observation is denoted by $\mathcal{D}$. The model parameters are denoted by θ, and M is the model. Then, $p(\theta|\mathcal{D}, M)$ is the posterior probability of the model parameters given the data. The likelihood, $p(\mathcal{D}|\theta, M)$, is the probability of the data given the parameters and the model. The initial knowledge about the model parameters is denoted by $p(\theta|M)$, and it is called the prior. The Bayes's theorem states that:

$$p(\theta|\mathcal{D}, M) = \frac{p(\mathcal{D}|\theta, M)p(\theta|M)}{p(\mathcal{D}|M)}. \tag{A.1}$$

The denominator is the probability of the data given the model, and it is called the evidence. The evidence normalizes the posterior probability and is independent of the model parameters. It is also computationally expensive to evaluate. Therefore, most sampling methods

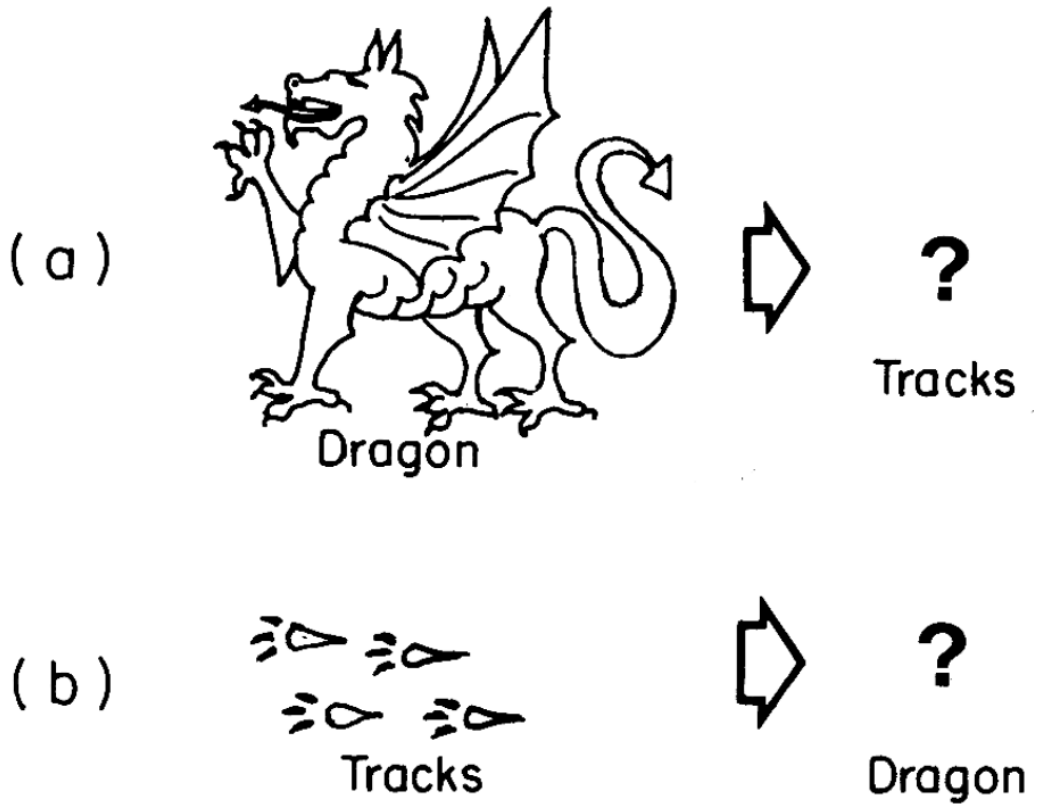


Figure A.1. (a) The direct problem: Describe the tracks of a dragon. (b) The inverse problem: Describe a dragon given its tracks. From Bohren and Huffman (1983).

used in parameter estimation problems – such as standard MCMC techniques (Metropolis et al., 1953; Hastings, 1970; Goodman and Weare, 2010) – do not compute the evidence and sample from the unnormalized posterior probability. The entire process described in Equation A.1 can be viewed as an update of the prior knowledge of the model parameters using the information from the data, resulting in the posterior probability. Then, the posterior can be marginalized over each parameter to obtain constraints on individual parameters.

A.2 Model comparison

The computation of the evidence is required to compare competing models and select the best model that describes the data. The evidence can be written as an integral over the parameter space:

$$p(\mathcal{D}|M) = \int p(\mathcal{D}|\theta, M)p(\theta|M)\mathrm{d}\theta. \tag{A.2}$$

Being the average of the likelihood over the prior, the evidence favors models with a compact parameter space over more complicated ones (Occam's razor), unless the likelihood is significantly higher for the more complicated model.

Two competing models can be compared using the Bayes factor, which is the ratio of the evidences of the two models:

$$\mathcal{B}_{12} = \frac{p(\mathcal{D}|M_1)}{p(\mathcal{D}|M_2)} \tag{A.3}$$

This comparison omits the factor $p(M_1)/p(M_2)$, which is the prior odds of the two models, often considered to be unity. The Bayes factor can be interpreted as the relative probability of the two models given the data. An interpretation of the Bayes ratio is shown in Table A.1.

Table A.1. Interpretation of the Bayes factor as given in Kass and Raftery (1995).

$\log(\mathcal{B}_{12})$	Interpretation
0 to 0.5	*No Evidence*
0.5 to 1	*Some Evidence*
1 to 2	*Strong Evidence*
> 2	*Decisive Evidence*

A.3 Nested sampling

Evaluating the evidence (Equation A.2) is a challenging task for standard MCMC techniques. To overcome this issue, Skilling (2004) introduced the nested sampling algorithm, which turned the problem on its head. Nested sampling is a Monte Carlo method that produces a direct estimate of the evidence, while posterior samples are an optional by-product.

The algorithm starts by randomly drawing N "active" (or "live") samples from the full prior distribution. The samples are then sorted by their likelihood, and the sample with the lowest likelihood is removed from the live set and replaced by a new sample drawn from the prior such that its likelihood is higher than the removed sample. At the next iteration, the sample with the lowest likelihood is removed and replaced in the same way. This process

results in ever-shrinking iso-likelihood contours and continues until a user-defined condition on the likelihood change is reached (evidence tolerance). Once converged, the posterior distribution can be built from the discarded live points ("trace"). The algorithm is illustrated in Figure A.2.

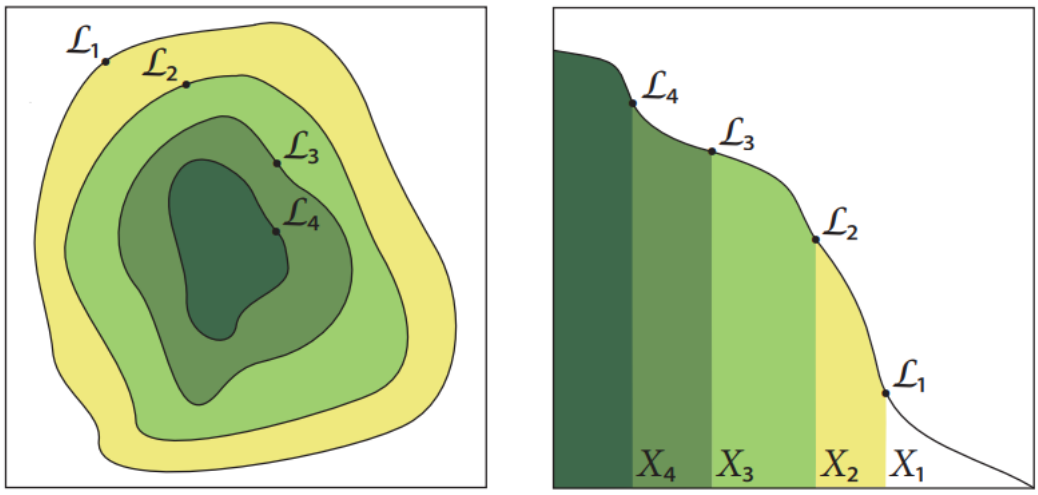


Figure A.2. Cartoon illustrating (a) the posterior of a two-dimensional problem; and (b) the transformed $\mathcal{L}(X)$ function where the prior volumes, X_i, are associated with each likelihood, $\mathcal{L}_i$. From Feroz et al. (2019).

An analogy is useful to explain the nested sampling algorithm. Imagine a group of people standing on a hill, where the people represent our "live" samples and their altitude on the hill is our "likelihood". Whoever is standing on the lowest point of the hill is removed and replaced by a new person on a higher point, randomly chosen from anywhere on the hill higher than where the removed person was. The position of the removed person is recorded. The next person to be removed will be at a higher altitude, until a point is reached where everyone will be standing near the top of the hill and the process will have to stop. Note that this would work even if the hill has two peaks, as everyone will eventually end up on the highest peak. By reconstructing the positions of the removed people, we can map the shape of the hill, our "posterior" distribution.

Mathematically, the nested sampling algorithm is based on the following transformation: instead of evaluating the computationally expensive integral in Equation A.2, we can write it as a 1-dimensional integral over the prior volume, X, as follows. The prior volume is defined by $\mathrm{d}X = p(\theta|M)\mathrm{d}\theta$, so that

$$X = \int_{\mathcal{L}(\theta)>\lambda} p(\theta|M)\mathrm{d}\theta \tag{A.4}$$

where we have used the symbol $\mathcal{L}$ to denote the likelihood. $\mathcal{L}(\theta) = \lambda$ defines the iso-likelihood contour. Assuming that $\mathcal{L}(X)$ (the inverse of Equation A.4) is a monotonically decreasing function, we can write the evidence, denoted by E, as follows:

$$E = \int_0^1 \mathcal{L}(X)\mathrm{d}X \tag{A.5}$$

The evidence can be approximated numerically using standard quadrature methods as the sum of the likelihoods, $\mathcal{L}(X_i) \equiv \mathcal{L}_i$, weighted by the $\mathrm{d}X_i$ intervals:

$$E \approx \sum_{i=1}^{N} \mathcal{L}_i \Delta X_i \tag{A.6}$$

Using the simple trapezium rule, the weights are $\Delta X_i = \frac{1}{2}(X_{i-1} - X_{i+1})$. With the evidence computed, the posterior distribution can be built by weighting the discarded live points by p_i:

$$p_i = \frac{\mathcal{L}_i \Delta X_i}{E} \tag{A.7}$$

and sample-based estimates of the posterior moments can be computed.

The MultiNest algorithm (Feroz and Hobson, 2008; Feroz et al., 2009, 2019) uses ellipsoids to approximate iso-likelihood contours. By drawing the new samples from multidimensional ellipsoids defined by the current live points instead of the full prior volume, the algorithm efficiently traverses the parameter space and multimodal posteriors can be explored. The algorithm is illustrated in Figure A.3.

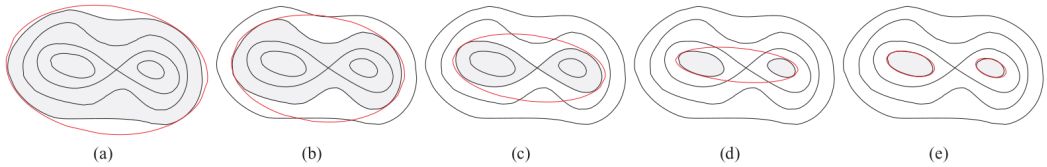


Figure A.3. Cartoon of ellipsoidal nested sampling from a simple bimodal distribution. In (a) we see that the ellipsoid represents a good bound to the active region. In (b)–(d), as we nest inwards we can see that the acceptance rate will rapidly decrease as the bound steadily worsens. (e) illustrates the increase in efficiency obtained by sampling from each clustered region separately. From Feroz and Hobson (2008).

Appendix B

PAOS – Complementary material

B.1 Wavefront propagation

In this section, we present how the Physical Optics Simulator (PAOS) implements Gaussian beams and other aspects of wavefront propagation in the Python class `WFO`. First, Section B.1.1 discusses the problem of aliasing and the strategy to minimize aliasing errors in PAOS. Then, Section B.1.2 to Section B.1.6 give an overview of Gaussian beams and the relevant concepts for the PAOS implementation, giving first the definition and, when appropriate, also an example script. Other important aspects of wavefront propagation in PAOS, such as propagators, paraxial phase shift, and the treatment of the apertures and stops, are discussed in Section B.1.7 to Section B.1.10.

B.1.1 Aliasing

To analyze wave propagation using the Fresnel approximation, a major issue to consider is aliasing errors (Lawrence et al., 1992). These errors result from the discrete nature of numerical calculations, which can distort the amplitude and phase of the optical wave.

An acceptable initial sampling condition is not sufficient to avoid aliasing. The beam spreads during propagation due to diffraction and can grow outside the bounds of the computer array. When that happens, the amplitude of the beam complex is folded back, creating aliasing errors. To prevent this, the propagator selection is crucial, as it affects the size and sampling of the array at each propagation step.

In the continuous case, the far- and near-field propagation expressions are mathematically equivalent; in the discrete case, they differ due to the evaluation of their quadratic phase factors. The near-field propagator has a slowly varying (well-behaved) quadratic phase factor at short propagation distances and a rapidly varying (aliasing) one at large propagation distances. The far-field propagator behaves oppositely: it aliases at short propagation distances but is well-behaved at long propagation distances. Therefore, either the near-field or the far-field propagator should be chosen depending on the system configuration and propagation distance to minimize aliasing errors.

However, some optical systems may not allow using the Fresnel number to define the near- and far-field regions. In such cases, a Gaussian pilot beam can be used as a reference for propagator selection (Lawrence et al., 1992). A Gaussian beam (see Section B.1.2)

has several properties, such as being analytically propagatable, that make it an effective surrogate for the actual beam.

Instead of tracking the Gaussian beam parameters during propagation to adjust the sampling exactly, which would change the units continuously, PAOS uses a region of constant sampling period near the Gaussian beam waist and a region of linearly increasing sampling period far from the waist. The switch is made at the Rayleigh range (see Section B.1.3), which serves as an alternative to the concepts of near-and far-field regions (Lawrence et al., 1992).

B.1.2 Gaussian beams

A Gaussian beam has an irradiance profile that follows an ideal Gaussian distribution (Self, 1983):

$$I(r) = I_0 e^{-\frac{2r^2}{w(z)^2}} = \frac{2P}{\pi w(z)^2} e^{-\frac{2r^2}{w(z)^2}} \tag{B.1}$$

where I_0 is the intensity of the beam on the axis, r is the radial distance, w is the radial distance where the intensity falls to I_0/e^2, or 13.5% of its value on-axis, and $w(z)$ is the semi-diameter of the beam and encompasses 86.5% of the total beam power, P.

Due to diffraction, a Gaussian beam will converge and diverge from the beam waist w_0, an area where the diameter of the beam reaches a minimum size. Therefore, the dependence of $w(z)$ on z, the longitudinal distance from the waist w_0 to the plane of $w(z)$. A Gaussian beam spreads out as (e.g. Self, 1983; Lawrence et al., 1992):

$$w(z)^2 = w_0^2 \left[1 + \left(\frac{\lambda z}{\pi w_0^2}\right)^2\right] = w_0^2 \left[1 + \left(\frac{z}{z_R}\right)^2\right] \tag{B.2}$$

where z_R is the Rayleigh distance. Thus, a Gaussian beam is defined by only three parameters: w_0, z_R, and the divergence angle θ. The complex amplitude of a Gaussian beam is of the form (e.g. Lawrence et al., 1992):

$$a(r, 0) = e^{-\frac{r^2}{w_0^2}} e^{-\frac{jkr^2}{R}} \tag{B.3}$$

where k is the wavenumber and R is the radius of the quadratic phase factor, henceforward *phase radius*. At the waist, where the wavefront is planar, this reduces to:

$$a(r, 0) = e^{-\frac{r^2}{w_0^2}} \tag{B.4}$$

B.1.3 Rayleigh distance

The value of z, where the cross-sectional area of the beam is double that at the waist, defines the Rayleigh distance of a Gaussian beam. This occurs when w(z) increases to $\sqrt{2}w_0$. Explicitly:

$$z_R = \frac{\pi w_0^2}{\lambda} \tag{B.5}$$

The physical significance of the Rayleigh distance is that it indicates the region where the curvature of the wavefront reaches a minimum value. Since

$$R(z) = z + \frac{z_R^2}{z} \tag{B.6}$$

the phase radius is $R = 2z_R$ in the Rayleigh range.

B.1.4 Gaussian beam propagation

To the accuracy of Fresnel diffraction, which implies (Lawrence et al., 1992):

1. aperture sizes significantly larger than the wavelength;
2. modest numerical apertures;
3. thin optical elements,

a Gaussian beam propagates as follows:

$$a(r,z) = e^{-j[kz-\theta(z)]} e^{-\frac{r^2}{w(z)^2}} e^{-\frac{jkr^2}{R(z)}} \tag{B.7}$$

where $\theta(z)$ is a piston term referred to as the *phase factor*, given by

$$\theta(z) = tan^{-1}\left(\frac{z_R}{z}\right) \tag{B.8}$$

which varies from π to $-\pi$ when propagating from $z = -\infty$ to $z = \infty$.

In PAOS, the propagation of the Gaussian beam is described using ABCD matrix optics. A complex radius of curvature $q(z)$, also called *complex beam parameter*, is defined as follows (Yariv, 1989):

$$\frac{1}{q(z)} = \frac{1}{R(z)} - \frac{j\lambda}{\pi n w(z)^2} \tag{B.9}$$

where n is the index of refraction. Propagating a Gaussian beam from some initial position (1) through an optical system (ABCD) to a final position (2) gives the following transformation (Yariv, 1989):

$$\frac{1}{q_2} = \frac{C + D/q_1}{A + B/q_1} \tag{B.10}$$

In PAOS, the initial sampling interval is proportional to D, the diameter of the input beam, and N, the size of the grid:

$$dx_i = dy_i = \frac{D \times \texttt{zoom}}{N} \tag{B.11}$$

where dx_i and dx_i are the initial sampling intervals along the sagittal and tangential directions, respectively. The `zoom` parameter is the reciprocal of the ratio between the initial beam width and the grid width. Note that N values that are not powers of 2 are considered invalid; this ensures that the FFT algorithm is used (Brigham and Morrow, 1967).

In Listing B.1 we report an example of how to use the Python class `WFO` to focus a non-apodized circular beam with a thin paraxial lens. Note that all lens units are in meters. The propagation distance is set equal to the focal length of the paraxial lens. The listing shows the values for the beam waist and the Rayleigh range at the focus position, as well as the distance to the focus before and after propagation[1].

[1]The interested reader may readily verify the results using the tool available at https://www.edmundoptics.com/knowledge-center/tech-tools/gaussian-beams/.

```python
from paos.classes.wfo import WFO
wfo = WFO(beam_diameter=1.0, wavelength=3.0e-6, grid_size=512, zoom=4)
wfo.lens(lens_fl = 1.0)
dtf = wfo.distancetofocus  # dtf = 1.0 m
wfo.propagate(dz = 1.0)
dtf = wfo.distancetofocus  # dtf = 0.0 m
w0 = wfo.w0  # w0 = 1.9 micron
zr = wfo.zr  # zr = 3.8 micron
```

Listing B.1. Gaussian beam propagation.

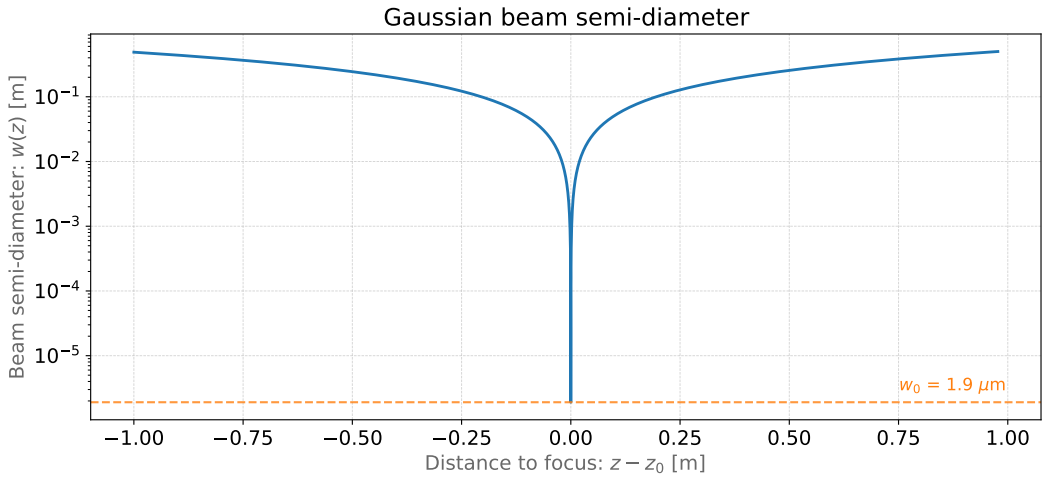


Figure B.1. Evolution of the Gaussian beam semi-diameter as a function of distance to focus, for the optical system in Listing B.1, where we continue the propagation after the focus.

B.1.5 Gaussian beam magnification

Gaussian beam magnification can also be described using ABCD matrix optics. From Equation B.10 and the definition of magnification change in the ABCD formalism given in Equation B.25, we have the following:

$$q_2 = M^2 q_1 \tag{B.12}$$

Using also Equation B.9, it follows that:

$$\begin{aligned} R_2 &= M^2 R_1 \\ w_2 &= M w_1 \\ z_{R,2} &= M^2 z_{R,1} \\ w_{0,2} &= M w_{0,1} \\ z_2 &= M^2 z_1 \end{aligned} \tag{B.13}$$

Note that in the current version of PAOS, we set the Gaussian beam width along x. So, only the sagittal magnification changes the properties of the Gaussian beam, and the tangential magnification changes only the curvature of the propagating wavefront.

Listing B.2 shows a practical example of how to use the Gaussian beam magnification implemented in the Python class `WFO`. The values of the waist and Rayleigh range are

```
from paos.classes.wfo import WFO
Ms, Mt = 1.0, 3.0
wfo = WFO(beam_diameter=1.0, wavelength=3.0e-6, grid_size=512, zoom=4)
w0 = wfo.w0  # w0 = 0.5
zr = wfo.zr  # zr = 2.62e5
wfo.Magnification(Ms, Mt)
w0 = wfo.w0  # w0 = 1.5
zr = wfo.zr  # zr = 2.36e6
```

Listing B.2. Gaussian beam magnification.

reported before and after magnification. They can be easily compared with the respective relations reported in Equation B.13.

B.1.6 Gaussian beam change of medium

A change of medium affects the complex radius of curvature of a Gaussian beam. Using Equation B.14, the complex radius of curvature changes as:

$$q_2 = q_1 n_2 / n_1 \tag{B.14}$$

From this relation and Equation B.9, it follows that:

$$\begin{aligned} R_2 &= R_1 n_2 / n_1 \\ w_2 &= w_1 \\ z_{R,2} &= z_{R,1} n_2 / n_1 \\ w_{0,2} &= w_{0,1} \\ z_2 &= z_1 n_2 / n_1 \end{aligned} \tag{B.15}$$

Moreover, since $\lambda_2 = \lambda_1 n_2 / n_1$, the focal ratio changes as:

$$f_{\#,2} = f_{\#,1} n_1 / n_2 \tag{B.16}$$

Listing B.3 illustrates a thin lens that modifies the curvature of an input beam, followed by a medium change. The reported Gaussian beam parameters vary as described in Equation B.15 and Equation B.16.

```
from paos.classes.wfo import WFO
wfo = WFO(beam_diameter=1.0, wavelength=3.0e-6, grid_size=512, zoom=4)
wfo.lens(lens_fl = 1.0)
w0 = wfo.w0  # w0 = 1.9 micron
fratio = wfo.fratio  # fratio = 1.0
n1, n2 = 1.0, 2.0
wfo.ChangeMedium(n1n2 = n1/n2)
w0 = wfo.w0  # w0 = 1.9 micron
fratio = wfo.fratio  # fratio = 2.0
```

Listing B.3. Gaussian beam change of medium.

B.1.7 Wavefront propagation

As discussed in the previous sections, PAOS propagates the pilot Gaussian beam through all the optical surfaces to calculate the beam width at all points in space. The Gaussian beam surrogates the actual beam, and the Gaussian beam parameters inform the POP simulation. In particular, the Rayleigh distance z_R is used to inform the choice of specific propagators. For a given point, the two inequalities:

$$\begin{aligned} &\text{inside} \leftrightarrow |z - z(w)| \leq z_R \\ &\text{outside} \leftrightarrow |z - z(w)| > z_R \end{aligned} \tag{B.17}$$

describe all four possibilities in moving from inside or outside to inside or outside the Rayleigh range ($\mathcal{R}$), defined as the region between $-z_R$ and z_R from the beam waist. Explicitly, these possibilities are (Lawrence et al., 1992, fig. 24):

(i) II(z_1, z_2): inside $\mathcal{R}$ to inside $\mathcal{R}$

(ii) IO(z_1, z_2): inside $\mathcal{R}$ to outside $\mathcal{R}$

(iii) OI(z_1, z_2): outside $\mathcal{R}$ to inside $\mathcal{R}$

(iv) OO(z_1, z_2): outside $\mathcal{R}$ to outside $\mathcal{R}$

Three primitive operators are needed to move from any point in space to any other: plane-to-plane (PTP), waist-to-spherical (WTS), and spherical-to-waist (STW). PAOS implements all four possible propagations combining these primitive operators.

B.1.8 Paraxial phase correction

For any optical element that can be modeled using its focal length f (e.g., mirrors, thin lenses, and refractive surfaces), the paraxial phase effect is given by:

$$t(x, y) = e^{jk(x^2+y^2)/2f} \tag{B.18}$$

where $t(x, y)$ is the complex transmission function. In other words, the element imposes a quadratic phase shift, which depends on the initial and final position compared to the Rayleigh range.

PAOS uses the Gaussian beam parameters to impose the phase shift. The code implementation consists of four steps:

1. estimate the Gaussian beam curvature after the element (object space);
2. check the initial position;
3. estimate the Gaussian beam curvature after the element (image space);
4. check the final position.

By combining the results of steps 2 and 4, PAOS selects the propagator, and the phase shift is imposed accordingly by defining a phase bias, as described in (Lawrence et al., 1992, p. 174).

B.1.9 Aperture

The actual wavefront propagated through an optical system intersects real optical elements (e.g., mirrors, lenses, slits) and can be obstructed, producing distinct diffraction patterns. PAOS supports circular, elliptical, and rectangular aperture shapes. For each, PAOS implements an appropriate aperture mask, projected on the plane orthogonal to the beam.

For instance, consider a wavefront normal to the sagittal direction and propagating with a slope angle in the tangential direction u_y. At some position z, the wavefront is centered in (x_c, y_c) and intersects an elliptical aperture centered in (x_a, y_a) with semi-axes (ϕ_x, ϕ_y). PAOS sets this aperture as:

$$\left(y_a - y_c, \phi_x, \frac{1}{\sqrt{u_y^2 + 1}} \phi_y \right) \tag{B.19}$$

Listing B.4 shows how to simulate the propagation of the wavefront through one such aperture. By changing the *obscuration* value to *True*, the same aperture can be used as an obscuration.

B.1.10 Stop

An aperture stop (or diaphragm) is an element of an optical system that determines how much light reaches the image plane. It is often the boundary of the primary mirror. An aperture stop can significantly affect the sizes of system aberrations, by preventing some of the severely aberrated rays from reaching the image (Born et al., 1999). PAOS implements a generic stop by normalizing the wavefront at the current position to unit energy.

Listing B.5 demonstrates how to use the class `WFO` to implement an aperture stop immediately after wavefront initialization.

```
import numpy as np
from paos.classes.wfo import WFO
phi_x = 0.5  # m
phi_y = 0.4
xa = 0
ya = 0.1
field = {"us": 0.0, "ut": 0.1}
vt = np.array([0.0, field["ut"]])
phi_y *= np.sqrt(1 / (vt[1] ** 2 + 1))
ya = ya - vt[0]
wfo = WFO(beam_diameter=1.0, wavelength=3.0e-6, grid_size=512, zoom=4)
aperture_shape = "elliptical"  # or "rectangular"
obscuration = False  # if True, applies obscuration
aperture = wfo.aperture(xa, ya, hx=phi_x, hy=phi_y, shape=aperture_shape,
    obscuration=obscuration)
```

Listing B.4. Elliptical, off-axis aperture.

```
from paos.classes.wfo import WFO
wfo = WFO(beam_diameter=1.0, wavelength=3.0e-6, grid_size=512, zoom=4)
wfo.make_stop()
E = (wfo.amplitude ** 2).sum()  # E = 1.0
```

Listing B.5. Aperture stop.

B.1.11 Paraxial ray-tracing

PAOS implements the method `raytrace` to perform diagnostic ray-tracing. This method receives as inputs the field slopes and the optical chain, then prints the ray positions and slopes in the tangential and sagittal planes for each surface of the optical chain. Listing B.6 reports an example of how to access this method. The next PAOS version may add support

```
from paos.core.raytrace import raytrace
raytrace(field={"us": 0.0, "ut": 0.0}, opt_chain=opt_chains[0])
```

Listing B.6. Paraxial ray-tracing.

for one of the existing Python codes that implement full ray-tracing (e.g., pyrate[2]). The main purpose is to obtain the expected maps of aberrations by the realistic elements of the optical chain, such as mirrors, without having to rely on external code.

B.2 ABCD matrices

In this section, we show how the ABCD matrix method implemented in the Python class `ABCD` can be used to model the effects of the propagation step (Section B.2.1), thin lens (Section B.2.2), dioptre (Section B.2.3), change of medium (Section B.2.4), thick lens (Section B.2.5), and magnification (Section B.2.6) on the input ray. The most common matrix

[2] https://salsa.debian.org/mess42/pyrate

operators are presented in an introductory manner, giving first the definition and then an example script using the class `ABCD`.

In each example, we consider a light ray propagating in the tangential plane (YZ), parameterized by the vector $\vec{v_t} = (y, u_y)$, where u_y is the slope, i.e., the tangent of the angle, which equals the angle in paraxial approximation. The same definitions also apply in the sagittal plane (XZ), modified when necessary when the cylindrical symmetry is violated; the relevant vector is $\vec{v_s} = (x, u_x)$.

B.2.1 Propagation

Either in free space or in a refractive medium, a propagation from point z_1 to point z_2 over a distance T is given by:

$$\begin{pmatrix} y_2 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ u_1 \end{pmatrix} = \hat{T} \begin{pmatrix} y_1 \\ u_1 \end{pmatrix} \tag{B.20}$$

i.e., a propagation changes only y.

Listing B.7 shows how to propagate a light ray $\vec{v_t} = (0, 1)$ over 50 mm. Note that PAOS considers a distance "positive" when the propagation is left → right.

```
from paos.classes.abcd import ABCD
thickness = 50.0  # mm
abcd = ABCD(thickness=thickness)
(A, B), (C, D) = abcd.ABCD  # A=1, B=50, C=0, D=1
vt = np.array([0, 1])
vt = abcd() @ vt  # vt = np.array([50, 1])
```

Listing B.7. ABCD propagation.

B.2.2 Thin lens

A thin lens changes only the slope angle u, and this is given by:

$$\begin{pmatrix} y_2 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\psi & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ u_1 \end{pmatrix} = \hat{L} \begin{pmatrix} y_1 \\ u_1 \end{pmatrix} \tag{B.21}$$

where $\psi = \frac{1}{f}$ is the lens optical power.

Listing B.8 shows how to simulate the effect of a thin lens with a radius of curvature $R = 20$ mm on a light ray $\vec{v_t} = (1, 0)$.

```
from paos.classes.abcd import ABCD
r = 20.0  # mm
abcd = ABCD(curvature=1.0/r)
(A, B), (C, D) = abcd.ABCD  # A=1, B=0, C=-0.05, D=1
vt = np.array([1, 0])
vt = abcd() @ vt  # vt = np.array([1, -0.05])
```

Listing B.8. ABCD thin lens.

B.2.3 Dioptre

When light propagates from a medium with refractive index n_1 into a dioptre of refractive index n_2, the slope angle u varies as follows:

$$\begin{pmatrix} y_2 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{\psi}{n_2} & \frac{n_1}{n_2} \end{pmatrix} \begin{pmatrix} y_1 \\ u_1 \end{pmatrix} = \hat{D} \begin{pmatrix} y_1 \\ u_1 \end{pmatrix} \tag{B.22}$$

The dioptre power is $\psi = \frac{n_2 - n_1}{R}$, where R is the dioptre radius of curvature. Following the standard convention, $R > 0$ if the center of curvature is at the right of the dioptre, and $R < 0$ if at the left.

Listing B.9 shows how to simulate the effect of a dioptre with refractive index $n = 1.25$ and a radius of curvature $R = 20$ mm on a light ray $\vec{v_t} = (1, 1)$.

```
from paos.classes.abcd import ABCD
n1, n2 = 1.0, 1.25
r = 20.0  # mm
abcd = ABCD(curvature = 1.0/r, n1 = n1, n2 = n2)
(A, B), (C, D) = abcd.ABCD  # A=1, B=0, C=-0.01, D=0.8
vt = np.array([1, 1])
vt = abcd() @ vt  # vt = np.array([1, 0.79])
n1/n2 = abcd.n1n2  # n1/n2 = 0.8
```

Listing B.9. ABCD dioptre.

B.2.4 Medium change

The limiting case of a dioptre with $R \to \infty$ represents a medium change.

$$\begin{pmatrix} y_2 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{n_1}{n_2} \end{pmatrix} \begin{pmatrix} y_1 \\ u_1 \end{pmatrix} = \hat{N} \begin{pmatrix} y_1 \\ u_1 \end{pmatrix} \tag{B.23}$$

B.2.5 Thick lens

A real (thick) lens is modeled as follows:

$$\begin{pmatrix} y_2 \\ u_2 \end{pmatrix} = \hat{D}_b \hat{T} \hat{D}_a \begin{pmatrix} y_1 \\ u_1 \end{pmatrix} \tag{B.24}$$

i.e., propagation through the dioptre D_a (first encountered by the ray), then propagation in the medium, followed by the exit dioptre D_b. Note that the limiting case in which the thickness of the dioptre, T, is negligible and can be set to zero gives back the thin lens. If a dioptre has $R \to \infty$, the lens is plano-concave or plano-convex, depending on the curvature of the other.

Listing B.10 shows how to simulate the effect of a bi-convex lens (R_a is positive and R_b is negative) made of ZnSe (Zinc selenide), modeled with a refractive index $n = 2.403$. This thick lens will cause the beam to converge[3].

[3]The interested reader may readily check this result using the tool provided at https://www.edmundoptics.com/knowledge-center/tech-tools/focal-length/.

```
from paos.classes.abcd import ABCD
r1, r2 = 20.0, -20.0  # mm
n_os, n_l, n_is = 1.0, 2.403, 1.0
center_thickness = 5.0
abcd = ABCD(curvature = 1.0/r1, n1 = n_os, n2 = n_l)
abcd = ABCD(thickness = center_thickness) * abcd
abcd = ABCD(curvature = 1.0/r2, n1 = n_l, n2 = n_is) * abcd
f_eff = abcd.f_eff  # f_eff = 7.454 mm
```

Listing B.10. ABCD thick lens.

B.2.6 Magnification

We can model a change in magnification as follows:

$$\begin{pmatrix} y_2 \\ u_2 \end{pmatrix} = \begin{pmatrix} M & 0 \\ 0 & 1/M \end{pmatrix} \begin{pmatrix} y_1 \\ u_1 \end{pmatrix} = \hat{M} \begin{pmatrix} y_1 \\ u_1 \end{pmatrix} \tag{B.25}$$

i.e., a magnification modifies both y and u.

Listing B.11 shows how to magnify a light ray $\vec{v_t} = (1, 1)$ by a factor of 2.

```
from paos.classes.abcd import ABCD
abcd = ABCD(M=2)
(A, B), (C, D) = abcd.ABCD  # A=2, B=0, C=0.5, D=0
vt = np.array([1, 1])
vt = abcd() @ vt  # vt = np.array([2, 0.5])
```

Listing B.11. ABCD magnification.

B.3 Refractive index

The following sections discuss how PAOS implements the refractive index of optical materials in the Python class `Material`. Section B.3.1 describes the chromatic dependence of the refractive index, while Section B.3.2 discusses the relevant modelization for the thermal and pressure dependencies.

B.3.1 Chromaticity of refractive index

The relation between the refractive index and wavelength can be modeled using an empirical relation such as the Sellmeier equation, a development in particular of Cauchy's work on classical dispersion theory. In its original form (Sellmeier, 1872), it is given as follows:

$$n^2(\lambda) = 1 + \sum_i \frac{K_i \lambda^2}{\lambda^2 - L_i} \tag{B.26}$$

where n is the refractive index, λ is the wavelength, and K_i and $\sqrt{L_i}$ are the Sellmeier coefficients, determined from precision measurements. Limitations in the accuracy of this model are present where the absorption is not negligible. Physically, each term of the

sum represents an absorption resonance of strength K_i at wavelength $\sqrt{L_i}$. Close to each absorption peak, a more precise dispersion model is required to avoid non-physical values.

PAOS implements the Sellmeier 1[4] equation to estimate the index of refraction relative to air for a particular optical glass at its reference temperature and pressure as indicated in the catalog, typically:

$$\begin{aligned} T_{ref} &= 20°C \\ P_{ref} &= 1 \text{ atm} \end{aligned} \tag{B.27}$$

The Sellmeier 1 form of the original equation consists of only three terms and is implemented as follows:

$$n^2(\lambda) = 1 + \frac{K_1\lambda^2}{\lambda^2 - L_1} + \frac{K_2\lambda^2}{\lambda^2 - L_2} + \frac{K_3\lambda^2}{\lambda^2 - L_3} \tag{B.28}$$

where the wavelength in micrometers has to be inserted.

Listing B.12 shows an example script to estimate the refraction index of BK7 for a range of wavelengths from visible to infrared using the class `Material`.

```
import numpy as np
from paos.util.material import Material
material = Material(wl=np.array([0.5, 1.5, 2.5]))
bk7 = material.materials["BK7"]
n = material.sellmeier(bk7["sellmeier"])  # n = [1.5214 1.5013 1.4860]
```

Listing B.12. BK7 index of refraction.

B.3.2 Temperature coefficient of refractive index

The refractive index of the dispersive medium depends not only on wavelength but also on temperature. The relationship between the change in refractive index and the change in temperature is called the temperature coefficient of the refractive index. This coefficient is defined as the deviation of the curve, dn/dT, and can have a positive or negative value. The values of the temperature coefficient can be given as absolute (measured under vacuum) and relative (measured under dry air at standard pressure).

PAOS defines the refractive index of air under standard conditions as the reference index; all other indices are relative. The air reference index of refraction is estimated as follows (Kohlrausch, 1986):

$$n_{air,ref} = 1.0 + 10^{-8} \cdot \left(6432.8 + \frac{2949810\,\lambda^2}{146\,\lambda^2 - 1} + 25540\frac{\lambda^2}{41\,\lambda^2 - 1}\right) \tag{B.29}$$

where λ is in micrometer units. Under different temperatures and pressures, this reference index is scaled as (Kohlrausch, 1986):

$$n_{air} = 1 + \frac{P\left(n_{air,ref} - 1\right)}{1.0 + 3.4785 \cdot 10^{-3}(T - 15)} \tag{B.30}$$

[4]Notation from `Zemax`.

PAOS calculates the refractive index of an optical material at a given pressure and temperature using the formula:

$$n(\Delta T) = \frac{n^2 - 1}{2n} D_0 \Delta T + n \tag{B.31}$$

where ΔT is the difference between the operative temperature of the material T_{oper} and the reference temperature T_{ref}, n is the refractive index estimated using Equation B.28, and D_0 is a temperature constant of the material.

Listing B.13 reports an example script to estimate the refraction index of SAPPHIRE for a given wavelength under reference and operating (ambient) temperature and pressure. Note that PAOS can easily model systems used in a vacuum by changing the air pressure to zero, and this can be set dynamically, as shown in the script.

```
from paos.util.material import Material
wl = 1.95  # micron
Tref, Tamb = 20.0, -223.0  # Celsius
Pamb = 1.0  # atm
material = Material(wl, Tambient=Tamb, Pambient=Pamb)
n_ref, n_amb = material.nmat(name="SAPPHIRE")  # n_ref=1.7392, n_amb
    =1.7367
material.Tambient = -223.0
material.Pambient = 0.0
n_ref, n_amb = material.nmat(name="SAPPHIRE")  # n_ref=1.7388, n_amb
    =1.7362
```

Listing B.13. Thermal dependence of the refractive index of SAPPHIRE.

Listing B.14 shows how to access the plotting routine in the `Material` class to plot the refractive index vs wavelength of supported optical materials at their operating and reference temperature.

```
import numpy as np
from paos.util.material import Material
mat = Material(wl=np.linspace(0.5, 2.5, 100), Tambient=-223, Pambient
    =1.0)
mat.plot_relative_index(material_list=['Caf2', 'Znse'])  # Plot selected
available_materials = list(mat.materials.keys())
mat.plot_relative_index(material_list=available_materials)  # Plot all
```

Listing B.14. Plot relative index of refraction.

Figure B.2 reports a plot of the relative index of refraction of two optical materials, CaF_2 and ZnSe[5], at their reference and operating temperatures, as can be accomplished using Listing B.14.

From this figure, it can be seen that the temperature coefficient is responsible for a wavelength-dependent vertical offset between the two curves; also, for different materials, this offset can be positive or negative.

[5]To compare the ZnSe curve with this plot, the reader may consult the online refractive index database at https://refractiveindex.info/?shelf=main&book=ZnSe&page=Marple.

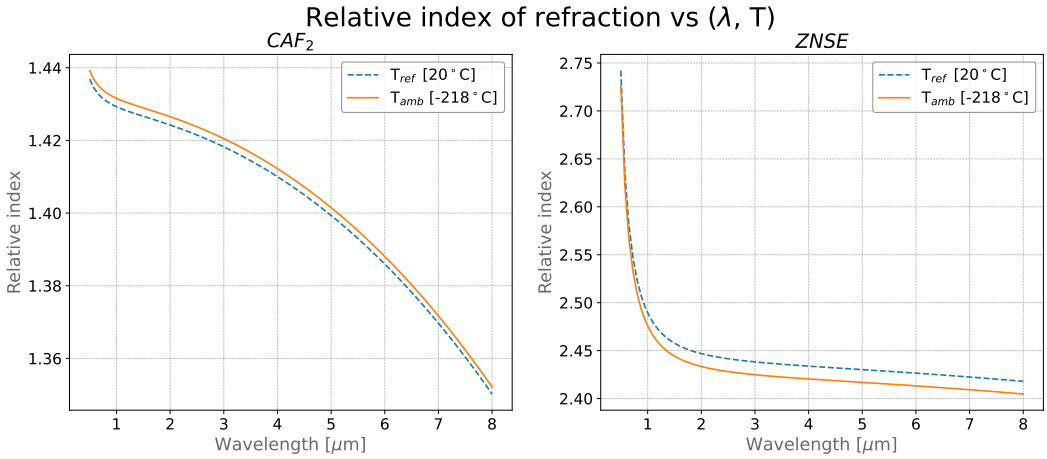


Figure B.2. Plots of the relative index of refraction as a function of wavelength for CaF_2 (left panel) and ZnSe (right panel) at the reference (dashed line) and an operating temperature (solid line). Standard pressure is assumed. The wavelength range shown roughly matches Ariel's spectral range.

Appendix C

Tier 1 – Complementary material

C.1 Analytical derivation of the molecular metric

As mentioned in Section 5.1.4, the metric here presented is (i) sensitive to the molecules, (ii) independent from the planet size, and (iii) independent from the scale height. To show it, we start by using the following notation: in transmission spectroscopy we are measuring

$$\frac{\Delta f}{f}(\lambda) = \frac{R_{pl}^2 + 2R_{pl} \cdot z(\lambda)}{R_\star^2} \tag{C.1}$$

where f is the measured flux from the star, Δf is the difference between the flux measured during the transit and the one measured out of transit, R_ρ and $R_\star$ are the planet and the star radii respectively; $z(\lambda)$ is the measured wavelength dependent transit depth. Now, applying the definition of $z(\lambda)$ from Lecavelier Des Etangs et al. (2008),

$$z(\lambda) = H \ln\left(\frac{\epsilon_{abs}\sigma_{abs}(\lambda)P_0}{\tau_{eq}}\sqrt{\frac{2\pi R_p H}{k_B^2 T_p^2}}\right) \tag{C.2}$$

where ϵ_{abs} and σ_{abs} are the abundance and cross section of the main absorbent abs at the λ wavelength. H is the scale height to which correspond the P_0 pressure and the τ_{eq} is the equivalent optical depth. Therefore, we have

$$\frac{\Delta f}{f}(\lambda) = \frac{R_{pl}^2 + 2R_{pl}H \ln\left(\frac{\epsilon_{abs}\sigma_{abs}(\lambda)P_0}{\tau_{eq}}\sqrt{\frac{2\pi R_p H}{k_B^2 T_p^2}}\right)}{R_\star^2} = \frac{R_{pl}^2 + 2R_{pl}H \cdot Z(\lambda)}{R_\star^2} \tag{C.3}$$

where, for simplicity, we called

$$Z(\lambda) = \ln\left(\frac{\epsilon_{abs}\sigma_{abs}(\lambda)P_0}{\tau_{eq}}\sqrt{\frac{2\pi R_p H}{k_B^2 T_p^2}}\right) \tag{C.4}$$

Therefore, to measure S_{band_i} in Equation 5.1 corresponds to computing the mean in the band:

$$S_{band_i} = \left(\frac{\Delta f}{f}\right)_{band_i} = \frac{R_{pl}^2}{R_\star^2} + \frac{2R_{pl}H}{R_\star^2} \cdot \frac{\sum_j^M Z_j}{M} = \frac{R_{pl}^2}{R_\star^2} + \frac{2R_{pl}H}{R_\star^2} \cdot Z_{band_i} \tag{C.5}$$

where Z_j is the equivalent of Equation C.4 in the j^{th} spectral bin and $Z_{band_i} = \frac{\sum_j^M Z_j}{M}$. Therefore, the dispersion of Equation 5.2 is computed as

$$\sigma_{band_i} = \frac{2R_{pl}H}{R_\star^2} \sqrt{\frac{\sum_j^M (Z_j - Z_{band_i})^2}{M}} = \frac{2R_{pl}H}{R_\star^2} \cdot \sigma_{Z_{band_i}} \tag{C.6}$$

where $\sigma_{Z_{band_i}} = \sqrt{\frac{\sum_j^M (Z_j - Z_{band_i})^2}{M}}$.

By combining the previous equations as done in Equation 5.3, we finally obtain

$$M_{mol} = \frac{1}{N} \sum_i^N \frac{\frac{2R_{pl}H}{R_\star^2} (Z_{band_i} - Z_{norm})}{\frac{2R_{pl}H}{R_\star^2} \sqrt{\sigma_{Z_{band_i}}^2 + \sigma_{Z_{norm}}^2}} = \frac{1}{N} \sum_i^N \frac{Z_{band_i} - Z_{norm}}{\sqrt{\sigma_{Z_{band_i}}^2 + \sigma_{Z_{norm}}^2}} \tag{C.7}$$

Therefore, we remove the planet and star radii dependence in the measurement. Similarly to what has been done in Désert et al. (2009), the subtraction between Z_{band_i} and Z_{norm} finally removes the scale height dependency as

$$Z_{band_i} - Z_{norm} = \ln\left(\frac{\epsilon_{abs,band_i}\sigma_{abs,band_i}}{\epsilon_{abs,norm}\sigma_{abs,norm}}\right) \tag{C.8}$$

where $\epsilon_{abs,band_i}\sigma_{abs,band_i}$ is the equivalent of $\epsilon_{abs}\sigma_{abs}(\lambda)$ in the band. This factor identifies the contribution of the main absorber in the band. Therefore, if we compare a band where a certain molecule has a strong feature, with one that is not supposed to give contributions to the spectrum, we can identify the molecular presence, compared to what is present in the second band.

So, finally M_{mol} becomes

$$M_{mol} = \frac{1}{N} \sum_i^N \frac{\ln\left(\frac{\epsilon_{abs,band_i}\sigma_{abs,band_i}}{\epsilon_{abs,norm}\sigma_{abs,norm}}\right)}{\sqrt{\sigma_{Z_{band_i}}^2 + \sigma_{Z_{norm}}^2}} \tag{C.9}$$

So, as promised, the metric is also sensitive to the molecular content.

To summarize, we removed the star, planet, and atmosphere size dependencies by subtracting the interesting feature bands for a normalization band and dividing the results by the combined dispersion. This results in a metric that is sensitive to the molecules contained in the atmosphere, but introduces a bias. In fact, the spectral dispersion $\sigma_{Z_{band_i}}$ depends on both the atmospheric feature dispersion and the observational noise.

C.2 P-statistic – Complementary figures

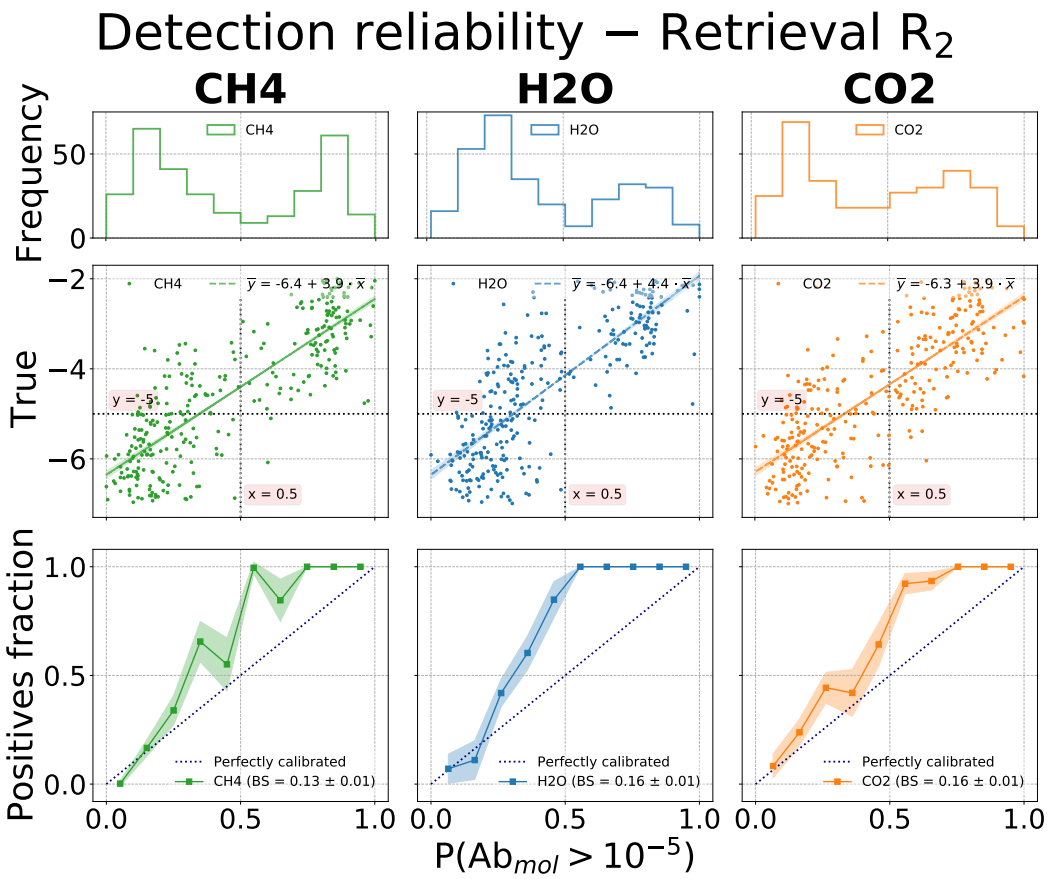


Figure C.1. Same as Figure 6.4. Detection reliability for the R_2 retrievals, that implement a model that is over-representative of the simulated atmospheres, by including CO, HCN, and H_2S as additional trace gases.

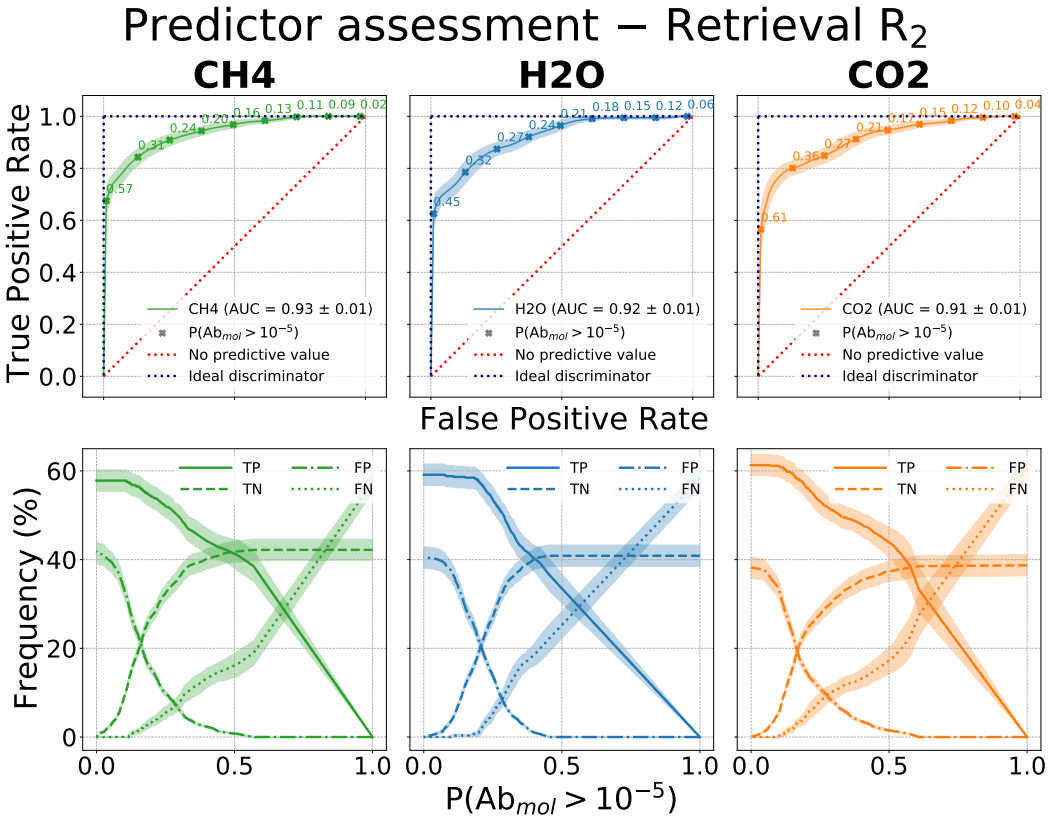


Figure C.2. Same as Figure 6.6. Predictor assessment for the R_2 retrievals, that implement a model that is over-representative of the simulated atmospheres, by including CO, HCN, and H_2S as additional trace gases.

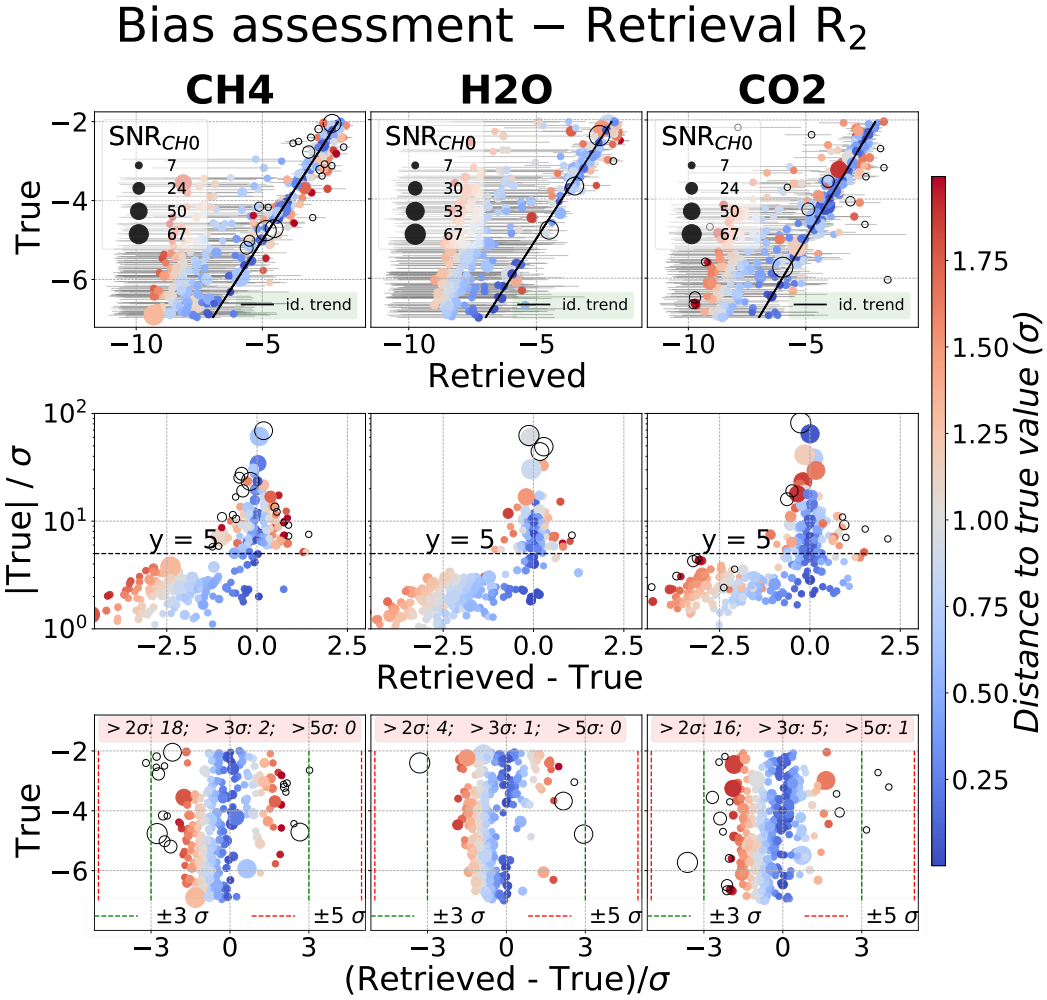


Figure C.3. Same as Figure 6.8 for the R_2 retrievals, that implement a model that is over-representative of the simulated atmospheres, by including CO, HCN, and H_2S as additional trace gases.

Appendix D

ARES VI – Complementary figures and tables

D.1 Planet-Stellar parameters

These parameters were used to simulate the transmission spectra and as input into *PandExo*.

Parameters	GJ1214 b	HD189733 b	WASP-121 b
R_* [Rsun]	0.21	0.75	1.46
T_* [K]	3250	5052	6459
Z_*	0.29	-0.02	0.13
log(g)	5.03	4.49	4.24
R_p [Rjup]	0.243	1.13	1.91
References	Charbonneau et al. (2009) Cloutier et al. (2021)	Stassun et al. (2017) Addison et al. (2019)	Delrez et al. (2016)

Table D.1. Planet-Stellar parameters.

D.2 Retrievals results

Tables of values retrieved by all retrieval models for all input configurations considering the best solution for each retrieval are shown here. For Free retrievals, we calculate the metallicity using TauREx 3 which includes it in the available derived parameters, while we compute the C/O ratio using formula 2 from Lee et al. (2013) and the marginalized posterior distributions of CO, CH_4, CO_2, and H_2O from each retrieval (best solution).

		JWST												Ariel					
		1D						3D											
		constant			equilibrium			constant			equilibrium			constant			equilibrium		
Parameters	**Input**	**Free**	ACE	Fast	Free	**ACE**	Fast	**Free**	ACE	Fast	Free	**ACE**	Fast	Free	**ACE**	Fast	Free	**ACE**	Fast
T_{surf} [K]		**970**	348	330	770	**1037**	1097	**1092**	1303	950	1055	**1165**	1422	918	**1084**	1162	1064	**1086**	1128
T_1 [K]		**639**	745	875	1035	**1345**	1341	**756**	305	306	407	**391**	317	684	**610**	325	621	**673**	533
T_2 [K]		**215**	236	189	331	**378**	334	**426**	560	602	306	**580**	674	443	**569**	597	331	**470**	579
T_{top} [K]		**198**	628	627	185	**212**	334	**536**	306	335	1234	**619**	532	524	**350**	367	983	**814**	833
$\log_{10}(P_1)$ [Pa]		**4.67**	2.25	2.63	3.95	**4.38**	4.24	**4.98**	2.15	2.15	2.44	**2.00**	2.09	4.72	**4.55**	2.44	4.59	**4.61**	4.06
$\log_{10}(P_2)$ [Pa]		**0.131**	1.58	2.08	2.24	**3.13**	2.91	**3.26**	1.20	1.27	-0.648	**1.31**	1.37	2.60	**2.61**	1.06	0.32	**2.52**	1.68
$\log_{10}(H_2O)$	**-1.00**	**-1.15**			-1.17			**-1.09**			-1.24			-1.13			-1.26		
$\log_{10}(CO)$	**-2.00**	**-2.24**			-8.62			**-2.11**			-4.86			-5.59			-7.48		
$\log_{10}(CH_4)$	**-1.52**	**-1.63**			-1.21			**-1.64**			-2.27			-1.61			-2.27		
$\log_{10}(CO_2)$	**-2.00**	**-2.10**			-9.34			**-2.06**			-1.57			-1.82			-1.46		
$\log_{10}(HCN)$	**-6.22**	**-8.16**			-8.35			**-8.37**			-8.21			-6.93			-7.29		
$\log_{10}(NH_3)$	**-3.30**	**-3.45**			-3.19			**-3.45**			-9.23			-3.57			-8.52		
Radius [R_{jup}]	**0.243**	**0.242**	0.248	0.247	0.241	**0.240**	0.238	**0.242**	0.241	0.244	0.245	**0.240**	0.240	0.242	**0.240**	0.241	0.247	**0.245**	0.247
C/O	**0.458**		0.459	0.442	0.90	**0.898**	0.807		0.255	0.145	0.29	**0.441**	0.313		**0.309**	0.169	0.33	**0.461**	0.371
$\log_{10}(Z)$	**2.00**		2.14	2.20	1.53	**1.91**	1.96		2.02	1.82	1.58	**2.01**	1.83		**2.06**	1.79	1.61	**2.06**	1.93
logE	**-**	**4642**	4623	4619	4631	**4640**	4633	**4619**	4585	4575	4547	**4557**	4504	736	**737**	736	731	**740**	735

Table D.2. Retrieval results of **GJ1214 b**. Best retrieval of each configuration is highlighted in bold.

		JWST															Ariel							
		1D								3D														
		constant				equilibrium				constant				equilibrium				constant				equilibrium			
Parameters	Input	**Free**	ACE	Fast	GGchem	Free	**ACE**	Fast	GGchem	**Free**	ACE	Fast	GGchem	Free	**ACE**	Fast	GGchem	**Free**	ACE	Fast	GGchem	Free	ACE	Fast	**GGchem**
T_{surf} [K]		**1352**	1554	508	503	626	**1760**	836	1610	**1313**	503	980	899	705	**1612**	729	1371	**1400**	1610	1038	1451	1641	1509	1103	**1433**
T_1 [K]		**731**	517	969	1643	1382	**834**	694	732	**745**	1898	1864	1541	1509	**879**	735	761	**1194**	1531	2265	1431	1557	1098	667	**1158**
T_2 [K]		**1140**	507	536	677	659	**1267**	1251	1322	**1308**	720	688	664	1069	**1367**	1175	1426	**1239**	555	587	578	1048	1175	1179	**1182**
T_{top} [K]		**931**	760	501	501	2299	**535**	1264	516	**1020**	501	502	501	1411	**548**	1748	522	**1311**	681	1653	655	1440	1509	1569	**1242**
$\log_{10}(P_1)$ [Pa]		**4.31**	0.228	3.53	3.06	2.65	**2.68**	2.00	2.28	**4.18**	3.12	2.69	2.90	3.63	**2.87**	2.01	2.35	**4.46**	4.12	3.00	4.20	4.49	4.18	3.29	**4.58**
$\log_{10}(P_2)$ [Pa]		**2.52**	-2.63	0.552	2.63	2.42	**2.39**	1.59	2.12	**2.13**	2.85	2.47	2.20	3.27	**2.60**	1.91	2.22	**1.61**	1.03	2.56	1.22	1.33	1.51	1.21	**1.64**
$\log_{10}(H_2O)$	-3.15	**-2.87**				-3.19				**-2.90**				-3.44				**-3.05**				-3.71			
$\log_{10}(CO)$	-4.00	**-7.86**				-3.12				**-7.92**				-3.30				**-7.92**				-3.92			
$\log_{10}(CH_4)$	-3.40	**-3.04**				-5.34				**-3.05**				-6.17				**-3.28**				-7.77			
$\log_{10}(CO_2)$	-7.70	**-9.16**				-5.89				**-9.15**				-6.58				**-9.37**				-7.05			
$\log_{10}(HCN)$	-7.00	**-9.19**				-9.32				**-9.09**				-7.99				**-8.41**				-9.11			
$\log_{10}(NH_3)$	-4.52	**-4.29**				-6.53				**-4.34**				-6.30				**-5.49**				-8.89			
$\log_{10}(FeH)$	-8.05	**-7.93**				-11.8				**-7.91**				-10.8				**-9.43**				-10.2			
$\log_{10}(SiO)$	-4.70	**-8.55**				-9.52				**-8.44**				-9.37				**-7.59**				-8.22			
$\log_{10}$(New Astron.)	-5.52	**-8.00**				-8.73				**-11.7**				-9.18				**-7.06**				-6.80			
$\log_{10}(K)$	-6.70	**-11.7**				-11.9				**-7.75**				-11.9				**-7.36**				-8.28			
$\log_{10}(TiO)$	-10.0	**-9.75**				-11.9				**-9.77**				-12.0				**-9.63**				-11.4			
$\log_{10}(VO)$	-9.00	**-8.68**				-11.9				**-8.70**				-11.9				**-9.97**				-11.2			
Radius [R_{jup}]	1.13	**1.13**	1.13	1.13	1.13	1.16	**1.12**	1.13	1.12	**1.12**	1.13	1.13	1.13	1.12	**1.12**	1.13	1.12	**1.12**	1.13	1.12	1.13	1.12	1.12	1.12	**1.12**
C/O	0.550		0.962	0.726	0.797	0.54	**0.677**	0.0377	0.418		0.921	2.00	0.874	0.58	**0.614**	0.0368	0.308		0.954	1.87	0.869	0.38	0.676	0.143	**0.544**
$\log_{10}(Z)$	0.00		0.00139	0.417	-0.047	0.02	**0.0864**	1.22	0.359		-0.444	-0.989	0.263	-0.18	**0.144**	1.34	0.541		0.0480	-0.858	-0.163	-0.44	0.150	1.03	**0.0782**
logE	-	**5818**	3282	-1089	3146	5545	**5747**	3390	5365	**5792**	2744	-3708	2102	5541	**5749**	3253	5354	**843**	809	791	809	840	848	840	**848**

Table D.3. Retrieval results of **HD189733 b**. Best retrieval of each configuration is highlighted in bold.

		JWST												Ariel					
		1D						3D											
		constant			equilibrium			constant			equilibrium			constant			equilibrium		
Parameters	**Input**	**Free**	Fast	GGchem	Free	Fast	**GGchem**	**Free**	Fast	GGchem	**Free**	Fast	GGchem	**Free**	Fast	GGchem	Free	Fast	**GGchem**
T_{surf} [K]		**2045**	3206	2122	2422	2590	**1830**	**880**	3342	475	**2736**	928	3277	**1805**	1273	3473	2372	1604	**1558**
T_1 [K]		**1519**	641	1652	3446	586	**2828**	**908**	709	465	**1282**	521	1081	**614**	622	779	1265	836	**1151**
T_2 [K]		**975**	2290	776	978	2658	**766**	**3745**	2704	3016	**2195**	2162	2394	**3671**	2799	2554	2048	2186	**2143**
T_{top} [K]		**2877**	596	2714	2737	468	**3596**	**3727**	784	466	**705**	1269	710	**3003**	558	837	2373	844	**2664**
$\log_{10}(P_1)$ [Pa]		**4.78**	2.01	2.53	3.63	2.02	**3.10**	**2.05**	2.51	3.52	**2.40**	3.81	2.10	**2.18**	2.84	2.46	3.78	4.30	**3.11**
$\log_{10}(P_2)$ [Pa]		**3.22**	0.31	2.43	3.21	-0.616	**2.37**	**1.68**	2.40	3.47	**1.56**	3.28	1.57	**1.33**	2.67	1.98	1.30	3.45	**0.989**
$\log_{10}(H_2O)$	**-3.30**	**-3.08**			-3.16			**-3.62**			**-4.12**			**-3.58**			-4.05		
$\log_{10}(CO)$	**-3.30**	**-3.24**			-3.18			**-3.97**			**-3.14**			**-4.07**			-3.16		
$\log_{10}(CH_4)$	**-5.00**	**-4.71**			-9.31			**-9.60**			**-9.81**			**-9.24**			-9.50		
$\log_{10}(CO_2)$	**-7.40**	**-7.98**			-6.73			**-9.36**			**-7.15**			**-9.37**			-7.69		
$\log_{10}(HCN)$	**-6.40**	**-8.52**			-9.29			**-9.23**			**-9.86**			**-8.87**			-9.35		
$\log_{10}(NH_3)$	**-5.22**	**-4.99**			-6.89			**-9.09**			**-9.64**			**-9.23**			-9.29		
$\log_{10}(FeH)$	**-5.70**	**-5.46**			-9.55			**-6.28**			**-9.42**			**-6.82**			-10.6		
$\log_{10}(SiO)$	**-4.22**	**-4.22**			-5.48			**-6.12**			**-5.53**			**-8.11**			-7.41		
$\log_{10}$(New Astron.)	**-5.52**	**-1.80**			-8.96			**-5.28**			**-7.22**			**-3.64**			-3.66		
$\log_{10}(K)$	**-6.70**	**-9.58**			-10.9			**-11.2**			**-10.6**			**-10.1**			-9.39		
$\log_{10}(TiO)$	**-7.00**	**-6.75**			-7.41			**-7.25**			**-7.93**			**-6.97**			-7.62		
$\log_{10}(VO)$	**-9.00**	**-8.90**			-11.2			**-11.5**			**-9.36**			**-11.0**			-10.9		
Radius [R_{jup}]	**1.91**	**1.95**	1.92	1.94	1.89	1.93	**1.89**	**1.98**	1.95	2.03	**1.92**	1.97	1.91	**1.98**	1.99	1.94	1.94	1.96	**1.96**
C/O	**0.550**		0.772	0.897	0.48	0.0365	**0.114**		0.862	0.883	**0.90**	0.792	0.266		0.924	0.669	0.88	0.883	**0.523**
$\log_{10}(Z)$	**0.00**		-0.222	0.132	0.02	1.23	**0.800**		-0.214	-0.804	**-0.14**	-0.708	-0.992		0.169	-0.308	0.20	-0.678	**-0.614**
logE	**-**	**4938**	4201	4504	4851	4541	**4922**	**4396**	2152	3086	**4872**	4550	4820	**641**	566	581	761	749	**768**

Table D.4. Retrieval results of **WASP-121 b**. Best retrieval of each configuration is highlighted in bold.

D.3 Temperature profiles

The input temperature profiles at the equator using the 1D and 3D thermal structure, over-plot with the temperature profiles of all the retrieval models are shown here.

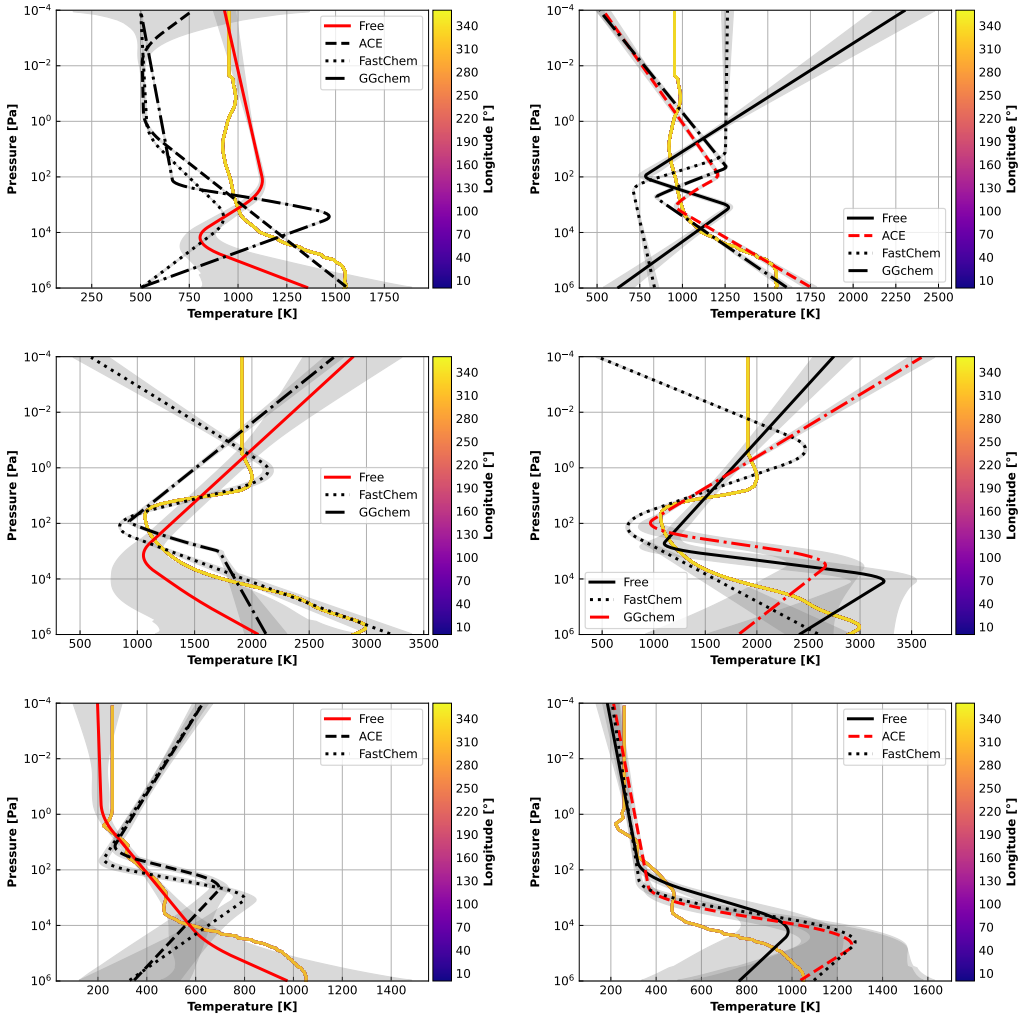


Figure D.1. JWST simulations. **Top to Bottom:** Temperature pressure profiles at the equator assuming 1D atmospheres for HD189733 b, WASP-121 b and GJ1214 b respectively. Substellar point at 180° longitude. **Left panels:** constant input chemistry. **Right:** equilibrium input chemistry. We over-plot the best TP-profiles of Free, ACE, FastChem and GGchem retrievals respectively in black solid, dashed, dotted and dashed-dotted lines. The retrieval having the highest Bayes factor is plotted in red.

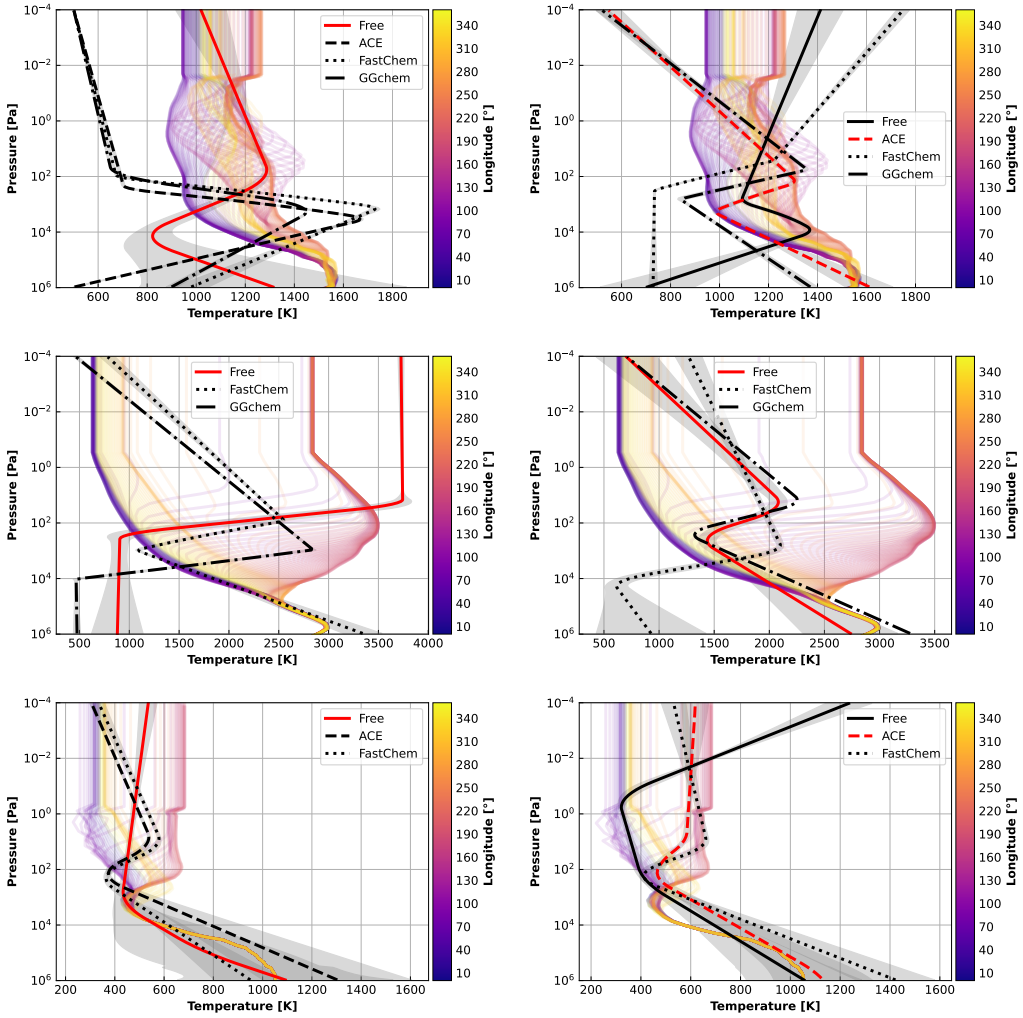


Figure D.2. Same as Figure D.1 using 3D thermal structure for the atmospheres.

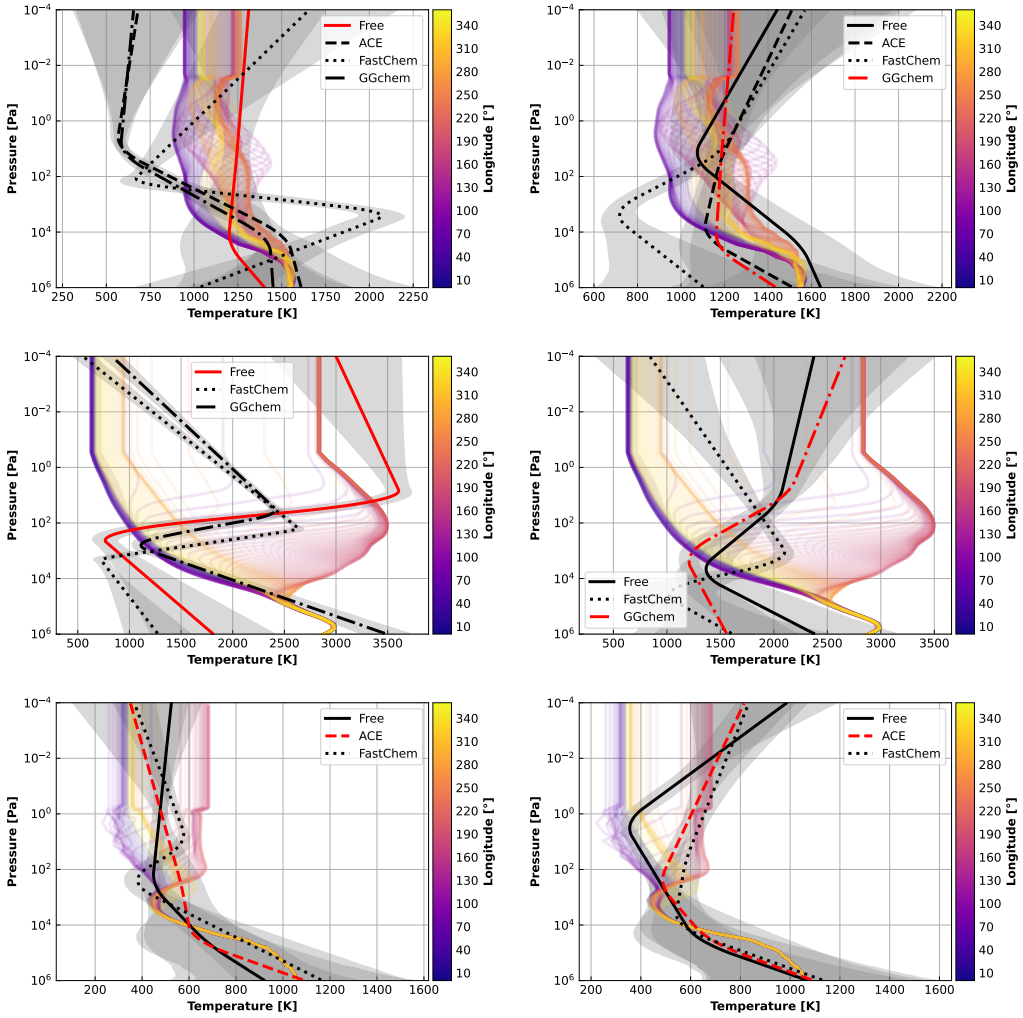


Figure D.3. Same as Figure D.2 for Ariel simulations.

D.4 Abundances profiles

The input abundances profiles at the equator using the 1D and 3D thermal structure, over-plot with the abundances profiles of all the retrieval models are shown here.

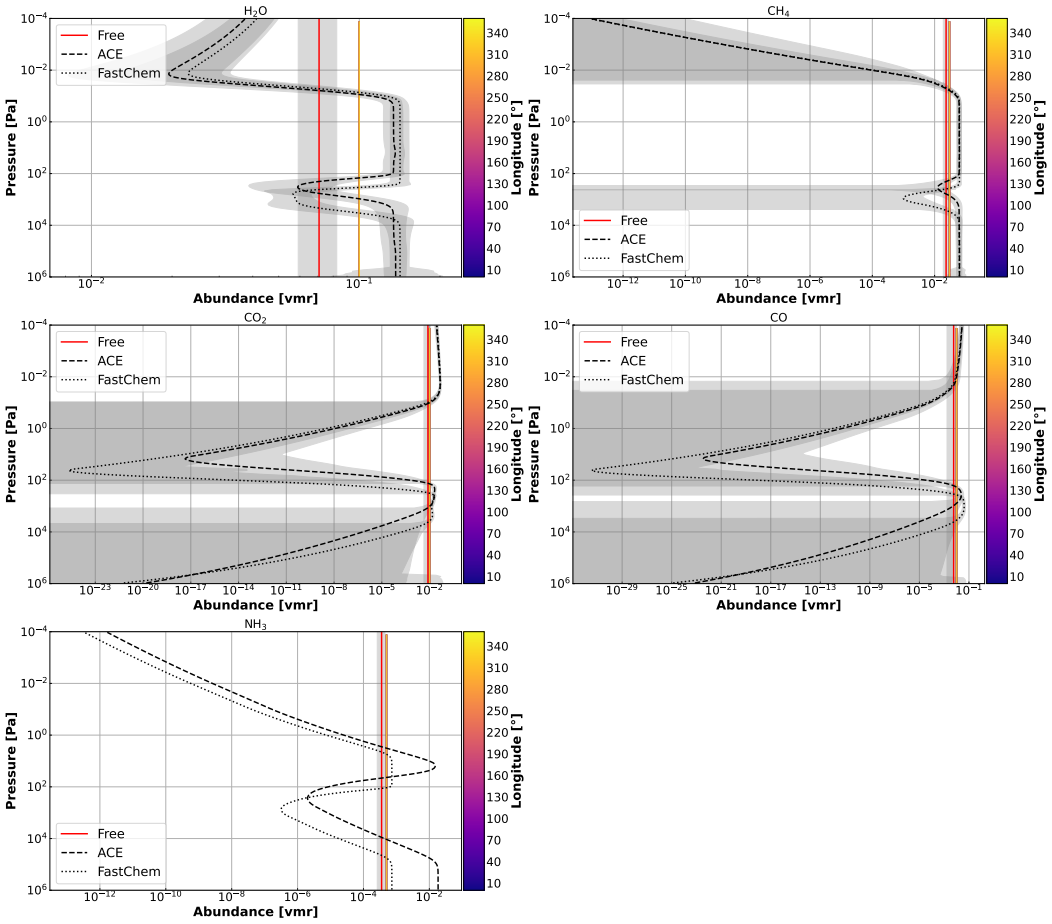


Figure D.4. JWST simulation with constant chemistry assuming 1D atmosphere for the thermal structure. Main absorbers volume mixing ratio (VMR) profiles at the equator on **GJ1214 b**. Substellar point at 180° longitude. We over-plot the best TP-profiles of Free, ACE, FastChem and GGchem retrievals respectively in black solid, dashed, dotted and dashed-dotted lines. The retrieval having the highest Bayes factor is plotted in red.

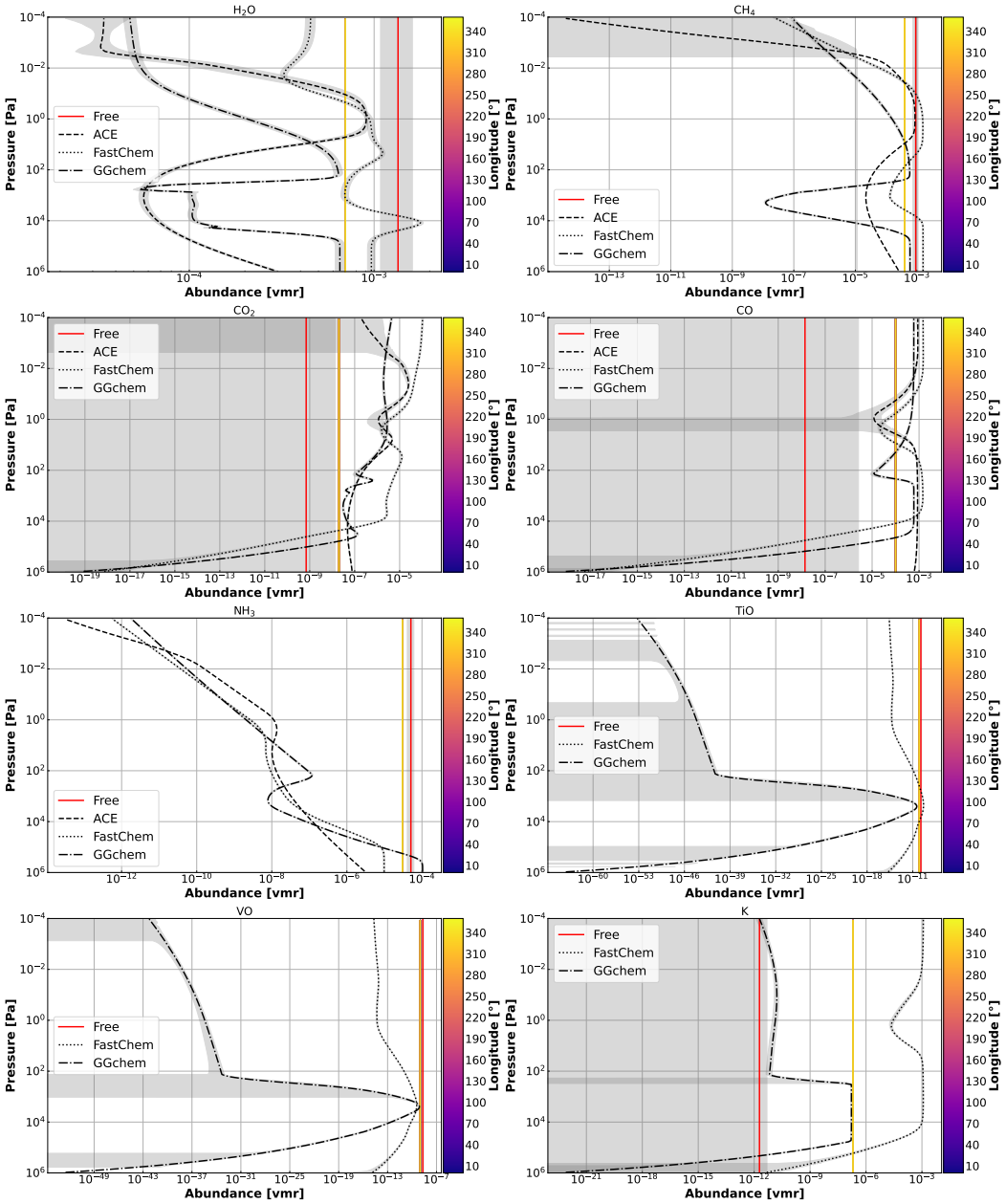


Figure D.5. Same as Figure D.4 for **HD189733 b**.

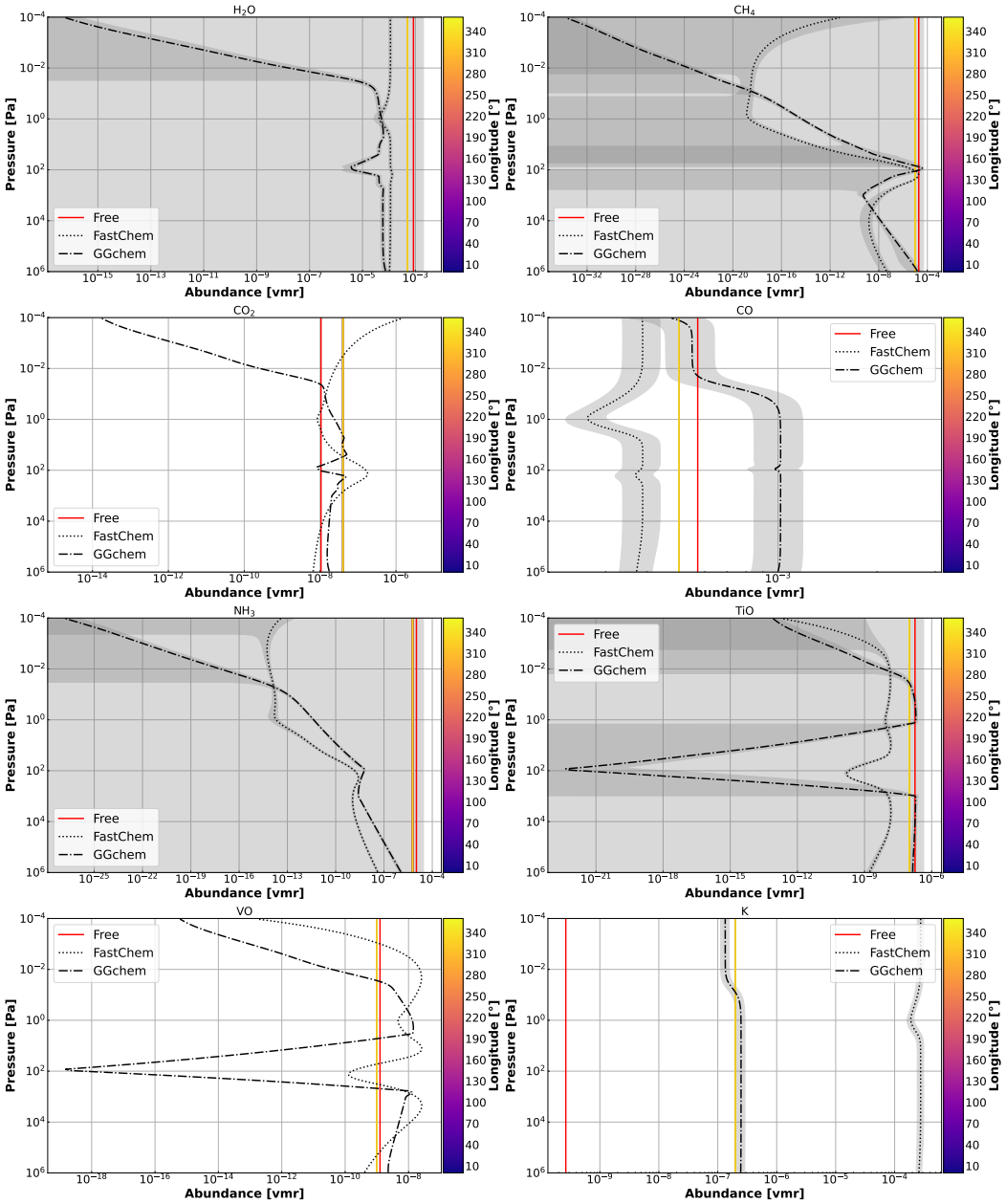


Figure D.6. Same as Figure D.4 for **WASP-121 b**

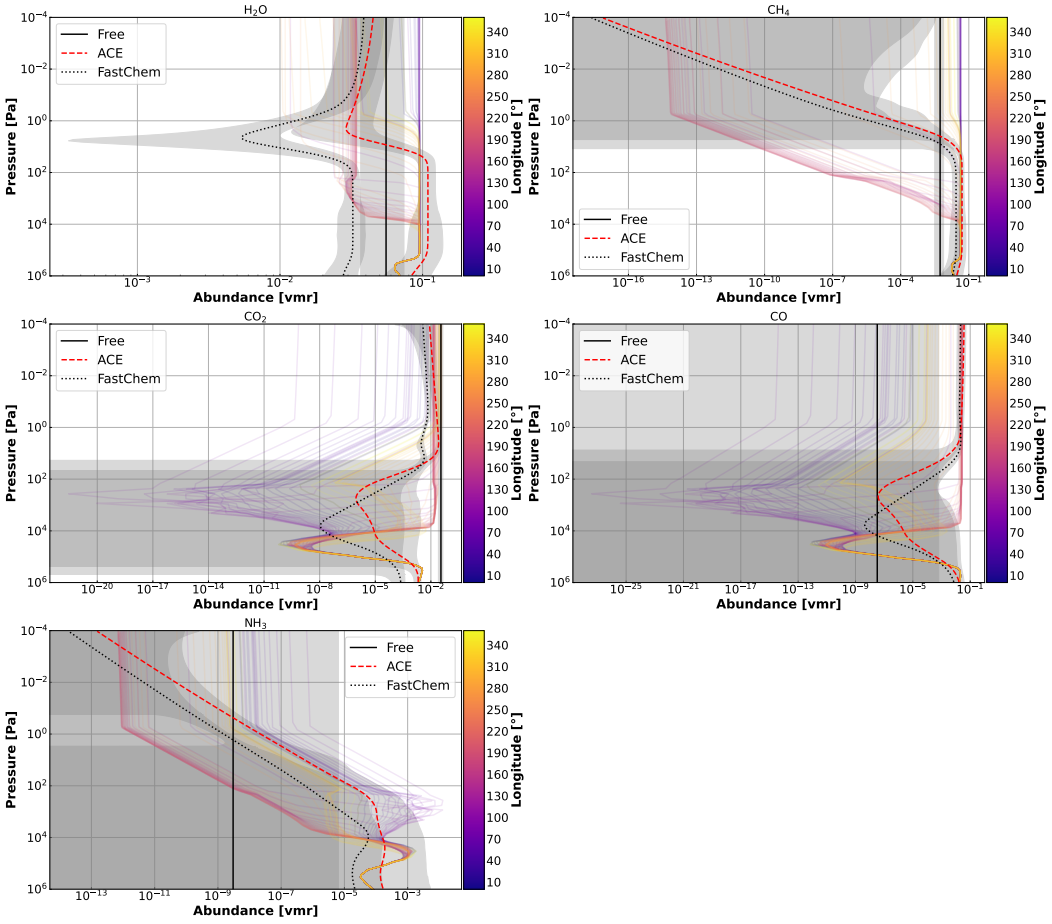


Figure D.7. Same as Figure D.4 for **GJ1214 b** with Ariel simulation assuming 3D atmosphere for the thermal structure.

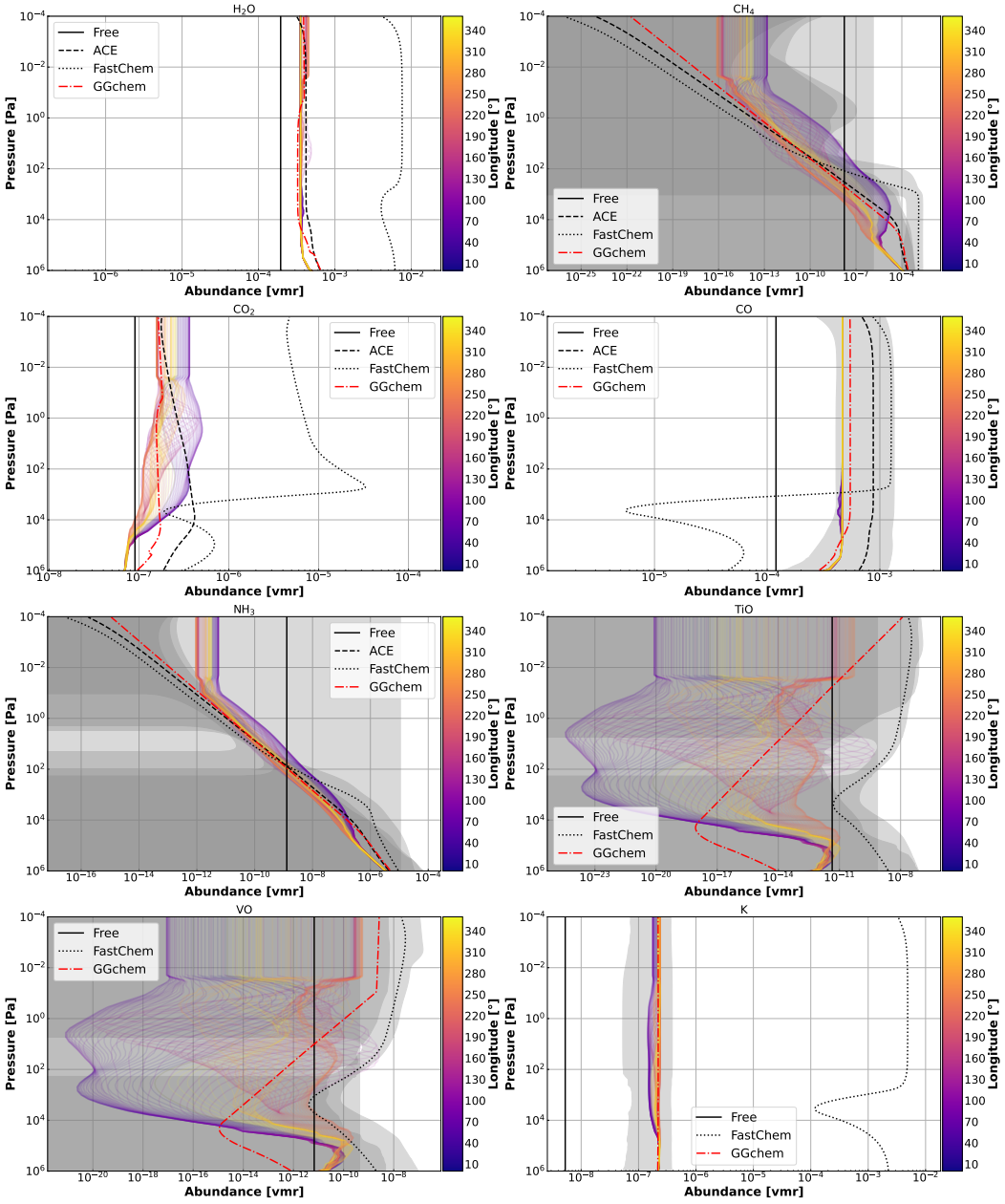


Figure D.8. Same as Figure D.4 for **HD189733 b** with Ariel simulation assuming 3D atmosphere for the thermal structure.

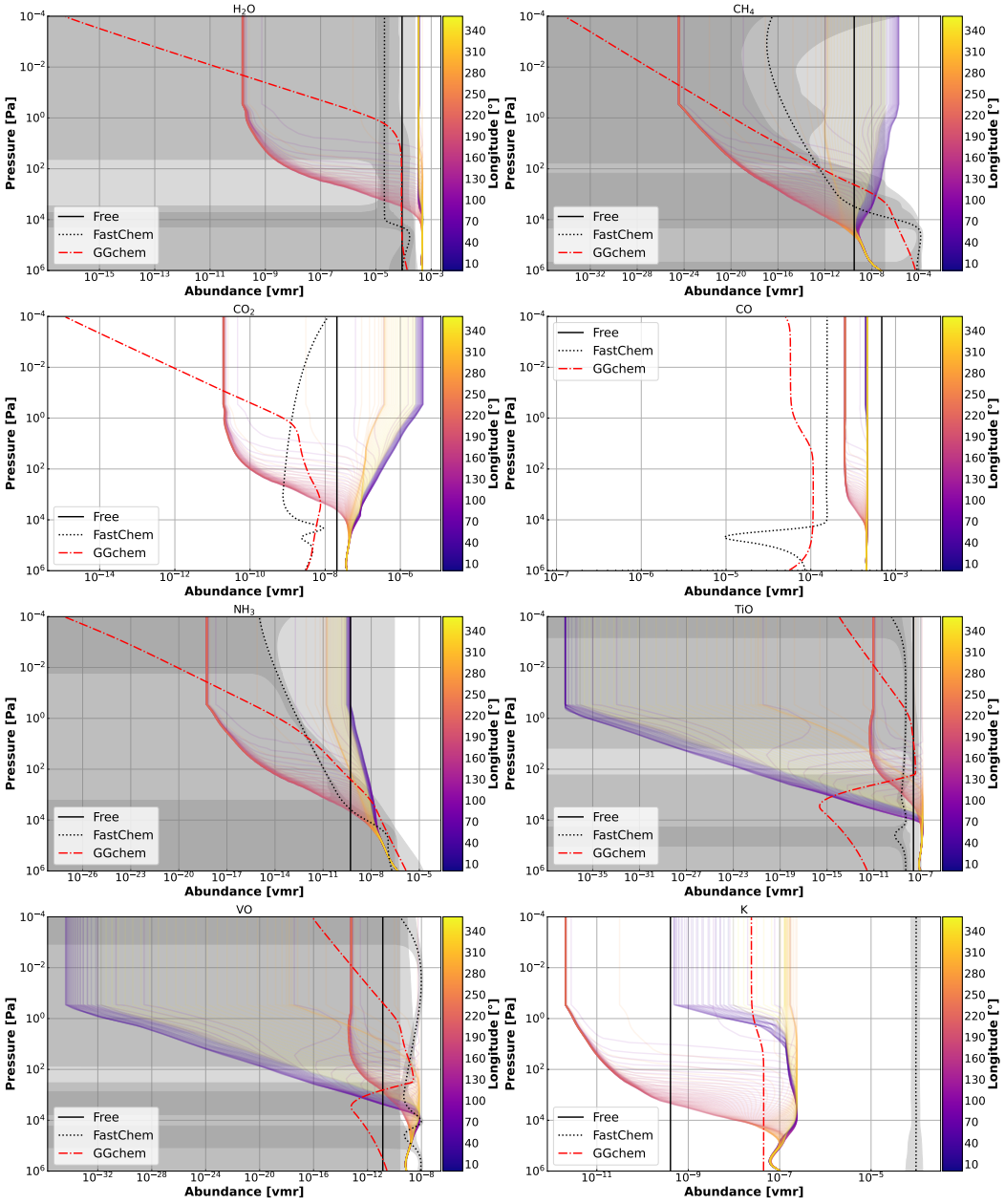


Figure D.9. Same as Figure D.4 for **WASP-121 b** with Ariel simulation assuming 3D atmosphere for the thermal structure.

D.5 Corner plots

The corner plot of the retrievals, over-plot with the input C/O ratio, metallicity (Z) and radius of the planet (R_p) for equilibrium chemistry are shown here.

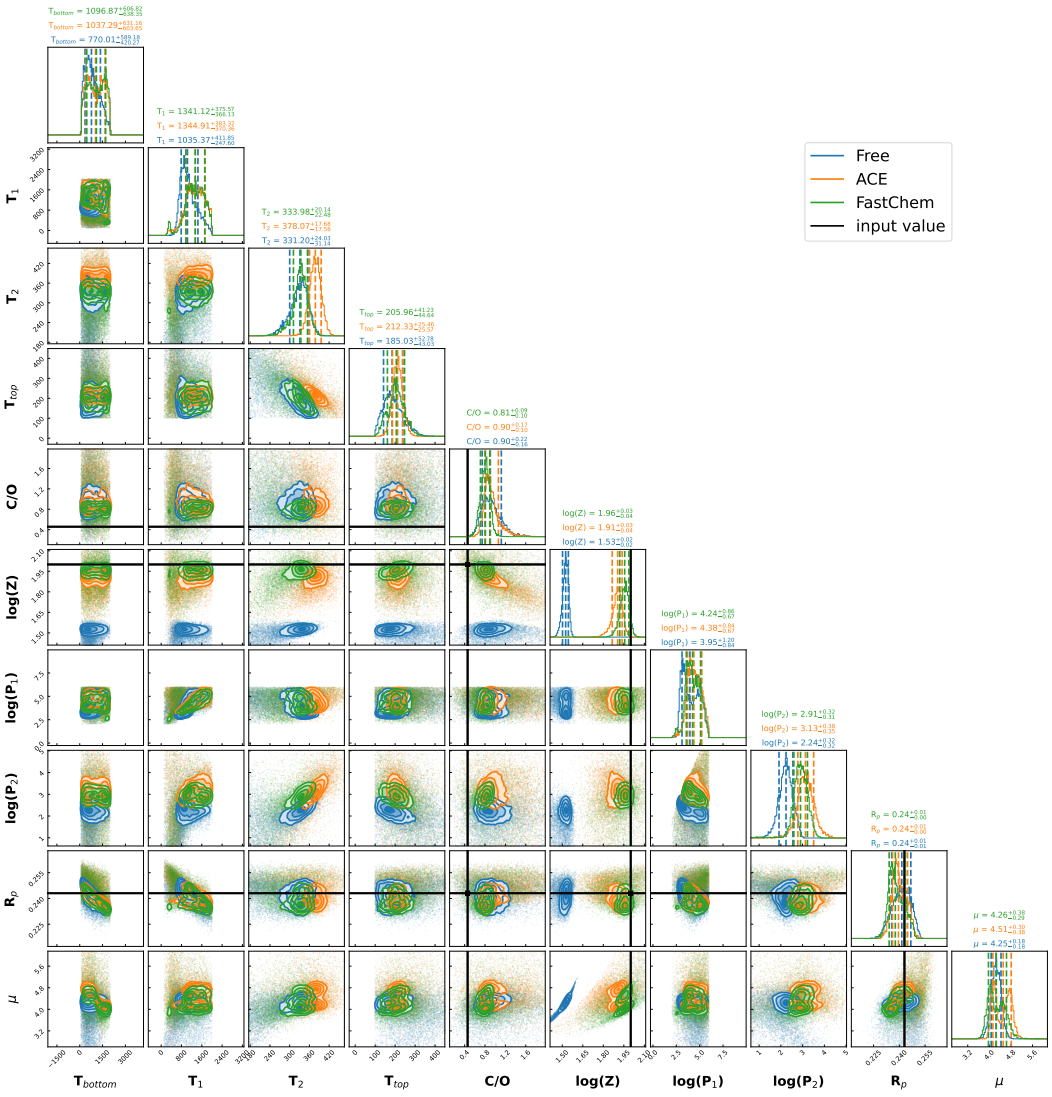


Figure D.10. JWST simulation with equilibrium chemistry on **GJ1214 b** assuming 1D atmosphere for the thermal structure. We over-plot Free, ACE, FastChem and/or GGchem retrievals respectively in blue, orange, green and red. The input value is indicated by the black line for the input C/O ratio, metallicity (Z) and radius of the planet (R_p) values.

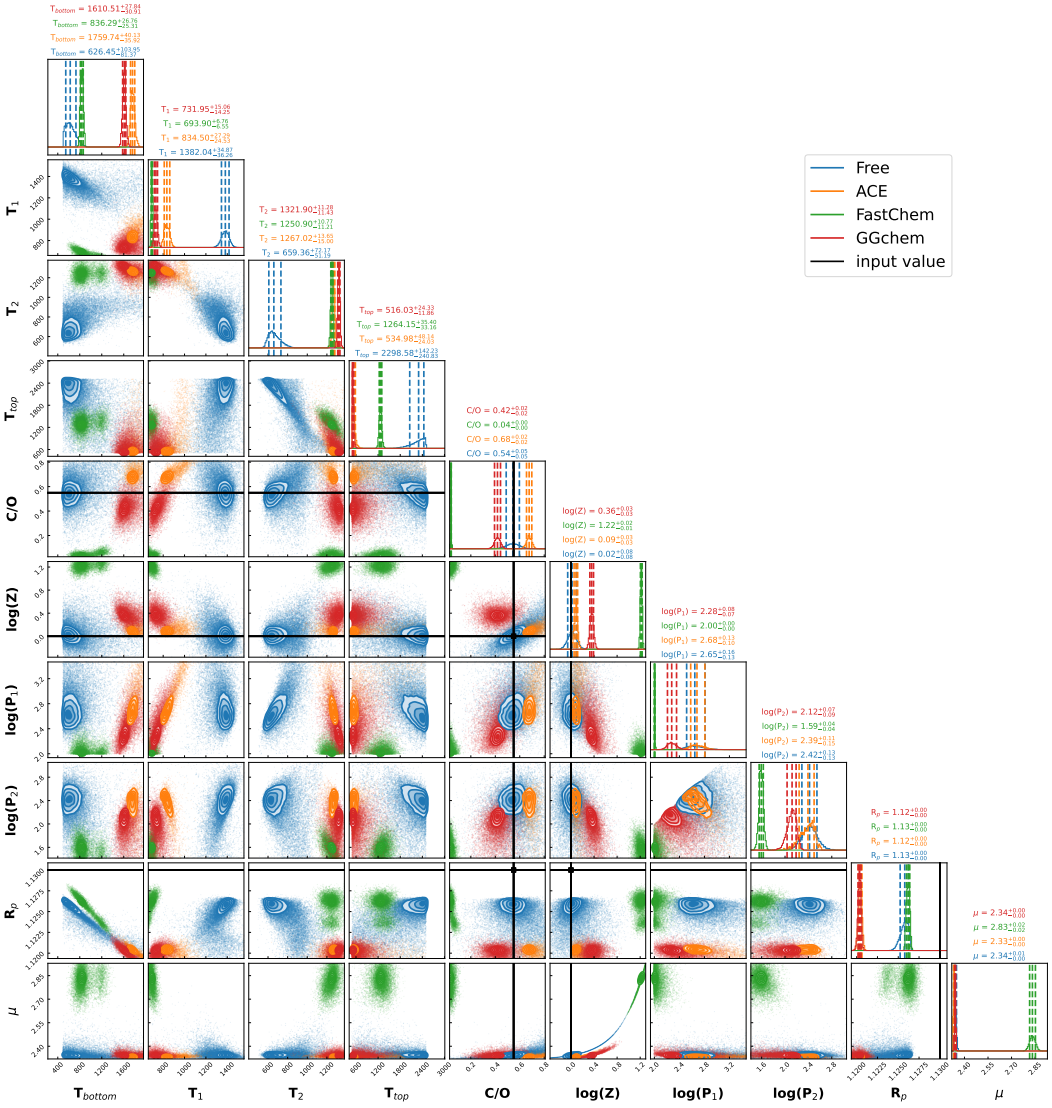


Figure D.11. Same as Figure D.10 for **HD189733 b**.

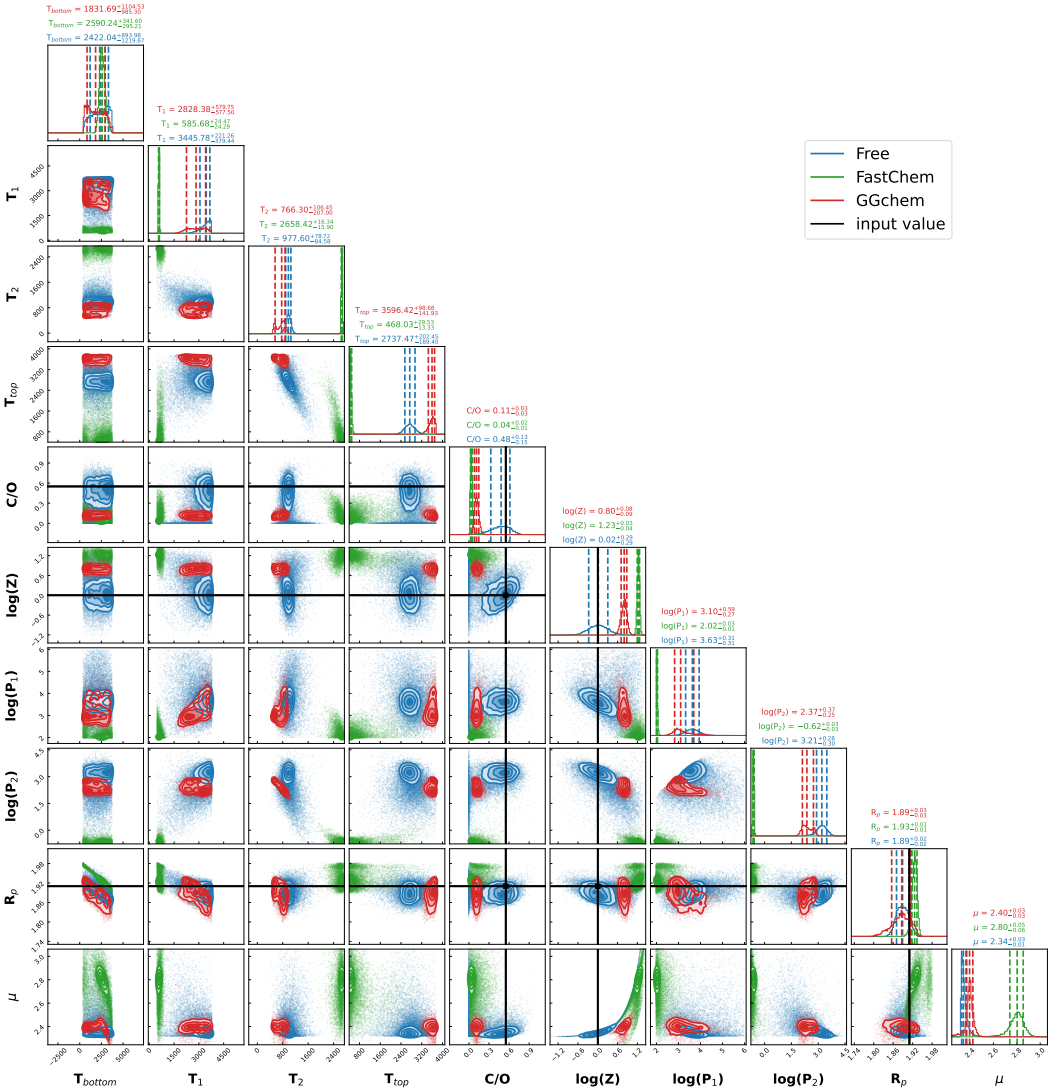


Figure D.12. Same as Figure D.10 for **WASP121 b**.

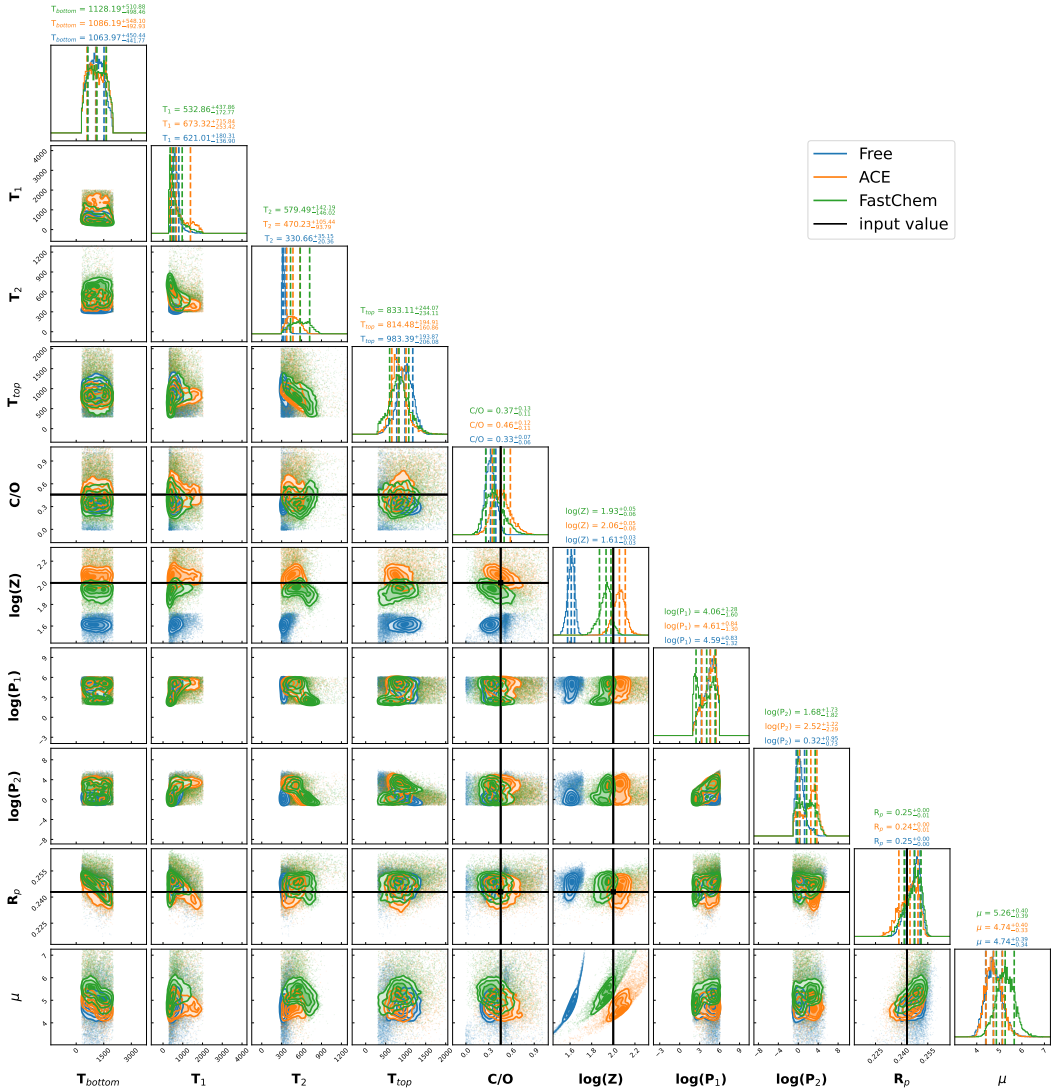


Figure D.13. Same as Figure D.10 for **GJ1214 b** with Ariel simulation assuming 3D atmosphere for the thermal structure.

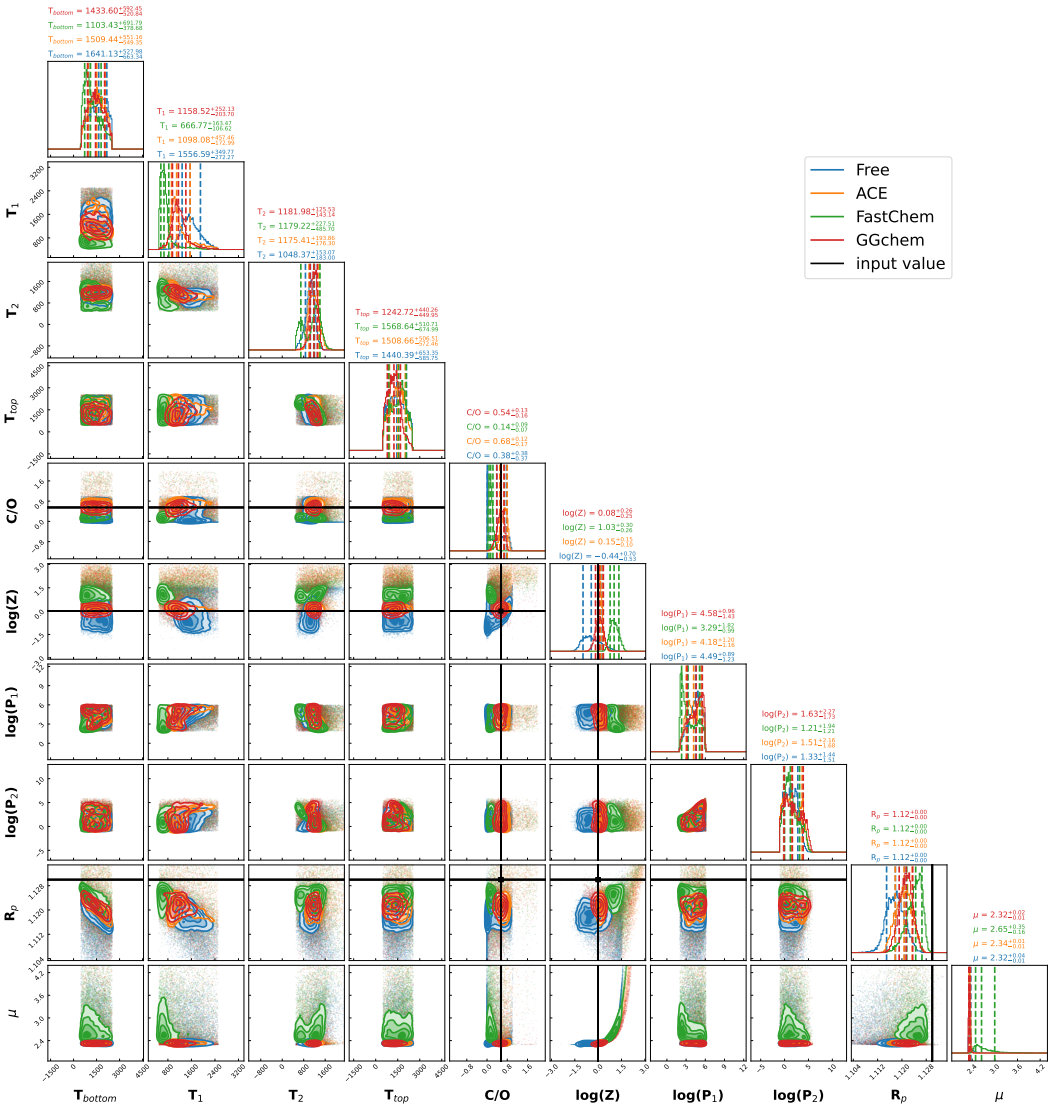


Figure D.14. Same as Figure D.10 for **HD189733 b** with Ariel simulation assuming 3D atmosphere for the thermal structure.

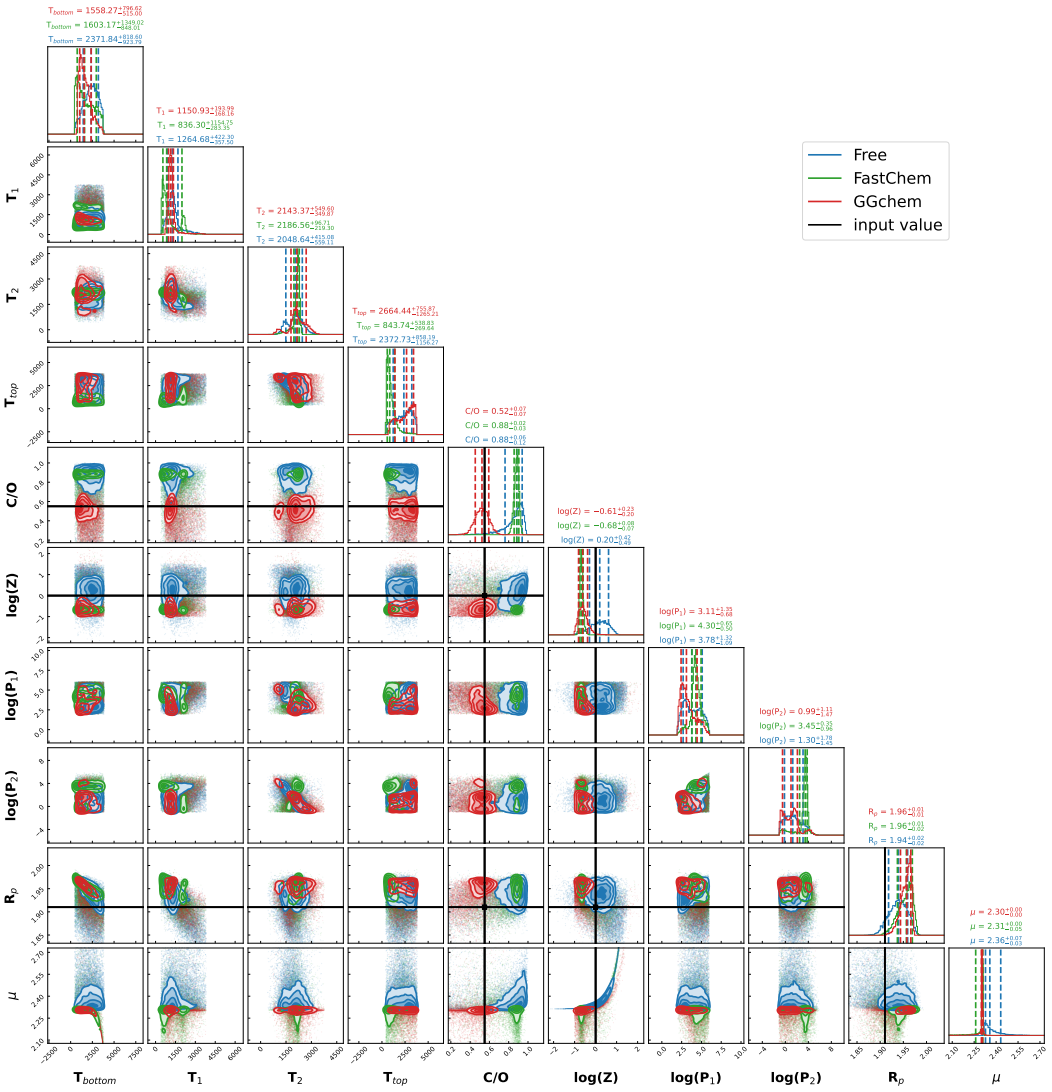


Figure D.15. Same as Figure D.10 for **WASP121 b** with Ariel simulation assuming 3D atmosphere for the thermal structure.

Bibliography

Optical Imaging and Aberrations, Part III: Wavefront Analysis, 2013.

M. Abel, L. Frommhold, X. Li, and et al. Collision-Induced Absorption by H2 Pairs: From Hundreds to Thousands of Kelvin. *Journal of Physical Chemistry A*, 115(25):6805–6812, June 2011. doi: 10.1021/jp109441f.

M. Abel, L. Frommhold, X. Li, and et al. Infrared absorption by collisional H_2-He complexes at temperatures up to 9000 K and frequencies from 0 to 20 000 cm^{-1}. *Journal of Chemical Physics*, 136(4):044319–044319, Jan. 2012. doi: 10.1063/1.3676405.

D. Adams, P. Gao, I. de Pater, and C. V. Morley. Aggregate Hazes in Exoplanet Atmospheres. *Astrophys. J.*, 874(1):61, Mar. 2019. doi: 10.3847/1538-4357/ab074c.

B. Addison, D. J. Wright, R. A. Wittenmyer, and et al. Minerva-Australis. I. Design, Commissioning, and First Photometric Results. *Publ. Astron. Soc. Pac.*, 131(1005): 115003, Nov. 2019. doi: 10.1088/1538-3873/ab03aa.

M. Agúndez, O. Venot, N. Iro, and et al. The impact of atmospheric circulation on the chemistry of the hot Jupiter HD 209458b. *Astron. Astrophys.*, 548:A73, Dec. 2012. doi: 10.1051/0004-6361/201220365.

M. Agúndez, O. Venot, F. Selsis, and et al. The Puzzling Chemical Composition of GJ 436b's Atmosphere: Influence of Tidal Heating on the Chemistry. *Astrophys. J.*, 781(2): 68, Feb. 2014. doi: 10.1088/0004-637X/781/2/68.

M. Agúndez, J. I. Martínez, P. L. de Andres, and et al. Chemical equilibrium in AGB atmospheres: successes, failures, and prospects for small molecules, clusters, and condensates. *Astron. Astrophys.*, 637:A59, May 2020. doi: 10.1051/0004-6361/202037496.

E.-M. Ahrer, K. B. Stevenson, M. Mansfield, and et al. Early Release Science of the exoplanet WASP-39b with JWST NIRCam. *Nature*, 614(7949):653–658, Feb. 2023. doi: 10.1038/s41586-022-05590-4.

A. Al-Refaie, Q. Changeat, O. Venot, and et al. TauREx 3.1 - Extending atmospheric retrieval with plugins. In *European Planetary Science Congress*, pages EPSC2020–669, Sept. 2020. doi: 10.5194/epsc2020-669.

A. F. Al-Refaie, Q. Changeat, I. P. Waldmann, and G. Tinetti. TauREx 3: A Fast, Dynamic, and Extendable Framework for Retrievals. *Astrophys. J.*, 917(1):37, Aug. 2021. doi: 10.3847/1538-4357/ac0252.

A. F. Al-Refaie, Q. Changeat, O. Venot, I. P. Waldmann, and G. Tinetti. A Comparison of Chemical Models of Exoplanet Atmospheres Enabled by TauREx 3.1. *Astrophys. J.*, 932 (2):123, June 2022a. doi: 10.3847/1538-4357/ac6dcd.

A. F. Al-Refaie, O. Venot, Q. Changeat, and et al. FRECKLL: Full and Reduced Exoplanet Chemical Kinetics distiLLed. *arXiv e-prints*, art. arXiv:2209.11203, Sept. 2022b. doi: 10.48550/arXiv.2209.11203.

S. Albrecht, J. N. Winn, J. A. Johnson, and et al. Obliquities of Hot Jupiter Host Stars: Evidence for Tidal Interactions and Primordial Misalignments. *Astrophys. J.*, 757(1):18, Sept. 2012. doi: 10.1088/0004-637X/757/1/18.

S. H. Albrecht, R. I. Dawson, and J. N. Winn. Stellar Obliquities in Exoplanetary Systems. *Publ. Astron. Soc. Pac.*, 134(1038):082001, Aug. 2022. doi: 10.1088/1538-3873/ac6c09.

L. Alderson, H. R. Wakeford, M. K. Alam, and et al. Early Release Science of the exoplanet WASP-39b with JWST NIRSpec G395H. *Nature*, 614(7949):664–669, Feb. 2023. doi: 10.1038/s41586-022-05591-3.

D. Allan. Historicity, strengths, and weaknesses of allan variances and their general applications. *Gyroscopy and Navigation*, 7:1–17, 01 2016. doi: 10.1134/S2075108716010028.

D. W. Allan. Should the classical variance be used as a basic measure in standards metrology? *IEEE Transactions on Instrumentation Measurement*, 36:646–654, June 1987. doi: 10.1109/TIM.1987.6312761.

R. Allart, V. Bourrier, C. Lovis, and et al. Spectrally resolved helium absorption from the extended atmosphere of a warm Neptune-mass exoplanet. *Science*, 362(6421):1384–1387, Dec. 2018. doi: 10.1126/science.aat5879.

D. S. Amundsen, N. J. Mayne, I. Baraffe, and et al. The UK Met Office global circulation model with a sophisticated radiation scheme applied to the hot Jupiter HD 209458b. *Astron. Astrophys.*, 595:A36, Oct. 2016. doi: 10.1051/0004-6361/201629183.

J. Arcangeli, J.-M. Désert, V. Parmentier, K. B. Stevenson, J. L. Bean, M. R. Line, L. Kreidberg, J. J. Fortney, and A. P. Showman. Climate of an ultra hot Jupiter. Spectroscopic phase curve of WASP-18b with HST/WFC3. *Astron. Astrophys.*, 625:A136, May 2019. doi: 10.1051/0004-6361/201834891.

P. J. Armitage and W. K. M. Rice. Planetary migration. *arXiv e-prints*, art. astro-ph/0507492, July 2005.

A. A. A. Azzam, J. Tennyson, S. N. Yurchenko, and et al. ExoMol molecular line lists - XVI. The rotation-vibration spectrum of hot H_2S. *Mon. Not. R. Astron. Soc.*, 460(4): 4063–4074, Aug. 2016. doi: 10.1093/mnras/stw1133.

A. Baglin, M. Auvergne, L. Boisnard, and et al. CoRoT: a high precision photometer for stellar ecolution and exoplanet finding. In *36th COSPAR Scientific Assembly*, volume 36, page 3749, Jan. 2006.

G. Á. Bakos. The HATNet and HATSouth Exoplanet Surveys. In H. J. Deeg and J. A. Belmonte, editors, *Handbook of Exoplanets*, page 111. 2018. doi: 10.1007/978-3-319-55333-7_111.

I. Baraffe, G. Chabrier, and T. Barman. The physical properties of extra-solar planets. *Reports on Progress in Physics*, 73(1):016901, Jan. 2010. doi: 10.1088/0034-4885/73/1/016901.

I. Baraffe, D. Homeier, F. Allard, and et al. New evolutionary models for pre-main sequence and main sequence low-mass stars down to the hydrogen-burning limit. *Astron. Astrophys.*, 577:A42, May 2015. doi: 10.1051/0004-6361/201425481.

K. Barbary. nestle: Nested sampling algorithms for evaluating Bayesian evidence. Astrophysics Source Code Library, record ascl:2103.022, Mar. 2021.

R. J. Barber, J. K. Strange, C. Hill, and et al. ExoMol line lists - III. An improved hot rotation-vibration line list for HCN and HNC. *Mon. Not. R. Astron. Soc.*, 437(2):1828–1835, Jan. 2014. doi: 10.1093/mnras/stt2011.

P. Barge and CoRoT Exoplanet Science Team. The first planets discovered from space with the CoRoT mission. In F. Pont, D. Sasselov, and M. J. Holman, editors, *Transiting Planets*, volume 253, pages 83–89, Feb. 2009. doi: 10.1017/S1743921308026264.

I. Barin and G. PLatzki. *Compilation of Thermochemical Data*, chapter 3, pages 33–34. John Wiley & Sons, Ltd, 1995. ISBN 9783527619825. doi: https://doi.org/10.1002/9783527619825.ch3. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527619825.ch3.

R. Barnes. *Formation and Evolution of Exoplanets*. 2010.

J. K. Barstow, Q. Changeat, R. Garland, and et al. A comparison of exoplanet spectroscopic retrieval tools. *Mon. Not. R. Astron. Soc.*, 493(4):4884–4909, Apr. 2020. doi: 10.1093/mnras/staa548.

E. J. Barton, S. N. Yurchenko, and J. Tennyson. Exomol line lists - ii. the ro-vibrational spectrum of sio. *Mon. Not. R. Astron. Soc.*, 434:1469–1475, Sept. 2013. doi: 10.1093/mnras/stt1105. URL https://ui.adsabs.harvard.edu/abs/2013MNRAS.434.1469B.

E. J. Barton, C. Chiu, S. Golpayegani, S. N. Yurchenko, J. Tennyson, D. J. Frohman, and P. F. Bernath. ExoMol molecular line lists V: the ro-vibrational spectra of NaCl and KCl. *Mon. Not. R. Astron. Soc.*, 442(2):1821–1829, Aug. 2014. doi: 10.1093/mnras/stu944.

E. J. Barton, C. Hill, S. N. Yurchenko, J. Tennyson, A. S. Dudaryonok, and N. N. Lavrentieva. Pressure-dependent water absorption cross sections for exoplanets and other atmospheres. *J. Quant. Spectrosc. Radiat. Transf.*, 187:453–460, Jan. 2017. doi: 10.1016/j.jqsrt.2016.10.024.

M. Bass, C. DeCusatis, J. Enoch, V. Lakshminarayanan, G. Li, C. Macdonald, V. Mahajan, and E. Van Stryland. *Handbook of Optics, Third Edition Volume II: Design, Fabrication and Testing, Sources and Detectors, Radiometry and Photometry*. McGraw-Hill, Inc., USA, 3 edition, 2009. ISBN 0071498907.

N. E. Batalha, A. Mandell, K. Pontoppidan, and et al. PandExo: A Community Tool for Transiting Exoplanet Science with JWST & HST. *Publ. Astron. Soc. Pac.*, 129(976): 064501, June 2017. doi: 10.1088/1538-3873/aa65b0.

N. M. Batalha. Exploring exoplanet populations with NASA's Kepler Mission. *Proceedings of the National Academy of Science*, 111(35):12647–12654, Sept. 2014. doi: 10.1073/pnas.1304196111.

C. Baxter, J.-M. Désert, S.-M. Tsai, and et al. 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. *Astron. Astrophys.*, 648:A127, Apr. 2021. doi: 10.1051/0004-6361/202039708.

J. L. Bean, K. B. Stevenson, N. M. Batalha, and et al. The Transiting Exoplanet Community Early Release Science Program for JWST. *Publ. Astron. Soc. Pac.*, 130(993):114402, Nov. 2018. doi: 10.1088/1538-3873/aadbf3.

B. Benneke. The Elemental Compositions and Cloud Properties of Hot Jupiters: A Comprehensive Atmospheric Retrieval Study of Hot Jupiter Transmission Spectra. In *American Astronomical Society Meeting Abstracts #225*, volume 225 of *American Astronomical Society Meeting Abstracts*, page 107.06, Jan. 2015.

B. Benneke and S. Seager. How to Distinguish between Cloudy Mini-Neptunes and Water/Volatile-dominated Super-Earths. *Astrophys. J.*, 778(2):153, Dec. 2013. doi: 10.1088/0004-637X/778/2/153.

B. Benneke, H. A. Knutson, J. Lothringer, and et al. 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.

A. G. Berezutsky, I. F. Shaikhislamov, M. S. Rumenskikh, and et al. On the transit spectroscopy features of warm Neptunes in the TOI-421 system, revealed with their 3D aeronomy simulations. *Mon. Not. R. Astron. Soc.*, 515(1):706–715, Sept. 2022. doi: 10.1093/mnras/stac1633.

S. M. Birkmann, P. Ferruit, G. Giardino, and et al. The Near-Infrared Spectrograph (NIRSpec) on the James Webb Space Telescope. IV. Capabilities and predicted performance for exoplanet characterization. *Astron. Astrophys.*, 661:A83, May 2022. doi: 10.1051/0004-6361/202142592.

A. Bocchieri, L. V. Mugnai, E. Pascale, Q. Changeat, and G. Tinetti. Detecting molecules in ariel low resolution transmission spectra. *Experimental Astronomy*, Sep 2023. ISSN 1572-9508. doi: 10.1007/s10686-023-09911-x. URL https://doi.org/10.1007/s10686-023-09911-x.

C. F. Bohren and D. R. Huffman. *Absorption and scattering of light by small particles*. 1983.

M. Born, E. Wolf, A. B. Bhatia, P. C. Clemmow, D. Gabor, A. R. Stokes, A. M. Taylor, P. A. Wayman, and W. L. Wilcock. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, Cambridge, 7 edition, 1999. doi: 10.1017/CBO9781139644181.

G. W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950. doi: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

E. O. Brigham and R. E. Morrow. The fast fourier transform. *IEEE Spectrum*, 4(12):63–70, 1967. doi: 10.1109/MSPEC.1967.5217220.

J. Brinkley, Stuart R. Calculation of the Equilibrium Composition of Systems of Many Constituents. *Journal of Chemical Physics*, 15(2):107–110, Feb. 1947. doi: 10.1063/1.1746420.

T. M. Brown. Transmission Spectra as Diagnostics of Extrasolar Giant Planet Atmospheres. *Astrophys. J.*, 553(2):1006–1026, June 2001. doi: 10.1086/320950.

G. Bruno. *De l'Infinito, Universo e Mondi*. 1584.

J. Buchner. Nested Sampling Methods. *arXiv e-prints*, art. arXiv:2101.09675, Jan. 2021a.

J. Buchner. UltraNest - a robust, general purpose Bayesian inference engine. *The Journal of Open Source Software*, 6(60):3001, Apr. 2021b. doi: 10.21105/joss.03001.

A. Burcat, B. Ruscic, Chemistry, and T. I. I. of Tech. Third millenium ideal gas and condensed phase thermochemical database for combustion (with update from active thermochemical tables). *Argonne National Laboratory Technical Note*, (ANL 05/20 TAE 960), 7 2005. doi: 10.2172/925269. URL https://www.osti.gov/biblio/925269.

A. Burrows, I. Hubeny, J. Budaj, and et al. Theoretical Spectral Models of the Planet HD 209458b with a Thermal Inversion and Water Emission Bands. *Astrophys. J. Lett.*, 668(2):L171–L174, Oct. 2007. doi: 10.1086/522834.

A. Burrows, J. Budaj, and I. Hubeny. Theoretical Spectra and Light Curves of Close-in Extrasolar Giant Planets and Comparison with Data. *Astrophys. J.*, 678(2):1436–1457, May 2008. doi: 10.1086/533518.

A. Caldas, J. Leconte, F. Selsis, and et al. Effects of a fully 3D atmospheric structure on exoplanet transmission spectra: retrieval biases due to day-night temperature gradients. *Astron. Astrophys.*, 623:A161, Mar. 2019. doi: 10.1051/0004-6361/201834384.

T. L. Campante, M. N. Lund, J. S. Kuszlewicz, and et al. Spin-Orbit Alignment of Exoplanet Systems: Ensemble Analysis Using Asteroseismology. *Astrophys. J.*, 819(1):85, Mar. 2016. doi: 10.3847/0004-637X/819/1/85.

H. Cao, M. K. Dougherty, G. J. Hunt, and et al. The landscape of Saturn's internal magnetic field from the Cassini Grand Finale. *Icarus*, 344:113541, July 2020. doi: 10.1016/j.icarus.2019.113541.

I. Carleo, S. Desidera, D. Nardiello, and et al. The GAPS Programme at TNG. XXVIII. A pair of hot-Neptunes orbiting the young star TOI-942. *Astron. Astrophys.*, 645:A71, Jan. 2021. doi: 10.1051/0004-6361/202039042.

K. M. S. Cartier, T. G. Beatty, M. Zhao, and et al. Near-infrared Emission Spectrum of WASP-103b Using Hubble Space Telescope/Wide Field Camera 3. *Astron. J.*, 153(1):34, Jan. 2017. doi: 10.3847/1538-3881/153/1/34.

A. Cassan, D. Kubas, J. P. Beaulieu, and et al. One or more bound planets per Milky Way star from microlensing observations. *Nature*, 481(7380):167–169, Jan. 2012. doi: 10.1038/nature10684.

H. M. Cegla, C. Lovis, V. Bourrier, and et al. The Rossiter-McLaughlin effect reloaded: Probing the 3D spin-orbit geometry, differential stellar rotation, and the spatially-resolved stellar spectrum of star-planet systems. *Astron. Astrophys.*, 588:A127, Apr. 2016. doi: 10.1051/0004-6361/201527794.

H. M. Cegla, C. A. Watson, S. Shelyag, and et al. Stellar Surface Magneto-convection as a Source of Astrophysical Noise. II. Center-to-limb Parameterization of Absorption Line Profiles and Comparison to Observations. *Astrophys. J.*, 866(1):55, Oct. 2018. doi: 10.3847/1538-4357/aaddfc.

V. Cessa, T. Beck, W. Benz, and et al. CHEOPS: a space telescope for ultra-high precision photometry of exoplanet transits. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 10563 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 105631L, Nov. 2017. doi: 10.1117/12.2304164.

Y. Chachan, H. A. Knutson, P. Gao, and et al. A Hubble PanCET Study of HAT-P-11b: A Cloudy Neptune with a Low Atmospheric Metallicity. *Astron. J.*, 158(6):244, Dec. 2019. doi: 10.3847/1538-3881/ab4e9a.

Q. Changeat. On Spectroscopic Phase-curve Retrievals: H_2 Dissociation and Thermal Inversion in the Atmosphere of the Ultrahot Jupiter WASP-103 b. *Astron. J.*, 163(3):106, Mar. 2022. doi: 10.3847/1538-3881/ac4475.

Q. Changeat, B. Edwards, I. P. Waldmann, and et al. Toward a More Complex Description of Chemical Profiles in Exoplanet Retrievals: A Two-layer Parameterization. *Astrophys. J.*, 886(1):39, Nov. 2019. doi: 10.3847/1538-4357/ab4a14.

Q. Changeat, A. Al-Refaie, L. V. Mugnai, and et al. Alfnoor: A Retrieval Simulation of the Ariel Target List. *Astron. J.*, 160(2):80, Aug. 2020a. doi: 10.3847/1538-3881/ab9a53.

Q. Changeat, L. Keyte, I. P. Waldmann, and G. Tinetti. Impact of Planetary Mass Uncertainties on Exoplanet Atmospheric Retrievals. *Astrophys. J.*, 896(2):107, June 2020b. doi: 10.3847/1538-4357/ab8f8b.

Q. Changeat, B. Edwards, A. F. Al-Refaie, and et al. Five Key Exoplanet Questions Answered via the Analysis of 25 Hot-Jupiter Atmospheres in Eclipse. *Astrophys. J. Suppl. Ser.*, 260(1):3, May 2022. doi: 10.3847/1538-4365/ac5cc2.

D. Charbonneau, T. M. Brown, R. W. Noyes, and et al. Detection of an Extrasolar Planet Atmosphere. *Astrophys. J.*, 568(1):377–384, Mar. 2002. doi: 10.1086/338770.

D. Charbonneau, L. E. Allen, S. T. Megeath, and et al. Detection of Thermal Emission from an Extrasolar Planet. *Astrophys. J.*, 626(1):523–529, June 2005. doi: 10.1086/429991.

D. Charbonneau, Z. K. Berta, J. Irwin, C. J. Burke, P. Nutzman, L. A. Buchhave, C. Lovis, X. Bonfils, D. W. Latham, S. Udry, R. A. Murray-Clay, M. J. Holman, E. E. Falco, J. N. Winn, D. Queloz, F. Pepe, M. Mayor, X. Delfosse, and T. Forveille. A super-Earth transiting a nearby low-mass star. *Nature*, 462(7275):891–894, Dec. 2009. doi: 10.1038/nature08679.

B. Charnay, V. Meadows, A. Misra, J. Leconte, and G. Arney. 3D Modeling of GJ1214b's Atmosphere: Formation of Inhomogeneous High Clouds and Observational Implications. *Astrophys. J. Lett.*, 813(1):L1, Nov. 2015. doi: 10.1088/2041-8205/813/1/L1.

M. Chase. *NIST-JANAF Thermochemical Tables, 4th Edition*. American Institute of Physics, -1, Aug. 1998.

S. Chatterjee, E. B. Ford, S. Matsumura, and et al. Dynamical Outcomes of Planet-Planet Scattering. *Astrophys. J.*, 686(1):580–602, Oct. 2008. doi: 10.1086/590227.

P. Chioetto. Optical and Opto-mechanical Analysis and Design of the Telescope for the Ariel Mission. *arXiv e-prints*, art. arXiv:2212.03686, Dec. 2022. doi: 10.48550/arXiv.2212.03686.

N. Choksi and E. Chiang. Sub-Neptune formation: the view from resonant planets. *Mon. Not. R. Astron. Soc.*, 495(4):4192–4209, July 2020. doi: 10.1093/mnras/staa1421.

E. Church and P. Takacs. Optimal estimation of finish parameters. *Ed. SPIE*, 1530, 12 1991. doi: 10.1117/12.50498.

E. L. Church, P. Z. Takacs, and T. A. Leonard. The Prediction Of BRDFs From Surface Profile Measurements. In J. C. Stover, editor, *Scatter from Optical Components*, volume 1165 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 136–150, Jan. 1990. doi: 10.1117/12.962842.

R. Cloutier, D. Charbonneau, D. Deming, X. Bonfils, and N. Astudillo-Defru. A More Precise Mass for GJ 1214 b and the Frequency of Multiplanet Systems Around Mid-M Dwarfs. *Astron. J.*, 162(5):174, Nov. 2021. doi: 10.3847/1538-3881/ac1584.

A. Collier Cameron, V. A. Bruce, G. R. M. Miller, and et al. Line-profile tomography of exoplanet transits - I. The Doppler shadow of HD 189733b. *Mon. Not. R. Astron. Soc.*, 403(1):151–158, Mar. 2010. doi: 10.1111/j.1365-2966.2009.16131.x.

S. Constantinou, N. Madhusudhan, and S. Gandhi. Early Insights for Atmospheric Retrievals of Exoplanets using JWST Transit Spectroscopy. *arXiv e-prints*, art. arXiv:2301.02564, Jan. 2023.

C. S. Cooper and A. P. Showman. Dynamics and Disequilibrium Carbon Chemistry in Hot Jupiter Atmospheres, with Application to HD 209458b. *Astrophys. J.*, 649(2):1048–1063, Oct. 2006. doi: 10.1086/506312.

G. Cracchiolo, G. Micela, G. Morello, and G. Peres. Correcting the effect of stellar spots on ARIEL transmission spectra - II. The limb-darkening effect. *Mon. Not. R. Astron. Soc.*, 507(4):6118–6131, Nov. 2021a. doi: 10.1093/mnras/stab2509.

G. Cracchiolo, G. Micela, and G. Peres. Correcting the effect of stellar spots on ARIEL transmission spectra. *Mon. Not. R. Astron. Soc.*, 501(2):1733–1747, Feb. 2021b. doi: 10.1093/mnras/staa3621.

A. J. Cridland, R. E. Pudritz, and M. Alessi. Composition of early planetary atmospheres - I. Connecting disc astrochemistry to the formation of planetary atmospheres. *Mon. Not. R. Astron. Soc.*, 461(3):3274–3295, Sept. 2016. doi: 10.1093/mnras/stw1511.

A. J. Cridland, E. F. van Dishoeck, M. Alessi, and et al. Connecting planet formation and astrochemistry. A main sequence for C/O in hot exoplanetary atmospheres. *Astron. Astrophys.*, 632:A63, Dec. 2019. doi: 10.1051/0004-6361/201936105.

T. J. David, L. A. Hillenbrand, E. A. Petigura, and et al. A Neptune-sized transiting planet closely orbiting a 5-10-million-year-old star. *Nature*, 534(7609):658–661, June 2016. doi: 10.1038/nature18293.

R. I. Dawson and J. A. Johnson. Origins of Hot Jupiters. *Annu. Rev. Astron. Astrophys.*, 56: 175–221, Sept. 2018. doi: 10.1146/annurev-astro-081817-051853.

H. J. Deeg and J. A. Belmonte. *Handbook of Exoplanets*. 2018. doi: 10.1007/978-3-319-55333-7.

L. Delrez, A. Santerne, J. M. Almenara, D. R. Anderson, A. Collier-Cameron, R. F. Díaz, M. Gillon, C. Hellier, E. Jehin, M. Lendl, P. F. L. Maxted, M. Neveu-VanMalle, F. Pepe, D. Pollacco, D. Queloz, D. Ségransan, B. Smalley, A. M. S. Smith, A. H. M. J. Triaud, S. Udry, V. Van Grootel, and R. G. West. WASP-121 b: a hot Jupiter close to tidal disruption transiting an active F star. *Mon. Not. R. Astron. Soc.*, 458(4):4025–4043, June 2016. doi: 10.1093/mnras/stw522.

O. D. S. Demangeon, F. Faedi, G. Hébrard, and et al. The discovery of WASP-151b, WASP-153b, WASP-156b: Insights on giant planet migration and the upper boundary of the Neptunian desert. *Astron. Astrophys.*, 610:A63, Mar. 2018. doi: 10.1051/0004-6361/201731735.

D. Deming, S. Seager, L. J. Richardson, and et al. Infrared radiation from an extrasolar planet. *Nature*, 434(7034):740–743, Mar. 2005. doi: 10.1038/nature03507.

L. D. Deming and S. Seager. Illusion and reality in the atmospheres of exoplanets. *Journal of Geophysical Research (Planets)*, 122(1):53–75, Jan. 2017. doi: 10.1002/2016JE005155.

J.-M. Désert, A. Lecavelier des Etangs, G. Hébrard, and et al. Search for Carbon Monoxide in the Atmosphere of the Transiting Exoplanet HD 189733b. *Astrophys. J.*, 699(1): 478–485, July 2009. doi: 10.1088/0004-637X/699/1/478.

B. Drummond, N. J. Mayne, J. Manners, I. Baraffe, J. Goyal, P. Tremblin, D. K. Sing, and K. Kohary. The 3D Thermal, Dynamical, and Chemical Structure of the Atmosphere of HD 189733b: Implications of Wind-driven Chemistry for the Emission Phase Curve. *Astrophys. J.*, 869(1):28, Dec. 2018. doi: 10.3847/1538-4357/aaeb28.

B. Edwards and G. Tinetti. The Ariel Target List: The Impact of TESS and the Potential for Characterizing Multiple Planets within a System. *Astron. J.*, 164(1):15, July 2022. doi: 10.3847/1538-3881/ac6bf9.

B. Edwards, L. Mugnai, G. Tinetti, E. Pascale, and S. Sarkar. An Updated Study of Potential Targets for Ariel. *Astron. J.*, 157(6):242, June 2019. doi: 10.3847/1538-3881/ab1cb9.

B. Edwards, Q. Changeat, R. Baeyens, and et al. ARES I: WASP-76 b, A Tale of Two HST Spectra. *Astron. J.*, 160(1):8, July 2020. doi: 10.3847/1538-3881/ab9225.

T. Encrenaz, G. Tinetti, M. Tessenyi, and et al. Transit spectroscopy of exoplanets from space: how to optimize the wavelength coverage and spectral resolving power. *Experimental Astronomy*, 40(2-3):523–543, Dec. 2015. doi: 10.1007/s10686-014-9415-0.

ESA. ESA's Cosmic Vision, 2015.

M. Ester, H.-P. Kriegel, J. Sander, and et al. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Second International Conference on Knowledge Discovery and Data Mining (KDD'96). Proceedings of a conference held August 2-4*, pages 226–331, Jan. 1996.

A. Falco, T. Zingales, W. Pluriel, and J. Leconte. Toward a multidimensional analysis of transmission spectroscopy. I. Computation of transmission spectra using a 1D, 2D, or 3D atmosphere structure. *Astron. Astrophys.*, 658:A41, Feb. 2022. doi: 10.1051/0004-6361/202141940.

F. Feroz and M. P. Hobson. Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *Mon. Not. R. Astron. Soc.*, 384(2):449–463, Feb. 2008. doi: 10.1111/j.1365-2966.2007.12353.x.

F. Feroz, M. P. Hobson, and M. Bridges. MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Mon. Not. R. Astron. Soc.*, 398(4):1601–1614, Oct. 2009. doi: 10.1111/j.1365-2966.2009.14548.x.

F. Feroz, M. P. Hobson, E. Cameron, and et al. Importance Nested Sampling and the MultiNest Algorithm. *The Open Journal of Astrophysics*, 2(1):10, Nov. 2019. doi: 10.21105/astro.1306.2144.

D. A. Fischer and J. Valenti. The Planet-Metallicity Correlation. *Astrophys. J.*, 622(2):1102–1117, Apr. 2005. doi: 10.1086/428383.

L. N. Fletcher, M. Gustafsson, and G. S. Orton. Hydrogen Dimers in Giant-planet Infrared Spectra. *Astrophys. J. Suppl. Ser.*, 235(1):24, Mar. 2018. doi: 10.3847/1538-4365/aaa07a.

A. Fortier, T. Beck, W. Benz, and et al. CHEOPS: a space telescope for ultra-high precision photometry of exoplanet transits. In J. Oschmann, Jacobus M., M. Clampin, G. G. Fazio, and et al., editors, *Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave*, volume 9143 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 91432J, Aug. 2014. doi: 10.1117/12.2056687.

J. J. Fortney, K. Lodders, M. S. Marley, and R. S. Freedman. A Unified Theory for the Atmospheres of the Hot and Very Hot Jupiters: Two Classes of Irradiated Atmospheres. *Astrophys. J.*, 678(2):1419–1435, May 2008. doi: 10.1086/528370.

J. J. Fortney, M. S. Marley, G. Laughlin, and et al. The Hunt for Planet Nine: Atmosphere, Spectra, Evolution, and Detectability. *Astrophys. J. Lett.*, 824(2):L25, June 2016. doi: 10.3847/2041-8205/824/2/L25.

J. J. Fortney, C. Visscher, M. S. Marley, and et al. Beyond Equilibrium Temperature: How the Atmosphere/Interior Connection Affects the Onset of Methane, Ammonia, and Clouds in Warm Transiting Giant Planets. *Astron. J.*, 160(6):288, Dec. 2020. doi: 10.3847/1538-3881/abc5bd.

K. France, R. O. P. Loyd, A. Youngblood, and et al. The MUSCLES Treasury Survey. I. Motivation and Overview. *Astrophys. J.*, 820(2):89, Apr. 2016. doi: 10.3847/0004-637X/820/2/89.

G. Fu, N. Espinoza, D. K. Sing, and et al. Water and an Escaping Helium Tail Detected in the Hazy and Methane-depleted Atmosphere of HAT-P-18b from JWST NIRISS/SOSS. *Astrophys. J. Lett.*, 940(2):L35, Dec. 2022. doi: 10.3847/2041-8213/ac9977.

B. J. Fulton and E. A. Petigura. The California-Kepler Survey. VII. Precise Planet Radii Leveraging Gaia DR2 Reveal the Stellar Mass Dependence of the Planet Radius Gap. *Astron. J.*, 156(6):264, Dec. 2018. doi: 10.3847/1538-3881/aae828.

B. J. Fulton, E. A. Petigura, A. W. Howard, and et al. The California-Kepler Survey. III. A Gap in the Radius Distribution of Small Planets. *Astron. J.*, 154(3):109, Sept. 2017. doi: 10.3847/1538-3881/aa80eb.

Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, and et al. The Gaia mission. *Astron. Astrophys.*, 595:A1, Nov. 2016. doi: 10.1051/0004-6361/201629272.

E. Galanti, Y. Kaspi, F. J. Simons, D. Durante, M. Parisi, and S. J. Bolton. Determining the Depth of Jupiter's Great Red Spot with Juno: A Slepian Approach. *Astrophys. J. Lett.*, 874(2):L24, Apr. 2019. doi: 10.3847/2041-8213/ab1086.

S. Gandhi and N. Madhusudhan. GENESIS: new self-consistent models of exoplanetary spectra. *mnras*, 472(2):2334–2355, Dec. 2017. doi: 10.1093/mnras/stx1601.

M. Gillon, A. H. M. J. Triaud, B.-O. Demory, and et al. Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1. *Nature*, 542(7642):456–460, Feb. 2017. doi: 10.1038/nature21360.

S. Ginzburg and R. Sari. Tidal heating of young super-Earth atmospheres. *Mon. Not. R. Astron. Soc.*, 464(4):3937–3944, Feb. 2017. doi: 10.1093/mnras/stw2637.

S. Ginzburg, H. E. Schlichting, and R. Sari. Core-powered mass-loss and the radius distribution of small exoplanets. *Mon. Not. R. Astron. Soc.*, 476(1):759–765, May 2018. doi: 10.1093/mnras/sty290.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

J. Goodman and J. Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80, Jan. 2010. doi: 10.2140/camcos.2010.5.65.

J. W. Goodman. *Introduction to Fourier optics*. 2005.

E. Goos, A. Burcat, and B. Ruscic. Extended third millenium ideal gas thermochemical database with updates from active thermochemical tables, 2016.

I. E. Gordon, L. S. Rothman, C. Hill, and et al. The HITRAN2016 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.*, 203:3–69, Dec. 2017. doi: 10.1016/j.jqsrt.2017.06.038.

S. Gordon and B. McBride. *Computer program for calculation of complex chemical equilibrium compositions and applications. Part 1 : Analysis*. NASA, 1994.

T. P. Greene, M. R. Line, C. Montero, and et al. Characterizing Transiting Exoplanet Atmospheres with JWST. *Astrophys. J.*, 817(1):17, Jan. 2016. doi: 10.3847/0004-637X/817/1/17.

S. L. Grimm, B.-O. Demory, M. Gillon, and et al. The nature of the TRAPPIST-1 exoplanets. *Astron. Astrophys.*, 613:A68, May 2018. doi: 10.1051/0004-6361/201732233.

A. Guzmán-Mesa, D. Kitzmann, C. Mordasini, and et al. Chemical diversity of the atmospheres and interiors of sub-Neptunes: a case study of GJ 436 b. *Mon. Not. R. Astron. Soc.*, 513(3):4015–4036, July 2022. doi: 10.1093/mnras/stac1066.

J. Harrington, M. D. Himes, P. E. Cubillos, and et al. The Bayesian Atmospheric Radiative Transfer (BART) Code: 3D Mapping and Machine Learning. In *American Astronomical Society Meeting Abstracts #235*, volume 235 of *American Astronomical Society Meeting Abstracts*, page 173.13, Jan. 2020.

Y. Hasegawa and R. E. Pudritz. Planet Traps and Planetary Cores: Origins of the Planet-Metallicity Correlation. *Astrophys. J.*, 794(1):25, Oct. 2014. doi: 10.1088/0004-637X/794/1/25.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York, New York, NY, second edi edition, 2009. ISBN 978-0-387-84857-0. doi: 10.1007/978-0-387-84858-7. URL http://link.springer.com/10.1007/978-0-387-84858-7.

W. K. Hastings. Monte Carlo Sampling Methods using Markov Chains and their Applications. *Biometrika*, 57(1):97–109, Apr. 1970. doi: 10.1093/biomet/57.1.97.

C. He, M. Radke, S. E. Moran, and et al. Optical Properties of Organic Hazes in Water-rich Exoplanet Atmospheres: Implications for Observations with JWST. *arXiv e-prints*, art. arXiv:2301.02745, Jan. 2023. doi: 10.48550/arXiv.2301.02745.

C. Helling. Exoplanet Clouds. *Annual Review of Earth and Planetary Sciences*, 47:583–606, May 2019. doi: 10.1146/annurev-earth-053018-060401.

C. Helling, Y. Kawashima, V. Graham, and et al. Mineral cloud and hydrocarbon haze particles in the atmosphere of the hot Jupiter JWST target WASP-43b. *Astron. Astrophys.*, 641:A178, Sept. 2020. doi: 10.1051/0004-6361/202037633.

C. Hill, S. N. Yurchenko, and J. Tennyson. Temperature-dependent molecular absorption cross sections for exoplanets and other atmospheres. *Icarus*, 226(2):1673–1677, Nov. 2013. doi: 10.1016/j.icarus.2012.07.028.

T. Hirano, Y. Suto, J. N. Winn, A. Taruya, N. Narita, S. Albrecht, and B. Sato. Improved Modeling of the Rossiter-McLaughlin Effect for Transiting Exoplanets. *Astrophys. J.*, 742(2):69, Dec. 2011. doi: 10.1088/0004-637X/742/2/69.

S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9 (8):1735–1780, 11 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL https://doi.org/10.1162/neco.1997.9.8.1735.

K. Hou Yip, Q. Changeat, N. Nikolaou, M. Morvan, B. Edwards, I. P. Waldmann, and G. Tinetti. Peeking inside the Black Box: Interpreting Deep Learning Models for Exoplanet Atmospheric Retrievals. *arXiv e-prints*, art. arXiv:2011.11284, Nov. 2020.

A. W. Howard, G. W. Marcy, J. A. Johnson, and et al. The Occurrence and Mass Distribution of Close-in Super-Earths, Neptunes, and Jupiters. *Science*, 330(6004):653, Oct. 2010. doi: 10.1126/science.1194854.

A. W. Howard, G. W. Marcy, S. T. Bryson, and et al. Planet Occurrence within 0.25 AU of Solar-type Stars from Kepler. *Astrophys. J. Suppl. Ser.*, 201(2):15, Aug. 2012. doi: 10.1088/0067-0049/201/2/15.

A. W. Howard, R. Sanchis-Ojeda, G. W. Marcy, and et al. A rocky composition for an Earth-sized exoplanet. *Nature*, 503(7476):381–384, Nov. 2013. doi: 10.1038/nature12767.

R. Hu, S. Seager, and Y. L. Yung. Helium Atmospheres on Warm Neptune- and Sub-Neptune-sized Exoplanets and Applications to GJ 436b. *Astrophys. J.*, 807(1):8, July 2015. doi: 10.1088/0004-637X/807/1/8.

M. Huerta, C. M. Johns-Krull, L. Prato, and et al. Starspot-Induced Radial Velocity Variability in LkCa 19. *Astrophys. J.*, 678(1):472–482, May 2008. doi: 10.1086/526415.

P. G. J. Irwin, N. A. Teanby, R. de Kok, and et al. The NEMESIS planetary atmosphere radiative transfer and retrieval tool. *J. Quant. Spectrosc. Radiat. Transf.*, 109:1136–1150, Apr. 2008. doi: 10.1016/j.jqsrt.2007.11.006.

A. R. Iyer, M. R. Swain, R. T. Zellem, and et al. A Characteristic Transmission Spectrum Dominated by H_2O Applies to the Majority of HST/WFC3 Exoplanet Observations. *Astrophys. J.*, 823(2):109, June 2016. doi: 10.3847/0004-637X/823/2/109.

C. R. Jenkins and J. A. Peacock. The power of Bayesian evidence in astronomy. *Monthly Notices of the Royal Astronomical Society*, 413(4):2895–2905, 05 2011. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.18361.x. URL https://doi.org/10.1111/j.1365-2966.2011.18361.x.

U. G. Jørgensen and M. Hundertmark. The Combined System of Microlensing Exoplanets and Their Host Stars. In H. J. Deeg and J. A. Belmonte, editors, *Handbook of Exoplanets*, page 182. 2018. doi: 10.1007/978-3-319-55333-7_182.

D. Juncher, Jørgensen, Uffe G., and Helling, Christiane. Self-consistent atmosphere modeling with cloud formation for low-mass stars and exoplanets. *A&A*, 608:A70, 2017. doi: 10.1051/0004-6361/201629977. URL https://doi.org/10.1051/0004-6361/201629977.

M. Jurić and S. Tremaine. Dynamical Origin of Extrasolar Planet Eccentricity Distribution. *Astrophys. J.*, 686(1):603–620, Oct. 2008. doi: 10.1086/590047.

L. Kaltenegger. How to Characterize Habitable Worlds and Signs of Life. *Annu. Rev. Astron. Astrophys.*, 55(1):433–485, Aug. 2017. doi: 10.1146/annurev-astro-082214-122238.

R. E. Kass and A. E. Raftery. Bayes Factor. *Journal of the American Statistical Association*, 90:430:773–795, 1995. doi: 10.1080/01621459.1995.10476572.

T. Kataria, A. P. Showman, N. K. Lewis, and et al. Three-dimensional Atmospheric Circulation of Hot Jupiters on Highly Eccentric Orbits. *Astrophys. J.*, 767(1):76, Apr. 2013. doi: 10.1088/0004-637X/767/1/76.

T. Kataria, A. P. Showman, J. J. Fortney, and et al. The Atmospheric Circulation of the Super Earth GJ 1214b: Dependence on Composition and Metallicity. *Astrophys. J.*, 785 (2):92, Apr. 2014. doi: 10.1088/0004-637X/785/2/92.

T. Kataria, A. P. Showman, J. J. Fortney, and et al. The Atmospheric Circulation of the Hot Jupiter WASP-43b: Comparing Three-dimensional Models to Spectrophotometric Data. *Astrophys. J.*, 801(2):86, Mar. 2015. doi: 10.1088/0004-637X/801/2/86.

Y. Kawashima and M. Ikoma. Theoretical Transmission Spectra of Exoplanet Atmospheres with Hydrocarbon Haze: Effect of Creation, Growth, and Settling of Haze Particles. II. Dependence on UV Irradiation Intensity, Metallicity, C/O Ratio, Eddy Diffusion Coefficient, and Temperature. *Astrophys. J.*, 877(2):109, June 2019. doi: 10.3847/1538-4357/ab1b1d.

Y. Kawashima, R. Hu, and M. Ikoma. Detectable Molecular Features above Hydrocarbon Haze via Transmission Spectroscopy with JWST: Case Studies of GJ 1214b-, GJ 436b-, HD 97658b-, and Kepler-51b-like Planets. *Astrophys. J. Lett.*, 876(1):L5, May 2019. doi: 10.3847/2041-8213/ab16f6.

D. Keating, N. B. Cowan, and L. Dang. Uniformly hot nightside temperatures on short-period gas giants. *Nature Astronomy*, 3:1092–1098, Aug. 2019. doi: 10.1038/s41550-019-0859-z.

E. M. R. Kempton, J. L. Bean, D. R. Louie, and et al. A Framework for Prioritizing the TESS Planetary Candidates Most Amenable to Atmospheric Characterization. *Publ. Astron. Soc. Pac.*, 130(993):114401, Nov. 2018. doi: 10.1088/1538-3873/aadf6f.

D. M. Kipping and E. Sandford. Observational biases for transiting planets. *Mon. Not. R. Astron. Soc.*, 463(2):1323–1331, Dec. 2016. doi: 10.1093/mnras/stw1926.

J. Kirk, M. K. Alam, M. López-Morales, and et al. Confirmation of WASP-107b's Extended Helium Atmosphere with Keck II/NIRSPEC. *Astron. J.*, 159(3):115, Mar. 2020. doi: 10.3847/1538-3881/ab6e66.

D. Kitzmann, K. Heng, M. Oreshenko, and et al. Helios-r2: A New Bayesian, Open-source Retrieval Model for Brown Dwarfs and Exoplanet Atmospheres. *Astrophys. J.*, 890(2): 174, Feb. 2020. doi: 10.3847/1538-4357/ab6d71.

D. Knox. Giordano Bruno. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2 edition, 2019.

D. Koch, W. Borucki, L. Webster, J. Jenkins, E. Dunham, S. Jordan, and Kepler Team. The Kepler Mission: A Search for Terrestrial Planets. In *Astronomische Gesellschaft Meeting Abstracts*, volume 18 of *Astronomische Gesellschaft Meeting Abstracts*, page MS 04 06, Jan. 2001.

F. Kohlrausch. *Praktische Physik. Zum Gebrauch fuer Unterricht, Forschung und Technik*. 1986.

T. D. Komacek, P. Gao, D. P. Thorngren, and et al. The Effect of Interior Heat Flux on the Atmospheric Circulation of Hot and Ultra-hot Jupiters. *Astrophys. J. Lett.*, 941(2):L40, Dec. 2022. doi: 10.3847/2041-8213/aca975.

L. Kreidberg. batman: BAsic Transit Model cAlculatioN in Python. *Publ. Astron. Soc. Pac.*, 127(957):1161, Nov. 2015. doi: 10.1086/683602.

L. Kreidberg. Exoplanet Atmosphere Measurements from Transmission Spectroscopy and Other Planet Star Combined Light Observations. In H. J. Deeg and J. A. Belmonte, editors, *Handbook of Exoplanets*, page 100. Springer International Publishing AG, 2018. doi: 10.1007/978-3-319-55333-7_100.

L. Kreidberg, J. L. Bean, J.-M. Désert, and et al. Clouds in the atmosphere of the super-Earth exoplanet GJ1214b. *Nature*, 505(7481):69–72, Jan. 2014. doi: 10.1038/nature12888.

J. Krissansen-Totton and J. J. Fortney. Predictions for Observable Atmospheres of Trappist-1 Planets from a Fully Coupled Atmosphere-Interior Evolution Model. *Astrophys. J.*, 933 (1):115, July 2022. doi: 10.3847/1538-4357/ac69cb.

J. E. Krist. PROPER: an optical propagation library for IDL. 6675:66750P, Sept. 2007. doi: 10.1117/12.731179.

D. Kubyshkina, A. A. Vidotto, C. Villarreal D'Angelo, and et al. Atmospheric mass-loss and stellar wind effects in young and old systems - I. Comparative 3D study of TOI-942 and TOI-421 systems. *Mon. Not. R. Astron. Soc.*, 510(2):2111–2126, Feb. 2022a. doi: 10.1093/mnras/stab3594.

D. Kubyshkina, A. A. Vidotto, C. Villarreal D'Angelo, and et al. Atmospheric mass loss and stellar wind effects in young and old systems - II. Is TOI-942 the past of TOI-421 system? *Mon. Not. R. Astron. Soc.*, 510(2):3039–3045, Feb. 2022b. doi: 10.1093/mnras/stab3620.

B. I. Lacy and A. Burrows. JWST Transit Spectra. I. Exploring Potential Biases and Opportunities in Retrievals of Tidally Locked Hot Jupiters with Clouds and Hazes. *Astrophys. J.*, 905(2):131, Dec. 2020. doi: 10.3847/1538-4357/abc01c.

V. Lakshminarayanan and A. Fleck. Zernike polynomials: a guide. *Journal of Modern Optics*, 58(18):1678–1678, Oct. 2011. doi: 10.1080/09500340.2011.633763.

H. Lammer, J. F. Kasting, E. Chassefière, and et al. Atmospheric Escape and Evolution of Terrestrial Planets and Satellites. *Space Sci. Rev.*, 139(1-4):399–436, Aug. 2008. doi: 10.1007/s11214-008-9413-5.

M. Lampón, M. López-Puertas, L. M. Lara, and et al. Modelling the He I triplet absorption at 10 830 Å in the atmosphere of HD 209458 b. *Astron. Astrophys.*, 636:A13, Apr. 2020. doi: 10.1051/0004-6361/201937175.

M. Lampón, M. López-Puertas, J. Sanz-Forcada, and et al. Modelling the He I triplet absorption at 10 830 Å in the atmospheres of HD 189733 b and GJ 3470 b. *Astron. Astrophys.*, 647:A129, Mar. 2021. doi: 10.1051/0004-6361/202039417.

M. Lampón, M. López-Puertas, J. Sanz-Forcada, and et al. Characterisation of the upper atmospheres of HAT-P-32 b, WASP-69 b, GJ 1214 b, and WASP-76 b through their He I triplet absorption. *Astron. Astrophys.*, 673:A140, May 2023. doi: 10.1051/0004-6361/202245649.

G. N. Lawrence, Lawrence, and G. N. Optical Modeling. 11:125, 1992. URL https://ui.adsabs.harvard.edu/abs/1992aooe...11..125L/abstract.

A. Lecavelier Des Etangs, F. Pont, A. Vidal-Madjar, and D. Sing. Rayleigh scattering in the transit spectrum of HD 189733b. *Astron. Astrophys.*, 481(2):L83–L86, Apr. 2008. doi: 10.1051/0004-6361:200809388.

J. Leconte, G. Chabrier, I. Baraffe, and et al. Is tidal heating sufficient to explain bloated exoplanets? Consistent calculations accounting for finite initial eccentricity. *Astron. Astrophys.*, 516:A64, June 2010. doi: 10.1051/0004-6361/201014337.

Leconte, J., Forget, F., Charnay, B., Wordsworth, R., Selsis, F., Millour, E., and Spiga, A. 3d climate modeling of close-in land planets: Circulation patterns, climate moist bistability, and habitability. *A&A*, 554:A69, 2013. doi: 10.1051/0004-6361/201321042. URL https://doi.org/10.1051/0004-6361/201321042.

Y. Lecun and Y. Bengio. *Convolutional networks for images, speech, and time-series*. MIT Press, 1995.

J.-M. Lee, K. Heng, and P. G. J. Irwin. Atmospheric Retrieval Analysis of the Directly Imaged Exoplanet HR 8799b. *Astrophys. J.*, 778(2):97, Dec. 2013. doi: 10.1088/0004-637X/778/2/97.

G. Li, I. E. Gordon, L. S. Rothman, Y. Tan, S.-M. Hu, S. Kassi, A. Campargue, and E. S. Medvedev. Rovibrational Line Lists for Nine Isotopologues of the CO Molecule in the X $^1\Sigma^+$ Ground Electronic State. *Astrophys. J. Suppl. Ser.*, 216(1):15, Jan. 2015. doi: 10.1088/0067-0049/216/1/15.

M. R. Line, A. S. Wolf, X. Zhang, and et al. A Systematic Retrieval Analysis of Secondary Eclipse Spectra. I. A Comparison of Atmospheric Retrieval Techniques. *Astrophys. J.*, 775(2):137, Oct. 2013. doi: 10.1088/0004-637X/775/2/137.

E. D. Lopez and J. J. Fortney. Understanding the Mass-Radius Relation for Sub-neptunes: Radius as a Proxy for Composition. *Astrophys. J.*, 792(1):1, Sept. 2014. doi: 10.1088/0004-637X/792/1/1.

R. J. MacDonald and N. Madhusudhan. Signatures of Nitrogen Chemistry in Hot Jupiter Atmospheres. *Astrophys. J. Lett.*, 850(1):L15, Nov. 2017. doi: 10.3847/2041-8213/aa97d4.

R. J. MacDonald, J. M. Goyal, and N. K. Lewis. Why Is it So Cold in Here? Explaining the Cold Temperatures Retrieved from Transmission Spectra of Exoplanet Atmospheres. *Astrophys. J. Lett.*, 893(2):L43, Apr. 2020. doi: 10.3847/2041-8213/ab8238.

N. Madhusudhan. Exoplanetary Atmospheres: Key Insights, Challenges, and Prospects. *Annu. Rev. Astron. Astrophys.*, 57:617–663, Aug. 2019. doi: 10.1146/annurev-astro-081817-051846.

N. Madhusudhan, A. Burrows, and T. Currie. Model Atmospheres for Massive Gas Giants with Thick Clouds: Application to the HR 8799 Planets and Predictions for Future Detections. *Astrophys. J.*, 737(1):34, Aug. 2011. doi: 10.1088/0004-637X/737/1/34.

N. Madhusudhan, M. Agúndez, J. I. Moses, and et al. Exoplanetary Atmospheres—Chemistry, Formation Conditions, and Habitability. *Space Sci. Rev.*, 205(1-4):285–348, Dec. 2016. doi: 10.1007/s11214-016-0254-3.

G. Marsaglia. Choosing a Point from the Surface of a Sphere. *The Annals of Mathematical Statistics*, 43(2):645 – 646, 1972. doi: 10.1214/aoms/1177692644. URL https://doi.org/10.1214/aoms/1177692644.

E. Martioli, G. Hébrard, A. C. M. Correia, and et al. New constraints on the planetary system around the young active star AU Mic. Two transiting warm Neptunes near mean-motion resonance. *Astron. Astrophys.*, 649:A177, May 2021. doi: 10.1051/0004-6361/202040235.

T. Matsakos and A. Königl. The Gravitational Interaction between Planets on Inclined Orbits and Protoplanetary Disks As the Origin of Primordial Spin-Orbit Misalignments. *Astron. J.*, 153(2):60, Feb. 2017. doi: 10.3847/1538-3881/153/2/60.

Y. Matsumoto and E. Kokubo. Formation of Close-in Super-Earths by Giant Impacts: Effects of Initial Eccentricities and Inclinations of Protoplanets. *Astron. J.*, 154(1):27, July 2017. doi: 10.3847/1538-3881/aa74c7.

M. Mayor and D. Queloz. A Jupiter-mass companion to a solar-type star. *Nature*, 378(6555):355–359, Nov. 1995. doi: 10.1038/378355a0.

M. Mayor, F. Pepe, D. Queloz, and et al. Setting New Standards with HARPS. *The Messenger*, 114:20–24, Dec. 2003.

T. Mazeh, T. Holczer, and S. Faigler. Dearth of short-period Neptunian exoplanets: A desert in period-mass and period-radius planes. *Astron. Astrophys.*, 589:A75, May 2016. doi: 10.1051/0004-6361/201528065.

B. J. McBride, Zehe, Michael J., and Gordon, Sanford. Nasa glenn coefficients for calculating thermodynamic properties of individual species. NASA Technical Report TP-2002-211556, NASA, 09 2002. Work of the US Gov. Public Use Permitted.

P. R. McCullough, N. Crouzet, D. Deming, and N. Madhusudhan. Water Vapor in the Spectrum of the Extrasolar Planet HD 189733b. I. The Transit. *Astrophys. J.*, 791(1):55, Aug. 2014. doi: 10.1088/0004-637X/791/1/55.

M. W. McElwain, L. D. Feinberg, M. D. Perrin, and et al. The James Webb Space Telescope Mission: Optical Telescope Element Design, Development, and Performance. *arXiv e-prints*, art. arXiv:2301.01779, Jan. 2023. doi: 10.48550/arXiv.2301.01779.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, and et al. Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21(6):1087–1092, June 1953. doi: 10.1063/1.1699114.

E. Miller-Ricci Kempton, K. Zahnle, and J. J. Fortney. The Atmospheric Chemistry of GJ 1214b: Photochemistry and Clouds. *Astrophys. J.*, 745(1):3, Jan. 2012. doi: 10.1088/0004-637X/745/1/3.

S. Millholland. Tidally Induced Radius Inflation of Sub-Neptunes. *Astrophys. J.*, 886(1):72, Nov. 2019. doi: 10.3847/1538-4357/ab4c3f.

S. Millholland, E. Petigura, and K. Batygin. Tidal Inflation Reconciles Low-density Sub-Saturns with Core Accretion. *Astrophys. J.*, 897(1):7, July 2020. doi: 10.3847/1538-4357/ab959c.

K. Molaverdikhani, T. Henning, and P. Mollière. From Cold to Hot Irradiated Gaseous Exoplanets: Fingerprints of Chemical Disequilibrium in Atmospheric Spectra. *Astrophys. J.*, 883(2):194, Oct. 2019. doi: 10.3847/1538-4357/ab3e30.

P. Mollière, J. P. Wardenier, R. van Boekel, and et al. petitRADTRANS. A Python radiative transfer package for exoplanet characterization and retrieval. *Astron. Astrophys.*, 627: A67, July 2019. doi: 10.1051/0004-6361/201935470.

J. I. Moses. Chemical kinetics on extrasolar planets. *Philosophical Transactions of the Royal Society of London Series A*, 372(2014):20130073–20130073, Mar. 2014. doi: 10.1098/rsta.2013.0073.

J. I. Moses, C. Visscher, J. J. Fortney, and et al. Disequilibrium Carbon, Oxygen, and Nitrogen Chemistry in the Atmospheres of HD 189733b and HD 209458b. *Astrophys. J.*, 737(1):15, Aug. 2011. doi: 10.1088/0004-637X/737/1/15.

J. I. Moses, N. Madhusudhan, C. Visscher, and et al. Chemical Consequences of the C/O Ratio on Hot Jupiters: Examples from WASP-12b, CoRoT-2b, XO-1b, and HD 189733b. *Astrophys. J.*, 763(1):25, Jan. 2013. doi: 10.1088/0004-637X/763/1/25.

J. I. Moses, P. Tremblin, O. Venot, and et al. Chemical variation with altitude and longitude on exo-Neptunes: Predictions for Ariel phase-curve observations. *Experimental Astronomy*, 53(2):279–322, Apr. 2022. doi: 10.1007/s10686-021-09749-1.

L. V. Mugnai, E. Pascale, B. Edwards, and et al. ArielRad: the Ariel radiometric model. *Experimental Astronomy*, 50(2-3):303–328, Oct. 2020. doi: 10.1007/s10686-020-09676-7.

L. V. Mugnai, A. Al-Refaie, A. Bocchieri, and et al. Alfnoor: Assessing the Information Content of Ariel's Low-resolution Spectra with Planetary Population Studies. *Astron. J.*, 162(6):288, Dec. 2021a. doi: 10.3847/1538-3881/ac2e92.

L. V. Mugnai, D. Modirrousta-Galian, B. Edwards, and et al. ARES. V. No Evidence For Molecular Absorption in the HST WFC3 Spectrum of GJ 1132 b. *Astron. J.*, 161(6):284, June 2021b. doi: 10.3847/1538-3881/abf3c3.

L. V. Mugnai, E. Pascale, A. F. Al-Refaie, and et al. ExoSim 2. The new time-domain simulator applied to the Ariel space mission. In *European Planetary Science Congress*, pages EPSC2022–370, Sept. 2022. doi: 10.5194/epsc2022-370.

L. V. Mugnai, A. Bocchieri, and E. Pascale. Exorad 2.0: The generic point source radiometric model. *Journal of Open Source Software*, 8(89):5348, 2023. doi: 10.21105/joss.05348. URL https://doi.org/10.21105/joss.05348.

S. Mukherjee, N. E. Batalha, J. J. Fortney, and M. S. Marley. PICASO 3.0: A One-dimensional Climate Model for Giant Planets and Brown Dwarfs. *Astrophys. J.*, 942(2): 71, Jan. 2023. doi: 10.3847/1538-4357/ac9f4810.48550/arXiv.2208.07836.

National Academies of Sciences. *Pathways to Discovery in Astronomy and Astrophysics for the 2020s*. 2021. doi: 10.17226/26141.

E. R. Newton, A. W. Mann, A. L. Kraus, and et al. TESS Hunt for Young and Maturing Exoplanets (THYME). IV. Three Small Planets Orbiting a 120 Myr Old Star in the Pisces-Eridanus Stream. *Astron. J.*, 161(2):65, Feb. 2021. doi: 10.3847/1538-3881/abccc6.

A. Niculescu-Mizil and R. Caruana. Predicting good probabilities with supervised learning. In *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning*, pages 625–632, 2005. doi: 10.1145/1102351.1102430.

L. Nortmann, E. Pallé, M. Salz, and et al. Ground-based detection of an extended helium atmosphere in the Saturn-mass exoplanet WASP-69b. *Science*, 362(6421):1388–1391, Dec. 2018. doi: 10.1126/science.aat5348.

A. Oklopčić. Helium Absorption at 1083 nm from Extended Exoplanet Atmospheres: Dependence on Stellar Radiation. *Astrophys. J.*, 881(2):133, Aug. 2019. doi: 10.3847/1538-4357/ab2f7f.

A. Oklopčić and C. M. Hirata. A New Window into Escaping Exoplanet Atmospheres: 10830 Å Line of Helium. *Astrophys. J. Lett.*, 855(1):L11, Mar. 2018. doi: 10.3847/2041-8213/aaada9.

M. Oreshenko, B. Lavie, S. L. Grimm, and et al. Retrieval Analysis of the Emission Spectrum of WASP-12b: Sensitivity of Outcomes to Prior Assumptions and Implications for Formation History. *Astrophys. J. Lett.*, 847(1):L3, Sept. 2017. doi: 10.3847/2041-8213/aa8acf.

J. E. Owen. Constraining the entropy of formation from young transiting planets. *Mon. Not. R. Astron. Soc.*, 498(4):5030–5040, Nov. 2020. doi: 10.1093/mnras/staa2784.

J. E. Owen and B. Campos Estrada. Testing exoplanet evaporation with multitransiting systems. *Mon. Not. R. Astron. Soc.*, 491(4):5287–5297, Feb. 2020. doi: 10.1093/mnras/stz3435.

J. E. Owen and Y. Wu. Kepler Planets: A Tale of Evaporation. *Astrophys. J.*, 775(2):105, Oct. 2013. doi: 10.1088/0004-637X/775/2/105.

E. Pace, A. Tozzi, M. A. Abreu, and et al. The telescope assembly of the Ariel space mission. In L. E. Coyle, S. Matsuura, and M. D. Perrin, editors, *Space Telescopes and Instrumentation 2022: Optical, Infrared, and Millimeter Wave*, volume 12180 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 1218011, Aug. 2022. doi: 10.1117/12.2629432.

E. Pacetti, D. Turrini, E. Schisano, and et al. Chemical Diversity in Protoplanetary Disks and Its Impact on the Formation History of Giant Planets. *Astrophys. J.*, 937(1):36, Sept. 2022. doi: 10.3847/1538-4357/ac8b11.

E. Palle, L. Nortmann, N. Casasayas-Barris, and et al. A He I upper atmosphere around the warm Neptune GJ 3470 b. *Astron. Astrophys.*, 638:A61, June 2020. doi: 10.1051/0004-6361/202037719.

E. Panek, J.-P. Beaulieu, P. Drossart, O. Venot, Q. Changeat, A. Al-Refaie, and A. Gressier. A re-analysis of equilibrium chemistry in five hot Jupiters. *arXiv e-prints*, art. arXiv:2306.10873, June 2023. doi: 10.48550/arXiv.2306.10873.

V. Parmentier, J. J. Fortney, A. P. Showman, C. Morley, and M. S. Marley. Transitions in the Cloud Composition of Hot Jupiters. *Astrophys. J.*, 828(1):22, Sept. 2016. doi: 10.3847/0004-637X/828/1/22.

V. Parmentier, M. R. Line, J. L. Bean, and et al. From thermal dissociation to condensation in the atmospheres of ultra hot Jupiters: WASP-121b in context. *Astron. Astrophys.*, 617: A110, Sept. 2018. doi: 10.1051/0004-6361/201833059.

V. Parmentier, A. P. Showman, and J. J. Fortney. The cloudy shape of hot Jupiter thermal phase curves. *Mon. Not. R. Astron. Soc.*, 501(1):78–108, Jan. 2021. doi: 10.1093/mnras/staa3418.

E. Pascale, N. Bezawada, J. Barstow, and et al. The ARIEL space mission. In M. Lystrup, H. A. MacEwen, G. G. Fazio, and et al., editors, *Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave*, volume 10698 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 106980H, July 2018. doi: 10.1117/12.2311838.

F. Pedregosa, G. Varoquaux, A. Gramfort, and et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, Oct. 2011. doi: 10.48550/arXiv.1201.0490.

J. Pepper, K. G. Stassun, and B. S. Gaudi. KELT: The Kilodegree Extremely Little Telescope, a Survey for Exoplanets Transiting Bright, Hot Stars. In H. J. Deeg and J. A. Belmonte, editors, *Handbook of Exoplanets*, page 128. 2018. doi: 10.1007/978-3-319-55333-7_128.

M. Perryman. *The Exoplanet Handbook*. 2018.

E. A. Petigura, G. W. Marcy, J. N. Winn, and et al. The California-Kepler Survey. IV. Metal-rich Stars Host a Greater Diversity of Planets. *Astron. J.*, 155(2):89, Feb. 2018. doi: 10.3847/1538-3881/aaa54c.

J. Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999.

W. Pluriel, N. Whiteford, B. Edwards, and et al. ARES. III. Unveiling the Two Faces of KELT-7 b with HST WFC3. *Astron. J.*, 160(3):112, Sept. 2020a. doi: 10.3847/1538-3881/aba000.

W. Pluriel, T. Zingales, J. Leconte, and V. Parmentier. Strong biases in retrieved atmospheric composition caused by day-night chemical heterogeneities. *Astron. Astrophys.*, 636:A66, Apr. 2020b. doi: 10.1051/0004-6361/202037678.

W. Pluriel, J. Leconte, V. Parmentier, and et al. Toward a multidimensional analysis of transmission spectroscopy. II. Day-night-induced biases in retrievals from hot to ultrahot Jupiters. *Astron. Astrophys.*, 658:A42, Feb. 2022. doi: 10.1051/0004-6361/202141943.

D. L. Pollacco, I. Skillen, A. Collier Cameron, and et al. The WASP Project and the SuperWASP Cameras. *Publ. Astron. Soc. Pac.*, 118(848):1407–1418, Oct. 2006. doi: 10.1086/508556.

O. L. Polyansky, A. A. Kyuberis, N. F. Zobov, and et al. ExoMol molecular line lists XXX: a complete high-accuracy line list for water. *Mon. Not. R. Astron. Soc.*, 480(2):2597–2608, Oct. 2018. doi: 10.1093/mnras/sty1877.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and et al. *Numerical recipes in C. The art of scientific computing*. 1992.

R. E. Pudritz, A. J. Cridland, and M. Alessi. Connecting Planetary Composition with Formation. In H. J. Deeg and J. A. Belmonte, editors, *Handbook of Exoplanets*, page 144. 2018. doi: 10.1007/978-3-319-55333-7_144.

D. Queloz, A. Eggenberger, M. Mayor, and et al. Detection of a spectroscopic transit by the planet orbiting the star HD209458. *Astron. Astrophys.*, 359:L13–L17, July 2000. doi: 10.48550/arXiv.astro-ph/0006213.

H. Rauer. The PLATO 2.0 mission. In *European Planetary Science Congress*, pages EPSC2013–707, Sept. 2013.

I. Ribas and J. Miralda-Escudé. The eccentricity-mass distribution of exoplanets: signatures of different formation mechanisms? *Astron. Astrophys.*, 464(2):779–785, Mar. 2007. doi: 10.1051/0004-6361:20065726.

I. Ribas, E. F. Guinan, M. Güdel, and et al. Evolution of the Solar Activity over Time and Effects on Planetary Atmospheres. I. High-Energy Irradiances (1-1700 Å). *Astrophys. J.*, 622(1):680–694, Mar. 2005. doi: 10.1086/427977.

G. R. Ricker, J. N. Winn, R. Vanderspek, and et al. Transiting Exoplanet Survey Satellite (TESS). *Journal of Astronomical Telescopes, Instruments, and Systems*, 1:014003, Jan. 2015. doi: 10.1117/1.JATIS.1.1.014003.

J. Rigby, M. Perrin, M. McElwain, and et al. The Science Performance of JWST as Characterized in Commissioning. *arXiv e-prints*, art. arXiv:2207.05632, July 2022. doi: 10.48550/arXiv.2207.05632.

W. Riley and D. Howe. Handbook of frequency stability analysis, July 2008. URL https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=50505.

M. Rocchetto, I. P. Waldmann, O. Venot, P. O. Lagage, and G. Tinetti. Exploring Biases of Atmospheric Retrievals in Simulated JWST Transmission Spectra of Hot Jupiters. *Astrophys. J.*, 833(1):120, Dec. 2016. doi: 10.3847/1538-4357/833/1/120.

W. D. Rogatto, J. S. Accetta, and D. L. Shumaker. The infrared & electro-optical systems handbook. electro-optical components, volume 3,. 1993.

L. A. Rogers, P. Bodenheimer, J. J. Lissauer, and et al. Formation and Structure of Low-density exo-Neptunes. *Astrophys. J.*, 738(1):59, Sept. 2011. doi: 10.1088/0004-637X/738/1/59.

T. S. Ross. Limitations and applicability of the maréchal approximation. *Appl. Opt.*, 48(10): 1812–1818, Apr. 2009. doi: 10.1364/AO.48.001812. URL https://opg.optica.org/ao/abstract.cfm?URI=ao-48-10-1812.

L. S. Rothman, I. E. Gordon, R. J. Barber, and et al. HITEMP, the high-temperature molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.*, 111:2139–2150, Oct. 2010. doi: 10.1016/j.jqsrt.2010.05.001.

G. M. Roudier, M. R. Swain, M. S. Gudipati, and et al. Disequilibrium Chemistry in Exoplanet Atmospheres Observed with the Hubble Space Telescope. *Astron. J.*, 162(2): 37, Aug. 2021. doi: 10.3847/1538-3881/abfdad.

H. N. Russell. Molecules in the Sun and Stars. *Astrophys. J.*, 79:317, Apr. 1934. doi: 10.1086/143539.

Z. Rustamkulov, D. K. Sing, R. Liu, and et al. Analysis of a JWST NIRSpec Lab Time Series: Characterizing Systematics, Recovering Exoplanet Transit Spectroscopy, and Constraining a Noise Floor. *Astrophys. J. Lett.*, 928(1):L7, Mar. 2022. doi: 10.3847/2041-8213/ac5b6f.

Z. Rustamkulov, D. K. Sing, S. Mukherjee, and et al. Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM. *Nature*, 614(7949):659–663, Feb. 2023. doi: 10.1038/s41586-022-05677-y.

F. Sanders. On Subjective Probability Forecasting. *Journal of Applied Meteorology and Climatology*, 2(2):191–201, 1963. doi: 10.1175/1520-0450(1963)002<0191:OSPF>2.0.CO;2.

S. Sarkar and N. Madhusudhan. JexoSim 2.0: end-to-end JWST simulator for exoplanet spectroscopy - implementation and case studies. *Mon. Not. R. Astron. Soc.*, 508(1): 433–452, Nov. 2021. doi: 10.1093/mnras/stab2472.

S. Sarkar, N. Madhusudhan, and A. Papageorgiou. JexoSim: a time-domain simulator of exoplanet transit spectroscopy with JWST. *Mon. Not. R. Astron. Soc.*, 491(1):378–397, Jan. 2020. doi: 10.1093/mnras/stz2958.

S. Sarkar, E. Pascale, A. Papageorgiou, and et al. ExoSim: the Exoplanet Observation Simulator. *Experimental Astronomy*, 51(2):287–317, Apr. 2021. doi: 10.1007/s10686-020-09690-9.

E. Schlawin, J. Leisenring, K. Misselt, and et al. JWST Noise Floor. I. Random Error Sources in JWST NIRCam Time Series. *Astron. J.*, 160(5):231, Nov. 2020. doi: 10.3847/1538-3881/abb811.

A. D. Schneider and B. Bitsch. How drifting and evaporating pebbles shape giant planets. I. Heavy element content and atmospheric C/O. *Astron. Astrophys.*, 654:A71, Oct. 2021. doi: 10.1051/0004-6361/202039640.

S. Seager and D. D. Sasselov. Theoretical Transmission Spectra during Extrasolar Giant Planet Transits. *Astrophys. J.*, 537(2):916–921, July 2000. doi: 10.1086/309088.

S. Seager, M. Kuchner, C. A. Hier-Majumder, and et al. Mass-Radius Relationships for Solid Exoplanets. *Astrophys. J.*, 669(2):1279–1297, Nov. 2007. doi: 10.1086/521346.

S. A. Self. Focusing of spherical gaussian beams. *Applied optics*, 22 5:658–61, 1983.

W. Sellmeier. Ueber die durch die aetherschwingungen erregten mitschwingungen der körpertheilchen und deren rückwirkung auf die ersteren, besonders zur erklärung der dispersion und ihrer anomalien. *Annalen der Physik*, 223(11):386–403, 1872. doi: https://doi.org/10.1002/andp.18722231105. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.18722231105.

C. J. Shallue and A. Vanderburg. Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90. *Astron. J.*, 155 (2):94, Feb. 2018. doi: 10.3847/1538-3881/aa9e09.

A. P. Showman, J. J. Fortney, Y. Lian, and et al. Atmospheric Circulation of Hot Jupiters: Coupled Radiative-Dynamical General Circulation Model Simulations of HD 189733b and HD 209458b. *Astrophys. J.*, 699(1):564–584, July 2009. doi: 10.1088/0004-637X/699/1/564.

A. P. Showman, X. Tan, and V. Parmentier. Atmospheric Dynamics of Hot Giant Planets and Brown Dwarfs. *Space Sci. Rev.*, 216(8):139, Dec. 2020. doi: 10.1007/s11214-020-00758-8.

D. K. Sing. Observational Techniques With Transiting Exoplanetary Atmospheres. *arXiv e-prints*, art. arXiv:1804.07357, Apr. 2018. doi: 10.48550/arXiv.1804.07357.

N. Skaf, M. F. Bieger, B. Edwards, and et al. ARES. II. Characterizing the Hot Jupiters WASP-127 b, WASP-79 b, and WASP-62b with the Hubble Space Telescope. *Astron. J.*, 160(3):109, Sept. 2020. doi: 10.3847/1538-3881/ab94a3.

J. Skilling. Nested Sampling. In R. Fischer, R. Preuss, and U. V. Toussaint, editors, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 735 of *American Institute of Physics Conference Series*, pages 395–405, Nov. 2004. doi: 10.1063/1.1835238.

A. M. S. Smith, J. S. Acton, D. R. Anderson, and et al. NGTS-14Ab: a Neptune-sized transiting planet in the desert. *Astron. Astrophys.*, 646:A183, Feb. 2021. doi: 10.1051/0004-6361/202039712.

W. J. Smith. *The Design of Optical Systems: General*. 2000. ISBN 0071593756.

D. R. Soderblom, D. K. Duncan, and D. R. H. Johnson. The Chromospheric Emission–Age Relation for Stars of the Lower Main Sequence and Its Implications for the Star Formation Rate. *Astrophys. J.*, 375:722, July 1991. doi: 10.1086/170238.

J. J. Spake, D. K. Sing, T. M. Evans, and et al. Helium in the eroding atmosphere of an exoplanet. *Nature*, 557(7703):68–70, May 2018. doi: 10.1038/s41586-018-0067-5.

K. G. Stassun, K. A. Collins, and B. S. Gaudi. Accurate Empirical Radii and Masses of Planets and Their Host Stars with Gaia Parallaxes. *Astron. J.*, 153(3):136, Mar. 2017. doi: 10.3847/1538-3881/aa5df3.

K. B. Stevenson, J. Harrington, S. Nymeyer, and et al. Possible thermochemical disequilibrium in the atmosphere of the exoplanet GJ 436b. *Nature*, 464(7292):1161–1164, Apr. 2010. doi: 10.1038/nature09013.

J. W. Stock, D. Kitzmann, A. B. C. Patzer, and et al. FastChem: A computer program for efficient complex chemical equilibrium calculations in the neutral/ionized gas phase with applications to stellar and planetary atmospheres. *Mon. Not. R. Astron. Soc.*, 479(1):865–874, Sept. 2018. doi: 10.1093/mnras/sty1531.

J. W. Stock, D. Kitzmann, and A. B. C. Patzer. FASTCHEM 2 : an improved computer program to determine the gas-phase chemical equilibrium composition for arbitrary element distributions. *Mon. Not. R. Astron. Soc.*, 517(3):4070–4080, Dec. 2022. doi: 10.1093/mnras/stac2623.

O. Struve. Proposal for a project of high-precision stellar radial velocity work. *The Observatory*, 72:199–200, Oct. 1952.

G. Sturrock, B. Manry, and S. Rafiqi. Machine Learning Pipeline for Exoplanet Classification. *SMU Data Science Review*, 2(1), 2019. URL https://scholar.smu.edu/datasciencereview/vol2/iss1/9.

D. C. Swift, J. H. Eggert, D. G. Hicks, and et al. Mass-Radius Relationships for Exoplanets. *Astrophys. J.*, 744(1):59, Jan. 2012. doi: 10.1088/0004-637X/744/1/5910.1086/141924.

G. Szabó M, S. Kálmán, L. Borsato, and et al. Sub-Jovian desert of exoplanets at its boundaries: Parameter dependence along the main sequence. *arXiv e-prints*, art. arXiv:2301.01065, Jan. 2023. doi: 10.48550/arXiv.2301.01065.

J.-P. Taché. Ray matrices for tilted interfaces in laser resonators. *Appl. Opt.*, 26(3):427–429, Feb. 1987. doi: 10.1364/AO.26.000427. URL https://opg.optica.org/ao/abstract.cfm?URI=ao-26-3-427.

X. Tan and T. D. Komacek. The Atmospheric Circulation of Ultra-hot Jupiters. *Astrophys. J.*, 886(1):26, Nov. 2019. doi: 10.3847/1538-4357/ab4a76.

J. Tennyson and S. N. Yurchenko. ExoMol: molecular line lists for exoplanet and other atmospheres. *Mon. Not. R. Astron. Soc.*, 425(1):21–33, Sept. 2012. doi: 10.1111/j.1365-2966.2012.21440.x.

J. Tennyson, S. N. Yurchenko, A. F. Al-Refaie, and et al. The ExoMol database: Molecular line lists for exoplanet and other hot atmospheres. *Journal of Molecular Spectroscopy*, 327:73–94, Sept. 2016. doi: 10.1016/j.jms.2016.05.002.

D. P. Thorngren, J. J. Fortney, R. A. Murray-Clay, and E. D. Lopez. The Mass-Metallicity Relation for Giant Planets. *Astrophys. J.*, 831(1):64, Nov. 2016. doi: 10.3847/0004-637X/831/1/64.

G. Tinetti, T. Encrenaz, and A. Coustenis. Spectroscopy of planetary atmospheres in our Galaxy. *Astron. Astrophys. Rev.*, 21:63, Oct. 2013. doi: 10.1007/s00159-013-0063-6.

G. Tinetti, P. Drossart, P. Eccleston, and et al. A chemical survey of exoplanets with ARIEL. *Experimental Astronomy*, 46(1):135–209, Nov. 2018. doi: 10.1007/s10686-018-9598-x.

G. Tinetti, P. Eccleston, C. Haswell, and et al. Ariel: Enabling planetary science across light-years. *arXiv e-prints*, art. arXiv:2104.04824, Apr. 2021. doi: 10.48550/arXiv.2104.04824.

R. Trotta. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49(2):71–104, Mar. 2008. doi: 10.1080/00107510802066753.

S.-M. Tsai, M. Malik, D. Kitzmann, and et al. A Comparative Study of Atmospheric Chemistry with VULCAN. *Astrophys. J.*, 923(2):264, Dec. 2021. doi: 10.3847/1538-4357/ac29bc.

S.-M. Tsai, E. K. H. Lee, and R. Pierrehumbert. A mini-chemical scheme with net reactions for 3D general circulation models. I. Thermochemical kinetics. *Astron. Astrophys.*, 664: A82, Aug. 2022. doi: 10.1051/0004-6361/202142816.

S.-M. Tsai, E. K. H. Lee, D. Powell, and et al. Photochemically produced SO_2 in the atmosphere of WASP-39b. *Nature*, 617(7961):483–487, May 2023a. doi: 10.1038/s41586-023-05902-2.

S.-M. Tsai, E. K. H. Lee, D. Powell, and et al. Photochemically produced SO_2 in the atmosphere of WASP-39b. *Nature*, 617(7961):483–487, May 2023b. doi: 10.1038/s41586-023-05902-2.

A. Tsiaras, I. P. Waldmann, T. Zingales, and et al. A Population Study of Gaseous Exoplanets. *Astron. J.*, 155(4):156, Apr. 2018. doi: 10.3847/1538-3881/aaaf75.

A. Tsiaras, I. P. Waldmann, G. Tinetti, and et al. Water vapour in the atmosphere of the habitable-zone eight-Earth-mass planet K2-18 b. *Nature Astronomy*, 3:1086–1091, Sept. 2019. doi: 10.1038/s41550-019-0878-9.

T. Tsuji. Molecular abundances in stellar atmospheres. II. *Astron. Astrophys.*, 23:411, Mar. 1973.

G. S. Tucker, P. Nagler, N. Butler, and et al. The Exoplanet Climate Infrared TElescope (EXCITE). In C. J. Evans, L. Simard, and H. Takami, editors, *Ground-based and Airborne Instrumentation for Astronomy VII*, volume 10702 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 107025G, July 2018. doi: 10.1117/12.2314225.

M. Tuomi. Evidence for nine planets in the HD 10180 system. *Astron. Astrophys.*, 543:A52, July 2012. doi: 10.1051/0004-6361/201118518.

D. Turrini, R. P. Nelson, and M. Barbieri. The role of planetary formation and evolution in shaping the composition of exoplanetary atmospheres. *Experimental Astronomy*, 40(2-3): 501–522, Dec. 2015. doi: 10.1007/s10686-014-9401-6.

D. Turrini, Y. Miguel, T. Zingales, and et al. The contribution of the ARIEL space mission to the study of planetary formation. *Experimental Astronomy*, 46(1):45–65, Nov. 2018. doi: 10.1007/s10686-017-9570-1.

D. Turrini, A. Zinzi, and J. A. Belinchon. Normalized angular momentum deficit: a tool for comparing the violence of the dynamical histories of planetary systems. *Astron. Astrophys.*, 636:A53, Apr. 2020. doi: 10.1051/0004-6361/201936301.

D. Turrini, E. Schisano, S. Fonte, and et al. Tracing the Formation History of Giant Planets in Protoplanetary Disks with Carbon, Oxygen, Nitrogen, and Sulfur. *Astrophys. J.*, 909 (1):40, Mar. 2021. doi: 10.3847/1538-4357/abd6e5.

A. Udalski, M. K. Szymański, and G. Szymański. OGLE-IV: Fourth Phase of the Optical Gravitational Lensing Experiment. *Acta Astron.*, 65(1):1–38, Mar. 2015. doi: 10.48550/arXiv.1504.05966.

S. Ulmer-Moll, N. C. Santos, P. Figueira, and et al. Beyond the exoplanet mass-radius relation. *Astron. Astrophys.*, 630:A135, Oct. 2019. doi: 10.1051/0004-6361/201936049.

D. Valencia, T. Guillot, V. Parmentier, and et al. Bulk Composition of GJ 1214b and Other Sub-Neptune Exoplanets. *Astrophys. J.*, 775(1):10, Sept. 2013. doi: 10.1088/0004-637X/775/1/10.

F. Valsecchi, S. Rappaport, F. A. Rasio, and et al. Tidally-driven Roche-lobe Overflow of Hot Jupiters with MESA. *Astrophys. J.*, 813(2):101, Nov. 2015. doi: 10.1088/0004-637X/813/2/101.

O. Venot, E. Hébrard, M. Agúndez, and et al. A chemical model for the atmosphere of hot Jupiters. *Astron. Astrophys.*, 546:A43, Oct. 2012. doi: 10.1051/0004-6361/201219310.

O. Venot, T. Cavalié, R. Bounaceur, and et al. New chemical scheme for giant planet thermochemistry. Update of the methanol chemistry and new reduced chemical scheme. *Astron. Astrophys.*, 634:A78, Feb. 2020a. doi: 10.1051/0004-6361/201936697.

O. Venot, V. Parmentier, J. Blecic, and et al. Global Chemistry and Thermal Structure Models for the Hot Jupiter WASP-43b and Predictions for JWST. *Astrophys. J.*, 890(2):176, Feb. 2020b. doi: 10.3847/1538-4357/ab6a94.

A. Vidal-Madjar, A. Lecavelier des Etangs, J. M. Désert, and et al. An extended upper atmosphere around the extrasolar planet HD209458b. *Nature*, 422(6928):143–146, Mar. 2003. doi: 10.1038/nature01448.

C. von Essen, M. Mallonn, L. Welbanks, and et al. An optical transmission spectrum of the ultra-hot Jupiter WASP-33 b. First indication of aluminum oxide in an exoplanet. *Astron. Astrophys.*, 622:A71, Feb. 2019. doi: 10.1051/0004-6361/201833837.

H. R. Wakeford and D. K. Sing. Transmission spectral properties of clouds for hot Jupiter exoplanets. *Astron. Astrophys.*, 573:A122, Jan. 2015. doi: 10.1051/0004-6361/201424207.

H. R. Wakeford, K. B. Stevenson, N. K. Lewis, and et al. HST PanCET program: A Cloudy Atmosphere for the Promising JWST Target WASP-101b. *Astrophys. J. Lett.*, 835(1):L12, Jan. 2017. doi: 10.3847/2041-8213/835/1/L12.

I. P. Waldmann, M. Rocchetto, G. Tinetti, E. J. Barton, S. N. Yurchenko, and J. Tennyson. Tau-REx II: Retrieval of Emission Spectra. *Astrophys. J.*, 813(1):13, Nov. 2015a. doi: 10.1088/0004-637X/813/1/13.

I. P. Waldmann, G. Tinetti, M. Rocchetto, and et al. Tau-REx I: A Next Generation Retrieval Code for Exoplanetary Atmospheres. *Astrophys. J.*, 802(2):107, Apr. 2015b. doi: 10.1088/0004-637X/802/2/107.

J. V. Wall and C. R. Jenkins. *Practical Statistics for Astronomers*. 2012.

J. P. Wardenier, V. Parmentier, and E. K. H. Lee. All along the line of sight: a closer look at opening angles and absorption regions in the atmospheres of transiting exoplanets. *Mon. Not. R. Astron. Soc.*, 510(1):620–629, Feb. 2022. doi: 10.1093/mnras/stab3432.

L. M. Weiss, S. C. Millholland, E. A. Petigura, and et al. Architectures of Compact Multi-planet Systems: Diversity and Uniformity. *arXiv e-prints*, art. arXiv:2203.10076, Mar. 2022. doi: 10.48550/arXiv.2203.10076.

L. Welbanks, N. Madhusudhan, N. F. Allard, and et al. Mass-Metallicity Trends in Transiting Exoplanets from Atmospheric Abundances of H_2O, Na, and K. *Astrophys. J. Lett.*, 887(1):L20, Dec. 2019. doi: 10.3847/2041-8213/ab5a89.

W. T. Welford. *Aberrations of Optical Systems*. Routledge, New York, 1st edition, 1986. doi: https://doi.org/10.1201/9781315136530.

W. F. Welsh and J. A. Orosz. Two Suns in the Sky: The Kepler Circumbinary Planets. In H. J. Deeg and J. A. Belmonte, editors, *Handbook of Exoplanets*, page 34. 2018. doi: 10.1007/978-3-319-55333-7_34.

P. J. Wheatley, R. G. West, M. R. Goad, and et al. The Next Generation Transit Survey (NGTS). *Mon. Not. R. Astron. Soc.*, 475(4):4476–4493, Apr. 2018. doi: 10.1093/mnras/stx2836.

W. B. White, S. M. Johnson, and G. B. Dantzig. Chemical Equilibrium in Complex Mixtures. *Journal of Chemical Physics*, 28(5):751–755, May 1958. doi: 10.1063/1.1744264.

D. S. Wilks. Chapter 9 - forecast verification. In D. S. Wilks, editor, *Statistical Methods in the Atmospheric Sciences (Fourth Edition)*, pages 369–483. Elsevier, fourth edition edition, 2019. ISBN 978-0-12-815823-4. doi: https://doi.org/10.1016/B978-0-12-815823-4.00009-2. URL https://www.sciencedirect.com/science/article/pii/B9780128158234000092.

J. N. Winn and D. C. Fabrycky. The Occurrence and Architecture of Exoplanetary Systems. *Annu. Rev. Astron. Astrophys.*, 53:409–447, Aug. 2015. doi: 10.1146/annurev-astro-082214-122246.

J. N. Winn, R. W. Noyes, M. J. Holman, and et al. Measurement of Spin-Orbit Alignment in an Extrasolar Planetary System. *Astrophys. J.*, 631(2):1215–1226, Oct. 2005. doi: 10.1086/432571.

C. P. Wirth, G. Zhou, S. N. Quinn, and et al. TOI-942b: A Prograde Neptune in a 60 Myr Old Multi-transiting System. *Astrophys. J. Lett.*, 917(2):L34, Aug. 2021. doi: 10.3847/2041-8213/ac13a9.

P. Woitke and C. Helling. Dust in brown dwarfs. III. Formation and structure of quasi-static cloud layers. *Astron. Astrophys.*, 414:335–350, Jan. 2004. doi: 10.1051/0004-6361:20031605.

P. Woitke, C. Helling, G. H. Hunter, and et al. Equilibrium chemistry down to 100 K. Impact of silicates and phyllosilicates on the carbon to oxygen ratio. *Astron. Astrophys.*, 614:A1, June 2018. doi: 10.1051/0004-6361/201732193.

A. Wolszczan and D. A. Frail. A planetary system around the millisecond pulsar PSR1257 + 12. *Nature*, 355(6356):145–147, Jan. 1992. doi: 10.1038/355145a0.

J. T. Wright. Radial Velocities as an Exoplanet Discovery Method. In H. J. Deeg and J. A. Belmonte, editors, *Handbook of Exoplanets*, page 4. 2018. doi: 10.1007/978-3-319-55333-7_4.

A. Wyttenbach, D. Ehrenreich, C. Lovis, and et al. Spectrally resolved detection of sodium in the atmosphere of HD 189733b with the HARPS spectrograph. *Astron. Astrophys.*, 577:A62, May 2015. doi: 10.1051/0004-6361/201525729.

F. Yan, E. Pallé, A. Reiners, and et al. A temperature inversion with atomic iron in the ultra-hot dayside atmosphere of WASP-189b. *Astron. Astrophys.*, 640:L5, Aug. 2020. doi: 10.1051/0004-6361/202038294.

A. Yariv. *Quantum electronics*. John Wiley and Sons, New York, 1989.

S. N. Yurchenko and J. Tennyson. ExoMol line lists - IV. The rotation-vibration spectrum of methane up to 1500 K. *Mon. Not. R. Astron. Soc.*, 440(2):1649–1661, May 2014. doi: 10.1093/mnras/stu326.

S. N. Yurchenko, R. J. Barber, and J. Tennyson. A variationally computed line list for hot NH_3. *Mon. Not. R. Astron. Soc.*, 413(3):1828–1834, May 2011. doi: 10.1111/j.1365-2966.2011.18261.x.

S. N. Yurchenko, J. Tennyson, and J. e. a. Bailey. Spectrum of hot methane in astronomical objects using a comprehensive computed line list. *Proceedings of the National Academy of Science*, 111:9379–9383, July 2014. doi: 10.1073/pnas.1324219111. URL https://ui.adsabs.harvard.edu/abs/2014PNAS..111.9379Y.

S. N. Yurchenko, A. F. Al-Refaie, and J. Tennyson. EXOCROSS: a general program for generating spectra from molecular line lists. *Astron. Astrophys.*, 614:A131, June 2018. doi: 10.1051/0004-6361/201732531.

B. Zadrozny and C. Elkan. Transforming classifier scores into accurate multiclass probability estimates. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 08 2002. doi: 10.1145/775047.775151.

R. T. Zellem, M. R. Swain, G. Roudier, and et al. Forecasting the Impact of Stellar Activity on Transiting Exoplanet Spectra. *Astrophys. J.*, 844(1):27, July 2017. doi: 10.3847/1538-4357/aa79f5.

R. T. Zellem, M. R. Swain, N. B. Cowan, and et al. Constraining Exoplanet Metallicities and Aerosols with the Contribution to ARIEL Spectroscopy of Exoplanets (CASE). *Publ. Astron. Soc. Pac.*, 131(1003):094401, Sept. 2019. doi: 10.1088/1538-3873/ab2d54.

M. Zhang, Y. Chachan, E. M. R. Kempton, and H. A. Knutson. Forward Modeling and Retrievals with PLATON, a Fast Open-source Tool. *Publ. Astron. Soc. Pac.*, 131(997):034501, Mar. 2019. doi: 10.1088/1538-3873/aaf5ad.

G. Zhou, S. N. Quinn, J. Irwin, and et al. Two Young Planetary Systems around Field Stars with Ages between 20 and 320 Myr from TESS. *Astron. J.*, 161(1):2, Jan. 2021. doi: 10.3847/1538-3881/abba22.

W. Zhu and S. Dong. Exoplanet Statistics and Theoretical Implications. *Annu. Rev. Astron. Astrophys.*, 59:291–336, Sept. 2021. doi: 10.1146/annurev-astro-112420-020055.

T. Zingales, A. Falco, W. Pluriel, and J. Leconte. Toward a multidimensional analysis of transmission spectroscopy. part iii: Modelling 2d effects in retrievals with taurex, 2022. URL https://arxiv.org/abs/2207.14247.

Acronyms

ADaRP Ariel Data Reduction Pipeline. 78, 90, 91

AIRS Ariel Infra-Red Spectrometer. 27, 28, 40, 68, 79, 80, 82, 85–87, 89, 119, 128, 135, 154, 168, 169, 222

AOCS Attitude and Orbit Control System. 78, 84, 85, 95, 99

APE Absolute Performance Error. 82, 91

ARES Ariel Retrieval of Exoplanets School. 176

Ariel Atmospheric Remote-Sensing Infrared Exoplanet Large-survey. 3, 17, 23, 25–28, 31–35, 39, 40, 43–45, 56, 59–61, 64, 67, 68, 70–72, 76–78, 80–82, 84–87, 89, 91, 94–96, 99, 102, 108, 109, 119, 124–130, 133, 136, 143, 146, 150–152, 154, 168, 176, 178, 180, 181, 220–223

ArielRad Ariel Radiometric Model. 33–35, 40, 44, 78–80, 86, 90, 124–127, 152, 154, 177, 180, 221, 222

CBE Current Best Estimate. 87

CDR Critical Design Review. 35, 76

CDS Correlated Double Sampling. 34, 82, 84, 99, 101, 108

CHEOPS CHaracterising ExOPlanet Satellite. 4

CNN Convolutional Neural Network. 134

CO Common Optics. 68

CoRoT COnvection ROtation and planetary Transits. 4

CRIRES CRyogenic high-resolution InfraRed Echelle Spectrograph. 200, 218, 219

EE Encircled (or Enclosed) Energy. 67–69, 71, 72

EoS Equation of State. 8

ERS Program 1366 JWST Transiting Exoplanet Community Early Release Science Program. 20

ESA European Space Agency. 25, 26, 76, 82

ESO European Southern Observatory. 200

ESPRESSO Echelle SPectrograph for Rocky Exoplanets and Stable Spectroscopic Observations. 200, 201, 213, 214

EUV Extreme Ultra-Violet. 11

EXCITE Exoplanet Climate Infrared TElescope. 40

ExoSim 2 The Exoplanet Observation Simulator. 31, 33, 34, 39–44, 78, 85, 90, 91, 94, 99, 108, 111, 114, 221, 222

FFT Fast Fourier Transform. 56

FGS Fine Guidance System. 27, 28, 40, 67–69, 71, 79, 80, 89, 91, 100, 101, 103, 105, 107, 109, 110, 112, 115, 117, 119, 128, 135, 221

FIREFLy Fast InfraRed Exoplanet Fitting for Lightcurves. 20, 22

FoM Figure of Merit. 81, 87, 88

FP Focal Plane Spectrum. 91

Gaia Global Astrometric Interferometer for Astrophysics. 4, 11

GCM Global Circulation Model. 176

GO General Observers. 200, 201, 206

GUI Graphical User Interface. 46, 52, 53, 221

HARPS High-Accuracy Radial velocity Planetary Searcher. 4

HARPS-N High Accuracy Radial velocity Planet Searcher for the Northern Hemisphere. 200, 201, 213

HATNet Hungarian-made Automated Telescope Network. 5

HST Hubble Space Telescope. 23, 38, 57, 58, 64–66, 76, 80, 82, 201

HTML HyperText Markup Language. 45

HWHM Half Width at Half Maximum. 100

IPRF intra-pixel response function. 96, 98, 99

IRAC Infrared Array Camera. 96, 108

IT In-Transit. 18

JexoSim The JWST Exoplanet Observation Simulator. 39

JT Joule-Thomson. 28

JWST James Webb Space Telescope. 20, 22, 38–40, 43, 80, 176–180, 182, 190, 191, 195–198, 200, 201, 203–206, 208, 209, 218, 220, 223

KELT Kilodegree Extremely Little Telescope. 5, 215

kNN k-Nearest Neighbors. 132–134, 136, 138, 140, 142, 144, 146, 149

LoS Line of Sight. 95, 96, 221

LSTM Long Short-Term Memory Network. 134

M1 Primary Mirror. 77, 221

MC Monte Carlo. 53

ML Deep and Machine Learning. 125, 128, 133, 134, 141, 142, 148, 149

MLP Multi-layer Perceptron. 134, 140, 142, 149

MPE Mean Performance Error. 82, 94, 95

MRS Mission Reference Sample. 86

MRS19 Mission Reference Sample 2019. 86–88, 152, 222, 223

NAMD Normalized Angular Momentum Deficit. 11, 210, 212

NASA National Aeronautics and Space Administration. 4, 7, 13, 126, 152, 184

NDR Non-Destructive Read. 40, 41, 84, 87, 99

NGTS Next Generation Transit Survey. 5

OGLE Optical Gravitational Lensing Experiment. 5

OGSE Optical Ground Support Equipment. 222

OOT Out-of-Transit. 18

OTF Optical Transfer Function. 92

PAOS Physical Optics Simulator. 31, 40, 44–49, 51–53, 55–59, 61–68, 71, 72, 84, 85, 89–91, 93, 94, 221, 222

PCE Photometric Conversion Efficiency. 80

PDE Performance Drift Error. 78, 82, 94, 95

PDF Portable Document Format. 45

pdf Probability Density Function. 70, 96, 174, 175

PDR Preliminary Design Review. 35, 76, 78, 79, 82, 91, 124, 221, 222

PICASO 3.0 The Planetary Intensity Code for Atmospheric Spectroscopy Observations. 20, 22

PLATO PLAnetary Transits and Oscillations of stars. 4

PLM Payload Module. 28

POP Physical Optics Propagation. 44–46, 52, 53, 56, 57, 61, 62, 71, 72, 84, 221

PPD Performance Parameters Database. 78, 79

PSD Power Spectral Density. 58, 70, 71, 94

PSF Point Spread Function. 33, 40, 44, 47, 48, 57–60, 64–68, 70, 71, 84–86, 89–94, 96, 98–102, 109, 221, 222

PyPI Python Package Index. 35, 36

QE Quantum Efficiency. 80, 81, 96, 98, 99, 101, 107, 108

rEE Radius of Encircled Energy. 68, 71, 92, 94, 100, 222

RFC Random Forest Classifier. 134, 141, 142, 149

RMS Root Mean Square. 59, 60, 65–67, 95, 106

ROC Receiver operating characteristic. 223

RPE Relative Performance Error. 78, 82, 94–96, 101, 102, 111

RW Reaction Wheel. 95

S/C Spacecraft. 28, 95

S2MD Simulators Software, Management, and Documentation. 94, 99, 119, 222

SD Surface Deviation. 70, 72

SED Spectral Energy Distribution. 78, 206

SFE Surface Form Error. 70, 71

SNR Signal-to-Noise Ratio. 29, 31, 34, 35, 86, 100, 111, 124, 151, 154, 160, 168–170, 174, 175, 180, 181, 215

SST Spitzer Space Telescope. 76, 80, 82, 96, 108

STOP Structural, Thermal, and Optical Performance. 222

SVC Support Vector Classifier. 134, 141, 142, 149

TA Telescope Assembly. 71, 72

TauREx 3 Tau Retrieval for Exoplanets. 24, 31, 35–39, 124–126, 146, 152, 154, 177, 181, 183, 220

TEP Telescope Exit Pupil. 68–71

TESS Transiting Exoplanet Survey Satellite. 4, 126, 152, 202, 210, 215

TIS Total Integrated Scatter. 71, 80

TRAPPIST TRAnsiting Planets and PlanetesImals Small Telescope. 11

TRL Technical Readiness Level. 76

UHJ ultra-Hot Jupiter. 23

VLT Very Large Telescope. 200, 218

WASP Wide Angle Search for Planets. 5, 215

WFC3 Wide Field Camera 3. 38

WFE Wavefront Error. 53, 59, 60, 65–68, 70–73

XUV Extreme Ultra-Violet. 11, 217

Glossary

Alfnoor The Thousand Lights Simulator: a tool for automated quantitative exoplanetary population studies. 33, 124, 125, 127, 152

ExoRad 2.0 The generic point-source radiometric simulator for exoplanet observations. 31, 33, 126, 180

Kepler The Kepler space telescope. 4, 5, 11, 13

L2 Sun-Earth Lagrange Point 2. 26, 27, 80

PROPER An Optical Propagation Library for IDL, MatLab, and Python. 56, 64–66