Dissertation

submitted to the

Combined Faculties of the Natural Sciences and Mathematics

of the Ruperto-Carola-University of Heidelberg. Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Matthias Severin Samland

born in Mannheim, Germany

Oral examination: 16. July 2019

High-Contrast Imaging Characterization of Exoplanets

Referees:

Prof. Dr. Thomas Henning

PD Dr. Sabine Reffert

Abstract

Direct imaging of exoplanetary systems and the spectral characterization of exoplanetary atmospheres are amongst the most challenging, as well as rapidly developing fields in astronomy, propelled by new technologies and observational strategies. In this thesis, I contributed to the atmospheric analysis of exoplanets, the development of new algorithms to find faint planet signatures in the data, and the improvement of the fidelity of obtained exoplanet spectra. I performed atmospheric analyses of directly imaged planets observed with the planet imaging instrument VLT/SPHERE. For this purpose, I wrote a statistical inference code (BACON, Bayesian Atmospheric CharacterizatiON), which uses self-consistently computed model atmospheres to derive atmospheric parameters. The planets I studied in this thesis are: 51 Eridani b, one the coldest methane-rich directly imaged planets; PDS 70 b, the first young planet discovered inside the gap of its host star's transition disk; HIP 65426 b, a planet of similar spectral type to PDS 70 b, but hotter and older; and GJ 504 b, a colder methane-rich companion which, depending on its age, could be a planet or brown dwarf. The new algorithm I developed to detect planets in high-contrast imaging data shifts the focus from an image analysis interpretation of the data, towards a time-domain analysis approach. I show that with this technique (TRAP, Temporal Reference Analysis for Exoplanets), an improvement of up to a factor of six in signal-to-noise can be achieved at very small angular separations between the planet and host star. Furthermore, I adapted the CHARIS instrument pipeline to use with SPHERE-IFS. This pipeline opens new possibilities for improving the quality of spectra obtained for exoplanets using SPHERE. Using this pipeline, I confirm the low flux emitted at around 1 micron previously obtained for 51 Eridani b, consistent with the absorption due to methane and water opacities predicted by models. Lastly, I discuss the future prospects for my work and how these approaches can be combined into a single framework.

Zusammenfassung

Das direkte Abbilden exoplanetarer Systeme und die spektrale Charakterisierung exoplanetarer Atmosphären gehören zu den herausforderndsten und sich am schnellsten entwickelnden Gebieten der Astronomie, die durch neue Technologien und Beobachtungsstrategien vorangetrieben werden. Ich habe zur atmosphärischen Analyse von Exoplaneten, zur Entwicklung neuer Algorithmen zum Auffinden schwacher Planetensignaturen in den Daten, sowie zur Verbesserung der Genauigkeit der erhaltenen Exoplanetenspektren beigetragen. Ich habe atmosphärische Analysen von direkt abgebildeten Planeten durchgeführt, die mit dem SPHERE Planetenabbildungsinstrument am VLT beobachtet wurden. Zu diesem Zweck habe ich einen Code für statistische Inferenz (BACON, Bayesian Atmospheric CharacterizatiON) geschrieben, der selbstkonsistent berechnete Modellatmosphären verwendet, um atmosphärische Parameter zu bestimmen. Die von mir in dieser Dissertation untersuchten Planeten sind: 51 Eridani b, einer der kältesten methanreichen direkt abgebildeten Planeten; PDS 70 b, der erste junge Planet, der in der Lücke einer zirkumstellaren Scheibe um seinen Zentralstern entdeckt wurde; HIP 65426 b, ein Planet von ähnlichem Spektraltyp wie PDS 70 b, jedoch heißer und älter; und GJ 504 b, ein kälterer methanreicher Begleiter, der je nach Alter des Systems ein Planet oder ein Brauner Zwerg seien könnte. Der neue Algorithmus, den ich zur Erkennung von Planeten in Bildsequenzdaten entwickelt habe, verlagert den Fokus von

einer Interpretation des Problems von der Bildanalyse der Daten hin zu einer Zeitserien-Interpretation. Ich zeige, dass mit dieser Technik (TRAP, Temporal Reference Analysis for Exoplanets) eine Verbesserung des Signal-Rausch-Verhältnisses um bis zu einem Faktor sechs bei sehr kleinen Winkelabständen zwischen Planet und Stern erzielt werden kann. Weiterhin habe ich die CHARIS Instrumenten-Pipeline für SPHERE-IFS angepasst. Diese Pipeline eröffnet neue Möglichkeiten zur Verbesserung der Qualität der mit SPHERE beobachteten Spektren von Exoplaneten. Mit dieser Pipeline bestätige ich den niedrigen Fluss der bei einer Wellenlänge von etwa 1.3 Mikrometer für 51 Eridani b gemessen wurde, und der mit der von Modellen vorhergesagten Absorption durch Methan- und Wasseropazitäten übereinstimmt. Zuletzt diskutiere ich die Zukunftsperspektiven meiner Arbeit und wie diese Ansätze in einem gemeinsamen Framework kombiniert werden können.

Contents

List of Figures

List of Tables

Chapter 1

"*Measure what is measurable, and make measurable what is not so.*"

by Galileo Galilei

---

1 Introduction

Since ancient times we humans have wondered about our place in the universe. One question in particular has always fascinated us: are there other worlds? If so, can we observe them? Learn about them? For most of human history these ideas and thoughts have been idle speculation.

Only at the dawn of modern astronomy, when Galileo Galilei pointed his telescope to the skies, did it become known that the planets are spheres – and indeed – other worlds. Worlds that even had their own moons. His work, *The Starry Messenger* (Sidereus Nuncius) was published in 1610. Since then people have gone on to wonder about the surfaces of these worlds, their atmospheres, even if they could possibly harbor life of their own.

However, for almost another 400 years, the planets in our own solar system would remain the only known worlds. But still people wondered: are there other worlds around the countless stars in the universe beside our Sun? Are there *extrasolar* planets? If they do exist, how common are they?

It took until the late 1980's for the first predictions and potential discoveries of such *exoplanets* to crop up. Only in 1995, with the discovery of 51 Pegasi b, was the first planet around a sun-like star convincingly shown.

The field of studying exoplanets is one of the youngest in astronomy, but also one of the most rapidly developing. From the initial discovery of one planet in 1995, we have come a long way with discoveries now numbering in the several thousands.

Most of these discoveries have been made by so-called *indirect* detection techniques, i.e. by inferring the existence of planets by their influence on the light detected from their host star.

However, this thesis is about what is arguably the most gratifying and oldest way of studying other worlds, taking images of them. It is also one of the most challenging problems in astronomy, technically, methodologically, and in terms of interpreting the results. Generally, this method of observing exoplanets is called *direct imaging* or *high-contrast imaging* (HCI), because there is huge brightness difference between the star and the planets. Even in the most favorable observed cases, planets are typically more than 100,000 times fainter than their star, but depending on wavelength, the type of planet, the contrast can be much worse. This situation is often likened to the problem of seeing a firefly next to a very bright street lamp. The faint signal of the firefly is drowned out by a halo of much brighter light.

1.1 Direct and Indirect Planet Detection Methods

Astronomers often split exoplanet detection techniques into two categories: *direct* detection and *indirect* detection methods. The former, which is also the main topic of this thesis, is the discipline of detecting photons "directly coming from the planet," which usually means either emission from the planet itself, or stellar light which is scattered by the planetary surface. Both emitted light as well as reflected light, carry information about the planetary atmosphere. With "indirect methods" on the other hand, we infer the existence of a planet based on measurements that *are influenced* by the existence of a planet, but we do not measure light from the planet itself. There are many ways in which a planet can make its presence known.

To set the stage, I first give a short overview of these so-called indirect methods. Indirect methods discovered the first exoplanets, as well as the vast majority of known exoplanets today. The "transit method" is discussed last, as it is the most closely related to direct detection methods since it allows us to learn about the atmosphere of planets. Some aspects usually discussed in the context of transits, such as the measurement of phase curves and secondary eclipse measurements are explained in the next Section 1.2 about direct detection of light emitted by exoplanets.

1.1.1 Radial Velocity of the host star

The radial velocity (RV) method is based on measuring the gravitational influence of a planet on its host star and vice-versa as shown in Fig. 1.1. Two massive objects both orbit around the common center of gravity, therefore a star "wobbles" due to the presence of a planet.

As measuring the stellar light is much easier than directly detecting the planet, measuring the radial motion of the star (its line-of-sight velocity) based on the Doppler shift of spectral lines is a powerful method for detecting exoplanets. The Doppler shift in the wavelength of light is given by (Einstein, 1905)

$$\lambda = \lambda_0 \cdot \frac{1 + \frac{1}{c} \cdot \mathbf{k} \cdot \mathbf{v}}{\sqrt{1 - \frac{\mathbf{v}^2}{c^2}}}, \tag{1.1}$$

where λ is the wavelength of a photon measured by an observer, λ_0 is the wavelength at which the photon is emitted, c is the speed of light, $\mathbf{k}$ is the unit vector point from the observer to the source of the photon and $\mathbf{v}$ is the velocity vector of the source at the time of emission, as seen from the observer, which means that by measuring the wavelength shift of lines we can measure the radial velocity of stars.

The achievable radial velocity precision σ_{RV} scales with the square-root of the number of lines of the star measured,

$$\sigma_{\mathrm{RV}} \propto \sqrt{N_{\mathrm{lines}}}. \tag{1.2}$$

In other words, high spectral resolution allows the co-addition of hundreds of thousand of individual line measurements, making the RV method a powerful tool for detecting even signals on the order of meters per second. At this precision, the main challenge is distin-

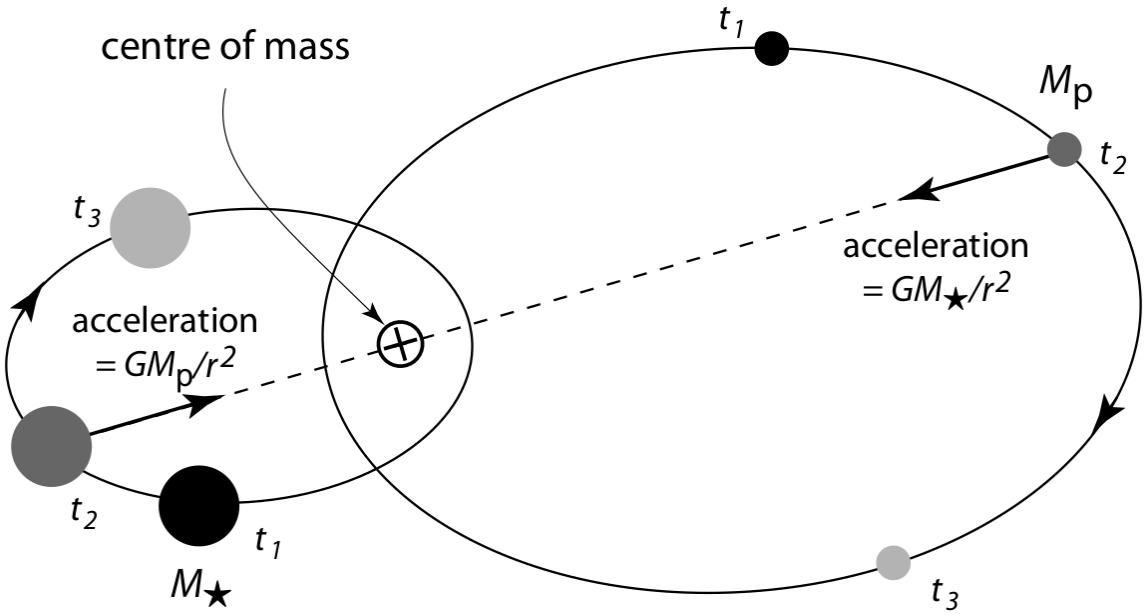


Figure 1.1: Schematic of the radial velocity detection method, which shows the gravitational interaction between the planet and its star. Source: Perryman (2011)

guishing a planet induced signal from intrinsic and systematic motion of the surface of the star, such as stellar activity or pulsation.

Astronomers usually quantify the radial velocity of a star in terms of the *semi-amplitude* K of the periodic radial velocity signal, which can be derived from Kepler's laws, and written as

$$K = \frac{28.433 \frac{\mathrm{m}}{\mathrm{s}}}{\sqrt{1-e^2}} \frac{m_{\mathrm{pl}} \cdot \sin i}{\mathrm{M_{Jup}}} \left(\frac{m_{\mathrm{pl}} + M_*}{\mathrm{M}_\odot} \right)^{-1/2} \left(\frac{P}{\mathrm{yr}} \right)^{-1/3} . \tag{1.3}$$

Here a planet of mass m_{pl} orbits a star of mass M_* with the eccentricity e, the inclination i, and a period of P. In this equation, we can see the effect of only measuring the radial component of the stellar motion: the planetary mass and the inclination of the orbit are degenerate, which is why we can generally only give a lower limit on the mass of a planet. If, however, we can independently determine the inclination i of the planet's orbit, for example, when it happens to transit in front of the star, we can break this degeneracy and determine the planet's mass. The combination of the transit method with radial velocity measurements is therefore a powerful tool and allows us to estimate both the radius (see explanation below), and mass, and therefore the average density of the planet. The average density can be used together with models of the interior of planets, to distinguish for example between gas dominated planets and rocky worlds.

1.1.2 Astrometry of the host star

The astrometric detection of exoplanets can be seen as the complement to the radial velocity method, in that it is based on the displacement of the star in the plane of the sky, i.e. perpendicular to the radial direction. By measuring this two dimensional motion, the orbit of the star can be solved unambiguously, and therefore the mass of the planet can be determined. Assuming the planet's mass to be much smaller than the stellar mass ($m_{\rm pl} \ll M_*$), the astrometric signal θ_A follows,

$$\theta_A = 3\,\mu\text{as}\,\frac{m_{\rm pl}}{\rm M_\oplus}\left(\frac{M_*}{\rm M_\odot}\right)^{-2/3}\left(\frac{P}{\rm yr}\right)^{2/3}\left(\frac{d}{\rm pc}\right)^{-1}, \tag{1.4}$$

for a planet on a circular orbit of period P, at a distance of d from the observer.

Due to the high-requirements on the astrometric precision, no new exoplanet has been *discovered* using this method so far, although attempts have been made to recover a signal of known exoplanets.

This should change with the astrometric precision achievable by the Gaia satellite, which is currently in the main phase of its mission. According to estimates by Sozzetti et al. (2014), the Gaia mission could reveal about 2600 exoplanets around M-dwarfs alone. This technique will be interesting, with regards to the direct detection of exoplanets in the future, as there is a potential overlap between the discovery space of Gaia and advanced high-contrast imaging facilities.

1.1.3 Microlensing of background objects

The microlensing method is probably the most indirect detection method for exoplanets. With this technique we do not detect light from the planet, nor do we usually detect light from the planet's host star. Microlensing is a consequence of Einstein's general theory of relativity, the fact that gravity curves space-time. An object that passes between us and a distant background object, can lead to a temporary "focusing" of the light of the background object over the duration of the alignment, making it appear brighter. The name "lensing" comes from this metaphor of the foreground object acting as a lens for the background.

In microlensing, not only does the brightness of the background star increase as a foreground star crosses in front, if the foreground star hosts planets at certain separations, the signature of these planets can be imprinted in the temporal behavior of the brightening/dimming of the background star. Microlensing is especially interesting from the point of view of population statistics of exoplanets, as it can be sensitive to Earth-mass planets. However, it does not allow us to characterize the discovered planets in detail, and it is not possible to follow-up on the observations. This method does not play any further role in my thesis and I refer the reader to Gaudi (2010) for more detailed information.

1.1.4 Gas kinematics of planet-forming disks

Detecting the presence of planets in very young systems that still host a planet-forming disk can be very difficult. The disk material often strongly attenuates the emission of the planet making it especially hard to detect. If a planet is still embedded in disk material the

difficulty of finding a planet and distinguishing it from, e.g. an overdensity in disk material, is even higher.

However, with the Atacama Large Millimeter/submillimeter Array (ALMA) observatory it has become possible to measure dust and gas distributions in planet-forming disks with high fidelity. Recently, the DSHARP survey (Andrews et al., 2018b) was published, showing many beautiful and informative dust continuum observations of transition disks, disks with gaps in their dust distribution, as is shown in Fig. 1.2.

While the gaps in these disks are highly suggestive of potential planets responsible for clearing these gaps, they are not proof, as other explanations are not completely ruled out. Alternative explanations include ice lines (e.g., Zhang et al., 2015), magneto-hydrodynamical effects (dead zones), or photoevaporation (e.g., Clarke et al., 2001). However, photoevaporation more likely explains holes rather than gaps, because an isolated inner disk would accrete quickly onto the star once a gap has opened. A detailed discussion of these mechanisms can be found in the DSHARP paper series.

However, another possibility to infer the existence of planets by measuring the gas velocity distribution inside the planet-forming disks has recently emerged (Teague et al., 2018; Pinte et al., 2018). In these works, the presence of a planet in the disk of HD 163296 is deduced by measuring the rotation curves of CO isotopologue emission to sub-percent precision with respect to the expected Keplerian rotation of the disk.

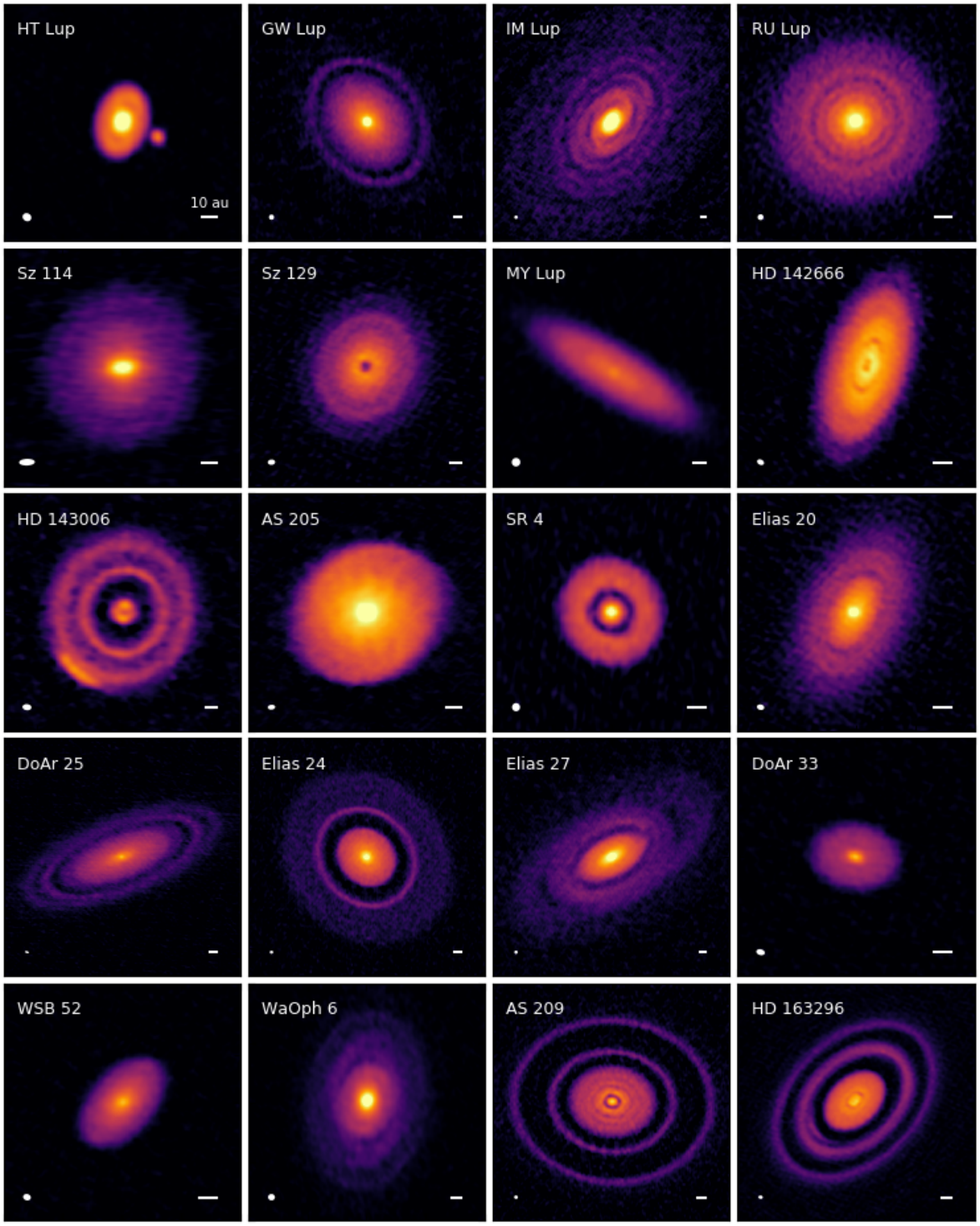


Figure 1.2: Gallery of 240 GHz (1.25 mm) continuum emission images for the disks in the DSHARP sample. Source: Figure 3 in Andrews et al. (2018b)

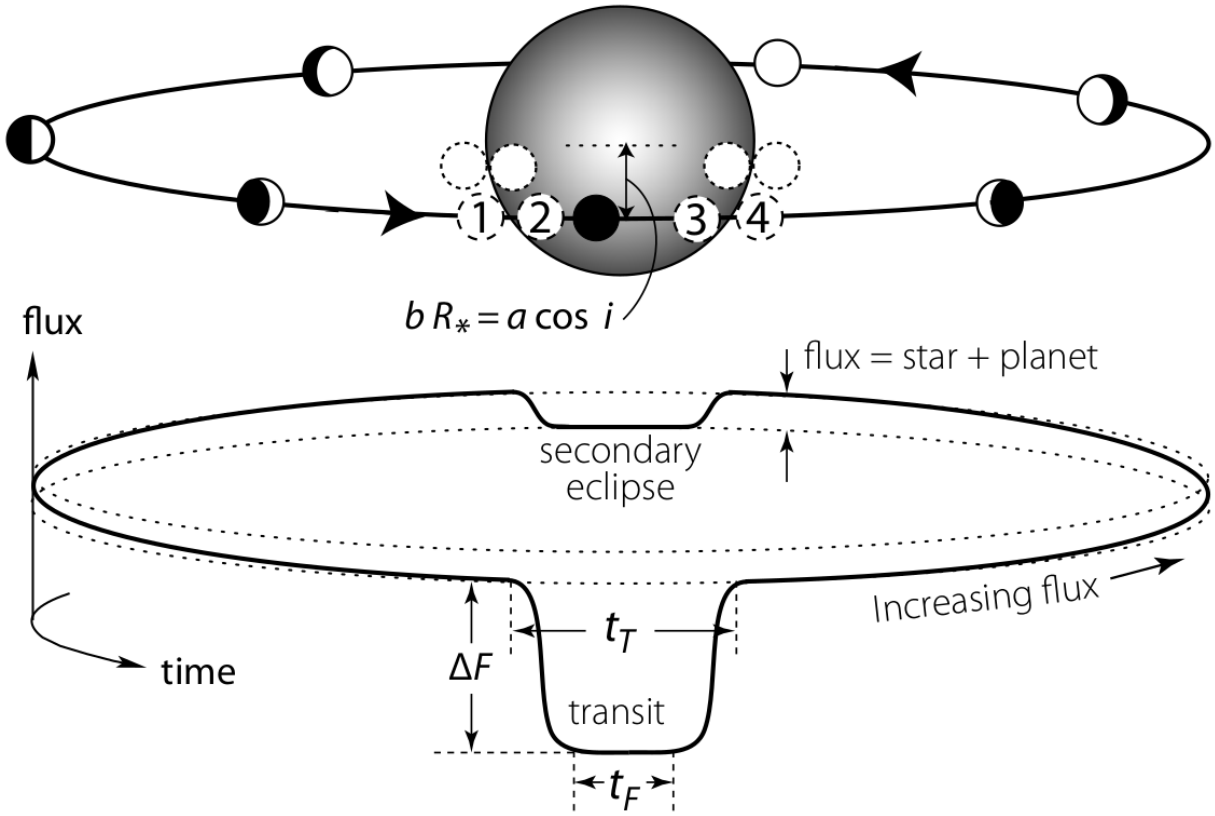


Figure 1.3: Schematic of the transit detection method showing observable quantities. As the planet orbits the star and the phase of the planet changes the integrated light signal of the system changes. When it passes directly in front or behind the star the overall flux drops. The temporal diversity between how much star and planet contribute to the overall flux enables the two signals to be teasing apart. Source: Perryman (2011)

1.1.5 Planetary Transit Method

The transit method is the most successful detection method by far, with thousands of discoveries to date. Figure 1.3 shows a schematic of how the transit method works. It involves measuring the integrated flux of the star-planet system with very high-precision and deducing information about the objects based on the temporal behavior of the signal. The transit method has both aspects of indirect and direct detection methods.

The primary transit

When the orbit of a planet happens to geometrically lie between us, the observer, and the star, it will block out a fraction of the starlight. This is called the *primary transit* and can be classified as an indirect method. The resulting dip in the flux of the star ΔF measured is related to the radius of the planet by $\Delta F = (\frac{R_{\text{planet}}}{R_{\text{star}}})^2$. As the stellar radius can be determined either from stellar evolutionary models or measured by interferometry, we can infer the planetary radius at a given wavelength. Based on the periodicity at which the transit occurs, we furthermore learn the orbital period of the planet. If there are multiple

planets in a system, in some cases, it is further possible to constrain their masses based on *transit timing variations*, because mutual interactions between the planet will change the timing at which each transit will occur. This depends on the precision with which the timing can be measured, the orbital architecture, and the masses of each body. It is also possible to detect exomoons – moons around exoplanets – using transit timing variations. The transit method also allows us to learn about the orbital alignment with respect to the rotation axis of the star using the Rositter–McLaughlin Effect. This effect occurs when the receding or approaching limb of the star is sequentially covered by the transiting planet, and thus rotationally red shifted and blue shifted components are blocked out.

Among the indirect methods, the transit method is special in that it also allows us to use different ways to gain information about the atmosphere of the planet, for example by the way it blocks the starlight at different wavelengths during the primary transit. The opacity differences in the atmosphere lead to the measurement of a different apparent planetary radius, from which we can construct a *transmission spectrum* of the planet. This spectrum shows the wavelength dependent effect of the atmosphere on the star light passing through it, as such it is strongly dependent on a lot of factors and usually quite difficult to interpret, as we need to have a good understanding of both the star (e.g. limb darkening laws, activity, star spots), and the planet. Language in the scientific community becomes a bit vague here, because if a "direct detection of a planet in reflected light" would be considered a direct method, why would a transmission spectrum not be considered the same? In this case, it is probably a question of arbitrarily requiring the additional condition of being able to spatially resolve the star and planet into two independent objects, but a clear classification is difficult. In part, this is likely due to how the usage of the words have evolved over time. Personally, I would argue that methods in which the photons we detect have interacted directly with the planet can be considered to be direct detection techniques.

However to be consistent, I discuss the more obviously direct detection aspects usually associated with the transit method – the measurement of phase curves and the secondary eclipse – together with other direct detection techniques in Section 1.2.3.

1.2 Direct Detection of Light emitted by Exoplanets

1.2.1 The problem of detecting planets directly

The primary challenge in directly detecting the light coming from an exoplanet is the extreme difference in intensity between the flux coming from a planet and its host star, the contrast F_{planet}/F_*. The contrast depends on a large number of factors, for example whether we observe the system in the visible in order to detect stellar light reflected off of a planet's atmosphere, or in the infrared to detect thermal emission from the planet. Detecting reflected light in visible light is especially difficult: for a system like Jupiter and the Sun (which if put at a distance of 10 pc would have an angular separation of 0.5"), the contrast between planet and star is $\sim 10^{-9}$. For an Earth–Sun system the contrast goes down to $\sim 10^{-10}$. Teasing out a signal less than a billionth the strength of the signal of the host star is a truly daunting task.

Figure 1.4 gives an overview of planet-to-star contrast ratios for different situations, assuming a Sun-like star. The left panel gives the model predicted contrast for a $1\,M_J$ planet in an old system (5 Gyr) over wavelength and as a function of separation from the star. The middle panel shows a $1\,M_J$ planet at 4 au, but at a number of different ages. The right panel shows the 5 Gyr old system, but with different planet masses.

This shows us several important things: the Rayleigh scattering component (reflected light) of a planet decreases with separation from the host star, the thermal emission of a planet decreases strongly with age, and higher mass planets are generally more detectable, due to lower planet-to-star contrast ratios.

Given the difficulty of detecting planets in reflected light (see Milli et al., 2013, for a discussion of detecting planets in the nearby Centauri B system), most current research focuses on near-infrared (NIR) emission of planets. The most favorable targets for direct detection are targets with strong emission (i.e. high temperature), either hot Jupiters, due to their proximity to their host star, or objects that are *young* and retain much of their energy from formation.

The first population – hot Jupiters – can be best detected and characterized using temporal and spectral diversity detection methods (see description of these methods in Section 1.2.4), because their angular separation is too small to meaningfully resolve them spatially.

The second population – hot, young planets, at separations of several astronomical units – is the main topic of this thesis, as these planets are particularly well suited for direct detection through imaging.

1.2.2 Exploiting diversity: how the planet differs from the star

In order to distinguish the light coming from a planet from light coming from its star, the light originating from the planet has to – in some way – be different from the light of the star. There are various different ways in which planet light can differ from star light. There are two levels to this distinction, one is an intrinsic difference between the stellar light and the planetary light, the other is diversity in the behavior of the systematic noise compared to the planet signal, i.e. a difference between the signals caused by how we observe the target and how the instrument interacts with the incoming signals. I will refer to these

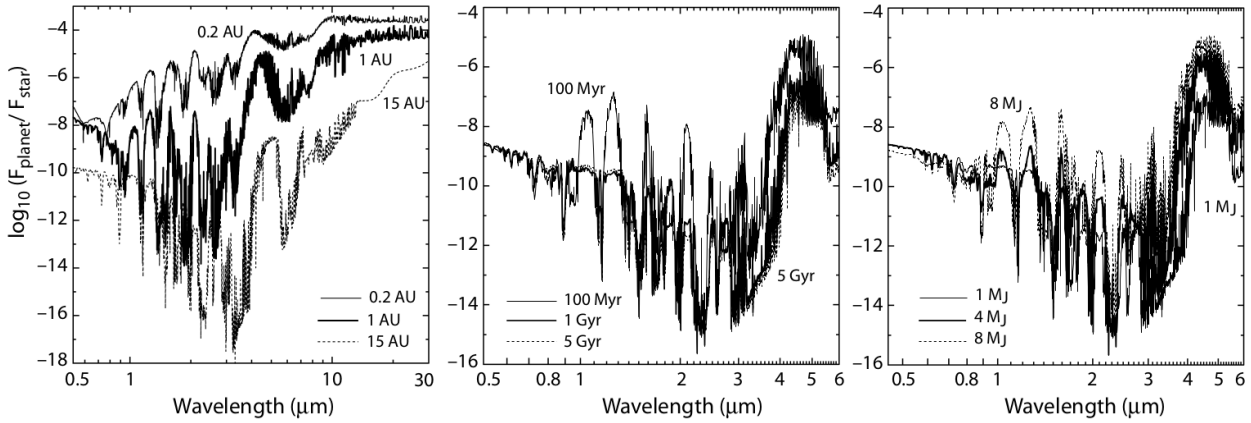


Figure 1.4: Model based planet-to-star contrast ratio for different wavelengths. The stellar model is assumed to be a sun-like star (G5V, solar metallicity), and the giant planet model includes H_2O and NH_3 clouds. Left: $1\,M_J$ planet with an age of 5 Gyr, assuming different orbital separations. Middle: $1\,M_J$ planet at 4 au, assuming different ages. Right: 5 Gyr old planet at 4 au, for a range of different planet masses. Source: Adaptation of Burrows et al. (2004) published in Perryman (2011)

as *intrinsic diversities* and *induced diversities*, respectively. In high-contrast imaging in particular, induced diversities play a major role in how we create a model of the systematic noise.

The most obvious diversity is spatial separation. In this scenario, the planet is located at a different position on the sky than the star. Thus, if light from the planet can be detected, we are able to use this spatial information to identify photons coming from the planet separately from those that originate from the star. We will refer to this as "spatial diversity", which is the basis for direct imaging. Similarly, the star and planet, due to their different temperature and composition, are different in their spectrum, which leads to intrinsic spectral diversity. As is shown in Fig. 1.3, describing the transit method, we can also exploit a difference in temporal behavior of the planet and stellar signal, i.e. "temporal diversity". While the light emitted by a star is generally unpolarized, it becomes polarized by scattering off a planet (or disk), resulting in "polarimetric diversity".

The technical problem we face is therefore the separation of the light of the star from the light of the planet, and numerous ways to achieve this have been developed. Usually a combination of the above diversities is used – and has to be used – to achieve this formidable goal of teasing out a very small signal next to a much stronger signal.

I discuss intrinsic diversities and how they can be exploited as well as general synergies that result from combining them with imaging (e.g. spectral imaging). However, it should be kept in mind that in principle all diversities can be combined with others and become more powerful as a result of such synergistic effects between them. In Chapter 2 for example, we combine induced spectral diversity with imaging. Chapter 5 deals in particular with how to improve the quality of the construction of spectral imaging data cubes for high-contrast imaging from raw data.

Lastly, in an ideal world, if we can perfectly spatially separate the light of the star from the light of the planet, we do not have to worry about unfavorable contrast and are purely

limited by the sensitivity of our instrument.

1.2.3 Temporal diversity

An intrinsic temporal difference between the behavior of the star and planet allows us to tease apart the flux components from these two sources. Figure 1.3 shows this for the transit method. We will now discuss the secondary eclipse and phase curve of planets, which are often grouped with the transit method, but involve the direct detection of emission from the planet based on temporal diversity. However, besides intrinsic temporal diversity, induced temporal diversity between the stellar and planetary signal caused by observational techniques can also be a powerful tool.

Spatially unresolved planet detection I

The secondary eclipse If a planet transits in front of its host star, it will eventually also pass behind the star. As it is partially or completely occulted by the star, the emissive contribution of the planet to the overall flux of the planet is reduced. By taking the difference of the flux measured during the secondary eclipse and the average flux before and after, we can obtain a more accurate measurement of the emission coming directly from the day-side of the planet, as it is easier to measure differences than absolute values.

The phase curve As Figure 1.3 shows, the planetary contribution to the integrated light of the planet–star system changes with the phase of the planet over the course of the planet's orbit. This is the case even if the planet is *not transiting* the star. The fraction of day-side vs. night-side contribution of the planet that we observe changes over time. Therefore, by observing whole or partial orbits of the planet, we can learn about the emission coming from the planet itself, based on the characteristic temporal behavior determined by the orbit of the planet. This can also be classified as "temporal diversity".

Phase curves have only been measured for hot Jupiters (e.g. Snellen et al., 2009; Knutson et al., 2012), massive planets close to their host stars, which are hot (> 1000 K) due to the stellar irradiation. There are multiple challenges associated with measuring the phase curve: a) the photometric precision needed to detect phase curves is very high, the hotter the planet, the easier it is to measure its contribution (i.e. a problem of contrast, as in direct imaging), b) short period orbits are favored, both because of the before mentioned reason of temperature, and because it is often required to stack multiple orbits to achieve high enough signal-to-noise, c) continuous observation and the required photometric precision and stability is easier to achieve from space, making observations expensive.

Using phase curves, we can learn about heat transport and winds in the atmospheres of detected planets (e.g. Knutson et al., 2009), as the phase change over time lets us scan the emission distribution of the surface of the planet, going so far as allowing the creation of coarse maps of the surface (e.g. de Wit et al., 2012). Furthermore, hot Jupiters are tidally locked to their host star, i.e. tidal forces have caused them to always face the same side to the star (in the same way that the same side of the Moon always faces the Earth). In a planetary atmosphere without wind that transports heat, we would expect the sub-stellar spot – the point of the planet surface facing the star – to be the hottest, due to it receiving the most direct irradiation. In reality, the measured location of this spot is displaced by some

amount due to transport processes in the atmosphere. For non-tidally locked planets, due to changes in flux as they rotate periodically, and recurrent patterns, we may in the future be able to disentangle latitudinally averaged surface structures, enabling us to distinguish between continents and oceans.

Induced temporal diversity in direct imaging

Direct imaging observations are usually performed in pupil tracking mode (see Chapter 1.4), in which the observed field-of-view rotates over the course of an observation sequence. This means off-axis companions move along an arc with respect to the on-axis primary source (host star), while the systematic noise pattern surrounding the star, caused by the atmosphere and the instrument optics, remains stationary, because the telescope pupil is fixed. We therefore have an induced spatial diversity over time, which can be interpreted as induced temporal diversity. This allows us to use similar systematic noise modeling techniques in high-contrast imaging as are being used in the field of transit light curve observations, where intrinsic temporal diversity is used. This interpretation of pupil tracking data as induced temporal diversity is explored in detail in Chapter 4 of this thesis, and bridges the methodological gap between the direct imaging and transit communities.

1.2.4 Spectral diversity

Similar to temporal diversity, we can distinguish between intrinsic spectral diversity between stellar and planetary spectra, e.g. due to their difference in temperature and composition, and induced diversity in the stellar and planetary signal we measure based on the wavelength dependence of the interaction of light with the atmosphere and telescope optics.

Spatially unresolved planet detection II

High dispersion spectroscopy (HDS) – spectroscopy with very high spectral resolution – has opened the door to distinguishing planet and star light by looking for line emission of molecules that do not occur in most stellar photospheres due to high temperatures. The first such success was the detection of carbon monoxide at 2.3 μm in the transmission spectrum of HD 209458 b (Snellen et al., 2010). Two years later, Brogi et al. (2012) managed to detect the same molecular species in the thermal spectrum of the *non-transiting* planet τ Boötis b, for which they determined its mass and orbital inclination. These methods usually do not solely rely on spectral diversity, but combine it with temporal diversity introduced by the orbital velocity of the planet around its star. HDS has made it possible to measure the spin of the brightest planets (Snellen et al., 2014).

Intrinsic and induced spectral diversity in direct imaging

Spectral diversity is especially powerful when combined with imaging. In spectral imaging one takes images at different wavelengths at the same time, for example using an integral field spectrograph (IFS), also known as an integral field unit (IFU). Images taken at different wavelengths at the same time, especially those wavelengths where the star and planet signal are intrinsically different (i.e. in a methane absorption band) can be used to subtract

the stellar halo without biasing the planet signal significantly. In the case of images in a methane absorption band this is because they do not contain much flux from the planet, due to absorption from methane, if it is present in the planet's atmosphere. This technique, which primarily aims to exploit intrinsic differences is referred to classical *spectral differential imaging* (e.g., Rosenthal et al., 1996; Racine et al., 1999a; Marois et al., 2005, SDI).

However, it was realized at a relatively early stage that the spectral diversity of the systematic noise induced by the wavelength dependent diffraction of light is a powerful tool in itself, even if no intrinsic spectral diversity exists (Sparks & Ford, 2002). The diffraction pattern of the telescope PSF (see, Eq. 1.5 and Fig. 1.5), speckles, and light scattering in the optics, all scale with wavelength. This means that the systematic pattern of light surrounding the bright star which obscures the planet, generally moves outwards with wavelength. Assuming further, that the aberrations are achromatic and that phase aberrations are small, the coronagraphic PSF image spatially scales linearly with wavelength (Pueyo & Kasdin, 2007), i.e. it becomes bigger with longer wavelength. In other words, by radially rescaling a PSF relative to a central source measured at $\lambda_2 > \lambda_1$ by a factor of λ_1/λ_2, we can spatially align the speckles to match the PSF at λ_1, thereby creating a systematics model for λ_1 (and vice-versa). This process of creating a systematics model based on the wavelength dependence of the diffraction/speckle pattern is sometimes called *spectral deconvolution*. The spectral diversity between the stellar signal (systematic noise pattern) and planetary signal is exploitable, because the position of the planet signal does not depend on wavelength, as it is an independent light source on the sky. This is one example of a synergistic effect between diversities, here the spatial and spectral diversity.

In practice it should be noted that there is a general trade-off to be made. Since spatially rescaling the PSF with wavelength is done using a first-order, linear approximation, and aberrations are generally not perfectly achromatic, using neighboring wavelength channels provides a more accurate systematics model than using widely separated spectral channels. In contrast, if the spectral channels are too close to each other, the radial displacement of the planet caused by rescaling the images is too small, especially at small angular separations between planet and host star. This will lead to contamination of the systematics model and self-subtraction of the planet signal, reducing the signal-to-noise ratio of the planet. Ideally, we would therefore want to use images that are separated enough in wavelength to have sufficient displacement, but not so much that the linear approximation breaks down. This trade-off is similar to the temporal exclusion criterion (i.e. protection angle) used in angular differential imaging (see Chapter 1.4), and could be referred to as a wavelength exclusion criterion. My new algorithm introduced in Chapter 4, may also partially address this issue when applied to spectral imaging, as we do not have to *explicitly* spatially rescale the images using the linear approximation in order to include their information on temporal systematics at different wavelengths in our regression model.

Modern direct imaging instruments are equipped with (low dispersion) integral field spectrographs that provide simultaneous images at a range of wavelengths (spectral image cubes) in order to make use of both intrinsic and induced spectral diversity (e.g., Macintosh et al., 2008; Beuzit et al., 2019; Groff et al., 2015). Work has also recently begun on combining HDS with high-contrast imaging by using *high-dispersion coronagraphy* (Snellen et al., 2015) and *molecular mapping* (Hoeijmakers et al., 2018). These are attempts at combining intrinsic spectral diversity on a line-by-line level, spectral variation of the lines introduced by

Doppler shifts caused by the planet's orbit and rotation (which adds a temporal diversity component), with the high stellar noise attenuation provided by spatially separating the sources and suppressing the light of the central star with coronagraphy. We can see that combining different diversities can add increasing layers of sophistication.

1.2.5 Polarimetric diversity

Starlight can, in most cases, be assumed to be unpolarized, whereas light reflected by a planet or disk will be polarized. As disks have a relatively large scattering surface area, this method of detection is particularly powerful for planet-forming disks and debris disks. For systems with extensive disks, it is possible to detect a general polarization fraction of the system induced by scattering of the disk, even without imaging and spatially resolving the disk structure.

However, as discussed above, the contrast for reflected light scattered by planets is very high (see Section 1.2.1). Planets have a relatively small surface and the amount of scattered light depends on the phase, i.e. the fraction of the planet's day-side visible to us. Furthermore, the polarization fraction is assumed to be on the order of a few percent (e.g., Milli et al., 2013; Jensen-Clem et al., 2016). As such, combining polarimetry with imaging (spatially separating the objects) becomes necessary to further suppress the amount of stellar light at the position of point sources and disks. This technique is called "polarimetric differential imaging" (e.g., Kuhn et al., 2001, PDI).

1.2.6 Other diversities

There are a number of other diversities that can be exploited in order to image planets. The discussion of these methods are beyond the scope of this thesis, however they include differentiating between light sources based on the coherence of light (e.g., Guyon, 2004, CDI) or arrival time statistics of photons using new detector designs that allow differentiating between individual photons (Meeker et al., 2018).

1.3 Direct Imaging (Spatial diversity)

One may naively expect that in astronomy, where the stars we observe are usually too small to resolve their diameter, and therefore register as *unresolved point sources*, it would be relatively easy to disentangle two such point sources next to each other.

However, the image of an unresolved object on a detector is not infinitely small, but determined by the optics of the telescope, the so called *point spread function* (PSF) that defines the diffraction pattern of the optics. This pattern is also called the Airy disk or Airy pattern. In the case of an ideal optical system the PSF is mainly determined by the observed wavelength of light λ, and the diameter D of the (circular) telescope aperture.

Two objects of similar brightness can be considered to be spatially resolved (distinguishable), according to the Rayleigh criterion if they are separated by

$$\theta(\text{radian}) = 1.22\,\lambda/D, \tag{1.5}$$

where D is the diameter of the aperture of the telescope and λ the wavelength of light. The factor 1.22 derives from the location of the first minimum of the diffraction pattern. Figure 1.5 shows two equally bright point sources (PSFs) observed with a circular aperture, at different separations. There are multiple Airy rings with decreasing intensity the further we move from the center of the source, therefore, even if we spatially resolve two objects, this does not imply that there is no stellar light at the location of a companion object. We often use λ/D (resolution elements) as a unit for angular separation in direct imaging, because of its convenience for interpreting results, as is done in Chapter 4 of this thesis.

Assuming the above criterion for an 8m–class telescope and a wavelength of 1 µm, we reach an angular resolution of about $0.03''$ or 30 mas (milli-arcsecond). At a distance of 30 pc this corresponds to slightly less than 1 au.

However, the above example applies for an ideal optical system, without any atmospheric disturbance, and two objects of the same apparent brightness. The reality is much more challenging. We have to correct the distorted wavefront of light after it passes through the atmosphere and strongly suppress the light of the star at the location of planets in order to make them detectable. The technologies developed to deal with these two problems are Adaptive Optics (AO) and Coronagraphy, respectively. The next two sections briefly introduce both subjects, however an exhaustive treatment of both topics is beyond the scope of this thesis.

The main focus of my work presented in the Chapters 4 and 5 is the step after these technological means to improve the data have been used. That is to say, this work focuses on answering the question of how can we extract the faint planetary signal out of the imperfect and noisy data and residual stellar light using data modeling and post-processing techniques.

1.3.1 Adaptive Optics: correcting the influence of the atmosphere

The light emitted from distant objects arrives at the Earth as a plane wavefront. The phase and amplitude distortions to this wavefront introduced as light passes through the Earth's atmosphere, are called "wavefront errors" and their correction is of paramount importance for high-contrast imaging. Figure 1.6 shows real images taken with the SPHERE instrument. The left and middle panel show a short exposure and long exposure taken without AO

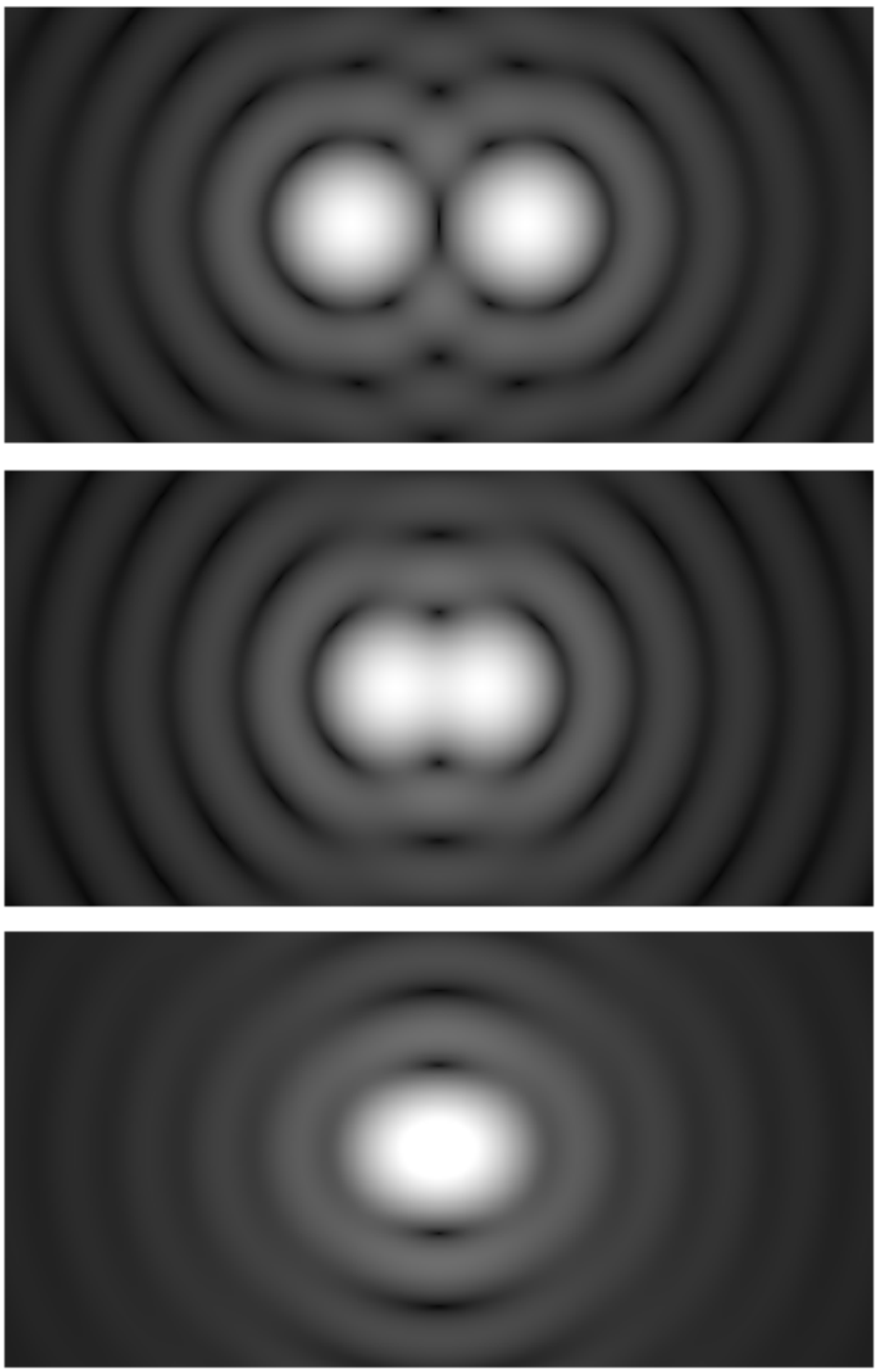


Figure 1.5: An Airy diffraction pattern generated by two equally bright point sources observed with a circular aperture, as is typical for most telescope mirrors. The two sources are shown clearly resolved (top), exactly meeting the Rayleigh criterion (middle) and below the Rayleigh criterion (bottom). Source: Wikimedia Commons, Public Domain

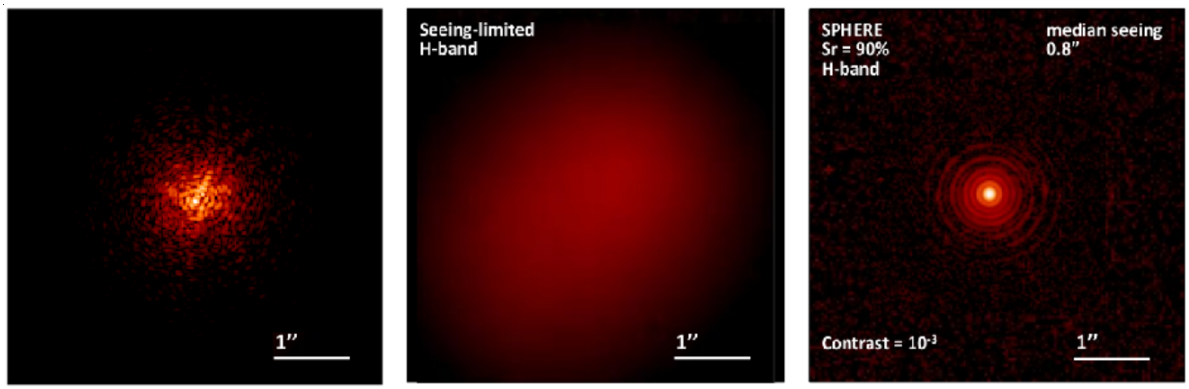


Figure 1.6: Example of images obtained of a point source with the VLT (circular $D = 8$m aperture) in H-band ($\lambda = 1.6\,\mu$m). Left: Short exposure without AO correction. Middle: Long exposure without AO correction. Right: Long exposure with SPHERE's SAXO extreme AO correction. Source: Cantalloube (2016)

correction, whereas the right image shows the PSF of the star after the wavefront has been corrected with the AO system. We can see that after removing the distortions, we can constrain 90% of the light in a *diffraction limited* PSF, i.e. achieve a Strehl ratio (SR) of 90%. Before the correction we are in the *seeing limited* regime.

Quantifying atmospheric turbulence

The effect of the atmosphere on the PSF is called "seeing". The seeing is caused by variations in the refractive index of the atmosphere, due to density and temperature variations. Distorted by the atmosphere, the incoming coherent light waves arrive at the ground with phase shifts and amplitude fluctuations.

There are three noteworthy parameters that are used to quantify the overall quality of observing conditions: the Fried parameter r_0, the isoplanatic angle θ_0, and the coherence time τ_0.

The *Fried parameter* (Fried, 1966) describes the average size of a turbulent cell in the atmosphere. According to Kolmogorov's turbulence theory (Kolmogorov, 1941), the energy distribution per turbulent eddy size follows a $-5/3$ slope, or in other words, more energy is in larger turbulent cells than in smaller ones. For the Fried parameter this means:

$$r_0[m] = 0.185\,\lambda^{6/5}\cos^{3/5} z \left(\int_0^\infty C_n^2(h)\,\mathrm{d}h\right)^{-3/5}, \tag{1.6}$$

where λ is the wavelength in meters, z is the zenith distance, h is the altitude in the atmosphere, and C_n^2 is the turbulence structure constant as a function of altitude. The $\lambda^{6/5}$ dependence of the Fried parameter means that turbulent cells will be larger for longer wavelengths, which reduces the influence of the atmosphere on observations. The closer the observed target is to the zenith the better the seeing as the path length through the atmosphere is reduced (i.e. lower airmass). A typical value for a good observing site is

approximately $r_0 = 75$ cm in the K-band.

A comparison of the diffraction limit of a telescope to the Fried parameter allows the estimation of when the seeing becomes dominant. For $D > r_0$ the seeing effect results in a simple shift of the PSF, for $D > 5\,r_0$ the image is dominated by the interference pattern of the distorted wavefronts, which will result in a "seeing halo", as in the left and middle panel of Fig. 1.6.

The *isoplanatic angle* corresponds to the size of a patch of the sky through which parallel light rays share the same characteristics. It is described by

$$\theta_0 = 0.31 \cos z \frac{r_0}{h_0} \propto \lambda^{6/5}, \tag{1.7}$$

where h_0 is the average altitude of the turbulence layer (about 5 km). It can be on the order of tens of arc-seconds for infrared observations ($10'' - 30''$ in K-band).

As most newer direct imaging instruments tend to have a smaller field-of-view (FoV) and operate in the near-infrared, the FoV does not exceed the isoplanatic angle. However, instruments with a larger FoV like NACO or IRDIS, experience some distortion at the edge of the FoV where the AO correction is not perfect. This effect is recognizable as a PSF distortion in radial direction, aligned with the bright central star, and can bias astrometry.

The *coherence time* also derives from the Fried parameter

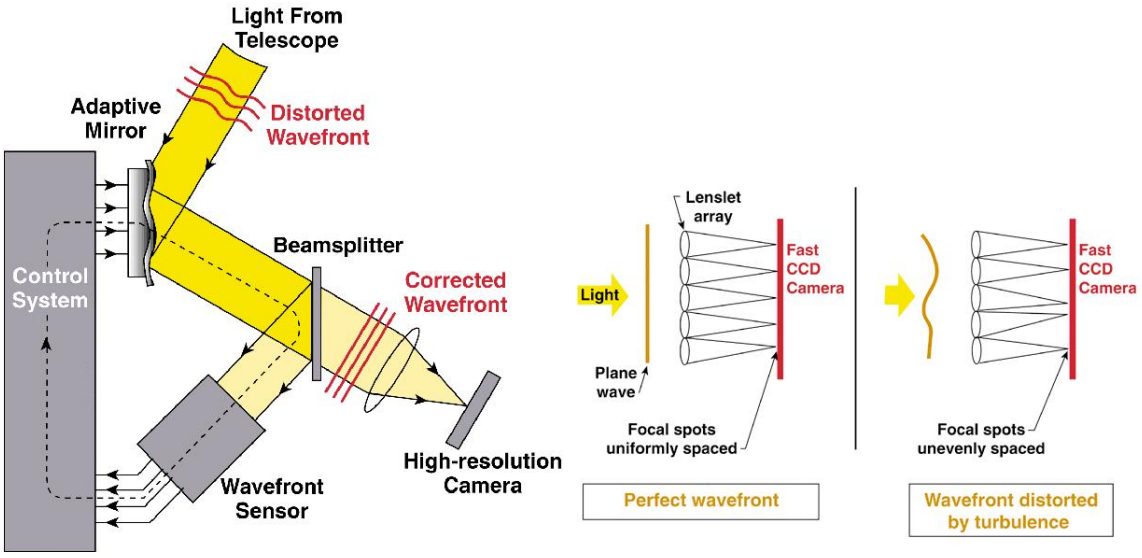
$$\tau_0 = 0.31 \frac{r_0}{v_0}, \tag{1.8}$$

where v_0 is the wind speed perpendicular to the line-of-sight. The coherence time at the Paranal observing site has a median τ_0(500nm) between 3 and 5 milliseconds. At K-band ($\lambda = 2.1\,\mu$m) this corresponds to 20–30 milliseconds. This quantity is extremely important, as it corresponds to the time it takes for one turbulence cell to be replaced by the next, and therefore, at what frequency the AO system has to perform to be able to correct the wavefront adequately. High wind speeds make it exceedingly difficult to obtain a stable feedback loop ("closed loop") for the adaptive optics system.

Principles of adaptive optics

An adaptive optics system consists of two main parts: a wavefront sensor (WFS) that measures the distortions of the incoming wavefront, and a deformable mirror (DM) that can be used to correct the wavefront errors. A schematic view of how an AO system works is shown in Fig. 1.7. A fraction of the light coming from the telescope is diverted by a beamsplitter into the wavefront sensor, which measures the deformation in the wavefront. These data are analysed by a real-time computer (RTC), which sends instructions to the deformable mirror, whose shape can be changed with small actuator-stamps attached to the backside, that is then used to correct the wavefront.

The basic principle of the Shack-Hartmann wavefront sensor (Shack, 1971, SH-WFS) is shown in the right panel. This is the most commonly used WFS concept. VLT/SPHERE, which has been used for the research contained in this thesis, and also the Gemini-GPI



(a) AO Loop **(b)** Shack-Hartmann Wavefront Sensor

Figure 1.7: Schematic view of how an AO system works. Courtesy: Claire Max, Center for Adaptive Optics, UCSC.

instrument, employ this WFS concept. The SH-WFS consists of a grid of small lenses (called a lenslet-array) which is used to determine the shape of the wavefront. If an undistorted plane wavefront enters the lenslet array, the resulting focal spots on the detector would be equally spaced. A distorted wavefront will result in relative shifts in the location of the focal spots. By measuring these shifts it is possible to calculate the wavefront errors and thus the shape of the deformable mirror necessary to compensate the distortion.

The pyramid wavefront sensor (Ragazzoni & Farinato, 1999, Py-WFS) is the second most commonly used WFS concept, and is installed at the LBT/FLAO, GMT/MagAO, and Subaru/SCExAO systems. It uses a pyramid element whose top is placed at the focal point of the beam, which acts as a two dimensional Foucault knife-edge test. The wavefront error can be reconstructed from the four resulting pupil images. This concept has the advantages of lower "aliasing" in the residual phase error budget, and requiring less photons to work. This is desirable as it increases the number of objects that can be observed without laser guide stars. Another advantage are the lower tip-tilt residuals, which typical means tha a separate tip-tilt sensor (as is also used in SAXO) is not necessary.

Future instruments like ELT/METIS will use a Py-WFS (Feldt et al., 2016).

Strehl ratio A commonly used quantity to encapsulate the performance of an AO system is the *Strehl ratio*. It is the ratio of the measured aberrated peak intensity of an imaged point source, compared to an ideal diffraction limited system. Therefore, a perfect optical system would obtain a Strehl ratio of 100% and correspond to the theoretical performance limit. The Strehl ratio for a non-adaptive ground-based observatory will typically not exceed 1% for long exposures in the optical, whereas basic AO systems increase this value to between 20% and 60% in the near-infrared (Rousset et al., 2003).

Another parameter that is important for quantifying AO systems is the outer working

angle (OWA) of the adaptive optics correction region. It is defined as

$$\mathrm{OWA} \propto \frac{\lambda}{D}\frac{\sqrt{n_{\mathrm{actuators}}}}{2}, \tag{1.9}$$

where λ is the wavelength, D is the diameter of the aperture, and $n_{\mathrm{actuators}}$. The number of spatial frequency modes that can be corrected is limited by the number of actuators. For the SPHERE SAXO adaptive optics system (Fusco et al., 2014) ($D = 8.2$ m, n$\sim$1400), the OWA is $\sim 20\lambda/D$ or ~ 0.60" (J-band), ~ 0.79" (H-band), and ~ 1.06" (K-band). This OWA can be clearly seen in images taken with SPHERE, as it corresponds to a ring of brighter speckles due to the decrease of AO performance at these separations (e.g., see Fig. 4.5 and 4.6).

The newest generation of *extreme adaptive optics* (XAO) systems are characterized by their fast real-time computer and deformable mirror with large actuator density. They can achieve Strehl ratios of $> 90\%$, as is the case in the right panel of Fig. 1.6 for the SPHERE instrument. A recent review of extreme adaptive optics can be found in Guyon (2018).

The limitations of AO systems

AO systems can never achieve a perfect correction. The imperfections in the wavefront that remain after correction are called *residual phase variance* σ^2_ϕ. The main contributors to σ^2_ϕ are: (Conan et al., 1994; Gendron & Lena, 1995):

$$\sigma^2_\phi = \sigma^2_{\mathrm{fitting}} + \sigma^2_{\mathrm{aliasing}} + \sigma^2_{\mathrm{servolag}} + \sigma^2_{\mathrm{noise}} + \sigma^2_{\mathrm{NCPA}} + \sigma^2_{\mathrm{calibrations}} + \sigma^2_{\mathrm{exogenous}} \tag{1.10}$$

- $\sigma^2_{\mathrm{fitting}}$: error due to not fitting all spatial frequencies (finite number of actuators).
- $\sigma^2_{\mathrm{aliasing}}$: error due to a finite number of sampling. elements in the WFS, preventing the sampling of high spatial frequencies. Pyramid wavefront sensors generally perform better in this term.
- $\sigma^2_{\mathrm{servolag}}$: error due to time delay between wavefront measurement and feedback to the deformable mirror. The AO sampling frequency for SPHERE is 1.2kHz.
- $\sigma^2_{\mathrm{noise}}$: error due to noisy measurements of the WFS (e.g. photon and read noise). This term is strongly influenced by the brightness of the star.
- σ^2_{NCPA}: error due to aberrations introduced after the WFS (see explanation below).
- $\sigma^2_{\mathrm{calibrations}}$: error due to mis-calibration of the system, e.g. imperfect translation of wavefront measurement to deformation command.
- $\sigma^2_{\mathrm{exogenous}}$: error due to the environment of the WFS (e.g. vibrations).

Any instrumental aberrations that occur after the beam-splitter that feeds into the WFS are by definition not corrected. These aberrations are called *non-common path aberrations* (NCPA). They have to be corrected by dedicated strategies, for SPHERE there are two such approaches: COFFEE (Sauvage et al., 2012; Paul et al., 2014) and ZELDA (N'Diaye et al., 2013, 2016; Vigan et al., 2018).

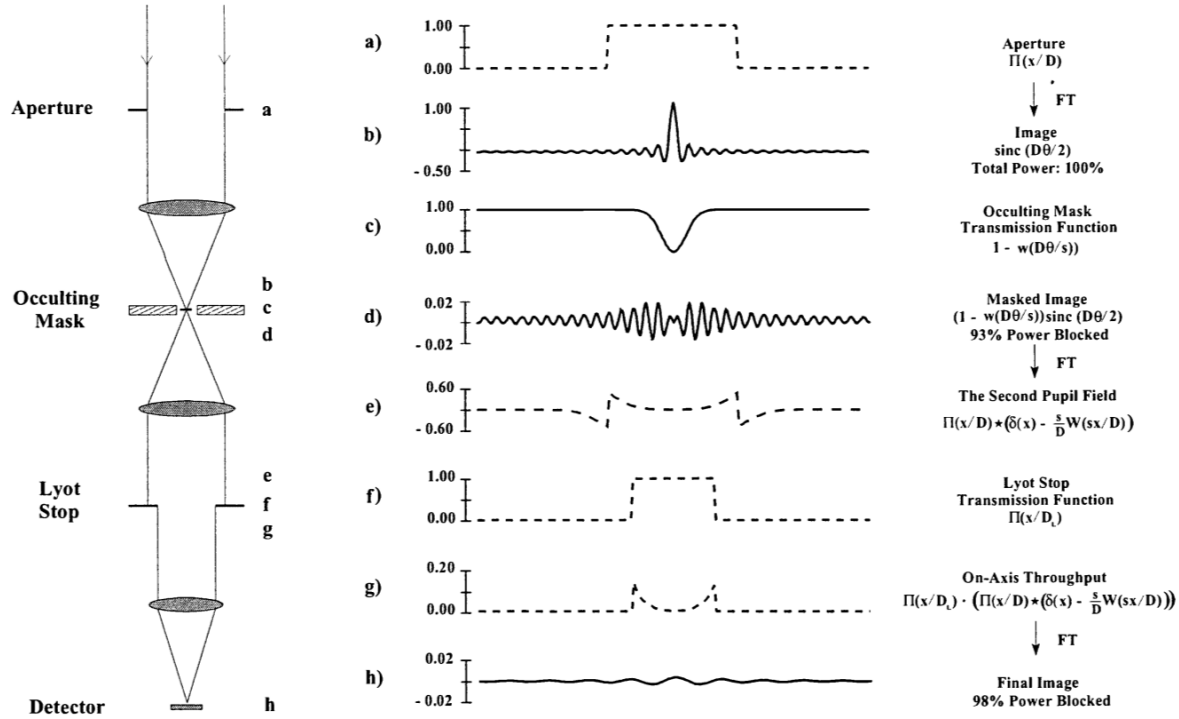


Figure 1.8: One-dimensional illustration of the working principle of a Lyot coronagraph, noting the location and profiles of optical elements and their influence. a) primary pupil for on-axis source, b) image before image plane stop, d) image after image plane stop, e) pupil before Lyot stop, f) Lyot stop, g) pupil after Lyot stop, and h) final on-axis image. The coronagraph of this example blocks 98% of on-axis light. Source: Lloyd et al. (2001)

1.3.2 Coronagrapy: suppressing on-axis stellar light

Once the effects of the atmosphere are corrected for and the PSF is close to diffraction limited and stable, the next objective is to block out the on-axis light coming from the bright star, while preserving flux coming from off-axis PSFs, such as sub-stellar companions or disks. To achieve this we need to remove, diffracted starlight from the region of interest, and the stellar PSF core to prevent detector saturation. For this purpose *coronagraphs* have been developed. The name stems from the first usage of such a device by Lyot in 1939 to observe the Sun's corona. Figure 1.8 explains the working principle of this kind of coronagraph.

For high-contrast imaging a coronagraph has to fulfill a number of constraints, that can be quantified by five main metrics:

1. **Raw Contrast:** The ratio of the unobscured stellar PSF peak to the starlight residual at the location of interest.

2. **Throughput:** The preserved flux fraction of an off-axis source.

3. **Inner Working Angle (IWA):** Angular separation at which the coronagraph throughput drops below 50%.

4. **Bandwidth:** The spectral range over which the coronagraph achieves it intended performance.

5. **Sensitivity:** Degradation of performance in the presence of aberrations.

Depending on whether a coronagraph is located in the focal plane(s) or pupil plane(s) of the telescope, we distinguish between Focal Plane Mask (FPM) coronagraphs, and pupil plane masks.

Furthermore, we can differentiate between coronagraphs that either alter the light wave's amplitude or phase. There are a huge variety of different coronagraph designs, stemming from combinations of these methods. A detailed overview of coronagraphic designs can be found in (Guyon et al., 2006), and a recent update in (Martinache, 2019).

In Chapter 4 of this thesis I present research aimed at improving the performance of algorithms for detecting exoplanets at very small angular separations. The deployment and development of new coronagraphs with small IWA is therefore complemented by my development of an algorithm that performs well at such small separations.

1.4 Modeling systematic noise in high-contrast imaging

Astronomical observations of isolated objects are constrained by the *sensitivity* of the instrument and longer integration times t generally correspond to higher signal-to-noise (SNR), scaling with SNR $\propto \sqrt{t}$. Therefore, in order to detect faint sources, we can increase the exposure time accordingly.

However, this is not the case for faint objects located next to bright stars. The instrumental aberrations result in a systematic pattern of stellar light surrounding the star that does not simply average out with longer exposure times. This pattern is referred to as a *speckle pattern* or *speckle halo*. For ground-based observatories like VLT/SPHERE, even using XAO and coronagraphs, the "raw contrast[1]" obtainable at 0.2–1″ is typically on the order of 10^{-4}. Beside the turbulent speckles, there is another regime of quasi-static speckles at a level of several 10^{-5} that evolve on time-scales from minutes to hours.

In order to achieve the required contrasts of $< 10^{-5}$ to detect planets in high-contrast imaging, modeling and removing this systematic noise pattern is essential. Chapter 4 of my thesis introduces an approach to modeling the systematic noise in HCI data that circumvents a limiting problem most algorithms face at very small angular separations, and that prevents the creation of models that take into account temporal evolution on second to minute time scales.

1.4.1 Speckle Noise

The systematic noise can be thought of as "leakage" of star light, predominantly consisting of a large number of "speckles", that by their nature, are of the size of a resolution element and often brighter than a planetary signal. The theoretical framework describing speckles and speckle noise was first presented in Goodman (1975) in the context of laser interferometry, and has been adapted to images obtained with AO systems by Cagigal & Canales (2000). Later, it was further extended to imaging employing both AO and coronagraphiy by Aime & Soummer (2004). Marois et al. (2008a) showed that mathematically the probability density function (PDF) describing the speckle intensity follows a modified Rician (MR) distribution. For one position it can be written as:

$$p_{\mathrm{MR}}(I) = \frac{1}{I_s} \exp\left(-\frac{I + I_c}{I_s}\right) I_0\left(\frac{2\sqrt{I \times I_c}}{I_s}\right), \tag{1.11}$$

where I is the integrated PSF intensity. It can be separated into the two terms $I = I_c + I_s$, where I_c is the local time-averaged static PSF contribution and I_s is the mean intensity of the speckle noise. Lastly, I_0 is the zeroth-order modified Bessel function of the first kind.

1.4.2 Temporal evolution of speckle noise

The speckles vary temporally on different time scales, from a few milliseconds for non-corrected atmospheric perturbations, to several minutes, hours or even days for quasi-static speckles. The quasi-static speckles vary with temperature, telescope flexures (due to the

[1] stellar light intensity at a given separation in raw data

changing gravity vector), and the movement of optical elements such as the atmospheric dispersion corrector (ADC) or derotator.

As these speckles are correlated in time, they do not average out over the observing sequence like other sources of stochastic noise (Macintosh et al., 2005). They can also interfere with the static pattern to create pinned speckles (Bloemhof et al., 2001). We therefore have to understand over what time scale regimes speckles evolve in order to model them correctly, which is essential for enabling us to detect planets.

The first detailed analysis of the temporal evolution of the instrumental speckle patterns was performed by Hinkley et al. (2007). They found two regimes for speckle decorrelation: a short time scale τ_{short} on the order of seconds, and a long time scale on the order of minutes to hours τ_{long}. Figure 1.9 shows the speckle lifetimes determined by measuring auto-correlations between flux values over an observation sequence. The top panel gives a longer baseline (Hinkley et al., 2007) and is based on 8s exposures, whereas the bottom panel shows a zoom in of the shortest time scales as determined for SPHERE observations taken at high-candence (1.6 Hz) (Milli et al., 2016). A similar analysis has recently been performed for SCExAO (Goebel et al., 2018), coming to similar conclusions.

This short time scale speckle component mentioned here is not to be confused with the residual atmospheric perturbations on milli-second time scales. Milli et al. (2016) argue that these exponential decaying ($\tau \sim 3.5$s) speckles are caused by turbulence inside the instrument, as they are seen both for observations with the telescope's internal lamp as well as on-sky observations. The longer, quasi-static component on minute to hour time scales is mostly dominated by moving optical components like the derotator and ADC.

These time scales are important to keep in mind when constructing models for the systematic noise and play an important role in the context of my time-domain based planet detection algorithm presented in Chapter 4.

1.4.3 The evolution of image processing for exoplanet imaging

Most systematic noise models for high-contrast imaging are based on "differential imaging", exploiting one or more diversities between the stellar and planetary signal (see Section 1.2.2). The classical concept behind differential imaging is the following: two images that share the same systematic noise (speckle halo), but differ in a planet signal, can be subtracted to attenuate the systematic noise. One image can therefore be used as a *model PSF*, or more generally, a *systematic model* of other frames. The reconstruction of the systematics model for a specific data point $(\lambda, x, t, ...)$ *without including* the off-axis signal of a companion object is made possible by exploiting the existing diversity in the signals. For example, in two images the off-axis signal can be at a different position, or, is only visible at one of the two observed wavelengths.

This is a very general mathematical problem of having a data set that contains two signals that we want to separate from each other based on their behavior (e.g. being static or changing with time). Therefore, numerous approaches have been developed to achieve this goal. While some vary in their fundamental principles, many differ in incremental steps and implementation details that have changed over time. As the field has organically and rapidly evolved in the past 15 years, the naming of certain concepts has not always been consistent. I therefore highlight the most important steps in the evolution of fundamental concepts in

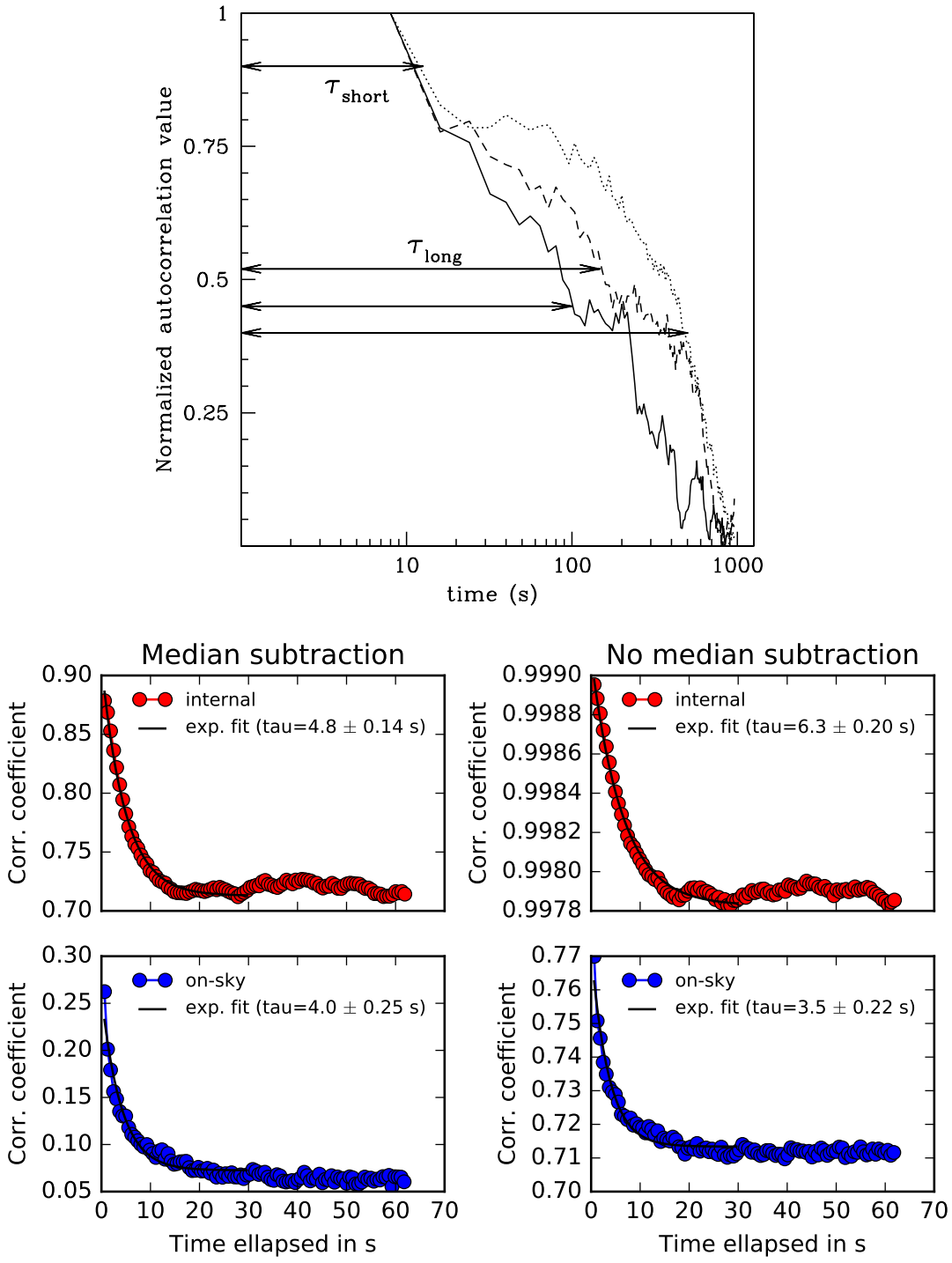


Figure 1.9: Top: Speckle evolution time scales as determined from an observation sequence comprised of 8s exposures and measured for three positions 0.19″ (solid line), 0.10″ (dashed line), 0.23″ (dotted line) (Hinkley et al., 2007, Fig. 4). Bottom: zoom in of short exponential decay speckle time scales as measured for SPHERE with an internal lamp and on-sky with and without median subtraction of static speckles (Milli et al., 2016, Fig. 11).

high-contrast image processing, focusing on mono-chromatic imaging data and methods that are based on the displacement of an off-axis signal in time and space. The same principles, sometimes in slightly modified forms, can be applied to spectral imaging and polarimetric imaging.

The major historical steps in the evolution of (ground-based) imaging techniques are:

- pupil tracking observations to produce a change in companion position over time (Marois et al., 2006)
- reference images for model PSF creation (Marois et al., 2006)
- optimized combination of reference images (Lafrenière et al., 2007)
- regularization to prevent over-fitting when optimizing (Pueyo et al., 2012)
- or using lesser dimensional representation of reference data to prevent over-fitting (Soummer et al., 2012; Amara & Quanz, 2012a)
- a stochastic, likelihood-based approach to forward model the companion signature (Cantalloube et al., 2015)

These developments build on each other and are not mutually exclusive. There are no dividing lines that prevent combining advances from these categories. For example, a likelihood-based approach that uses dimensionality reduction before an optimized combination of the data is possible. In the following sections, I discuss these developments in more detail.

Roll deconvolution The first image processing technique that is similar to most modern approaches is called the *roll deconvolution* (or *roll angle subtraction*) method for space-based observatories (Müller & Weigelt, 1985). It was used to improve the image quality of the Hubble Space Telescope (HST) observations with its initially flawed optics. Rotating an on-axis spacecraft along its axis keeps the telescope pupil fixed with respect to the detector. This means that the diffraction pattern does not change as the telescope moves to different *roll angles*, but the sky and off-axis sources move with respect to the diffraction pattern. Subtracting two such images from each other would therefore remove the systematic patterns and reveal fainter objects nearby.

The pupil tracking mode The logical extension of the roll angle subtraction approach was its application to ground-based observatories: using the *pupil tracking mode* (also called *pupil stabilized mode*) for high-contrast observations. This mode is often referred to as *angular differential imaging* (ADI, Marois et al., 2006), but I will refrain from using the two synonymously, because ADI is strongly associated with a specific *method* of using pupil tracking observations (see below). This mode of observation has proven especially useful, because it can be combined with other techniques.

While it is not practical to rotate a ground-based telescope around its optical axis, this is not necessary. Observations using alt-az mounted telescopes, as opposed to telescopes with an equatorial mounting, experience field rotation in their focal plane due to the rotation of the Earth while tracking a celestial source. The relative angle position between the observed

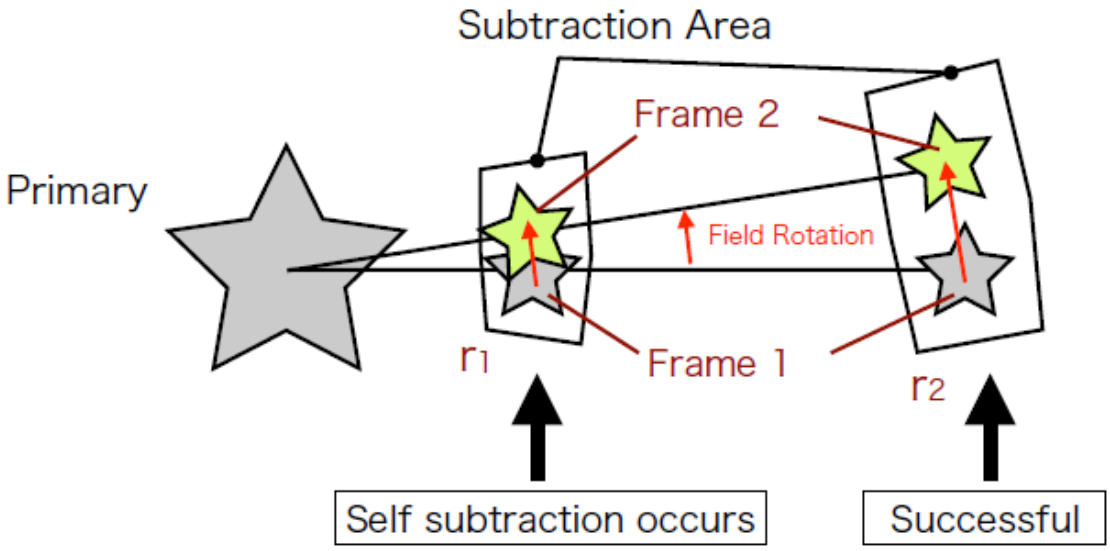


Figure 1.10: Illustration of self-subtraction of the planet signal due to insufficient field-of-view rotation in angular differential imaging and its dependence on angular separation from the primary star. Courtesy to Ryo Kandori

field's orientation with respect to a fixed direction on the detector is called the "parallactic angle", and it depends on the declination and hour angle of the observed object and on the latitude of the observatory. Objects whose *meridian passage* (highest point above the horizon) is closer to the local zenith experience more rapid FoV-rotation.

For most astronomical observations this rotation is compensated by an optical derotator or rotation of the instrument's focal plane, which is the so-called *field-stabilized* mode. For pupil tracking observations, in contrast, the main telescope pupil is stabilized in the instrumental focal plane, such that in each subsequent exposure the image orientation of the sky is slightly rotated, whereas the orientation of the systematic noise pattern does not change, as the speckle halo is mostly determined by the telescope optics[2], which remain static.

Classical ADI Marois et al. (2006) proposed an easy and effective way of building a systematic noise model using data obtained in pupil tracking mode. This method is now known as "classical ADI". The basic idea is that the speckle pattern is predominantly quasi-static, while an off-axis companion signal changes its position over time. This means that given sufficient rotation of the images over the course of the observations, simply taking the temporal median of the images leads to the small and changing off-axis signal of a companion being removed so that a "clean" systematics model can be obtained. Subtracting the systematics model from all frames, rotating each frame to the same North orientation, and combining them can then increase the achieved contrast by a factor of 100 or more compared to the raw contrast. However, this method only removes the most stable speckles and does not ef-

[2]the main mirrors M1 and M2 (including the telescope spiders), but also the deformable mirror itself. For SPHERE for example the fabrication pattern of the piezo actuators is imprinted in the speckle pattern at the correction radius.

ficiently account for frame-to-frame differences in the speckle field. There are some varieties of classical ADI implementations to alleviate some of these problems:

1. Smart ADI (s-ADI Marois et al., 2006) builds one systematics model for each frame, using only frames that are displaced in time enough that an off-axis source has moved at least $1\lambda/D$, with respect to the frame in question. This is a commonly used temporal exclusion criterion called *protection angle*, which is used for most contemporary algorithms. The effect of self-subtraction is shown in Fig. 1.10. The optimal value for the exclusion criterion depends strongly on the speckle correlation time scales explored in Chapter 1.4.2.

2. Radial ADI (r-ADI Marois et al., 2006) takes into account that the displacement on the detector per rotation angle depends on the angular separation from the host star, and therefore divides the image into annuli. Each annulus has a different protection angle, corresponding to a constant displacement value.

3. Image rotation and subtraction (IRS Ren et al., 2012) creates the systematics model from 180° rotated images, exploiting point symmetries in the speckle field.

Locally optimized combination of images (LOCI) The LOCI method introduced by Lafrenière et al. (2007) builds a local PSF model for each (sub)image based on a linear combination of *other* (sub)images in the observation sequence. The linear coefficient (weight) for each image is chosen such that the linear combination of all images (the model PSF) minimizes the residuals when it is subtracted. Usually the linear coefficients are computed in *optimization zones*, which are different from the *subtraction zone* (in size, location, or topology) in order to reduce *self-subtraction*, i.e. over-fitting caused by including actual signal in the systematics model. Generally a protection angle is used. After subtracting the optimized PSF model from each frame, the steps are the same as for classical ADI, derotating (North aligning) and combining the images. Again several variants of this basic idea exist:

1. Damped LOCI (d-LOCI Pueyo et al., 2012): given the problem of over-fitting the signal, the cost function is modified to include a regularization term (dampening). It also introduces a positivity constraint on the coefficients.

2. Adaptive LOCI (a-LOCI Currie et al., 2012): pre-selection of reference frames based on their correlation with a specific frame and subsequent optimization of the coefficients to maximize the SNR of point sources.

3. Template LOCI (t-LOCI Marois et al., 2014): build on d-LOCI and a-LOCI components, but includes an input spectrum and unsaturated, non-coronagraphic reference PSF for the planet model, to reduce self-subtraction in the optimization.

4. Matched LOCI (m-LOCI Wahhaj et al., 2015): inject synthetic signals to find linear combinations that maximize the SNR.

5. optimized IRS (OIRS Dou et al., 2015): the same method as the IRS approach (see above), but using optimized combination of images similar to LOCI.

Dimensionality reduction of reference data (PCA) The above described LOCI method faces the problem that the images used to build the systematics model share many common features. They do not constitute an orthogonal set of basis vectors and solving for the optimal coefficients is an ill-posed problem[3] (the problem of co-linearity). A commonly used technique to reduce co-linearity and transform the data into an orthogonal set of basis vectors is the use of *principle component decomposition* (PCA). The reference frames are decomposed into a set of principle component images (sometimes called eigenimages[4]). The resulting basis vectors are sorted by the amount of variance they explain, therefore truncating the series removes the least correlated component structures first. The quasi-static structures of the systematics can therefore be reconstructed in a lower-dimensional subspace, that is less likely to contain spurious signals and is mathematically well-behaved. This approach was proposed for high-contrast imaging by Amara & Quanz (2012a) and (Soummer et al., 2012) independently.

After decomposing the reference frames into principle components, the weights of each component have to be determined to minimize the distance between the data and the systematics model. This is the same problem as in LOCI (see Savransky, 2015).

There are a number of other image processing approaches related to dimensionality reduction techniques, such as *Low-rank plus sparse decomposition* (LLSG Gomez Gonzalez et al., 2016).

Stochastic, likelihood-based approach to companion signal forward modeling (ANDROMEDA) The approach that has been used by previous methods was to subtract the systematics model and subsequently measure the signal of interest in the noise-attenuated frames. However, this whole process did not explicitly take into account our knowledge about the signal. Cantalloube et al. (2015) proposed the maximum likelihood-based ANDROMEDA algorithm, in which the signature that a planet would leave in the data if it were present, is computed and explicitly fitted for. The algorithm is based on the pair-wise subtraction of images and comparison of the residuals with the expected characteristic signal of a planet (a positive-negative pattern) that would result from the subtraction of two displaced planet PSFs.

In Chapter 4 I introduce a new algorithm called TRAP (Temporal Reference Analysis for Exoplanets), that provides a novel paradigm: interpreting the data as a "set of time-series at different locations" instead of a "set of images at different times". This view opens up interesting new possibilities in terms of reference data selection and exclusion. It allows us to forego any kind of temporal exclusion criteria and therefore we can incorporate speckle systematics on all temporal timescales (see Section 1.4.2) into our systematics model, limited only by the length of the exposure time. The algorithm also uses PCA and a forward model of the off-axis planet signal.

[3]Note that one way to address ill-posed problems is by introducing a regularization to impose the desired behavior on the coefficients in the absence of a clear solution

[4]The name depends on which dimension of the variance of the data is decomposed into eigenvectors. In Chapter 4 we use eigenlightcurves. Eigenspectra are used often as well.

1.5 Directly imaged sub-stellar objects

While direct imaging can be used to detect binary stars and multiple stellar systems, what we are most interested in is the detection of *sub-stellar objects*. Stars are defined by hydrogen burning in their core, which at solar-metallicity starts at about $0.072\,M_\odot$, and constitutes the bottom of the main-sequence. Below this is the regime of sub-stellar objects. The more massive family of sub-stellar objects is called *brown dwarfs*, whereas the low mass end is called *planets*. The dividing line is often drawn at $0.013\,M_\odot$ or about $13\,M_{Jup}$, above which deuterium burning can occur (Spiegel et al., 2011); although whether this really constitutes a meaningful definition of what is a planet has been debated, as the deuterium-burning limit is model dependent and hard to experimentally determine in practice. Another approach is to distinguish objects by their formation mechanism, rather than a delimiting mass (e.g. Caballero, 2018). There is another regime in terms of nuclear burning that starts in the brown dwarf regime at about $0.06\,M_\odot$, the lithium burning limit (Chabrier et al., 2000).

Similar to normal stars, sub-stellar objects can also be classified based on their observed spectra. Figure 1.11 shows a color-magnitude diagram of most known directly imaged sub-stellar companions[5], as well as the sequence of M-stars and brown dwarfs. It should be noted that the H2-band shown is centered on the methane/water window, whereas the H3-band is in the methane/water absorption band. Sub-stellar objects follow a relatively well defined track in the color-magnitude diagram, which, similar to the Hertzsprung-Russel (HR) diagram, is related to the effective temperature T_{eff} of the objects. We can distinguish different spectral types. Below the M-type is the L-type regime in which, as the absolute brightness decreases, the objects continue to become redder, until relatively suddenly, objects of even lower brightness become bluer again. This transition is called the L–T transition and happens between $T_{\text{eff}} \sim 1200$ – 1400 K. After this transition begins the T-spectral type, which finally ends with the Y-type for objects below about 400 K.

Sub-stellar objects are not hot enough to burn hydrogen, and their supply of deuterium is limited, causing their temperature and radii, and therefore their spectral type, to evolve with age. These quantities are further dependent on the formation mechanism (e.g. Baraffe et al., 2002; Spiegel & Burrows, 2012; Mordasini, 2013; Mordasini et al., 2017). An example for such evolutionary models (assuming a *hot start*, i.e. retaining all of their entropy at formation), is shown in Fig. 1.12. Overplotted are known sub-stellar objects. The colored lines denote tracks of equal mass, where blue and green are planets and brown dwarfs respectively, according to the deuterium-burning criterion. The classification shown in this plot should therefore be taken with grain of salt, as a lot of uncertainty exists in both observation and modeling.

As mentioned, the radii of giant planets also evolve over time. However, at ages younger than several hundred million years they are not expected to be smaller than about $1.1\,R_{Jup}$ regardless of the formation mechanism.

[5]For which data in the H2- and H3-bands exists.

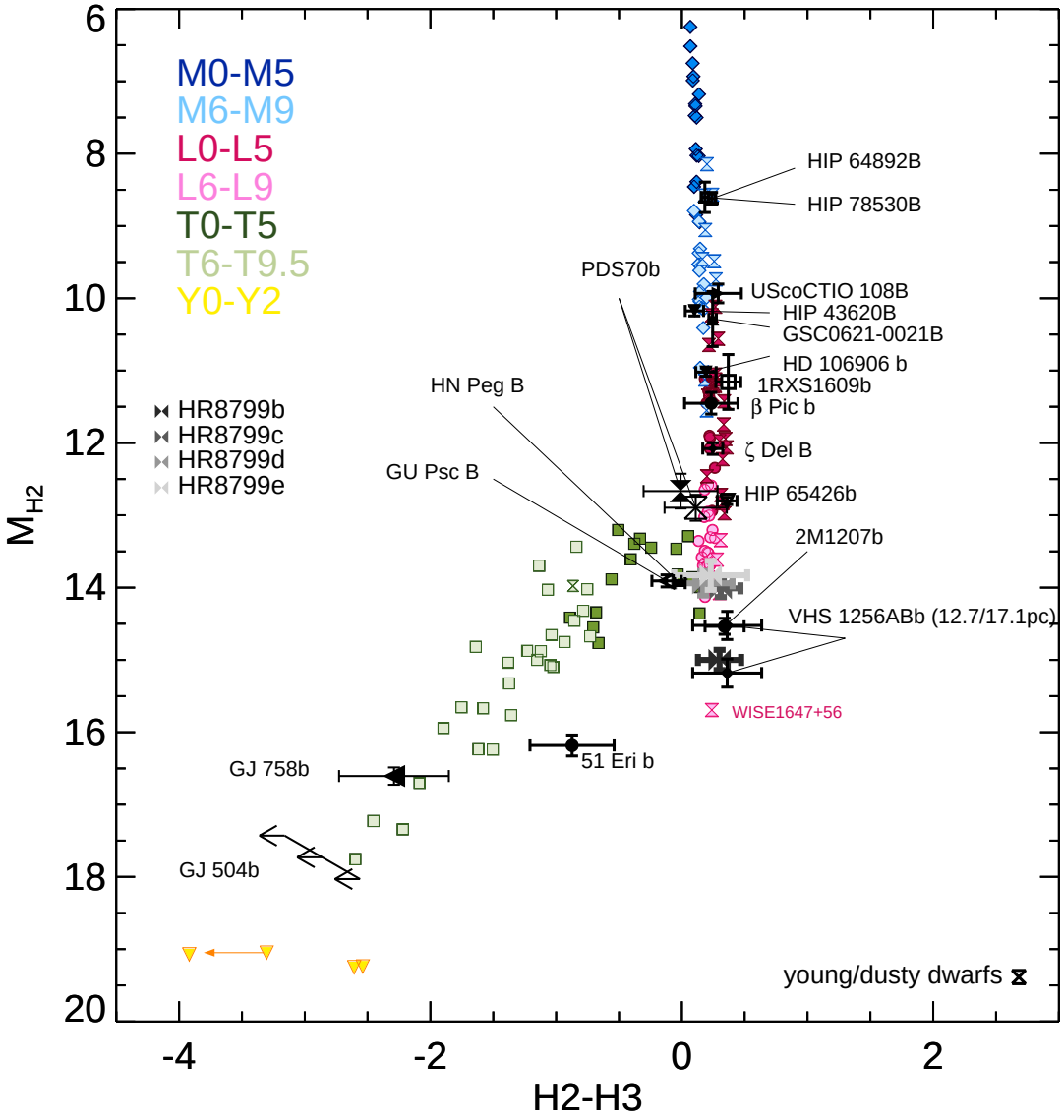


Figure 1.11: Color-magnitude diagram (CMD) containing most known sub-stellar companions. A detailed description can be found in Chapter 2.4.1. Updated with respect to newer detections, courtesy of Mickaël Bonnefoy.

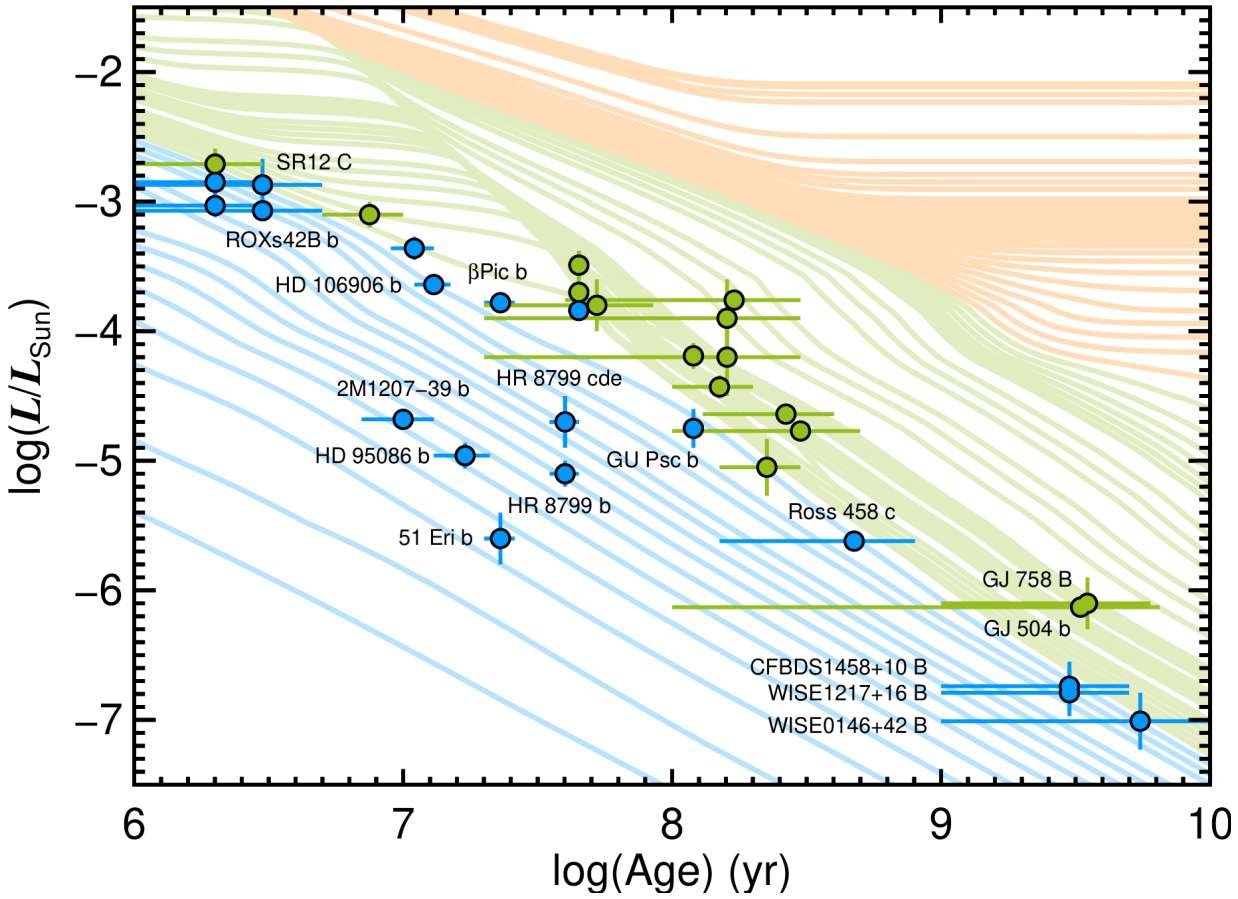


Figure 1.12: Hot-start evolutionary models of bolometric luminosity over age (Burrows et al., 1997). Plotted are different mass regimes: orange (stars), green (brown dwarfs), and blue (planetary) tracks denote masses of $> 80\,M_{\rm Jup}$, 16 – 80 $M_{\rm Jup}$, and $< 14\,M_{\rm Jup}$, respectively. Overplotted are known sub-stellar companions. Source: Bowler (2016)

1.5.1 L–T spectral types

As brown dwarfs and planets age they radiate and cool down. At early times, when the object is hot, refractory condensates such as iron and silicates, exist in the atmosphere and form clouds in the upper layers of the atmosphere. As the object cools these clouds become thicker and more opaque, resulting in redder near-infrared colors. Cooling down, the cloud layer moves to progressively higher pressures, deeper into the atmosphere, and eventually drops below the visible photosphere. As this happens, grain settling continues to clear the visible atmosphere above the cloud deck removing opacities and further cooling down the upper atmosphere. At these cooler temperatures CH_4 starts dominating over CO and the removal of cloud opacities allows flux to escape from deeper layers in the water window, leading to a brightening of the J-band at the L–T transition at around $T_{\rm eff} \sim 1200 - 1400$ K (Marley et al., 2012). The L–T transition exemplifies the importance of cloud physics and cloud models in the study of sub-stellar objects. The processes described above can occur for both brown dwarfs and planets, depending on temperature. However, as planets are less massive than brown dwarfs, their surface gravity is lower and this can lead to noticeable differences. Low surface gravities are for example associated with redder colors (Liu et al.,

2016).

T-type objects, like the planet 51 Eridani b studied in detail in Chapter 2, continue to be dominated by methane and water, and the importance of water (and water clouds) increases at lower temperatures, especially for the Y-type regime.

1.6 What we can learn about atmospheres from data: self-consistent models vs. free retrieval

Once we have obtained observations of the spectrum of an object, we want to know what we can learn about the properties of its atmosphere. At zeroth order, we can identify known molecular features in the spectrum, but usually our questions are more detailed. We want to quantify the physical properties of a planet's atmosphere, like its temperature and radius. We may also want to ask more complicated questions about the abundance of certain molecules and the uncertainty associated with this value.

In order to answer such questions we have to compare the spectral information we have obtained on planets with theoretical models. This involves using statistical inference and model comparison to retrieve the models and model parameters which provide the best explanation for the observed data. This process is called *atmospheric retrieval* and is generally done using Bayesian statistics (as in Chapter 2). There are two main approaches to do this.

One option is to pre-compute an extensive grid of *self-consistent atmospheric models* based on our physical understanding of the atmospheres of planets and compare the data to these models. This approach has been followed in Chapters 2 and3 of this thesis. The advantage of this approach is that the physics is self-consistent and reflects our best current understanding. The disadvantage is that it is computationally expensive, and assumes that the model is a complete and accurate representation of the observed object. If the models significantly deviate from the physical reality of the observed objects, so will our conclusions.

The other option is to attempt a so-called *free retrieval* using parametrized models. In this approach each model evaluated is computed "on the fly" inside the statistical sampler (e.g., MCMC, nested-sampling), therefore no pre-computed model grid is required. This is achieved by sacrificing some elements of self-consistency in order to reduce computational time. The pressure-temperature structure is parameterized, and so are chemical abundances in the atmosphere. The parameters are allowed to vary as free-parameters (minus prior constraints). At each evaluation of a parameter combination a spectrum is computed for the given atmospheric structure and abundances using a radiative transfer code. The advantage of this approach is that it is unnecessary to compute a grid in advance, as the sampling will converge to the part of parameter space constrained by the data. If the data is good enough, it may further allow us to deduce physical properties of the atmosphere that more constraining self-consistent models could not explain and thereby contribute to improving our understanding. The disadvantage is that the lack of self-consistency can also lead to unphysical solutions, especially when the available data is of bad quality or contains residual systematic noise that is interpretated as a physical property of the atmosphere by the retrieval code. Parametrized retrievals have been used extensively in the literature for both transmission and emission spectra of planets (e.g. Line et al., 2012; Benneke & Seager, 2012; Line et al., 2015; Waldmann et al., 2015; Line et al., 2016; MacDonald & Madhusudhan,

2017).

In either approach, sampling the complete model parameter space, gives us the set of possible parameter combinations that result in spectra consistent with the observed spectrum.

1.7 Overview of directly imaged planets

The number of directly imaged companions that are in the planetary mass regime is still low. Table 1.1 gives an overview of all planetary mass[6] companions discovered to date.

Only 10 such planets at orbital separations below 100 au are known, and four of these are in the same system (HR 8799). If we include bound objects at larger separations this number increases to about 20 objects. Of these objects, 8 to 9 have a small enough angular separation to fit into the field-of-view of a typical integral field spectrograph assuming the star is at the center of the detector ($<1''$).

The majority of the directly imaged planetary mass companions are relatively hot L-type objects. The only clear planetary mass T-type companion on a close-in orbit is 51 Eridani b, which is studied in detail in Chapter 2. The ages of the discovered objects span from a few million years, such as PDS 70 b, which is found in a gap of its host star's disk (similar to the transition disks shown in the gallery in Fig. 1.2), to tens of million years for close-in planets. Planetary mass objects on wide-orbits are sometimes hundreds of million, or even billions of years old. This shows the bias in the direct imaging detection technique that works better for young objects. At small separations, it is much more difficult to detect planetary mass objects that are older.

In Chapter 3 I give an overview about contributions that I have made to characterization studies. Together with the object studied in Chapter 2, this includes two L-type objects of different age, and two T-type objects of different age, and gives a feeling for the diversity of objects.

[6] According to hot-start models.

Table 1.1: Overview of directly imaged planetary-mass objects, based on a compilation by Bowler (2016). Updated with new discoveries: PDS 70 b (Keppler et al., 2018; Müller et al., 2018), HIP 65426 b (Chauvin et al., 2017; Cheetham et al., 2019), and updated values for 51 Eri b (Samland et al., 2017).

Planetary-mass companion on close-in orbits (<100 au)									
Name	Mass ($M_{\rm J}$)	Luminosity ($\log(L_{\rm bol}/L_{\odot})$)	Age (Myr)	Sep. (")	Sep. (au)	NIR SpT	Orbital Motion?	Pri. Mult.	Pri. Mass ($M_{\odot}$)
51 Eri b	3±1	–5.4±0.07	23±3	0.45	13	T4.5–T6	Yes	S	1.75
HIP 65426 b	8–11	–4.05±0.03	14±4	0.83	>80	mid L	Yes	S	1.96
PDS 70 b	5–9	–3.86±0.2	5.4±1	0.2	22	mid L	Yes	S	0.76
HD 95086 b	5±2	–4.96±0.10	17±4	0.6	56	L/T:	Yes	S	1.6
HR8799b	5±1	–5.1±0.1	40±5	1.7	68	∼L/Tpec	Yes	S	1.5
(LkCa 15 b)	6±4	...	2±1	0.08	20	...	Yes	S	1
HR8799c	7±2	–4.7±0.1	40±5	0.95	38	∼L/Tpec	Yes	S	1.5
HR8799d	7±2	–4.7±0.2	40±5	0.62	24	∼L7pec	Yes	S	1.5
HR8799e	7±2	–4.7±0.2	40±5	0.38	14	∼L7pec	Yes	S	1.5
β Pic b	12.7±0.3	–3.78±0.03	23±3	0.4	9	L1	Yes	S	1.6
Planetary-mass Companions on Wide Orbits (>100 au)									
WD0806–661 b	7.5±1.5	...	2000±500	130	2500	Y?	No	S	2
Ross 458c	9±3	–5.62±0.03	150–800	102	1190	T8.5pec	No	B	0.6, 0.09
ROXs42B b	10±4	–3.07±0.07	3±2	1.2	140	L1	Yes	B	0.89, 0.36
HD106906 b	11±2	–3.64±0.08	13±2	7.1	650	L2.5	No	B	1.5
GU Psc b	11±2	–4.75±0.15	120±10	42	2000	T3.5	No	S	0.3
CHXR 73 b	13±6	–2.85±0.14	2±1	1.3	210	M9.5	No	S	0.3
SR12 C	13±2	–2.87±0.20	3±2	8.7	1100	M9.0	No	B	1.0, 0.5
TYC9486–927–1 b	12–15	...	10–45	217	6900	L3	No	S	0.4

1.8 Purpose and outlook of this thesis

The purpose and goal of my thesis is improving the direct detection and characterization of exoplanets, their atmospheres, and all the steps entailed in this endeavor, from the *detailed study of atmospheric properties* using the spectrum of the planet and statistical tools, to the *development of data analysis techniques for detecting planets*. The direct detection of exoplanets is very challenging, and often operates at the very limit of what we can measure. Because of this it is necessary to carefully examine both, the data, and the models, to draw valid conclusions. Therefore, my thesis can be roughly divided into two parts.

In the first part I describe my work on the *atmospheric characterization* of exoplanets using their spectra and statistical tools that I have developed to infer their properties using physical models. I derive and discuss the physical properties for a range of directly imaged planets, from cold, methane-rich objects like 51 Eridani b and GJ 504 b, to hot and young objects, like the newly formed planet PDS 70 b, which is still interacting with its host star's transition disk, and the even hotter and slightly older object HIP 65426 b.

In the second part, I introduces a novel method to detect and characterize exoplanets. This method has the potential to significantly improve performance, especially at very small angular separations, which is the most interesting and challenging regime for direct imaging of planets. I further introduce a way to independently confirm and improve spectral results obtained with the SPHERE-IFS instrument. This second part of the thesis highlights both, the untapped potential available in the data, as well as ways forward to make ever more challenging detections possible.

In the Conclusion and Outlook of my thesis (Chapter 6), I discuss the main areas in which high-contrast imaging has the most potential for improvement and where development is necessary, and how I plan to address these questions. I further discuss how the different aspects, i.e. the atmospheric modeling, and the modeling of systematics in the data, can be combined in one unified framework, instead of considering them as separate steps.
The detailed outline is as follows:

Part 1: Atmospheric characterization of directly imaged exoplanets
In Chapter 2, I present my work on one of the coldest methane-rich directly imaged planets, 51 Eridani b, published in Samland et al. (2017). This study encompassed the data reduction, an empirical comparison to other known objects, and atmospheric parameter inference using the code I wrote for this purpose, called BACON (Bayesian Atmospheric CharacterizatiON). The atmospheric modeling of the atmosphere constitutes the main part of the paper, whereas the data reduction is a snap-shot of the state-of-the-art and "good practice"-approaches at the time: using synthetic photometry with an accurate stellar model for flux calibration, the empirical measurement of spectral covariance from the extracted spectrum, and a homogeneous application of the same algorithm to all data sets. In Section 2.7, I discuss the literature on 51 Eridani b after the publication of Samland et al. (2017). Of particular interest is Rajan et al. (2017), which was published shortly after my work by an independent research group. I argue in detail why the results of their study, which at first seem contradictory to mine, do not challenge my conclusions.

In Chapter 3, I provide an overview of the atmospheric characterization work I have contributed to a number of published studies, based on the BACON-tool developed in Chap-

ter 2. The objects studied are: the first planet discovered inside the gap of a transition disk PDS 70 b (Müller et al., 2018), the first planet discovered with the SPHERE instrument HIP 65426 b (Cheetham et al., 2019), and a new analysis of GJ 504 b (Cheetham et al., 2019), which depending on its age could either be a planet or brown dwarf.

Part 2: Development of high-contrast imaging detection methods
In Chapter 4, I describe a novel method I developed for detecting planets in pupil-stabilized high-contrast imaging data that improves the achieved contrast especially at very small angular separations. I make use of a temporal, non-local systematics model, which does not require the temporal exclusion criterion, which usually ensures that the planet signal has been displaced far enough on the detector to prevent over-fitting and self-subtraction of the signal. With my implementation of this algorithm, called TRAP (Temporal Reference Analysis for Exoplanets) an improvement of up to a factor of six in contrast is possible at the closest separations ($< 3\lambda/D$).

In Chapter 5, I describe the current status of my adaptation of the CHARIS pipeline for extracting microspectra from raw data to the SPHERE-IFS instrument. With this pipeline we have a method of confirming the spectral properties of exoplanets imaged with the SPHERE-IFS, independent of the official ESO pipeline. I have used this code to extract the image cubes from raw IFS data of 51 Eridani b paper as used in Chapter 2, and used the same post-processing methodology. I qualitatively confirm the spectrum obtained in Chapter 2.

Lastly, I provide a summary of the results of my work in Chapter 6. I discuss the future perspective of my work and it can continue to contribute to the field of direct detection of exoplanets, both in terms of detecting new objects as well as better understanding them.

Chapter 2

"*In space there are countless constellations, suns and planets; we see only the suns because they give light; the planets remain invisible, for they are small and dark. There are also numberless earths circling around their suns.*" (Despre infinit univers si lumi)

by Giordano Bruno (1548–1600)

---

2 Spectral and atmospheric characterization of 51 Eridani b using VLT/SPHERE

This chapter was published in a refereed article "Spectral and atmospheric characterization of 51 Eridani b using VLT/SPHERE", Samland et al. (2017), for which I am the lead author and which has been adapted for this thesis.

Overview: In this chapter, I perform an in-depth analysis of the exoplanet 51 Eridani b based on data obtained with the extreme adaptive optics instrument VLT/SPHERE. 51 Eridani b is an exoplanet around a young (20 Myr) nearby (29 pc) F0-type star, which was recently discovered by direct imaging. It is one of the closest direct imaging planets in angular and physical separation ($\sim$0.5″, $\sim$13 au) and is well suited for spectroscopic analysis using integral field spectrographs. The aim of this analysis was to refine the atmospheric properties of the known giant planet, and to constrain the architecture of the system further by searching for additional companions. Both dual-band imaging with IRDIS and integral field spectra with IFS were used, extending the spectral coverage of the planet to the complete Y- to H-band range and providing additional photometry in the K12-bands (2.11, 2.25 µm). The object is compared to other known cool and peculiar dwarfs. The posterior probability distributions for parameters of cloudy and clear atmospheric models are explored using MCMC. I verified the methods used by determining atmospheric parameters for the two benchmark brown dwarfs Gl 570D and HD 3651B. We further used archival VLT-NACO (L') Sparse Aperture Masking data to probe the innermost region for additional companions. Lastly, I give a summary of studies published after my 51 Eridani b study, and show that my results remain unchallenged.

Results: We obtained the first spectrophotometric measurements in the Y and K bands for the planet and revise its J-band flux to values 40% fainter than previous measurements. Cloudy models with uniform cloud coverage provide a good match to the data. I derived the temperature, radius, surface gravity, metallicity, and cloud sedimentation parameter $f_{\rm sed}$. I find that the atmosphere is super-solar ([Fe/H] $= 1.0 \pm 0.1$ dex), and the low $f_{sed} = 1.26^{+0.36}_{-0.29}$ value is indicative of a vertically extended, optically thick cloud cover with small sized particles. The model radius and surface gravity estimates suggest higher planetary masses of $M_{\rm gravity} = 9.1^{+4.9}_{-3.3}\,{\rm M_J}$. The evolutionary model only provides a lower mass limit of $> 2\,{\rm M_J}$ (for pure hot-start). The cold-start model is ruled out by the luminosity of the planet. The SPHERE and NACO/SAM detection limits probe the 51 Eri system at solar system scales and exclude brown-dwarf companions more massive than 20 $\rm M_J$ beyond separations of $\sim$2.5 au and giant planets more massive than 2 $\rm M_J$ beyond 9 au.

2.1 Introduction

The number of extrasolar giant planets found with ground-based high-contrast imaging techniques is growing steadily (e.g., Marois et al., 2008b, 2010b; Lagrange et al., 2010; Rameau et al., 2013; Bailey et al., 2014; Wagner et al., 2016) and the advent of dedicated high-contrast imaging instruments such as SPHERE (Spectro-Polarimetric High-contrast Exoplanet REsearch; Beuzit et al., 2008) and GPI (Gemini Planet Imager; Macintosh et al., 2014) has made it possible to study and characterize these planets and substellar companions in detail with low- to mid-resolution spectrometry and/or narrowband photometry (e.g., Ingraham et al., 2014; Chilcote et al., 2015; Apai et al., 2016; Maire et al., 2016a; Vigan et al., 2016a; Zurlo et al., 2016). At the same time, modeling of giant planet and brown dwarf atmospheres has made important progress with the development of cloudy models for colder objects (e.g., Allard et al., 2012; Morley et al., 2012; Baudino et al., 2015).

51 Eridani b is the first discovered planet with the GPI instrument (Macintosh et al., 2015) and was characterized using J- and H-band spectra taken with GPI and Keck/NIRC2 photmetry in the L' band. This object occupies a unique place in parameter space as a young, low-mass ($M < 10\,M_{\rm J}$), methane-rich, cold (∼700 K), but seemingly cloudy planet. This peculiar object is located at an angular separation from its host star ($\rho \sim 0.5''$), which is suited for spectroscopic characterization within the small field of view (FoV) of integral field spectrographs (IFS), but far enough away to achieve good signal-to-noise ratio (S/N) despite its contrast. Given these characteristics, it will become a benchmark object for current and future instruments as well as for the calibration of atmospheric models.
Its host star is part of a multiple system together with an M-dwarf binary (Montet et al., 2015) and is located in the well-studied β Pictoris moving group (Zuckerman et al., 2001). The age estimates range from 12 to 23 million years (Myr) (e.g., Simon & Schaefer, 2011; Binks & Jeffries, 2014; Mamajek & Bell, 2014; Bell et al., 2015), and we follow the adopted age of the discovery paper as 20 ± 6 Myr for all components of the system. A recent dynamical mass estimate of the distant binary M-dwarf companion GJ 3305 predicts an older age of the GJ 3305 AB system of 37 ± 9 Myr. An astrometric follow-up paper by De Rosa et al. (2015) has confirmed that the planet is co-moving with 51 Eri. The tentative orbital solutions (semimajor axis $a = 14^{+7}_{-3}$ au, orbital period $T = 41^{+35}_{-13}$ years, inclination $i = 138^{+15}_{-13}$) suggest that the planet does not share the inclination of the distant M-dwarf companion GJ 3305 (Montet et al., 2015).
The host star also has an infrared excess that can be modeled by two components corresponding to a warm belt of debris at 5.5 AU and another colder one at 82 AU (Patel et al., 2014; Riviere-Marichalar et al., 2014). As such, the architecture of 51 Eri is reminiscent of our solar system and of other benchmark systems such as HR 8799 and HD 95086.
In this work we present new near-infrared (NIR) spectra and photometric data obtained with the SPHERE instrument at the Very Large Telescope (VLT) in Chile, as part of the consortium guaranteed-time exoplanet imaging survey SHINE (SpHere INfrared survey for Exoplanets; Chauvin et al. 2016, in prep). The SPHERE observations are described in Sect. 2.2 and the data reduction in Sect. 2.3. The spectrophotometric analysis of 51 Eri b is discussed in detail in Sect. 2.4.
Finally, we present sensitivity limits to additional closer-in companions in Sect. 2.5, extended

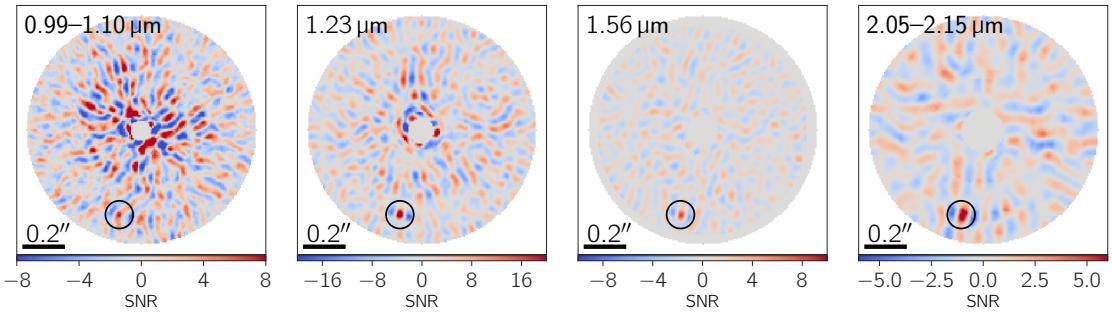


Figure 2.1: Shown are S/N maps created by ANDROMEDA for IFS and IRDIS K1. The maps are in order of ascending wavelength. The first two are extracted from the YJ-IFS data and the third from the YH-IFS data. The Y-band image (left) shows the median combined map between 0.99 – 1.10 µm as S/N is low, whereas the second and third image, which correspond to the peak in J and H, show single channels. The right panel shows the IRDIS K1 filter. Standard astronomical orientation is used, where up is north and left is east. The black circle indicates the position of the planet. The azimuthal negative wings around the planet signal is the characteristic planet signature that ANDROMEDA is fitting for in ADI data and not undesirable self-subtraction as in the PCA/LOCI approach (Cantalloube et al., 2015).

to the innermost region by archival Sparse Aperture Masking (SAM) data taken with the VLT-NACO instrument in the L' band, and end with our summary and conclusions in Sect. 2.6. The astrometric analysis of the planet is deferred to a future paper.

Table 2.1: Observing log

UT date	Instr. Mode	IRDIS Filter	IFS band	IRDIS DIT[a] (sec, #)	IFS DIT[a] (sec, #)	T_{exp}[b] (min)	Field Rot.[c] (deg)	Sr[d]
2015-09-25	IRDIFS_EXT	K12	YH	16×256	16×256	68.3	41.66	0.80 – 0.90
2015-09-26	IRDIFS	H	YJ	4×918	64×64	68.3	43.64	0.80 – 0.90
2015-12-26	IRDIFS	H23	YJ	16×256	32×128	68.3 (34.7[b])	37.19	0.75 – 0.85
2016-01-16	IRDIFS	H23	YJ	16×256	64×64	68.3	41.76	0.75 – 0.90

[a] Detector integration time [b] Exposure time after bad frame removal; about half the frames are unusable. [c] All observation were centered on the meridian passage of the target with an airmass between 1.08 and 1.10. [d] Strehl ratio measured at 1.6 µm as measured by AO system.

2.2 Observations

The SPHERE (Beuzit et al., 2008) instrument is an extreme adaptive optics system (SAXO; Fusco et al., 2014) that feeds three science instruments: the infrared dual-band imager and spectrograph (IRDIS; Dohlen et al., 2008), an integral field spectrograph (IFS; Claudi et al., 2008), and the visible light imaging polarimeter (ZIMPOL; Thalmann et al., 2008). We observed 51 Eri four times between September 2015 and January 2016 as part of the SPHERE GTO program using IRDIS in the dual-imaging mode (DBI; Vigan et al., 2010) and the IFS operating simultaneously (IRDIFS and IRDIFS_EXT modes, see Table 2.1). The observations were obtained with an apodized pupil Lyot coronagraph (Soummer, 2005; Boccaletti et al., 2008), consisting of a focal mask with a diameter of 185 milli-arcsec. The pupil stabilized mode was used close to meridian passage to exploit angular differential imaging (ADI) post-processing (Marois et al., 2006) with the goal of attenuating residual speckle noise. The usual SPHERE survey observation strategy was employed: 1) Photometric calibration: imaging of star offset from coronagraph mask to obtain PSF for relative photometric calibration at the beginning and end of the observation sequence; 2) Centering: imaging with the star behind the coronagraphic mask with four artificially induced satellite spots using the deformable mirror (Langlois et al., 2013) for deriving the star center location directly before and after the science sequence; 3) Coronagraphic sequence; 4) Sky background observation using same configuration as coronagraphic sequence. Finally, north angle offset and pixel scale were determined using astrometric calibrators as part of the SPHERE GTO survey for each run (Maire et al., 2016b). All the other calibration files (e.g., dark, flat, and spectral calibration) are obtained during the day following the observation via the instrument internal calibration hardware. Four IRDIS observations were conducted in three different filter setups: once in broadband H (BB_H), twice in dual-band H23 (H2 $\lambda_c = 1589\,\mathrm{nm}$, FWHM $= 53\,\mathrm{nm}$; H3 $\lambda_c = 1667\,\mathrm{nm}$, FWHM $= 56\,\mathrm{nm}$), and once in dual-band K12 (K1 $\lambda_c = 2103\,\mathrm{nm}$, FWHM $= 102\,\mathrm{nm}$; K2 $\lambda_c = 2255\,\mathrm{nm}$, FWHM $= 109\,\mathrm{nm}$, see also Table 2.2). The YJ setup (YJ: 0.95 – 1.35 μm, spectral resolution $R \sim 54$) was used three times and the YH mode (YH: 0.95 – 1.65 μm, spectral resolution $R \sim 33$) once. Observing conditions were variable for the two September data sets, but yielded the best data quality. Both December and January observations were conducted in bad seeing conditions with a strong jet stream that caused saturation near the edge of the coronagraphic mask when using the standard exposure times.

2.3 Data reduction and spectrophotometric extraction

Basic reduction of both the IRDIS and IFS data (background subtraction, flat fielding, bad pixel removal, centering, and spectral calibration for IFS) was performed using the pipeline of the SPHERE data center hosted at the Institut de Planétologie et d'Astrophysique de Grenoble (IPAG) using the SPHERE Data Reduction Handling (DRH) pipeline (version 15.0; Pavlov et al., 2008). The calibrated output consists of data cubes for each waveband, recentered onto a common origin using the satellite spot reference. The unsaturated stellar PSF frames taken before and after the coronagraphic sequence were reduced via the same

routines and also corrected for the neutral density filter transmission[7]. The variation of the stellar flux measurement of the host star ($\lesssim$5% for all used data) is propagated in the uncertainties of the companion flux measurement.

2.3.1 IFS data reduction and spectra extraction

In addition to the DRH pipeline, custom IDL routines (Mesa et al., 2015) have been used for the basic reduction following Maire et al. (2016a) with an additional step added to further refine the wavelength calibration using the shift in satellite spot position. The analysis of DBI and/or IFS data with aggressive spectral differential imaging (SDI; Racine et al., 1999b) algorithms, such as algorithms that include all other spectral channels as reference to model the speckle pattern, may lead to wavelength-dependent biases in the signal of a planet that cannot be modeled in a straightforward way (Maire et al., 2014). In order to avoid biasing the extracted spectrum while still retaining a good S/N, we opted for a non-aggressive method for the removal of the speckle noise in two steps. We first reduced the data via ADI post-processing and noted the channels in which, because of the peak in methane and water absorption, flux is neither expected nor observed. We then went back to the cosmetically reduced data cubes and used these selected channels as reference for a first classical SDI (cSDI) step, i.e., scaled to the same λ/D and mean flux for each channel, and we subtract these from all other channels. For the YJ spectrum, we used the channels between the Y and J band (1.11 – 1.17 μm) as reference. For the YH spectrum, we used the 1.14 μm channel to reduce all shorter wavelengths (Y band) and the 1.41 μm channel for the rest of the spectrum (J and H band). Because the YH spectrum spans a big wavelength range, we used two different reference channels, depending on the wavelength, to ensure that the effect of chromatic aberration on the speckle subtraction was minimized. These SDI preprocessed data cubes with attenuated speckle noise were then used as input for the following ADI reduction via various algorithms. We used the Specal pipeline (R. Galicher, priv. comm.) developed for the SHINE survey as a first-quick look reduction. For the spectral extraction we tested three different reduction approaches: PCA (Soummer et al., 2012; Amara & Quanz, 2012b, Specal implementation used), TLOCI (Marois et al., 2014, Specal implementation used), and ANDROMEDA (Cantalloube et al., 2015). We chose to focus our analysis on the spectra extracted with the ANDROMEDA algorithm, which was used for the first time on SPHERE/IFS data. This algorithm provides robust YJ and YH spectra and has a number of advantages compared to other reduction methods. In ANDROMEDA the signal is explicitly modeled, therefore no post-processing is necessary to extract an unbiased planetary signal, S/N, and detection limits, i.e., no self-subtraction correction by injection of artificial signals (see, e.g., Lagrange et al., 2010; Marois et al., 2010a) is needed. Furthermore, in contrast to other methods, ANDROMEDA has only one tunable parameter N_{smooth} (set to 8 pixels) and it only marginally affects the determined noise level at close separations (thus the S/N of a detection) and could affect the astrometry, but not the signal itself. We confirmed that 51 Eri b is located far enough from the center for none of this to be the case. Cantalloube et al. (2015) lists additional parameters, but these are either set directly by the wavelength of the observation or can be set to default owing to the much higher stability of SPHERE compared to NACO. As such, the ANDROMEDA reduction is very reproducible in the sense

[7] https://www.eso.org/sci/facilities/paranal/instruments/sphere/doc.html

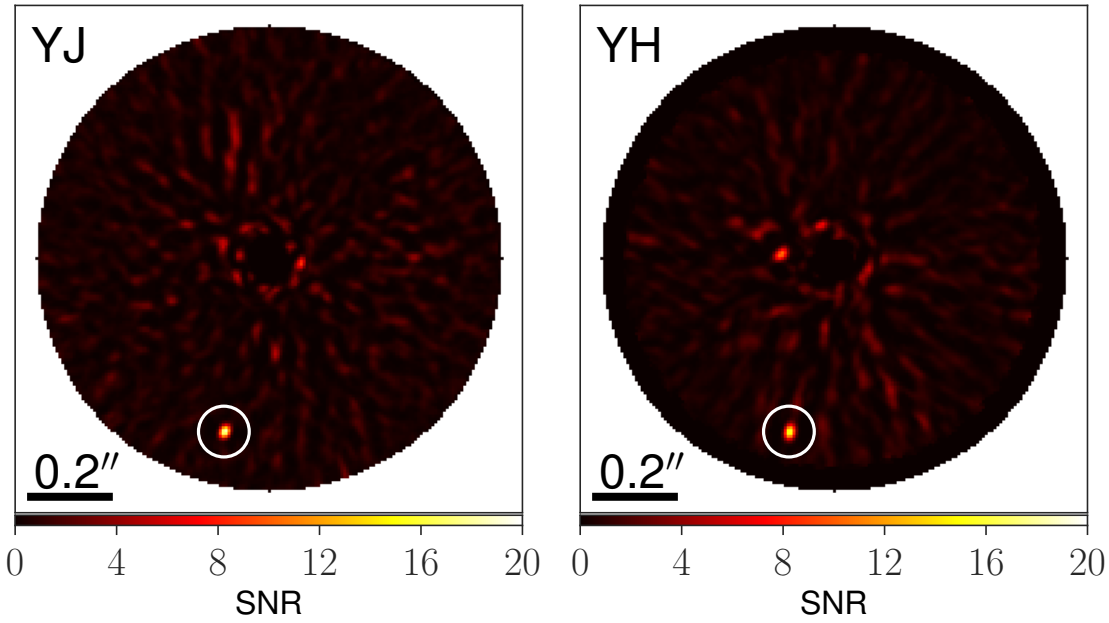


Figure 2.2: Shown are the detection images made using the method described in Thiébaut et al. (2016) with the S/N maps created by ANDROMEDA for IFS-YJ and YH data sets. The left image shows the collapsed image for the YJ band, the right image YH band. The circle white indicates the position of 51 Eri b.

that it is less prone to subjective choices of parameters that influence the data reduction. Figure 2.1 shows the planet in the ANDROMEDA reduction at four different wavelengths, corresponding to the Y, J, H, and K1 feature.
In addition to the reductions for every spectral channel, we also produced a collapsed "detection image" (see Fig. 2.2) to measure precisely the position of the planet at high S/N and to look for additional companions. For these images, instead of median combining all spectral channels, we follow the method introduced by Thiébaut et al. (2016). We first produce S/N maps for all spectral channels, which are a by-product of the ANDROMEDA algorithm, threshold them above zero, and sum up the squares of the thresholded S/N maps to make clean, collapsed images. In addition to getting the precise position for a planet, the advantage of this approach is that it is not necessary to make any assumptions about the exact spectral shape of a potential point source and it is suitable for visual inspection for further potential candidates.
The spectra and photometry extracted using ANDROMEDA and used for the atmospheric characterization are shown in Fig. 2.4 and discussed in Sect. 2.3.4. Reductions with alternative algorithms are shown and further discussed in Appendix 2.9.1.

2.3.2 Broad and dual-band imaging

In addition to the basic reduction the cubes are corrected for distortion and for the north angle offset determined from the astrometric calibrations (Maire et al., 2016b). For the unsaturated calibration frames of IRDIS, we used a custom routine that does not interpolate bad pixels in the PSF from the surrounding pixels, but replaces the respective bad pixels

with the value obtained by fitting a Moffat function to the PSF in all frames.
The cosmetically cleaned and centered data cubes were then used as input for further ADI/SDI post-processing pipelines. We again used three different approaches to reduce the data: PCA, MLOCI (Wahhaj et al., 2015), and ANDROMEDA. All three data reductions were consistent within their respective uncertainties. However, we chose to use ANDROMEDA for the final reduction of the IRDIS photometry presented here to be consistent with the IFS reduction.
Additionally we include the broadband L' photometric data point observed with the W. M. Keck Observatory Near Infrared Camera 2 (NIRC2; L' band, $\lambda_c = 3780,\mathrm{nm}$, FWHM = 700 nm) reported in Macintosh et al. (2015). The absolute magnitude $\mathrm{L}' = 13.82 \pm 0.27$ mag was converted to flux $f_{L'} = (1.82 \pm 0.45) \cdot 10^{-17}\,\mathrm{Wm^{-2}\mu m^{-1}}$ with the same distance as for the rest of the analysis ($29.4 \pm 0.3\,\mathrm{pc}$).

2.3.3 Conversion of the planet contrasts to physical fluxes

In order to convert the measured star to planet contrast in IFS and IRDIS data to physical fluxes we used a synthetic photometry approach. This can be summarized in three steps:

1. We built the spectral energy distribution (SED) of the star (see Fig. 2.3) from Tycho $\mathrm{B_T}$, $\mathrm{V_T}$ (Hoeg et al., 1997), Johnson filter U, V, B (Mermilliod, 2006), WISE W3 photometry (Cutri et al., 2013), and IRAS 12 μm photometry (Helou & Walker, 1988). The 2MASS J, H, $\mathrm{K_s}$ (Cutri et al., 2003) as well as W1-W2 photometry could not be used because of saturation of the central region of the star. The 2MASS $\mathrm{K_s}$ band is not flagged as saturated in the catalog, but can clearly be seen to be saturated in the individual 2MASS $\mathrm{K_s}$-band images. On the other hand, W4 had to be excluded owing to noticeable infrared excess.

2. We scaled a BT-NextGen model (Allard et al., 2012) with $T_{\mathrm{eff}} = 7200$ K, $\log \mathrm{g} = 4.0$ dex, and $[\mathrm{Fe/H}] = 0.0$ dex, to fit the above mentioned flux values via χ^2 minimization. The chosen model parameters are close to those determined from high-resolution spectra of the star ($T_{\mathrm{eff}} = 7256$ K, $\log \mathrm{g} = 4.13$ dex, and [Fe/H] = 0.0 dex; Prugniel et al., 2007)

3. We determined the mean stellar flux in the used SPHERE/IRDIS bandpasses and IFS bins (YJ: 0.95 – 1.35 μm, spectral resolution $R \sim 54$; YH: 0.95 – 1.65 μm, spectral resolution $R \sim 33$), by applying the spectral response curve of the instrument, i.e., the normalized wavelength dependent end-to-end transmission including optical elements (e.g., beam-splitters and coronagraph) and filters to the flux-calibrated synthetic spectra. We used the whole spectral response curve for the IRDIS bands. For IFS, because the spectral response is almost flat inside each respective spectral channel, we approximated the spectral response as a Gaussian of a width corresponding to the resolution of the spectrograph in the respective mode.

Our approach differs from that taken in Macintosh et al. (2015), in that we use a stellar atmosphere model for the flux calibration SED and not a blackbody spectrum. Comparing the two approaches over the NIR wavelength range of interest, we observe deviations due to spectral features on the order of $\sim 3\%$.

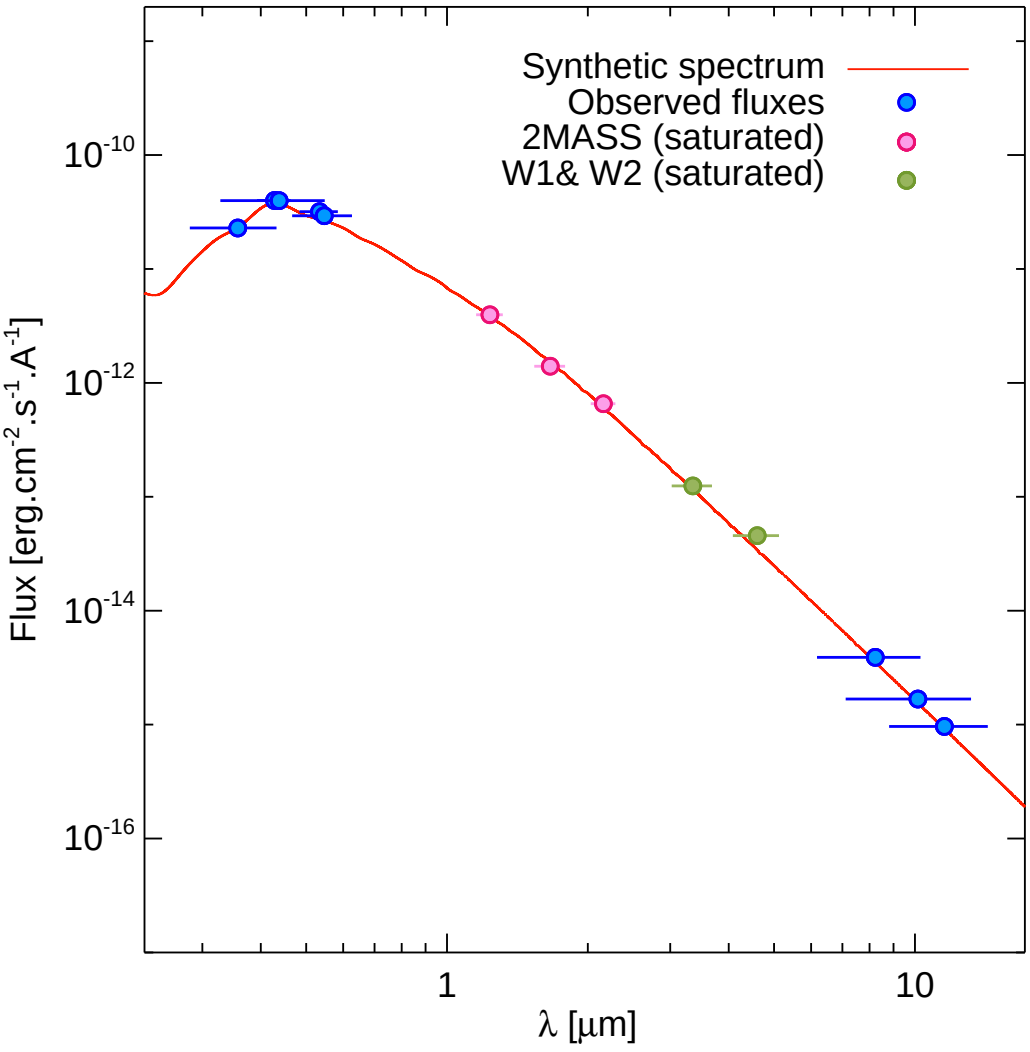


Figure 2.3: BT-NextGen synthetic spectrum of host star 51 Eri, scaled to match SED of optical and mid-infrared photometry. Owing to saturation of the star in these bands, the 2MASS J, H, and K_s, as well as W1-W2 were excluded from the fit.

Table 2.2: IRDIS photometry

Filter	λ (µm)	Width (µm)	Contrast	err. Contrast	App. Flux ($\mathrm{Wm^{-2}\mu m^{-1}}$)	Err. Flux ($\mathrm{Wm^{-2}\mu m^{-1}}$)	Abs. Magnitude[a]
ADI							
BB_H	1.626	0.291	$1.20 \cdot 10^{-6}$	$3.70 \cdot 10^{-7}$	$2.00 \cdot 10^{-17}$	$6.17 \cdot 10^{-18}$	17.11 ± 0.29
H2	1.589	0.048	$3.10 \cdot 10^{-6}$	$8.39 \cdot 10^{-7}$	$5.39 \cdot 10^{-17}$	$1.46 \cdot 10^{-17}$	16.07 ± 0.26
H3[b]	1.667	0.056	$4.25 \cdot 10^{-7}$	$3.36 \cdot 10^{-7}$	$6.42 \cdot 10^{-18}$	$5.09 \cdot 10^{-18}$	> 17.59
K1	2.103	0.102	$6.73 \cdot 10^{-6}$	$9.02 \cdot 10^{-7}$	$4.46 \cdot 10^{-17}$	$5.97 \cdot 10^{-18}$	15.21 ± 0.14
K2[b]	2.255	0.109	$2.08 \cdot 10^{-6}$	$1.64 \cdot 10^{-6}$	$1.10 \cdot 10^{-17}$	$9.84 \cdot 10^{-18}$	> 15.79
SDI+ADI							
H2 - H3	1.589	0.048	$2.32 \cdot 10^{-6}$	$3.99 \cdot 10^{-7}$	$4.04 \cdot 10^{-17}$	$6.94 \cdot 10^{-18}$	16.39 ± 0.17
K1 - K2	2.103	0.102	$5.90 \cdot 10^{-6}$	$5.04 \cdot 10^{-7}$	$3.91 \cdot 10^{-17}$	$2.67 \cdot 10^{-18}$	15.35 ± 0.08

Values under header ADI are obtained using ADI processing only. Values under SDI+ADI are obtained using SDI followed by ADI. Uncertainties are given as 1σ. The contrast uncertainties include speckle noise as dominant noise term and the variation in measured host star flux based the two unsaturated stellar images as a minor contribution. [(a)] With distance modulus $\mu = m - M = 2.34$ using Vega magnitude system. Distance uncertainty is negligible in magnitude measurement. [b] Forced photometry: magnitude 1σ upper limits obtained by adding the respective flux measurement and uncertainty.

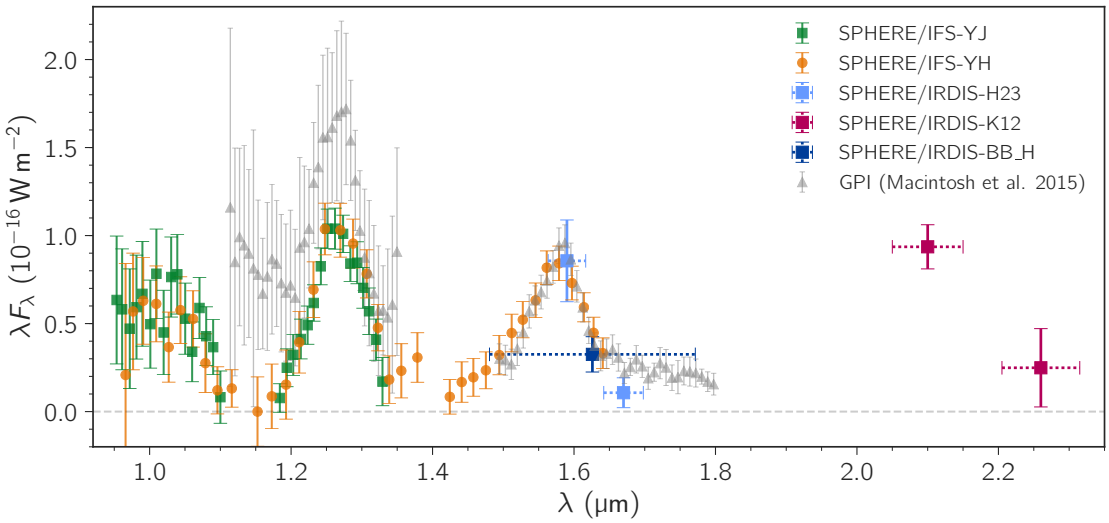


Figure 2.4: Spectral energy distribution for 51 Eri b constructed from our YJ-, YH-IFS spectra and IRDIS photometry in H23, BB_H, and K12. Channels used as reference for classical SDI were removed as they are biased. In addition to the SPHERE data we plot the two GPI spectra in the J and H band, respectively. The flux in the J band is consistent between the two SPHERE spectra, but significantly different compared to GPI. Uncertainties are given as 1σ and are assumed to be Gaussian.

2.3.4 Spectrum of 51 Eridani b

The SED of 51 Eri b showing all of our observations is presented in Fig. 2.4. Our IRDIS photometry is summarized in Table 2.2, where values are given for ADI and SDI plus ADI data reduction. For completeness we also plot the GPI spectra published in the discovery paper (Macintosh et al., 2015). With the SPHERE data, we extended the spectral coverage of the atmosphere to the Y band, provided the first photometry in the K band, and substantially improved the S/N in the J band. All of these are of paramount importance for deriving atmospheric parameters and cloud characteristics, as is discussed in Sect. 2.4.1 and 2.4.2. In further analysis we use both IFS spectra, the four ADI-only narrowband photometric data points in the H and K band. Additionally we also use the GPI-H spectrum, as it extends the wavelength coverage of the H band toward longer wavelengths, as well as the L' band photometry of Macintosh et al. (2015). We are not using the broadband H-band observation in our later analysis because it does not further constrain the spectral shape. The SPHERE YH spectrum is in excellent agreement in the overlapping part with the IRDIS H2 and BB_H photometry taken on different dates, as well as the previous high S/N H-band spectrum obtained by GPI. The only discrepancy we see is between the J-band flux reported in Macintosh et al. (2015) and our data. The question presenting itself is therefore the origin of the difference observed in the J band for which there are two possibilities: first, strong (∼40%) atmospheric variability in the atmosphere of the planet; or second, systematic offsets in the absolute calibration between the data sets. The J band is known to be more sensitive to temporal amplitude changes in the atmosphere of L/T-type objects than the H and K bands (e.g., Radigan et al., 2012; Biller et al., 2015), but even so, given that we see no significant difference in the H band, we think it is unlikely that 40% variability in

the J band is in agreement with our consistent values for the H-band flux. We therefore believe that the difference in the J band between our observations and previous observations is a result of systematics. The reduction of the YH spectrum via different algorithms shows consistent results (Fig. 2.17) giving us confidence in the overall reliability of the data, data reduction, and calibration.
While we use the ANDROMEDA method in this paper, (Macintosh et al., 2015) uses TLOCI. We also compared different reduction methods in Appendix 2.9.1 and notice a difference in J-band flux between our two data sets depending on the reduction method; for example, using ANDROMEDA, the flux measured in the two data sets is within 10%, which is consistent in their respective uncertainties, while for TLOCI this differs by $\sim 40\%$. It is possible that absolute calibration is more difficult for the other algorithms compared here because additional steps to account for algorithm throughput are necessary.

2.4 Spectrophotometric analysis

2.4.1 Empirical comparison to known objects

Our SPHERE YJ and YH spectra of 51 Eri b confirm the presence of several deep water and methane absorption bands typical of T-dwarfs from 1.1 to 1.2, 1.3 to 1.5, and longward of 1.6 μm. We compared the YH spectrum and K1 photometry of 51 Eri b to that of L- and T-type objects from the SpeXPrism library (Burgasser, 2014) completed with spectra from Mace et al. (2013) and Best et al. (2015). The comparison spectra were smoothed to the resolution of the IFS YH spectrum and their flux was integrated within the wavelength intervals covered by each channel of the IFS and the IRDIS K1 filter. We used the G goodness of fit indicator for the comparison (Cushing et al., 2008), with the implementation following Vigan et al. (2016b), accounting for the filter widths and the uncertainty on the 51 Eri b spectrophotometry.

The results are shown in Fig. 2.5. The best fits are obtained for late-L and early-T objects, in agreement with the placement of the planet in color-color and color-magnitude diagrams (see below). The best fitting object is PSO J207.7496+29.4240, a peculiar T0 object, and possibly an unresolved binary, from the Best et al. (2015) library. A visual inspection of the fit reveals that while the object is able to reproduce the overall JHK spectral slope, 51 Eri b has deeper methane plus water absorptions. The fit of the YH+K1 spectrophotometry is influenced by the strong overluminosity of the K1 band caused by the reduced collision-induced absorption (CIA) of H_2. The Y-band flux is also known to be modulated by the surface gravity and metallicity (Burgasser et al., 2006a; Liu et al., 2007). To mitigate this in the comparison to higher log g objects, we decided to rerun the fit on the YJ spectrum and on the part of the YH spectrum excluding the Y band (hereafter JH spectrum). The results are shown in Fig. 2.6.

The T7–T8 objects represent the best match to the planet YJ spectrum only. Among the sample of T5.5–T7 objects, the brown dwarfs SDSSpJ111010.01+011613.1, 2MASSIJ0243137-245329, 2MASSJ12373919+6526148, and 2MASSIJ1553022+153236 are minimizing G and therefore represent the best fits to the YJ spectrum. SDSSpJ111010.01+011613.1 and 2MASSIJ0243137-245329 belong to the growing class

Table 2.3: IFS photometry.

Filter	λ (μm)	Width (μm)	Contrast (10^{-6})	Abs. Magnitude[a]
J	1.245	0.240	1.03 ± 0.67	17.40 ± 0.71
J3	1.273	0.051	2.22 ± 0.53	16.52 ± 0.26
H2	1.593	0.052	2.70 ± 0.70	16.22 ± 0.28

Photometric magnitudes for IRDIS filters derived from IFS spectra. Uncertainties are given as 1σ. [(a)] With distance modulus $\mu = m - M = 2.34$ using Vega magnitude system. Distance uncertainty is negligible in magnitude measurement.

of red T dwarfs (Gagné et al., 2014; Stephens et al., 2009). SDSSpJ111010.01+011613.1 has been proposed as a member of the AB Doradus moving group (Gagné et al., 2015). The two other objects are a binary (unresolved in the SpeX slit; Burgasser et al., 2006c) and a magnetically active object with strong H_α emission, respectively; these display some variability in the J band (Burgasser et al., 2000; Artigau et al., 2003; Kao et al., 2016).

The JH spectrum is best represented by SDSSJ141530.05+572428.7, a T3 dwarf from the SpeXPrism libraries, which is again a candidate unresolved binary (Burgasser et al., 2010). Therefore, there appears to be a correlation between the spectral type of the best fit template found and the maximum wavelength of the photometric points included in the fit. We interpret this as, first, the consequence of the unusual red slope of the near-infrared SED of the planet compared to the templates and, second, the fit limited to the shortest wavelengths becomes more sensitive to the $CH_4 + H_2O$ absorption from 1.1 to 1.2 μm, which is characteristics of late-T dwarfs. To conclude, the planet SED is often reproduced by candidate unresolved binaries, which is a class of objects that was also found to provide a good fit to the HR8799b and c planets (Bonnefoy et al., 2016).

We took advantage of the SPHERE spectra to generate synthetic photometry for the narrowband filters of SPHERE overlapping with the wavelength range of the IFS spectra (assuming simple top-hat profile): J ($\lambda_c = 1245\,$nm, FWHM $= 240\,$nm), J3 ($\lambda_c = 1273,$nm, FWHM $= 51\,$nm), and H2 ($\lambda_c = 1593\,$nm, FWHM $= 52\,$nm). Photometric magnitudes in these bands enable a homogeneous comparison of the planet properties with those of known reference objects.

The photometry was obtained considering a flux calibrated spectrum of Vega (Bohlin, 2007) and ESO `Skycalc` web application[8] (Noll et al., 2012; Jones et al., 2013). We find J $= 19.74 \pm 0.71$ mag, J3 $= 18.86 \pm 0.26$ mag, and H2 $= 18.56 \pm 0.28$ mag (Table 2.3). We combined this synthetic photometry with that obtained in K1 (17.55 ± 0.14 mag) to show the position of the planet in color-color and color-magnitude diagrams (Fig. 2.7 and 2.8). The CMD are built with low-resolution spectra taken from the literature and published parallaxes. For the field dwarfs, we used the spectra from Leggett et al. (2000) and from the SpeXPrism library (Burgasser, 2014). The SpeXPrism spectra were calibrated in flux using the H-band 2MASS photometry of the targets. We used parallaxes from the literature (mostly from Monet et al., 1992; Faherty et al., 2012) and newly revised values from Liu et al. (2016) where applicable. We repeated this procedure for young, low-gravity and/or dusty M, L, and T dwarfs (spectra taken for the most part from Allers & Liu (2013) and

[8]http://www.eso.org/observing/etc/bin/gen/form?INS.MODE=swspectr+INS.NAME=SKYCALC

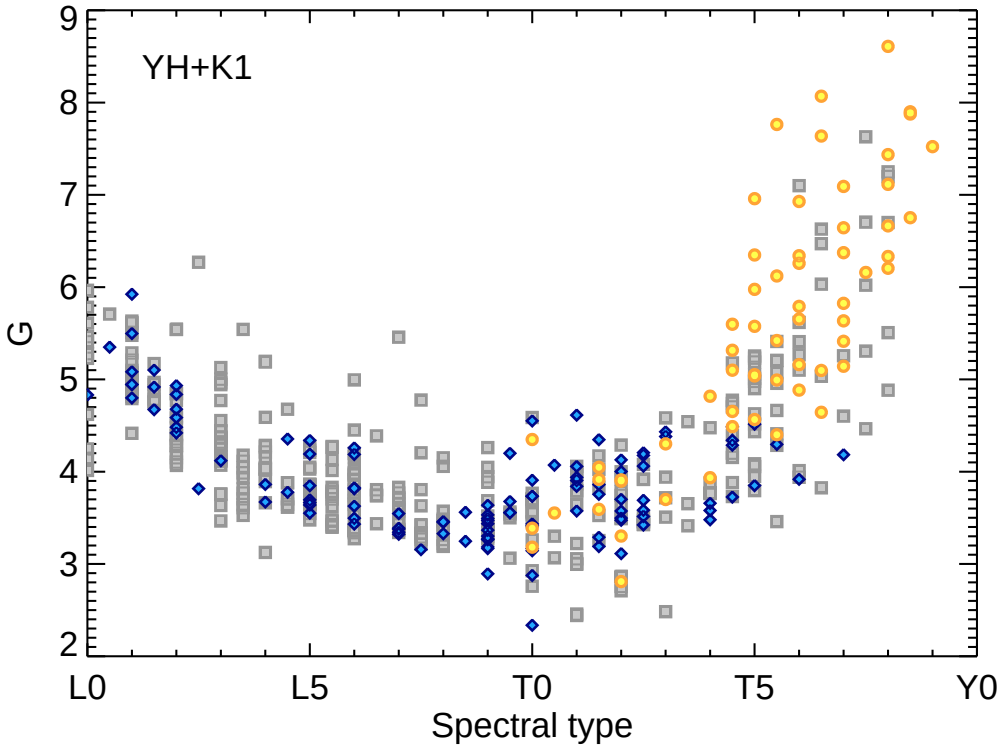


Figure 2.5: Goodness of fit G for the comparison of 51 Eri b YH+K1 spectrophotometry with that of template spectra of L and T dwarfs from the SpeXPrism (gray squares), Mace et al. (2013) yellow circles, and Best et al. (2015) (blue diamonds) libraries.

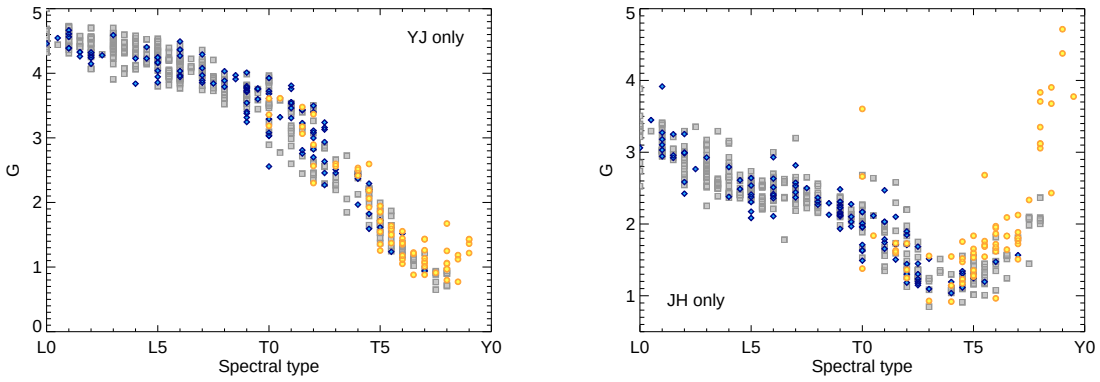


Figure 2.6: Same as Fig. 2.5 but considering the YJ spectrum (left) and JH spectrum (right) of 51 Eri b only.

parallaxes from Faherty et al. (2012) and Zapatero Osorio et al. (2014)). We added the known T-type companions (and the isolated object CFHTBD2149; Delorme et al., 2017a) with known distances and with some knowledge of their metallicity either from the primary star [Fe/H]* or from the companion spectrum [Fe/H]^c. Bonnefoy et al. 2017 (in prep., and ref. therein) provide a full description.

The planet has the luminosity of T6–T8 dwarfs but much redder colors that are consistent with those of late-L dwarfs in J3/J3-K1 and H2/H2-K1 color-magnitude diagrams (CMD). In these diagrams, the benchmark T6.5–T8.5 objects suspected to be metal rich and/or younger than the field (CFBDSIRJ214947.2-040308.9, GJ 758b, ROSS 458C, SDSSJ175805.46+463311.9/G 204-39B; Delorme et al., 2012; Vigan et al., 2016a; Burningham et al., 2011; Faherty et al., 2010) also have redder colors than the sequence of field dwarfs. Although they are not as red as those of 51 Eri b. In color-color diagrams (Figure 2.8), 51 Eri b falls at the location of the L/T transition objects in color-color diagrams, although the planet luminosity and the presence of a methane bands in its spectrum is inconsistent with this object being at the L/T transition. Instead, it suggests that the object has a color deviation that is similar to the color deviation seen for young and/or dusty late-L dwarfs (green stars in Fig. 2.8) with respect to regular late-L dwarfs. The peculiar T7 dwarf CFBDSIRJ214947.2-040308.9 is also deviating from the sequence of T dwarfs but to a lower extent. We interpret the deviation for 51 Eri b as a consequence of the reduced opacities caused by CIA of H_2 (Borysow, 1991), which occurs in low gravity and metal-enriched atmospheres and that primarily affects the K band (Allard et al., 2001). We cannot exclude that it could also be caused by a haze of submicron-sized particles as proposed for low-gravity L/T transition objects (see Marocco et al., 2014; Bonnefoy et al., 2016; Hiranaka et al., 2016) and consistent with our atmospheric modeling analysis in Sect 2.4.2.

We compare in Fig. 2.9 the spectrophotometry of 51 Eri b to those of extra T-type objects known to be younger than the field or dusty objects (Burgasser et al., 2011; Naud et al., 2014). No object could simultaneously reproduce the YH-band features and the K1 flux of 51 Eri b in agreement with the previous analysis. We also note a strong departure of the 0.95–1.05 μm flux of the planet whose origin is unclear, but further discussed in detail in Sect. 2.4.2. The depth of the 1.1–1.2 and 1.3–1.5 μm bands of 51 Eri b is only reproduced by those of objects later than T7.

In summary, the empirical approach: 1) confirms the peculiarity of 51 Eri b, 2) further suggests that the planet shares the properties of late-T dwarfs, 3) suggests that some of the properties of the planet are related to low-surface gravity and young age or super-solar metallicity, and 4) is limited by the lack of objects from clusters and young moving groups with spectral types later than T5. These findings are in good agreement with our atmospheric modeling as described in the next section.

2.4.2 Atmospheric modeling with petitCODE

In order to characterize 51 Eri b we carried out dedicated calculations with *petitCODE*, which is a self-consistent 1D radiative-convective equilibrium code, solving for the atmospheric temperature structure and abundances, assuming equilibrium chemistry. For every converged structure the *petitCODE* calculates an emission and transmission spectrum, where the latter is of no importance for studying 51 Eri b, given that the planet is not transiting. The first

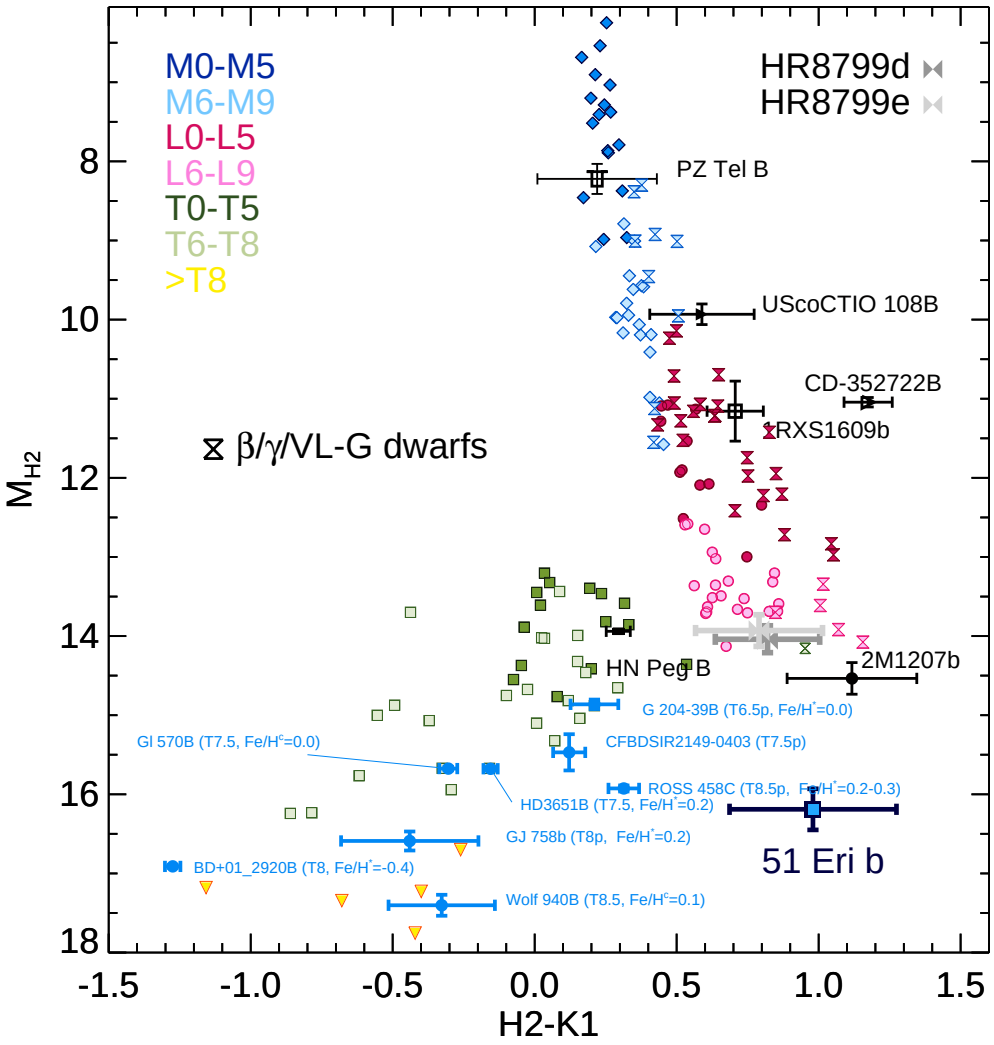


Figure 2.7: Placement of 51 Eri b in color-magnitude diagram. All magnitudes except for K1 are derived from the IFS spectra.

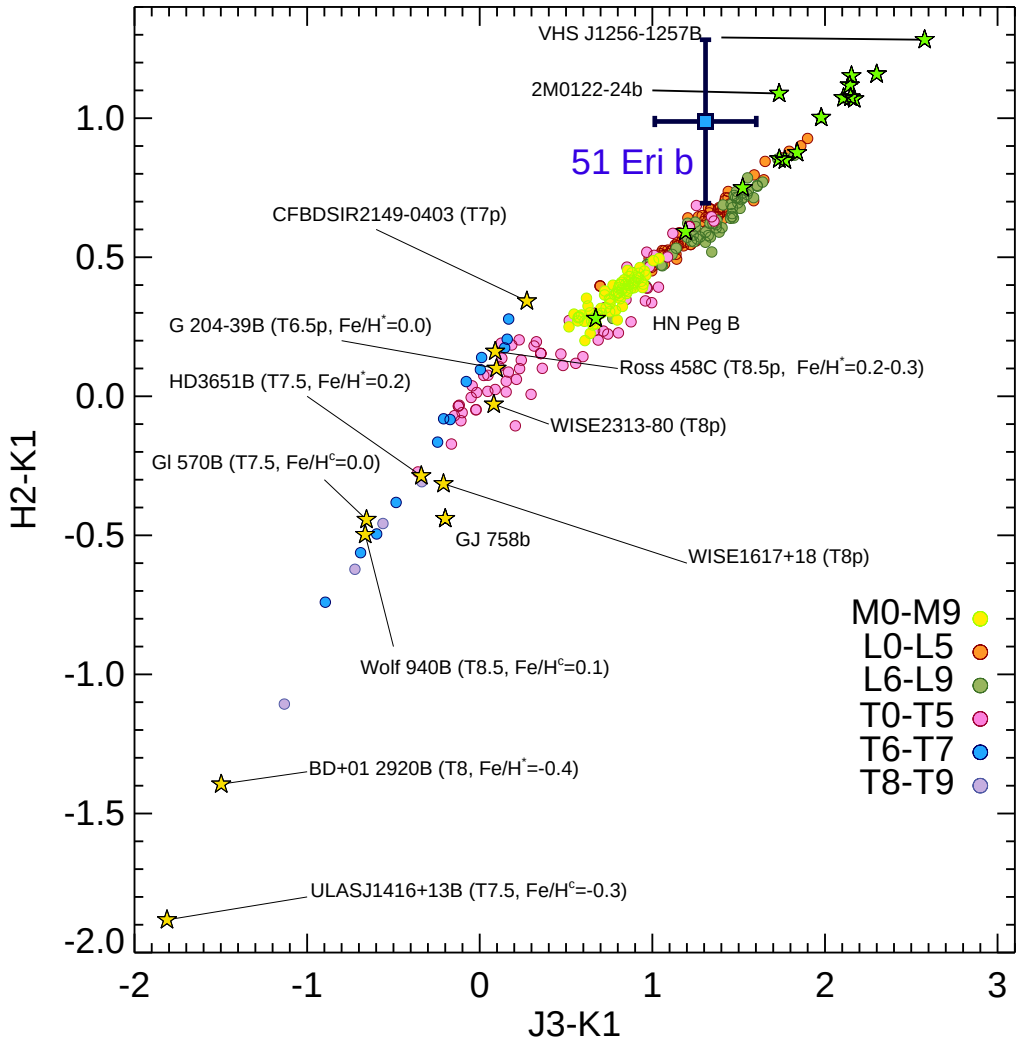


Figure 2.8: Placement of 51 Eri b in color-color diagram. All magnitudes except for K1 are derived from the IFS spectra.

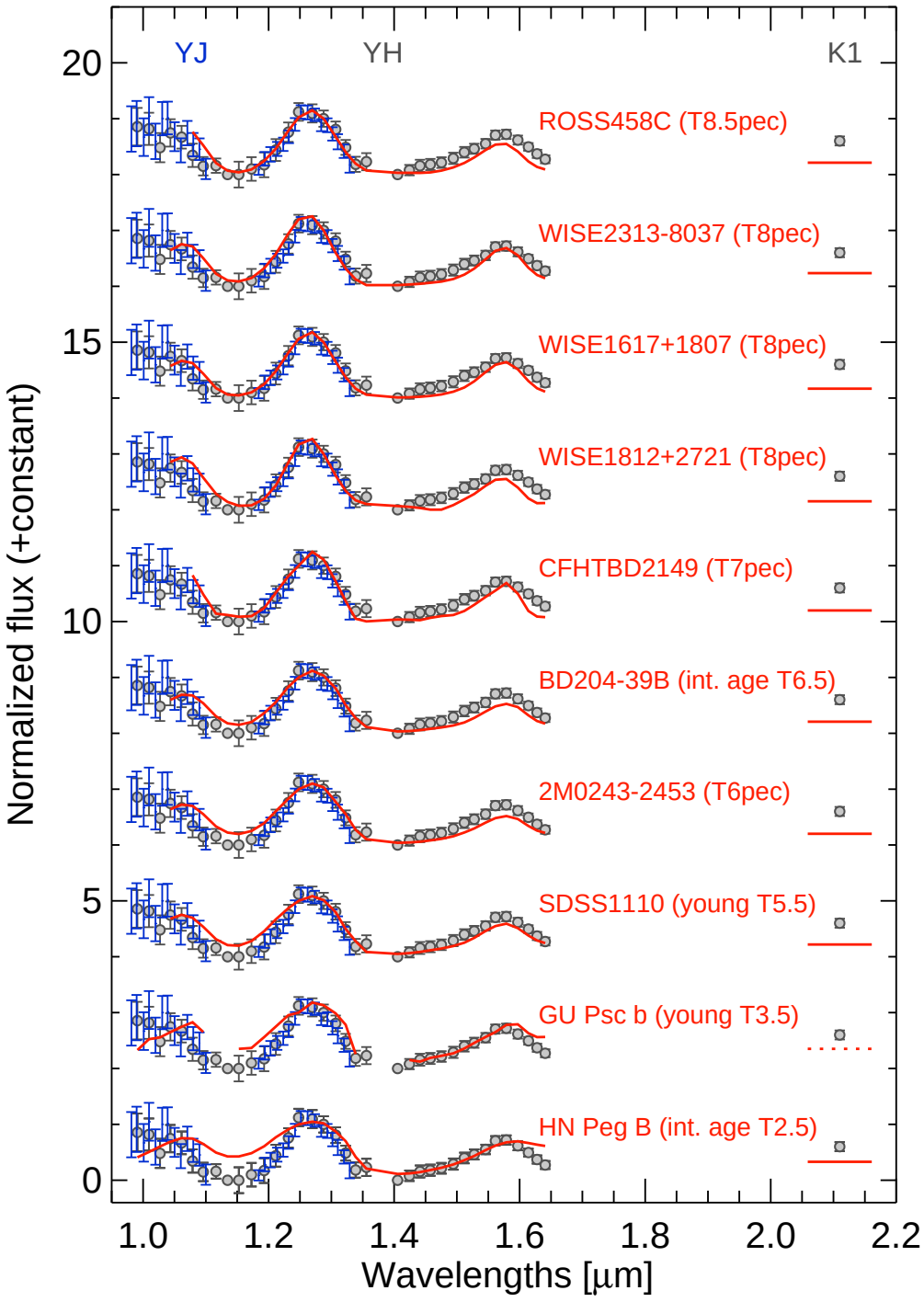


Figure 2.9: Comparison of the flux density F_λ of 51 Eri b to those of selected peculiar T-type objects whose flux density has been normalized to match that of the planet between 1.2 and 1.3 μm.

version of the code has been reported on in Mollière et al. (2015a) and updates have been shortly described in Mancini et al. (2016a,b). The current version of the code is described in detail in Mollière et al. (2017).

In its current form the code includes molecular and atomic line and continuum (CIA) opacities and an implementation of the cloud model by Ackerman & Marley (2001). The petitCODE treats the non-ideal line shapes of Na and K atoms using the wing profiles by Nicole Allard; see Mollière et al. (2015a) for a more detailed description. As possible cloud species $MgAl_2O_4$, $MgSiO_3$, Mg_2SiO_4, Fe, KCl, and Na_2S can be included with the optical constants taken from Palik (2012) for $MgAl_2O_4$, Scott & Duley (1996); Jaeger et al. (1998) for $MgSiO_3$, Servoin & Piriou (1973) for Mg_2SiO_4, Henning & Stognienko (1996) for Fe, Palik (2012) for KCl, and Morley et al. (2012) for Na_2S. Finally, the implementation of the Ackerman & Marley (2001) cloud model deviates from the description in the original paper in the sense that the mixing length is set equal to the atmospheric pressure scale height in all cases. This is different from the Ackerman & Marley (2001) description, where the mixing length can be up to 10 times smaller than the pressure scale height in the radiative regions. In the regions above the cloud deck the cloud mass fraction is proportional to $P^{f_{\rm sed}/\alpha}$, where $f_{\rm sed}$ is the ratio of the mass averaged settling velocity of the cloud particles and the atmospheric mixing speed and α is the ratio between the mixing length of the eddy diffusion process and the atmospheric scale height. In our implementation of the Ackerman & Marley (2001) model it holds that $\alpha = 1$. The given power law can be derived from solving the homogeneous part of the differential equation for the condensate density (Equation 4; Ackerman & Marley, 2001). Therefore, for a given $f_{\rm sed}$ value, clouds in the *petitCODE* implementation are more extended than in the Ackerman & Marley (2001) description. Further, the atmospheric mixing speed is equal to K_{zz}/L, where K_{zz} is the atmospheric eddy diffusion coefficient and L is the associated mixing length, or mean free path, of the mixing process. Because the *petitCODE* implementation sets $L = H_P$, where H_P is the pressure scale height, the mixing velocity is smaller than in the Ackerman & Marley (2001) description, which favors smaller cloud particles at a given $f_{\rm sed}$ value. Therefore, adopting $L = H_P$ results in effectively smaller $f_{\rm sed}$ values when comparing cloud properties of the original Ackerman & Marley (2001) description to the *petitCODE* at the same $f_{\rm sed}$ value.

Two dedicated grids were calculated for 51 Eri b (see Table 2.4). The first grid is a clear grid (subsequently PTC-Clear, i.e., cloud free), assuming scaled solar compositions for the planetary abundances. We varied the effective temperature $T_{\rm eff}$ between 500 and 1700 K, the surface gravity by assuming $\log g$ values between 3 and 6 (with g in cgs units), and the metallicities [Fe/H] between -1.0 and 1.4.

The second grid is a cloudy grid (PTC-C, "cloudy") for which we assumed a mixing coefficient $K_{zz} = 10^{7.5}$, which is similar to the value used in Macintosh et al. (2015). Here the varied grid parameters are $T_{\rm eff}$ =500–850 K, $\log g$ =3-5, [Fe/H] = 0.0–1.4, and $f_{\rm sed}$ = 0.5–2.0 (1–5 for initial exploration). Following Morley et al. (2012), the opacities of $MgAl_2O_4$, $MgSiO_3$, Mg_2SiO_4, and Fe were neglected for this cool grid, such that for the clouds only KCl and Na_2S opacities were considered.

Finally, our calculations were carried out assuming equilibrium chemistry for the gas composition and for identifying the cloud deck locations within the atmospheres. It is well known that for planets, compared to the higher mass brown dwarfs, disequilibrium effects, and the associated quenching of CH_4 and NH_3 abundances may be more important (see, e.g.,

Zahnle & Marley, 2014). Because we clearly detect methane in the atmosphere of 51 Eri b and we find best fit $\log g > 4$ (see Sect. 2.4.2), we conclude that CH_4 quenching is not very strong in this object; this is in agreement with the results presented in Zahnle & Marley (2014) for higher $\log g$ objects.

In addition to the two grids outlined above we compare our results with the cloudy model atmospheres described in Morley et al. (2012). However, as the grid does not include super-solar metallicity the resulting parameters are skewed (Appendix Fig. 2.23 and Fig. 2.27). We focus our discussion on the *petitCODE* models.
A summary of the used grids can be found in Table 2.4 and our *petitCODE* model grids will be made available online.

Determination of the spectral covariance matrices

When comparing the spectrum obtained with an IFS instrument with a model, taking into account the spectral covariance of the residual speckle noise has been shown to be of great importance for assessing the uncertainty of the fitted atmospheric model parameters (Greco & Brandt, 2016). Following the methods presented by these authors, we determine the mean spectral correlation between all spectral channels within an annulus of width $1.5\lambda/D$ at the separation of the planet (with the planet masked out by a $2\lambda/D$ radius mask),

$$\psi_{ij} = \frac{\langle I_i I_j \rangle}{\sqrt{\langle I_i^2 \rangle \langle I_j^2 \rangle}}, \tag{2.1}$$

where $\langle I_i \rangle$ is the average intensity inside the annulus at wavelength λ_i. The correlation matrix can then be used to obtain the covariance matrix C, which is used in computing the Gaussian log-likelihood $\ln L$ (or χ^2) for the MCMC model fit according to

$$-2 \ln L \equiv \chi^2 = (S-F)^T C^{-1} (S-F), \tag{2.2}$$

where S is the observed spectrum and F the model spectrum. In the case of uncorrelated noise C is equal to the unity matrix and χ^2 reduces to the more familiar sum over the residuals squared, which is not correct for correlated IFS data. The correlation matrix for the YH spectrum is shown in Fig. 2.10. We can see that each channel at the separation of 51 Eri b is strongly correlated with three to four of its adjacent channels in both directions. Contrary to Greco & Brandt (2016), we note that there are also anti-correlations present, which are due to the use of classical SDI and the larger spectral coverage available with the SPHERE IFS. The SHERE spectral coverage, unlike GPI spectra, spans multiple bands and band gaps.
As we do not have access to the reduced GPI data of 51 Eri b, we assume the fiducial model for a GPI-H spectrum reduced using simultaneous SDI and ADI as given in Greco & Brandt (2016) to calculate the correlation matrix at the angular separation of 51 Eri b.

Markov Chain Monte Carlo exploration of atmospheric models and parameters

We use the python implementation of the affine-invariant Markov Chain Monte Carlo (MCMC) algorithm *emcee* (Goodman & Weare, 2010; Foreman-Mackey et al., 2013) to

Table 2.4: Model grids used as input for MCMC exploration.

Model	$T_{\rm eff}$ (K)	ΔT (K)	$\log g$ $\log_{10}$(cgs)	$\Delta \log g$ $\log_{10}$(cgs)	[Fe/H] (dex)	Δ[Fe/H] (dex)	$f_{\rm sed}$	$\Delta f_{\rm sed}$
petitCODE (clear)	500 – 1700	50	3.0 – 6.0	0.5	-1.0 – 1.4	0.2	n/a	n/a
petitCODE (cloudy)	500 – 850	50	3.0 – 5.0	0.5	0.0 – 1.4	0.2	0.5 – 2.0[a]	0.5[a]
Morley+12	600 – 800	100	4.0 – 5.5	0.5	0.0	n/a	2 – 5	1

The radius of the planet was included as an additional analytic fit-parameter regardless of the model ranging from $0.1\,R_{\rm J}$ to $2\,R_{\rm J}$. [(a)] For the initial exploration a wider grid between 1 and 3 was used with a step size of 1. Smaller values were consistently preferred, leading to the final grid values.

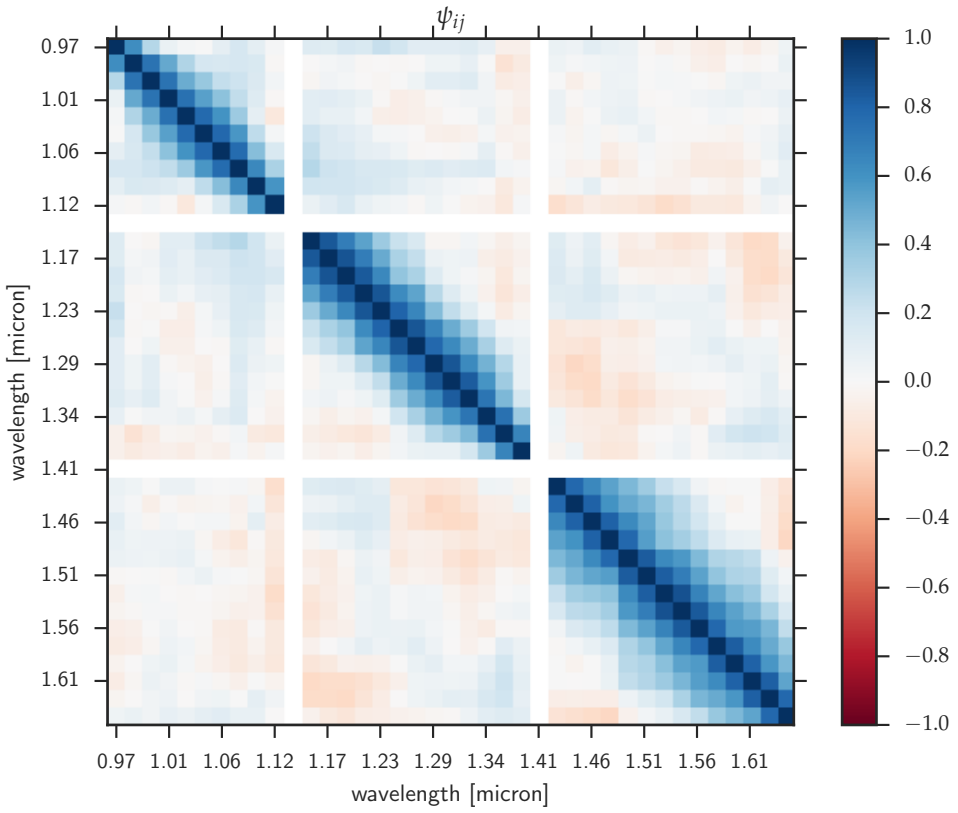


Figure 2.10: Correlation matrix ψ_{ij} showing the correlation between each pair of spectral channels (1: completely correlated; -1: completely anti-correlated; 0: uncorrelated). The 1.14 µm channel was used as reference for SDI for all wavelength channels shorter than this, whereas the 1.41 µm channel was used as reference for all other channels.

explore the posterior probability distribution of model parameters for various atmospheric model grids (see Table 2.4). Our custom procedure can handle model grids of an arbitrary number of parameters, number of photometric data points and/or spectra, as well as their respective covariance matrices. The only restriction is that we require the model grid to be regularly spaced in each individual parameter to allow for efficient N-dimensional linear interpolation, where N is the number of free atmospheric model grid parameters. As the atmosphere of 51 Eri b is not well characterized yet, we use flat priors over parameter ranges listed in Table 2.4. Planetary radii are fitted as a separate analytic parameter. The log-likelihood for each spectrum with their complete covariance matrix and each photometric data point are evaluated separately. These values are then summed to obtain the overall log-likelihood of the model given the data, and there is no statistical weighting between the data sets. Rather than defining a wavelength dependent weighting scheme, this is more properly taken into account by using the real covariances between the data. This effectively down-weights the relative importance of the many spectral data points with respect to the fewer, yet independent, photometric data points. Uncertainties are assumed to be uncorrelated between the separate data sets. The likelihood evaluation is carried out in luminosity space, taking into account the additional uncertainty of the systems distance (29.4 ± 0.3 pc),

which can be important for the radius uncertainty of the planet, which otherwise would be slightly underestimated. In this case, owing to the proximity and brightness of the host star, the distance uncertainty is only on the order of 1%, and thus does not impact the radius uncertainty much, but for objects at larger distance this can become a significant factor.

We follow a different approach when treating upper limits compared to many previous studies. We treat data points "below the detection limit" not as non-detection or upper limits in the fit because one does not look for a previously undiscovered source. We know where to measure the flux. Knowing the position of the source contains strong prior information, and even data points that are below the formal detection limit contain useful information; this can be seen by the fact that even data points that are technically not 3σ detections follow the model predictions very well. These non-detection points can still contain significant flux and in the "worst case" are consistent with negligible flux within their 1σ uncertainties. We use this approach of "forced photometry" (Lang et al., 2016), which is a method successfully used in other fields of astronomy, such as the study of faint galaxies and quasars (e.g., Venemans et al., 2015), to replace the more common practice of simply excluding data points below the classical detection threshold for point sources of unknown position because this would mean mixing two unrelated statistical quantities in an unjustified way and would effectively lead to throwing away informative data. Also replacing these measurements with an upper limit as is common—while seemingly the conservative choice—is not necessarily the optimal choice. In direct imaging, reporting only the upper limit is equivalent to just reporting the uncertainty for the measurement without reporting the measurement itself. Applying forced photometry for all measurements means consistently reporting both measurement and uncertainty. This has the advantage that all the data are treated uniformly and no arbitrary choice about a cutoff value for "detection" has to be chosen. The problem is illustrated by Fig. 2.1, where one would not claim the discovery of an unknown planet at the position of 51 Eri b given only the Y-band image, but the fact that the occurrence of a clear excess in flux is located at the exact position of the planet visible in the other bands is much more informative and illustrates the importance of prior knowledge of the position of a planet for characterization.
However, to put the importance of the Y-band measurement in this particular case into perspective, it should be pointed out that while it is true that all points included for forced photometry contain some information on the spectrum of the planet owing to their low S/N and because other parts of the spectrum already put very strong constraints on the Y-band model fluxes, they do not impact the derived atmospheric parameters significantly. This is especially true for cases in which the rest of the spectrum poorly constrains the spectral shape at wavelengths where the flux measurements are not above the detection limit; for example, models comparisons that seek to distinguish between the presence or absence of a physical model component, such as thermal inversion or significant disequilibrium chemistry.

For all of these reasons, we also include the measured flux in the methane "non-detection" bands H3 and K2 (see Table 2.2). The only data that we do not include in the fit are the spectral channels that were used as a reference in the SDI step of data reduction, as these are biased, and the first three IFS channels as they are most affected by degrading overall system performance and telluric lines.

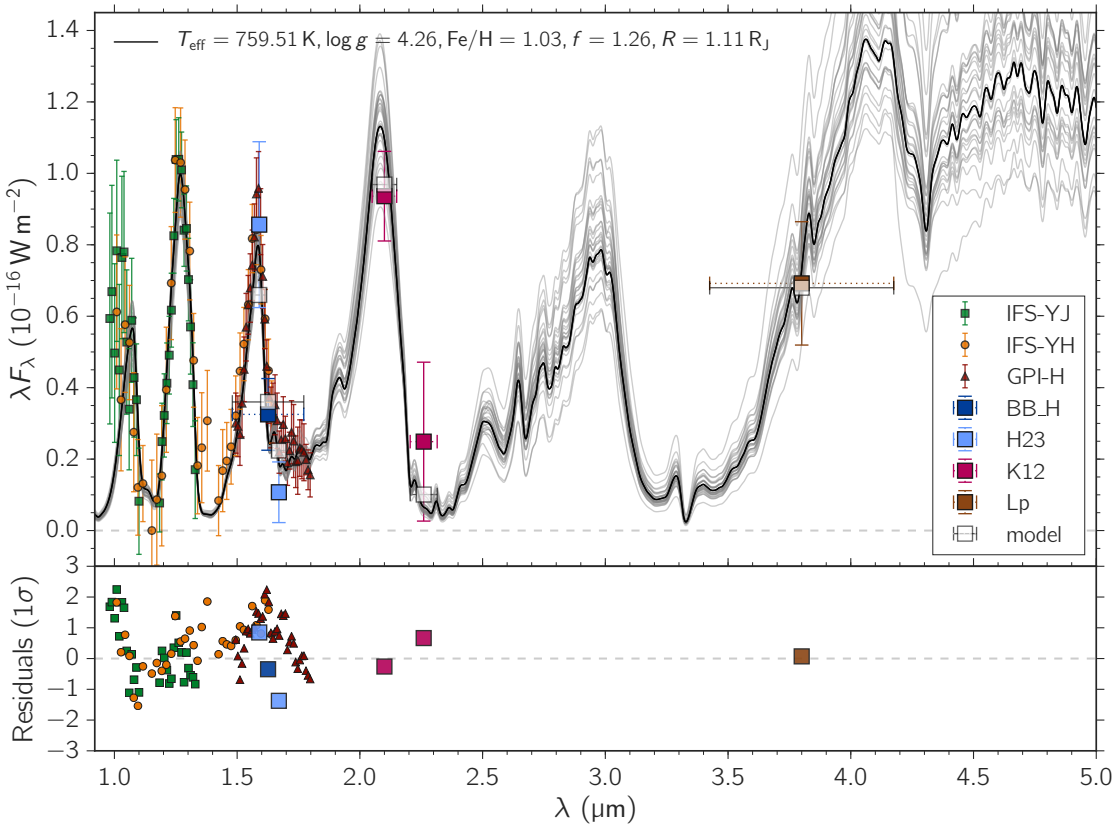


Figure 2.11: Plot shows the *petitCODE* cloudy model interpolated to the parameters best describing the data according to the posterior probability distribution (black line), as well as the SPHERE spectrophotometric data, GPI H-band spectrum, and L' data point from Macintosh et al. (2015). For photometric data points, the x-error bar reflects the filter width rather than uncertainties. The gray lines represent 32 randomly drawn samples from the posterior probability distribution to reflect the spread of plausible model parameter combinations that fit the data. Photometric points describe the average flux in the respective filter, whereas the orange points describe the average flux in the respective filter for the best fitting model. The residuals are shown in multiples of 1σ uncertainties of the data.

Discussion of physical parameters

The best fitting models for the cloudy model grid (PTC-C) are shown in Fig. 2.11, where the black line represents the best fit and the gray lines showing the spectrum for 16 randomly drawn parameter combinations from the posterior probability distribution. As the most extensive model of the three, the posterior probability distribution of the PTC-C model is shown for each of the model parameters along with their marginalized values in Fig. 2.12. Cloud-free models are incapable of explaining all of the observed spectral features simultaneously: models that explain the Y, J, and H peaks are not able to explain the K1- and L'-band data (see Fig. 2.26). They also result in model predictions that are unphysical for young giant planets, for example, high $\log g = 5.35^{+0.15}_{-0.12}$ and very low radius $R = 0.40 \pm 0.02\,R_J$ (see Fig. 2.26). Cloudy models vastly improve the consistency with the data over the whole spectral range for which data are available. Our discussion below centers on the results obtained on the *petitCODE* cloudy models. These results cover the complete parameter space relevant

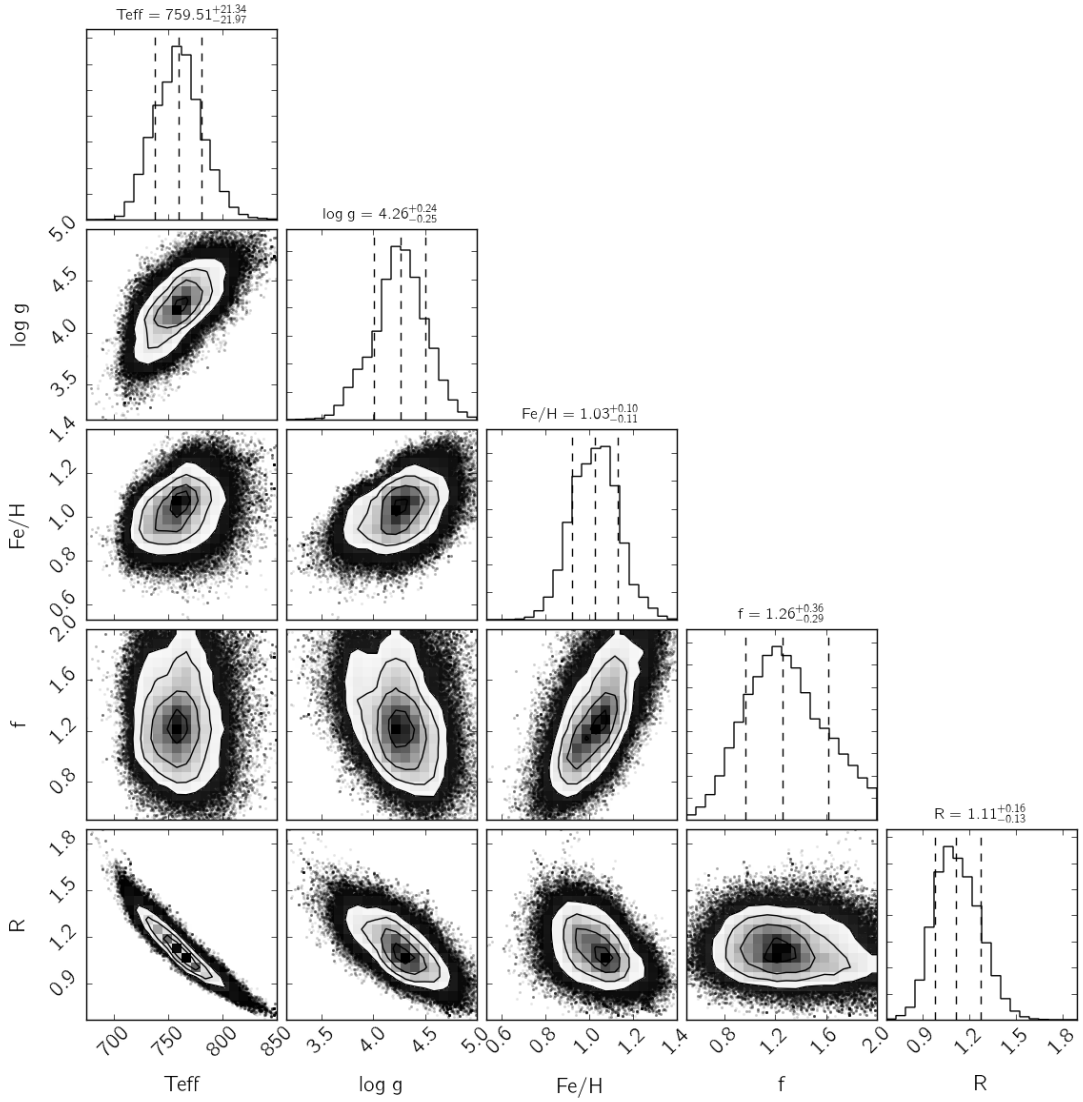


Figure 2.12: Corner plot showing the posterior probability distribution of the cloudy *petitCODE* grid with respect to each of its parameter pairs as well as the marginalized distribution for each parameter. The uncertainties are given as 16% to 84% percentiles as is common for multivariate MCMC results.

for 51 Eri b, including metallicity and cloud sedimentation values ($f_{\rm sed}$). The results of all tested models are summarized in Table 2.5.

Temperature, radius, and surface gravity We obtain a value of $T_{\rm eff} = 760 \pm 20\,{\rm K}$ for the effective temperature, $R = 1.11^{+0.16}_{-0.14}\,{\rm R_J}$ for the radius, and $\log g = 4.26 \pm 0.25$ (cgs-units) for the surface gravity of 51 Eri b. The effective temperature and radius of the planet are expectedly correlated ($T_{\rm eff} \propto R^{-0.5}$ for black bodies) as they both relate to the luminosity of the planet. With a temperature that is likely above 700 K, it appears to be above the temperature for which sulfur chemistry becomes an important factor for 51 Eri b as discussed in Zahnle et al. (2016). The radius is consistent with the radius of Jupiter and may be slightly

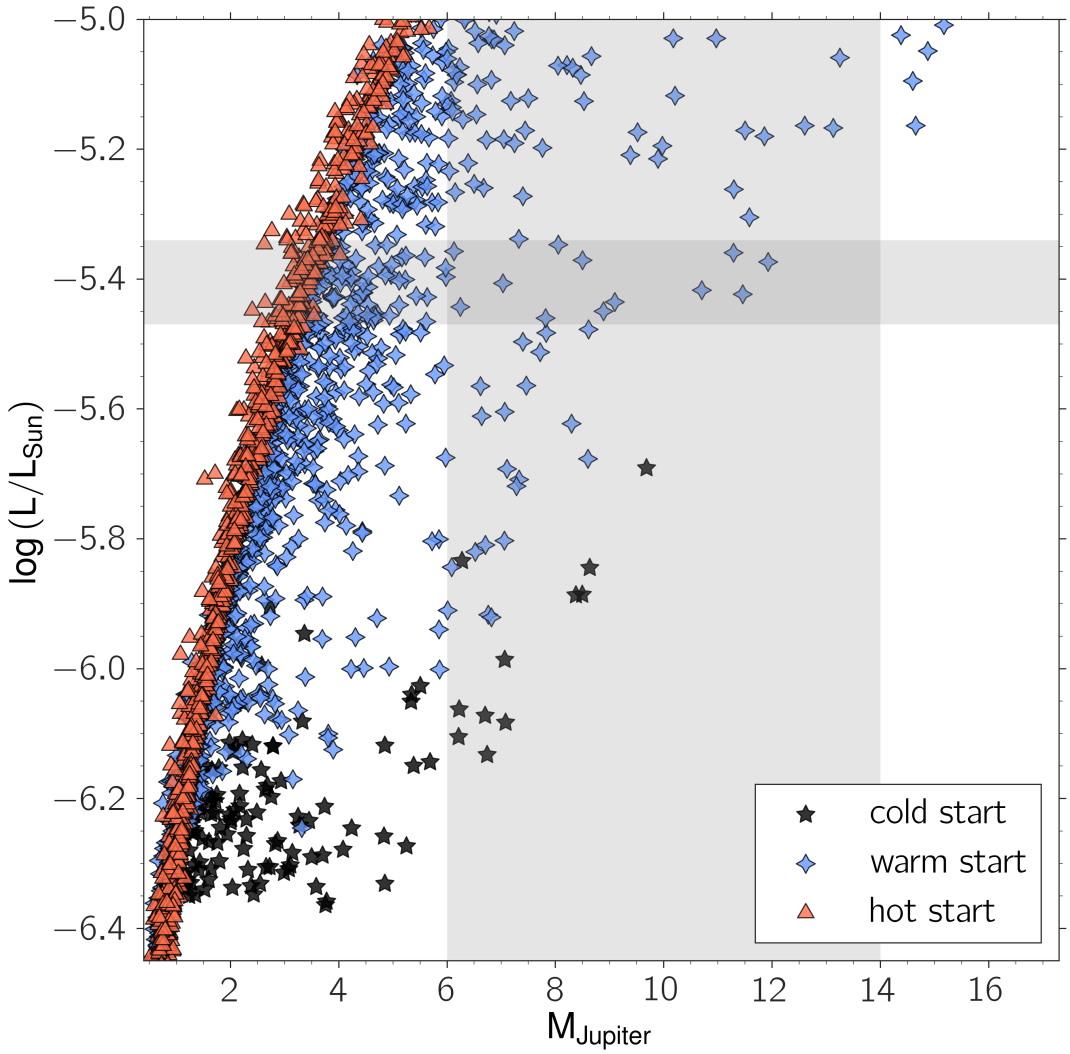


Figure 2.13: Luminosity-mass relationship from core-accretion population synthesis model at 20 Myr. Populations using different entropy assumptions are plotted, corresponding to what is traditionally referred to as cold (black) and hot start (red). The warm start (blue) model corresponds to a cold gas accretion model, but allowing for higher core masses than $17\,M_{Earth}$. Gray shaded regions correspond to the luminosity of 51 Eri b derived in this work and the mass range determined from surface gravity and radius.

larger as expected for young objects (Chabrier et al., 2009; Mordasini et al., 2012a).

Mass To check the consistency of the best fitting solution with our physical understanding we can use multiple approaches to derive the mass of the planet. Using the posterior sampling of surface gravities and planet radii, according to $M = g/\mathrm{g_J} \cdot (R/\mathrm{R_J})^2$, where $\mathrm{g_J} = 24.79\ \mathrm{m\ s^{-2}}$ and $\mathrm{R_J} = 6.9911 \times 10^7$ m are the surface gravity and (volumetric mean) radius of Jupiter, respectively, we get a mass estimate of $M_{\mathrm{gravity}} = 9.1^{+4.9}_{-3.3}\,\mathrm{M_J}$. Another approach is to use the luminosity of the planet derived similarly from the posterior sampling of the radius and effective temperature according to $L/\mathrm{L_\odot} \sim (R/\mathrm{R_\odot})^2 \cdot (T/\mathrm{T_\odot})^4$, where $R_\odot = 9.728\,\mathrm{R_J}$ and $T_\odot = 5772$ K are the radius and effective temperature of the Sun, we get a luminosity of $L/\mathrm{L_\odot} \sim (3.94^{+0.66}_{-0.55}) \times 10^{-6}$ or $\log(L/\mathrm{L_\odot})$ between -5.470 and -5.338 (compared to -5.4 and -5.8, Macintosh et al., 2015), which can be converted to a mass assuming a formation (initial entropy) model. Figure 2.13 shows the luminosity–mass relationship derived from a complete core accretion population synthesis model (Mordasini et al., 2012b,a). The shaded region corresponds to the above-mentioned luminosity and (surface gravity derived) mass range. Three populations at 20 Myr using different assumptions are shown, where two populations correspond to the traditional hot-start and cold-start populations, and one corresponds to an intermediate warm-start population. In the hot-start case, energy from the accretion shock is not radiated away efficiently and is deposited in the planet. In the cold-start case, all energy from gas accretion is radiated away, furthermore core masses are restricted to $< 17\,\mathrm{M_{Earth}}$ to mimic the traditional cold-start model by Marley et al. (2007). The warm-start population is similar to the cold-start population in terms of accretion physics, but allows higher core masses, which in turn leads to more energy deposition during the growth of the planet (Mordasini, 2013). A similar result would be achieved by allowing for a spread in the radiation efficiency of the accretion shock (Marleau et al., 2017). We can see that the observed luminosity range excludes the traditional cold-start model (consistent with theoretical predictions Marleau & Cumming, 2014), but includes both objects of the hot- and warm-start case with a large spread in masses. Small masses between 2.4 and $5\,\mathrm{M_J}$ are preferred in the hot-start case, whereas a big spread of masses between 2.4 and $12\,\mathrm{M_J}$ are possible in the warm-start case. While in this synthesis model objects with smaller mass are more common in this luminosity range, big masses are not excluded from the point of view of the formation model.
The above discussion and Fig. 2.13 exemplify the problem of determining the mass of directly imaged exoplanets in the absence of low-mass and cool benchmark objects with independent mass measurement. In the two approaches, using the measured luminosity together with evolutionary models gives a statistical picture of the distribution of planets resulting from the planet formation synthesis modeling approach, and allows for a wide range of masses for the given age and luminosity, depending strongly on the accretion physics assumed. In principle, the mass derived from the surface gravity and radius are more constraining, but depend strongly on the atmospheric model assumptions, which in the case of cold and cloudy objects still have many uncertainties. Assuming the physics model represents the real nature of the planet, the determined log g and radius is more consistent with a more massive planet that would be expected based on hot-start evolution models. It should be mentioned, however, that we can rule out the brown dwarf regime, as brown dwarfs at this age we would expect to see a significantly larger radius because they have deuterium burning

as an additional heat source.
In conclusion it can be said that both mass estimates are highly model dependent and there are multiple big sources of uncertainty in both approaches. For atmospheric models it is possible that the cloud physics are not sufficiently well modeled, leading to big uncertainties in the surface gravity determination. The surface gravity determination is also most strongly impacted by the J-band flux for which some amount of variability cannot be completely ruled out at this point. On the other hand, evolutionary models have a big intrinsic spread as they reflect the statistics of populations rather than single objects (e.g., different core-mass fractions). Initial conditions for planet formation and evolution are not well constrained. Another aspect that deserves further research is the current lack of evolutionary models that consider super-solar metallicity objects. While the composition is reflected in the core mass of the models used here, the thermal evolution does not include the increase in opacity caused by high metallicity. This is also an issue for all other evolution models currently available.

Metallicity The metallicity $[\mathrm{Fe/H}] = 1.0 \pm 0.1$ dex is super-solar, and significantly above that of the solar metallicity host star. This is similar, but even more pronounced than what has previously been observed in the cool object GJ 504 b ($\mathrm{Fe/H} = 0.60 \pm 0.12$ around a slightly metal-rich host star, Skemer et al., 2016). The metallicity determined here is in good agreement with predictions of bulk composition for giant planets formed by core accretion (e.g., Mordasini et al., 2014). Other studies of massive directly imaged exoplanets also suggest super-solar metallicities, for example HR 8799 b (Lee et al., 2013).

Increasing the planetary metallicity strongly enhances the K-band brightness, redistributing a part of the flux from shorter to longer wavelengths. The reason for the metallicity-dependent K-band brightness is that a change in metallicity shifts the position of the planetary photosphere within the atmosphere because in hydrostatic equilibrium it holds that $d\tau = (\kappa/g)\,\mathrm{d}P$, where τ is the optical depth, κ the opacity, g the planetary surface gravity, and P the pressure. If the pressure dependence of κ is neglected, an increase in κ, resulting from an increase in metallicity, shifts the photosphere ($\tau_{\mathrm{Phot}} \sim 2/3$) to lower pressures. Nonetheless, it is critical to note that the opacity varies as a function of pressure: the strength of pressure broadened molecular and atomic line wings is proportional to the pressure P, but for the many lines of water and methane the effect in the K band is of secondary importance. More importantly, the continuum opacity due to CIA of H_2–H_2 and H_2–He pairs is linear in pressure for all wavelengths. Further, CIA exhibits a peak in opacity at 2.3 μm, i.e., very close to the K band, whereas the CIA opacity in the neighboring H band is lower by a factor of 100. Consequently, as the photosphere is shifted to lower pressures, owing to an increased metallicity, the contribution of CIA to the total opacity in the K band diminishes, such that the opacity minimum resulting from the scissor-like crossover of the water and methane opacities becomes visible as an emission feature (Allard et al., 2001). Owing to the steep decrease of the CIA opacity toward smaller wavelengths the H band is unaffected by the increase in metallicity. As a final test we carried out runs neglecting the CIA opacities and were unable to reproduce the effect of the metallicity-dependent K band.

The strong influence of metallicity on key spectral features shows the importance of having a broad wavelength coverage of all features present in the near- to mid-infrared and using model grids that include non-solar metallicity as a free parameter. Finally, to make sure our

methodology and models are not systematically biased toward providing high metallicity results, we analyzed two benchmark brown dwarfs (Gl 570D and HD 3651B; similar to Line et al., 2015) and confirmed that the metallicities derived are reasonable. The details of this analysis can be found in Appendix 2.9.2.

Clouds For the cloud sedimentation parameter, we derive a value of $f_{sed} = 1.26^{+0.36}_{-0.29}$. A lower $f_{\rm sed}$ results in more vertically extended optically thicker clouds with smaller particle sizes. While the slight differences in atmospheric model implementations make it difficult to compare this result exactly with previous research, $f_{\rm sed}$ as low as this (< 2) are unusual for self-luminous substellar objects of low temperature, especially considering that our implementation of $f_{\rm sed}$ would result in a lower value in the Ackerman & Marley (2001) implementation (see model discussion in Sec. 2.4.2). Values of $f_{\rm sed}$ between 3 and 5 are usually reported, for example, for GJ 504 b (Skemer et al., 2016) and GJ 758 B (5, Vigan et al., 2016a). Cushing et al. (2008) report values between 1 and 4 for their sample of L and T brown dwarfs, but all of these objects are significantly hotter than 51 Eri b and only models of solar metallicity are considered. The lack of similar objects and detailed analyses including metallicity as a free parameter make a real comparison difficult. Lack of metallicity as free parameter in the model can significantly alter the cloud parameter because it tends to compensate for the lack or overabundance of heavy elements in the spectrum. We encourage modelers to include low $f_{\rm sed}$-values as well as metallicity in their consideration for future model grids.
The $f_{sed} = 1.26^{+0.36}_{-0.29}$ we obtain for 51 Eri b reflects a particle size distribution with mean values of 1 µm, and slightly below in the upper regions (below 10^{-2} bar). Owing to the width of $\sigma = 2$ of the lognormal size distribution, however, the opacities are dominated by the larger particles.

Patchy cloud models

As variability has been observed in a number of brown dwarfs, the idea that for cool substellar objects cloud coverage may be less than 100% should not be excluded a priori. Macintosh et al. (2015) used such a patchy cloud model, which can be expressed as a linear combination of a clear and a cloudy atmosphere. They also included disequilibrium chemistry in the cloudless model. We tested this idea with the following simple composite model using *petitCODE*

$$F_{\rm patchy} = CF \cdot F_{\rm cloudy} + (1 - CF) \cdot F_{\rm clear}, \tag{2.3}$$

where CF is the cloud fraction and $F_{\rm cloudy}$ and $F_{\rm clear}$ are the flux of the cloudy and clear model, respectively. Under this model we have the following MCMC fitting parameters $\theta = (T_{\rm cloudy}, T_{\rm clear}, CF, \log g, [{\rm Fe/H}], f_{\rm sed}, R)$, i.e., we now fit for the cloud fraction and allow the two models to have different temperatures, as the cloudy and clear model fluxes probe different temperatures. Because both models must, however, describe the same physical planet, we keep the metallicity, as well as the surface gravity and radius, the same for both models. Furthermore, we impose $T_{\rm cloudy} < T_{\rm clear}$ as a prior, as the cloudy model flux is supposed to come from higher in the atmosphere than the clear flux, which in this model

Table 2.5: Summary of model results

Model	$T_{\rm eff}$ (K)	$\log g$ $\log_{10}$(cgs)	[Fe/H] (dex)	$f_{\rm sed}$	R (R_J)	L ($10^{-6}\,L_\odot$)	$\log L$ $\log_{10}(L_\odot)$	$M_{\rm gravity}$ (M_J)
PTC-C	760^{+21}_{-22}	$4.26^{+0.24}_{-0.25}$	$1.03^{+0.10}_{-0.11}$	$1.26^{+0.36}_{-0.29}$	$1.11^{+0.16}_{-0.13}$	$3.94^{+0.66}_{-0.55}$	-5.470 to -5.338	$9.1^{+4.9}_{-3.3}$
PTC-C w.o. Y-band	754^{+23}_{-23}	$4.25^{+0.32}_{-0.37}$	$1.04^{+0.11}_{-0.12}$	$1.33^{+0.38}_{-0.33}$	$1.13^{+0.17}_{-0.15}$	$3.92^{+0.68}_{-0.61}$	-5.479 to -5.337	$9.1^{+7.3}_{-4.3}$
PTC-Patchy[a]	757^{+24}_{-24}	$4.47^{+0.24}_{-0.26}$	$1.25^{+0.10}_{-0.16}$	$1.07^{+0.36}_{-0.31}$	$1.11^{+0.16}_{-0.14}$	$3.84^{+0.63}_{-0.56}$	-5.484 to -5.350	$14.5^{+8.7}_{-5.6}$
PTC-Clear	982^{+18}_{-15}	$5.35^{+0.15}_{-0.12}$	$1.36^{+0.03}_{-0.06}$	-	$0.40^{+0.02}_{-0.02}$	$1.43^{+0.06}_{-0.06}$	-5.863 to -5.827	$14.5^{+4.7}_{-3.1}$
Morley+12	684^{+16}_{-20}	$5.19^{+0.10}_{-0.11}$	-	$4.16^{+0.52}_{-0.67}$	$1.01^{+0.07}_{-0.06}$	$2.12^{+0.14}_{-0.13}$	-5.700 to -5.645	$64.9^{+19.1}_{-15.6}$
PTC-C (Macintosh et al[b])	785^{+11}_{-17}	$3.35^{+0.29}_{-0.21}$	$0.83^{+013}_{-0.12}$	$2.54^{+0.32}_{-0.47}$	$1.12^{+0.08}_{-0.05}$	$4.55^{+0.34}_{-0.30}$	-5.372 to -5.311	$1.2^{+1.0}_{-0.4}$

See Appendix 2.9.4 for corner plots of the posterior probability distributions of the model parameters for each model. [(a)] Effective temperature calculated by $T_{\rm eff} = \left[CF \cdot T^4_{\rm cloudy} + (1-CF)\cdot T^4_{\rm clear})\right]^{1/4}$ with $T_{\rm cloudy} = 751^{+24}_{-25}$, $T_{\rm clear} = 813^{+67}_{-40}$ and cloud fraction $CF = 0.91 \pm 0.05$. [(b)] Using our cloudy model and the same data as Macintosh et al. (2015) without covariance treatment.

corresponds to holes in the clouds.
However, the result of this test shows no significant improvement of the fit for the resulting composite model spectrum as cloud coverage tends toward $> 90\%$ (see Appendix Fig. 2.28 for corner plot). As the resulting spectra (Appendix Fig. 2.24) look almost the same, we conclude that a patchy cloud model may not be necessary to explain the observed spectrum, and at least at this point, the increase in model complexity is not justified. It should be pointed out that using the patchy cloud model improves the fit marginally when data from Macintosh et al. (2015) are used exclusively. This may be attributed to the higher J-band flux in GPI and resulting bluer spectrum.
To be clear, we do not wish to say that patchy cloud models in general do not work or should be avoided, but that for this particular planet, data set, and model, cloudy models alone seem to be capable of fitting the data very well. It may well be that inclusion of more physics (e.g., disequilibrium chemistry) improves the results. It is also important to keep in mind that a simple linear combination of the clear and cloudy models, as carried out here, is not self-consistent and strictly speaking not physically correct (Marley et al., 2010). A detailed model comparison with a more rigorous patchy cloud model should be performed in the future to test whether further increasing the model complexity is justified by the gain in fit quality, for example, by using Bayesian evidence (e.g., nested sampling).

Unexplained spectral features

A number of features in the spectrum of 51 Eri b exist that cannot be explained with the current model, either pointing to unaccounted systematic effects in the data or the need for more sophisticated atmospheric models.

1. The Y-band peak in the data is stronger and seems to extend to smaller wavelengths than predicted by the model. The Y band is difficult to observe with good S/N, mostly because overall instrument performance degrades toward shorter wavelengths (e.g., worse AO correction and end-to-end instrument throughput). It is also subject to some unresolved telluric features at short wavelengths ($\lesssim 1$ µm) in the Earth's atmosphere. There are multiple plausible scenarios for the perceived discrepancy: a) residual speckle flux at planet position at these wavelengths, b) a genuine instrument systematic effect (e.g., unaccounted variations in system transmission), and c) a real physical phenomenon or improper treatment of potassium wings or abundances in model. If there is residual speckle flux (i.e., speckle noise) and the noise is spectrally correlated (as it is, the treatment of which is described in Sect. 2.4.2 and taken into account), we expect it to affect at least half of the Y-band channels consistently (as about six neighboring channels are correlated). Seeing visually that a number of points scatter "systematically" higher or lower than the model is actually what we should expect in this case at low S/N. It is important to remember that we can only plot 1D error bars, which looks like we have a systematic effect if we work under the assumption that the measurements are uncorrelated and should scatter randomly around the true values. The proper treatment of IFS covariances is a relatively new practice in this field and needs to be kept in mind. As such, the method of forced photometry should only be used in conjecture with proper covariance treatment. On the other hand, it makes it challenging to distinguish residual speckle flux from other instrument systematics

that may only be present in certain wavelength regions. "Systematic deviations" that conform to the correlation length are not much of an issue for the parameter fitting because we already take this effect into account. To confirm this, we performed the same fit without including the Y-band data at all, which only marginally changes the results. This test shows that the relatively large and correlated uncertainties in the Y-band data are not very constraining at this point, which strengthens our confidence in the robustness of the analysis and our treatment of the noise; see Figure 2.25 for posterior distribution in the case in which Y-band data are excluded. Resolving the issue of which of the three scenarios (or mixture of them) is dominant will require us to obtain more high S/N Y-band measurements. If the elevated flux level in the observation are shown to be persistent and significant in upcoming observations, this raises the possibility that the model treatment of potassium wings or abundances needs to be reconsidered and improved (e.g., better alkaline profiles and a mechanism for depletion of alkaline species).

2. We observe an emission feature at $\sim$1.35 μm that is not explained by the model. While it is possible that this is caused by instrument systematics or the fact that it is in a region of strong telluric absorption, it is striking that both the GPI and SPHERE observations show an increase in flux. A very similar feature in the deep water bands between the J and H peaks at 1.35 – 1.40 μm has been observed and discussed by King et al. (2010) in the ϵ Indi Ba and Bb brown dwarf binary members. These authors also list objects with descriptions of similar features, for example, the T1 spectral standard SDSS0151+1244 (Burgasser et al., 2006b), the T8.5 and T9 dwarfs ULAS1238 and ULAS1335 (Burningham et al., 2008), and some L dwarfs (e.g., 2MASS J1507–1627 (L5) Burgasser, 2007). King et al. (2010) argue that this feature is due to the structure of the strongest part of the water absorption bands in the higher levels of the atmosphere, which may be the result of an underestimated local temperature in this region of the atmosphere. According to their toy model, raising the temperature, and therefore changing the temperature gradient can reconcile the modeled and observed flux levels, although they could not point to a reasonable physical mechanism, such as back-warming owing to an additional opacity source. If this feature is indeed a real feature, in the case of 51 Eri b, it may be related to its atmospheric cloud structure, but at this point this is very speculative. The fact that both the target planet and Earth's atmosphere contain complex telluric features at these wavelengths makes it difficult to draw strong conclusions.

3. The H-band feature has a broad tail toward shorter wavelengths and an extended wing toward longer wavelengths, which has a profound impact on the model fit. In general the H-band wings strongly favor models with higher log g and lower metallicity, whereas the rest of the spectrum favors lower log g and higher metallicity (especially the need for high metallicity to produce the K1-peak). Excluding the GPI H-band spectrum from the fit allows the PTC-C model as well as the Morley et al. (2012) models to fit the strength of all of the observed features well (except for the width and height of the Y-band peak and the width of the H-band peak). Including H-band wings in the fit puts strong weight on these features and the resulting best model is a compromise between fitting the H-band wings and the amplitude of the peak. While this spectrum

fits the overall shape of the H band well, it does not match the absolute amplitude of the H peak. A zoom in on the wavelength range covered by IFS data is shown in Fig. 2.21. This trade off shows how important extensive coverage of the spectral bands is for drawing physical conclusions.

2.5 Constraints on additional companions

2.5.1 VLT-NACO: Sparse aperture masking with L' filter

To constrain the presence of any potential companions at smaller separations, we processed and analyzed archival sparse aperture masking (SAM) data taken with the VLT-NACO instrument. The observations were made on 2009-12-26 using the L' filter and the 7-hole aperture mask. The calibrator stars HIP 22226, HIP 30034, HIP 24947, and HIP 32435 were used, and the conditions were between median and bad. Single exposures had detector integration times (DITs) of 0.2s with a total of 3200 frames (NDIT = 200, 16 cubes). The calibrators had DITs of 0.25 with the same number of frames. Data were processed using the IDL aperture masking pipeline developed at the University of Sydney. The data processing steps are described in Tuthill et al. (2000); Kraus et al. (2008) and the references therein. Briefly, the images were sky subtracted, flat fielded, cleaned of bad pixels and cosmic rays, and then windowed with a super-Gaussian function. The closure phases were then measured from the Fourier transforms of the resulting cleaned cubes. The closure phases were calibrated by subtracting the average of those measured on several unresolved calibrator stars observed during the same night with the same instrument configuration. To estimate the detection limits of the SAM data, a Monte Carlo simulation was performed. Using a Gaussian distribution, we generated 10,000 simulated data sets consistent with the measured uncertainties. For each point on a grid of separation, position angle and contrast, we defined our detection limits to be the point at which 99.9% of the simulated data sets were fit better by a point source model than the binary model. These 3.3σ detection limits were then scaled to 5σ to simplify the comparison with the results from SPHERE. No additional point sources are detected.

2.5.2 SPHERE

We detect no additional point sources in the SPHERE data. Contrast curves for the more extended FoV achievable with IFS and IRDIS were compiled with different reduction methods. The methodology for deriving the contrast was kept as similar as possible between the algorithms. For the LOCI and PCA reductions the contrast curves correspond to the azimuthal 5σ self-subtraction corrected variance in the respective separation bin, whereas ANDROMEDA inherently models the detectable signal contrast and does not need self-subtraction correction. The effect of small-sample statistics at small separations (Mawet et al., 2014) and the coronagraphic throughput (A. Boccaletti, priv. comm.) have been accounted for. The IFS-YJ PCA reduction was performed with the more aggressive simultaneous ADI plus SDI algorithm for detection and the median combination of all channels is shown. The top panel of Fig. 2.14 shows the achieved contrast with both the innermost region explored by small aperture masking and the exploration region of SPHERE. It is not

straightforward to convert the "detection images" shown in Fig. 2.2 into quantitative contrast curves and detection limits. As such they are not used for this purpose in this paper, but serve as a qualitative probe for additional candidates. However, no obvious candidates are seen.

For the conversion from contrast to mass limits, we used the same JHKL' magnitudes, distance, and age for the host star as in Macintosh et al. (2015). We use evolutionary tracks of Baraffe et al. (2003) together with the atmosphere model of Baraffe et al. (2015a) for the SPHERE data and BT-Settl models of Allard et al. (2012) for the NaCo data because Baraffe et al. (2015a) does not include NaCo L' filters. The mass limit derived from the IFS data assumes a companion-to-star contrast, which is constant with wavelength as a conservative choice for which we already run into the lower mass limits of the used models. The mass limits (see bottom panel of Fig. 2.14) constrain the presence of additional components in the system very well. The SAM data reject $> 20\,\mathrm{M_J}$ companions at $\sim 2-4$ au, while the IFS data are sensitive to planets more massive than $4\,\mathrm{M_J}$ beyond 4.5 au and $2\,\mathrm{M_J}$ beyond 9 au.

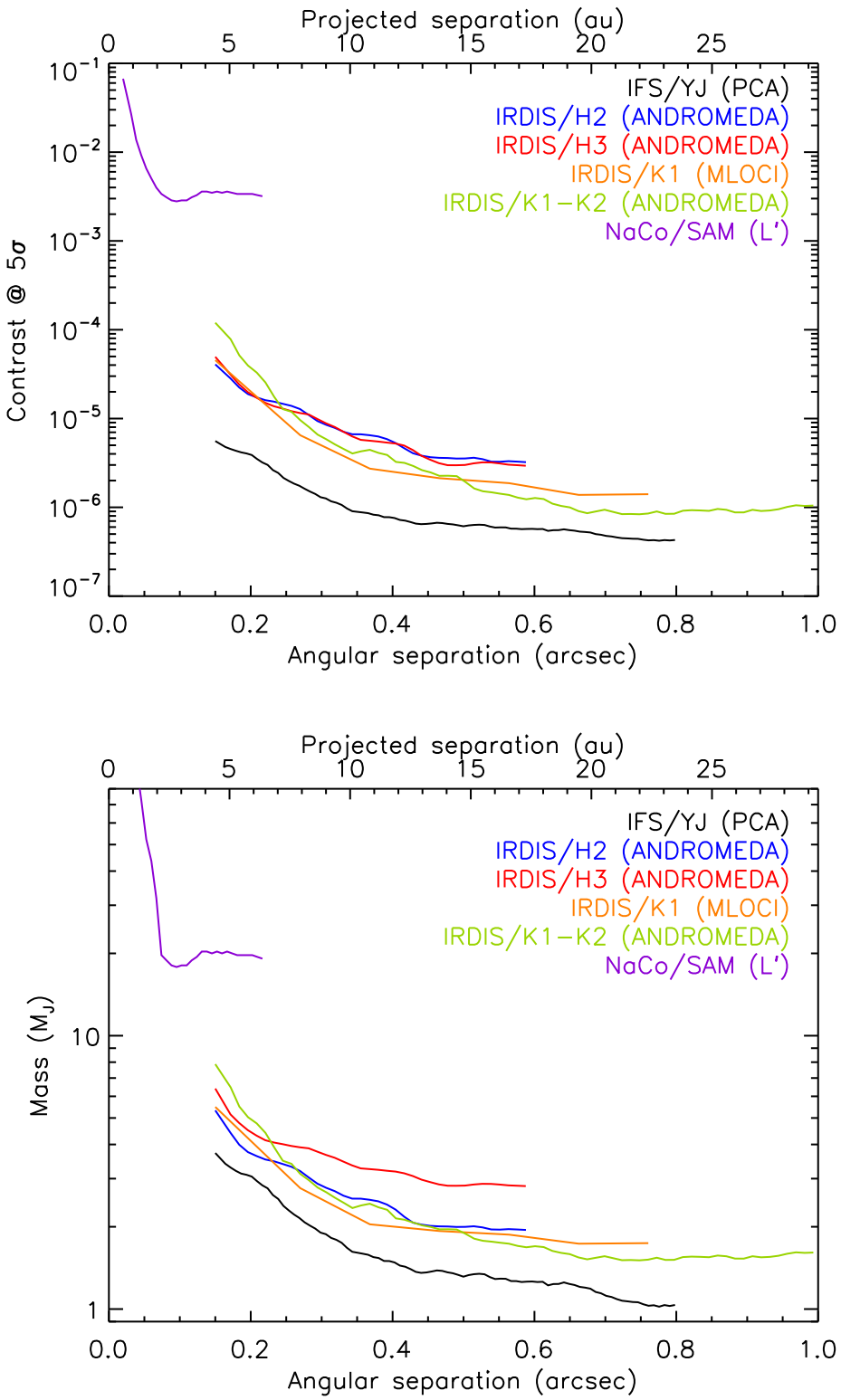


Figure 2.14: 5σ-contrast (top) and mass (bottom) is plotted for NACO/SAM (L') as well as the observations with the best quality for IRDIS and IFS, respectively.

2.6 Summary and conclusions

Our new study of 51 Eri b provides new and improved spectra and photometry and allow us to revise the previous flux measurements and to explore new wavelength bands, especially the Y and K band. The photometric measurements obtained with SPHERE are $J = 19.74 \pm 0.71$ mag, $J3 = 18.86 \pm 0.26$ mag, $H2 = 18.56 \pm 0.28$ mag, and $K1 = 17.55 \pm 0.14$ mag. The broad wavelength coverage was made possible by combining data sets from the Y band up to the L' band, allowing us to take a comprehensive view of the planet for the first time; this showed how important thorough knowledge of all features is for understanding and modeling the system. Given separation of $\sim 0.5''$ of 51 Eri b, it is very suitable for high-contrast spectral observations and will become a benchmark object for current and future atmospheric models, especially once further spectra and photometry at longer wavelengths are obtained and all NIR features are mapped in detail. The models produced in this work provide strong predictions on the expected flux and shape of these features and validation or rejection of these predictions will further improve our understanding.
In this study, for the first time for SPHERE data, we combined the use of the recently developed ANDROMEDA algorithm to extract an unbiased planetary spectrum with a proper treatment of the spectral covariance and forced photometry with a state-of-the-art atmospheric model including clouds and varying metallicities combined with a detailed MCMC analysis.
We would like to advocate the use of forced photometry, together with the proper treatment of the noise covariance in the direct imaging community, to use all fluxes obtained at the known position of a planet; such data points can contain information even if the flux values obtained are below the detection limit, which is a quantity related to the probability of a previously unknown point source to be a real detection and cannot be directly applied as a cutoff threshold for an already known object. Furthermore, the usage of empirical covariances for IFS spectra can be used to take care of the relative weighting of spectral and photometric data, without having to rely on ad hoc weighting factors to artificially down-weight the spectral data with respect to independent photometric data. Our best fitting planet parameters for the cloudy models are $T_{\rm eff} = 760 \pm 20\,{\rm K}$, $R = 1.11^{+0.16}_{-0.14}\,{\rm R_J}$, $\log g = 4.26 \pm 0.25$ (cgs-units), highly super-solar metallicity $[{\rm Fe/H}] = 1.0 \pm 0.1$ dex, and $f_{sed} = 1.26^{+0.36}_{-0.29}$, indicating the presence of a vertically extended, optically thick cloud cover with small particle size. According to our models the planet seems to have an effective temperature above 700 K and thus sulfur chemistry, as discussed in Zahnle et al. (2016), probably does not play a major role. We note that the effective temperature is in general higher compared to Macintosh et al. (2015). The new parameters are suggestive of a higher mass for the planet than previously thought. The high surface gravity at a radius slightly bigger than that of Jupiter is consistent with a high-mass planet $M_{\rm gravity} = 9.1^{+4.9}_{-3.3}\,{\rm M_J}$, whereas the formation model that we consider is compatible with a wide range of masses depending on the initial conditions and does not strongly constrain the mass. Assuming the model atmosphere derived mass would mean that we can reject pure hot- and pure cold-start models. However, if 51 Eri b were in the brown dwarf mass regime we would expect to see a higher radius of 51 Eri b owing to deuterium burning, which makes this scenario unlikely. This discussion highlights the immense difficulty of precise mass determinations using direct imaging.

Tests performed for patchy cloud models showed that they do not improve the result significantly for the data used in comparison with a model of uniform cloud coverage and at this point do not seem to justify the increase in model complexity, which comes with the linear combination of clear and cloudy models. Further tests should be performed to explore patchiness in detail, for example using Bayesian evidence in the model comparison to account for overfitting and complexity. The consistency of the H-band flux over three independent measurements, speaks against strong variability in the J band, but to answer this question conclusively additional data is necessary. If there truly is variability, a more complex model, such as a patchy cloud model becomes necessary to explain the data. There is a strong need to consider super-solar metallicities in models of exoplanet atmospheres, beyond what is currently available. Beyond the characterization of the planet itself, neglecting super-solar metallicity will impact thermal evolution models of exoplanet and consequently limits placed on the occurrence rate of planets of certain mass through direct imaging. This impact will be especially noticeable if observations are performed in the K band.
The empirical comparison to other substellar objects confirmed the peculiarity of 51 Eri b. It is located in a unique place in the color-color and color-magnitude diagrams, which may be related to low-surface gravity and/or young age effects, but also shares common properties with other late-T dwarfs. The empirical characterization approach is limited by a lack of comparable objects from clusters and young moving groups with spectral type later than T5. Finally, no additional point sources were detected in the data. However, the SPHERE/IFS observations together with the NACO/SAM data provide strong constraints on the existence of additional objects in the system, rejecting $> 20\,\mathrm{M_J}$ companions at $\sim 2-4$ au and planets more massive than $4\,\mathrm{M_J}$ beyond 4.5 au and $2\,\mathrm{M_J}$ beyond 9 au.

Future IFS observations in the Y , K, and L bands, with existing and upcoming instruments (e.g., SPHERE, GPI, ALES, and CHARIS), as well as photometric measurements with JWST/NIRCAM and MIRI at mid-infrared wavelengths, can significantly improve the constraints on the atmospheric parameters. This will make 51 Eri b one of the planets with the best spectral coverage available and can thereby serve as a benchmark for atmospheric model development.

2.7 Newer publications on 51 Eri b and their relation to my work

After the publication of Samland et al. (2017) two more relevant studies on 51 Eri b have been published. One is the study by Rajan et al. (2017), who did a similar atmospheric analysis based on GPI and KECK data, and Maire et al. (2019) who published the astrometric analysis of 51 Eri b based on SPHERE data that was omitted from my work, because the baseline was too short to provide any significant improvement over previous results by De Rosa et al. (2015).

The comparison with results by Rajan et al. (2017) is especially important, as it appeared very shortly after Samland et al. (2017). Therefore, neither paper was published at the respective time when the papers were submitted, and they exist as independent works. Disentangling and discussing the compatible and incompatible results is important, as this has not yet been done in the published literature, and will lead to confusion for a reader coming at this topic.

In the next section I will attempt to shed light on, what at first glance, seem like contradictory results. I will also discuss a number of weaknesses that are present in the Rajan et al. (2017) study, that cast doubt on their conclusions, and argue why the conclusions of Samland et al. (2017) remain, at this point, unchallenged.

2.7.1 Similarities and differences between both studies

Let us first focus on the similarities between both studies. Both studies used high-contrast imaging data, with our study centering around SPHERE data, and the other study centering around GPI (Macintosh et al., 2008) data. SPHERE uses YJ- and YH-band spectra, i.e. spanning two or three bands simultaneously, whereas GPI takes J-, H-, K-band spectra in separate observations. Both used L-band data obtained with KECK, and Rajan et al. (2017) adds another point at M-band, while they do not cover the Y-band.

Both studies perform empirical comparisons to other sub-stellar objects, and compare the data to atmospheric models to study the atmospheric properties of 51 Eri b.

The J-band difference

Before we discuss the difference in analysis and interpretation of the spectra, first a note on the data underlying both papers. As described and discussed in Samland et al. (2017) (see Section 2.3.4), we obtain a J-band flux that is about 40% lower than in the original discovery paper Macintosh et al. (2015). Rajan et al. (2017) uses the same J- and H-band data and data reductions as the discovery paper, which means that they also use a higher J-band value which may impact some of their results.

The same arguments as to why we think the J-band flux in Samland et al. (2017) is more reliable still apply, i.e. the consistent result in J-band flux in both YJ and YH-spectra, and the consistency with the GPI H-band flux of the SPHERE YH-spectrum. There is no reason to assume the J-band to be biased by the reduction whereas the H-band is consistent, as the spectrum was taken simultaneously. On the other hand, the GPI data in H-band and J-band

were obtained on different nights (H-band: 2014-12-18, J-band: 2015-01-30). Additionally, the wings of the J-band GPI data do not return to low values in the absorption bands, indicating an overall offset or bias. Recently, we have independently reduced the GPI J-band data using ANDROMEDA, and the band integrated SNR achieved with ANDROMEDA is only about five, confirming that the conditions under which the data were taken were not very good (W. Brandner, priv. communication). We further note that, the higher J-band flux is incompatible with theoretical models: in Rajan et al. (2017) (Fig. 10, 13), even including the high J-band fluxes, their best-fit cloudy atmosphere models consistently show lower fluxes in the J-band, comparable to our extracted spectrum.

In the meantime we have obtained newer 51 Eri b observations in excellent and stable conditions, that are consistent with our published results, that will be published in a later work.

2.7.2 Analysis and interpretation of atmospheric properties

Both studies performed comparisons with other empirical objects and characterized the atmosphere of 51 Eri b. The empirical comparison comes to similar conclusions putting 51 Eri b closest to mid-to-late L-type objects, but concludes that only limited information can be gained by the comparison, because of the paucity of similar comparison objects.

However, the atmospheric modelling results come to very different conclusions. In the abstract of Rajan et al. (2017) they write:
"[...] The model fits suggest that 51 Eri b has an effective temperature ranging between 605 and 737 K, a solar metallicity, and a surface gravity of $log(g) = 3.5 - 4.0$ *dex, and the atmosphere requires a patchy cloud atmosphere to model the spectral energy distribution (SED). From the model atmospheres, we infer a luminosity for the planet of* -5.83 *to* -5.93 *(*$\log L/L_\odot$*), leaving 51 Eri b in the unique position of being one of the only directly imaged planets consistent with having formed via a cold-start scenario."*

On the other hand, our study concluded that the planet is ranging between 735 and 781 K, has a super-solar metallicity of $[\mathrm{M/H}] = 1.0 \pm 0.1$ dex, a surface gravity of $\log(g) = 4.0 - 4.5$ dex, an atmosphere that does not require patchy clouds, a luminosity for the planet of -5.34 to -5.47 ($\log L/L_\odot$), clearly ruling out the cold-start scenario.

These results appear to be in direct contradiction to each other. While some of these differences indeed need to be explored, some appear simply to be due to overemphasis in the abstract. Let us take a closer look at the methodological difference between the studies and general problems I see with how some results in Rajan et al. (2017) were obtained and/or interpreted.

2.7.3 Methodological problems in Rajan et al.

Model grids

The results of Rajan et al. (2017) are based on multiple model grids that are very heterogeneous in their scope and properties. They used three model grids, a) iron/silicate clouds, b) sulfide/salt clouds, c) cloudless models. The last – as also shown in our study – cannot reproduce the data. Their study therefore also focuses on the cloudy models.

The sulfide/salt grid used by Rajan et al. (2017) is an extended version of the Morley et al. (2012) grid that we also used as an test case in our study, and is similar to petitCODE with minor qualitative differences (see Section 2.4.2). When comparing the two studies, we should therefore focus on the sulfide/salt cloud model grids, that also provide the best-fitting spectra. They used existing and already available model grids, which comes with multiple drawbacks: 1) partially insufficient extent of parameter space coverage for 51 Eri b, 2) coarse grid resolution in parameter space, 3) missing models. In our study, we computed a new model grid using petitCODE, designed to cover all relevant parameters for 51 Eri b with sufficient and regularly spaced grid points. In some cases, in which fitting the models to the data showed that one or more parameters seemed to have their posterior likelihood peak outside of the grid, we extended the grid with new models, such that the best-fit parameters are firmly within our grid. The petitCODE model grids include only sulfide/salt species, because the iron/silicate opacities are expected to be unimportant, due to the pressure-temperature (PT) structure of the atmosphere of a planet at 51 Eri b's effective temperature, and the resulting equilibrium chemistry for cloud condensation. The iron/silicate species are expected to be significantly below the photosphere of the planet that we can probe (e.g., Morley et al., 2012).

A major problem with the iron/silicate cloud grid used in Rajan et al. (2017), is that it does not explore the surface gravity, metallicity, and cloud sedimentation parameter ($f_{\rm sed}$). These parameters where fixed to $\log(g) = 3.25$, solar metallicity, and $f_{\rm sed} = 2$. The only free parameters were the effective temperature, the cloud (or hole) fraction, and the radius. The choices of these fixed parameters were not further justified. Because of the degenerate nature of atmospheric models at the given wavelength coverage and spectral resolution of the data (multiple parameter combinations can result in similar spectra), the obtained fitted parameters cannot be interpreted. For these iron/silicate models the resulting best-fit radius furthermore is unrealistically small at $\sim 0.68\,\mathrm{R_J}$, highlighting that this assumed model, overall, does not capture the physics of the object appropriately. Furthermore, the $f_{\rm sed} = 2$ sets the settling/mixing ratios in the atmosphere to one fixed value, which is not further physically motivated. This allows iron/silicate clouds to be transported to higher altitude layers, than, e.g. self-consistent cloud models would predict.

Statistical inference of atmospheric parameters

In our study, the statistical inference of atmospheric parameters was based around using sufficiently finely spaced model grids, combined with an N-dimensional linear interpolator in model parameter space. The complete model parameter space was then sampled using a Markov chain Monte Carlo approach to obtain the posterior probability distribution of all parameters (Section 2.4.2). The N-dimensional linear interpolation was confirmed to be sufficient for the grid spacing by computing a number of atmospheric models between grid points and comparing them to the respective interpolated models, and confirming the resulting differences to be insignificant compared to the data uncertainties.

In Rajan et al. (2017), the model grid spacing, with the exception of surface gravity and temperature, is below what we considered justifiable for our study ($\Delta[\mathrm{M/H}] = 0.5$, $\Delta f_{\rm sed} = 1$). Furthermore, they only make comparisons to the grid points themselves without interpolating the models. Considering the coarse spacing of the grid and the *curse of dimensionality*, i.e. the fact that there is an exponential increase in the volume of parameter space associated

with adding extra dimensions, this leaves them without a reliable way to map from data space to model parameter space, as only a nearest neighbor approximation is used in this sparsely sampled space.

For complex physical models such as these atmospheric models that span a 4-dimensional parameter space (excluding radius), there exist a large quantity of possible parameter combinations, that do not fall on a grid point, that can result in very similar looking spectra (model degeneracies). Such an approach is prone to miss the real peak in the likelihood landscape.

While Rajan et al. (2017) talk about "posterior distributions" and "normalized posterior distributions" that they obtained for the model parameters, they do not perform Bayesian inference, and they are not sampling the model posterior distribution space, as such these distributions do not correspond to "posterior probability distributions". The detailed method they use is not explained in detail in the paper, but appears to be a form of bootstrapping using variations of the input data within their uncertainties with subsequent mapping to the best-fit corresponding grid point, as well as tallying up the number of best-fit correspondences in a normalized histogram. In the limit of an infinitely finely spaced model grid, this would be similar to a frequentist approach to getting a parameter distribution (flat priors in all parameters), where the data given the model, rather than the model given the data is explored.

Considering all of the above, it becomes clear that the parameters inferred using this approach are not *only* limited by the data quality, but also the coarseness and (ir)regularity of the model grid itself. The obtained distributions are not well resolved, not well-behaved (non-Gaussian), discretely sampled, undersampled, and hard to interpret using quantities like the median of the distribution and 16%-84%-percentile intervals.

To contrast this with our approach, we obtain close to Gaussian marginalized distributions and their respective covariance. While we test hundreds of thousands of model parameter combinations to explore the posterior distribution of the parameter space to create a nearly continuous distribution, the approach used in Rajan et al. (2017) is ultimately not reflecting the real statistics of the model parameters (posterior distribution) and results in a discrete, binned distribution, that is too coarse to reliably use summary statistics. Bootstrapping is a powerful tool if one has a good understanding of the model and wants to estimate the uncertainties of the data, but it is not an appropriate approach for model comparison.

It should further be pointed out, that the histograms shown in the corner plots in Rajan et al. (2017) (e.g., Fig. 11, 14) show the 16th, 50th, 84th percentile values of the respective parameters. In a valid posterior probability distribution, and if the distribution is Gaussian, this corresponds to the $1-\sigma$ uncertainty interval (16th–84th percentile interval) and the median value (50th percentile). However, the uncertainties quoted in the text of the paper are smaller than these values by about a factor of two.

They write: *"For the surface gravity and metallicity posterior distributions, we present the median values and error bar assuming a Gaussian distribution, though they may not be Gaussian."* The same seems to have been done for the temperature, because the percentile interval does not correspond to the 1-σ uncertainties quoted in the text. Statistically, this approach of fitting a Gaussian to a non-Gaussian distribution is ad-hoc and unjustified.

In the following I will look at the different parameters studied and their conclusions on them.

Temperature The temperature range of 605 to 737 K quoted in their abstract corresponds to the median temperature of the sulfide/salt cloud models (605 K), and the median value of the iron/silicate cloud models (737 K), and does not include the uncertainty range derived from the parameter distributions. In fact, both, the best-fitting spectra of the iron/silicate cloud grid (Fig. 10) at 900 K, as well as the two best-fitting spectra for the sulfide-salt cloud model (Fig. 13, right panel) at 725 and 775 K, have temperatures that are higher than this range, and are not located close to the median of the distribution. This suggests that the obtained "posteriors" do not accurately reflect the real distribution of the parameters in likelihood landscape. While it is true, that the best-fitting model (according to a chosen criterion, in Rajan et al. (2017), the maximum likelihood), does not **necessarily** have to coincide with the median of the posterior probability distribution (in a proper Bayesian inference, due to the impact of priors), in a non-Bayesian inference such as performed here, the peak should correspond to the maximum likelihood solution, which in turn should correspond to the median or mean, if the distribution is Gaussian. Proper care has to be taken in the interpretation here, especially because the median of the distribution refers to a value inside a parameter continuum, but the maximum likelihood solutions computed by Rajan et al. (2017) correspond to discrete grid points. Lastly, the temperatures of their maximum-likelihood grid point for the sulfide/salt cloud model are close to our result of $T_{\rm eff} = 760 \pm 20\,{\rm K}$.

Surface gravity The iron/silicate cloud grid does not vary in surface gravity, so we cannot draw conclusions about the surface gravity from this grid. Let us therefore look at the sulfide/salt grid. Here we see that the distribution shown in Fig. 14 of Rajan et al. (2017), does not seem to have its peak inside the bounds of the range of surface gravities explored by the grids. At the very least using any kind of Gaussian approximation for this distribution is unjustified. They write: *"The surface gravity PDF suggests that the planet has high surface gravity. However, 51 Eri b is clearly a low-mass companion, indicating that the data do not constrain the gravity. A prior might help constrain the distribution, but there are currently no physically motivated priors available for the surface gravity of young planets."*

How the authors conclude that "51 Eri b is clearly a low-mass companion" or what this means in terms of surface gravity is not further explained. We show in Fig. 2.13 that our values for the surface gravity ($\log g = 4.26 \pm 0.25$), combined with the posterior distribution of modelled planetary radii, still result in masses of $M_{\rm gravity} = 9.1^{+4.9}_{-3.3}\,{\rm M_J}$ which are in the planetary mass regime and compatible with warm start evolutionary models.

Metallicity In their abstract they write that the "model fits suggest (...) a solar metallicity", but they also write in their conclusion, that *"(...) The surface gravity and metallicity both appear to be unconstrained by the data, but empirical fits to young T-dwarfs suggest that the planet has lower surface gravity".* As such, they themselves conclude that the metallicity is unconstrained by their modelling. As they make no further argument in the paper, as to why solar metallicity is suggested by the models, it indicates a too strong wording in the abstract. At most they can conclude that the data is consistent with solar metallicity. However, their sulfide/salt cloud models (the only ones that include metallicity as a free parameter) shown in Fig. 14 of Rajan et al. (2017) would indicate a small preference for higher metallicities of $[{\rm M/H}] = 0.5$.

I would argue that the above phrase *"unconstrained by the data"* is furthermore not quite accurate in this context, as the parameters clearly *are* constrained by the data in our analysis (e.g., Fig 2.12), and the amount and quality of data is at least on a comparable level. In this case the limiting factor is not only the data uncertainty itself, but lies with the statistical inference approach which limits the detail in which the parameter space is explored (as outlined above).

Cloud coverage One major criticism that can be made is that Rajan et al. (2017) in their abtract and conclusion do not clearly separate which results were derived from which models, and how the respective result can be interpreted in the light of other models. For example, while most of their stated conclusions are based on the sulfide/salt cloud models, the claim that a patchy cloud model is required is based entirely on the fitted cloud (hole) fraction in the iron/silicate models, which otherwise does not fit the overall spectrum very well. There is the additional caveat described above, that fixed model parameters were not further justified or explored.

The motivation behind the iron/silicate cloud model is that the cloud composition might still be influenced by the deep iron/silicate condensate grains, as 51 Eri b is close to the L-T transition, such that a patchy model with emission from this deeper iron/silicate dominated part of the atmosphere, may be used to explain the spectrum.

While this is a valid idea, there are multiple potential problems. For one, there are physical reasons to expect that iron/silicate condensates are located below the photosphere at these temperatures (Morley et al., 2012). Furthermore, they include only iron/silicate species, they miss opacities from other species that are likely to occur at the temperatures of 51 Eri b, raising the question how representative of a real atmosphere this test case is. Lastly, there are the above concerns about fixing other parameters in the grid to values that are not further motivated (e.g., the surface gravity of 3.5 is not consistent with the surface gravity of the sulfide/salt clouds) and not exploring $f_{\rm sed}$.

In order to test the patchy cloud model hypothesis in our study, we fit a linear combination of cloudy and clear models, coupled using a cloud fraction parameter (Section 2.4.2). As we use a Bayesian framework, we could further enforce the prior condition that $T_{\rm cloudy} < T_{\rm clear}$, as holes in the clouds ("clear models") would correspond to deeper layers in the atmosphere and should therefore have a higher temperature. We also tied the surface gravity together for both models, because even the holes in cloud coverage should not expose layers so deep as to be sensitive to a significantly different surface gravity. The exploration of this combined parameter space of models yielded a best-fitting cloud coverage of $> 90\%$ (Fig. 2.28), and the best-fit spectrum was visually almost indistinguishable from the best-fit pure cloudy model (Fig. 2.24). We therefore concluded that there is not sufficient justification for a patchy cloud model, as the simpler model also explains the data. As a caveat, one could argue that a more realistic approach to this problem would be using only one overall PT-structure for the atmosphere and computing the radiative transport with and without clouds, instead of having two independent cloudy and cloud-free atmospheres with separate PT-structures. This approach should be explored in future work. However, this does not change our conclusion in so-far as a cloudy atmosphere alone is sufficient to explain the data as currently available.

Planetary radius While I already have argued why the distributions derived by Rajan et al. (2017) might be biased, it is still puzzling that the distributions for radii derived all peak below $1\,\mathrm{R_J}$, at odds with the our understanding of young massive planets. Furthermore the overall lower radii values are coupled with a distribution of temperatures that is significantly below temperatures we derive, whereas their best-fit spectrum (solar-metallicity) is about 120 K hotter than their median value and at roughly $1\,\mathrm{R_J}$ (Fig. 13).

Luminosity and formation In the abstract they wrote: *"From the model atmospheres, we infer a luminosity for the planet of* -5.83 *to* -5.93 *(*$\log L/L_\odot$*), leaving 51 Eri b in the unique position of being one of the only directly imaged planets consistent with having formed via a cold-start scenario"*, which quite dramatically differs from our inferred luminosity of -5.34 to -5.47. This also leads to the opposite conclusions, in which our results clearly rule out cold start models (see Fig. 2.13), and Rajan et al. (2017) conclude that 51 Eri b is special in that it is consistent with a cold-start scenario.

As discussed in Section 2.4.2 the luminosity of the planet can be derived from the radius and effective temperature of the planet according to $L/\mathrm{L_\odot} \sim (R/\mathrm{R_\odot})^2 \cdot (T/\mathrm{T_\odot})^4$, where $R_\odot = 9.728\,\mathrm{R_J}$ and $T_\odot = 5772$ K are the radius and effective temperature of the Sun. We computed the distribution of planetary luminosities, directly from the sampled posterior radius and temperature distribution of the planet. When interpreting the luminosity results by Rajan et al. (2017), therefore, the criticism regarding the determined temperatures and radii, as described above, applies. In addition, it should be noted that the luminosity distributions as shown in Fig. 11 and 14 of Rajan et al. (2017) are skewed, which means that the median value and intervals quoted are offset from the actual peak of the distribution towards lower luminosities. The peak of the distributions is closer to -5.6, which would be hard too justify with a cold-start model (see Fig. 2.13). A critical discourse about the suitability of the median for this distribution is missing.

Finally, the distributions of luminosities shown in Rajan et al. (2017) are skewed toward higher luminosities. However, the temperature and radii distributions both appear to be skewed towards smaller values. How both can be true at the same time is unclear.

I conclude that the statistically rigorously derived parameter distributions published in Samland et al. (2017) are still the most reliable to date.

2.8 Inclusion of new L- and M-band data in fit and Outlook

Lastly, I included the updated L-band, and novel M-band photometric point from Rajan et al. (2017) in my fit.

Figure 2.15 shows the updated best-fit plot, analogue to Fig. 2.11. The impact on the parameters is negligible. However, we can clearly see that the M-band data deviates strongly from the models by more than 4-σ, which poses two questions: 1) Is the M-band data reliable? 2) If the M-band data is reliable, what implications does this have for the models?

M-band data is notoriously difficult to reduce, because the high sky brightness at $> 4\,\mu\mathrm{m}$ and low transmissivity of the atmosphere at the M-band reduces the achievable SNR. The

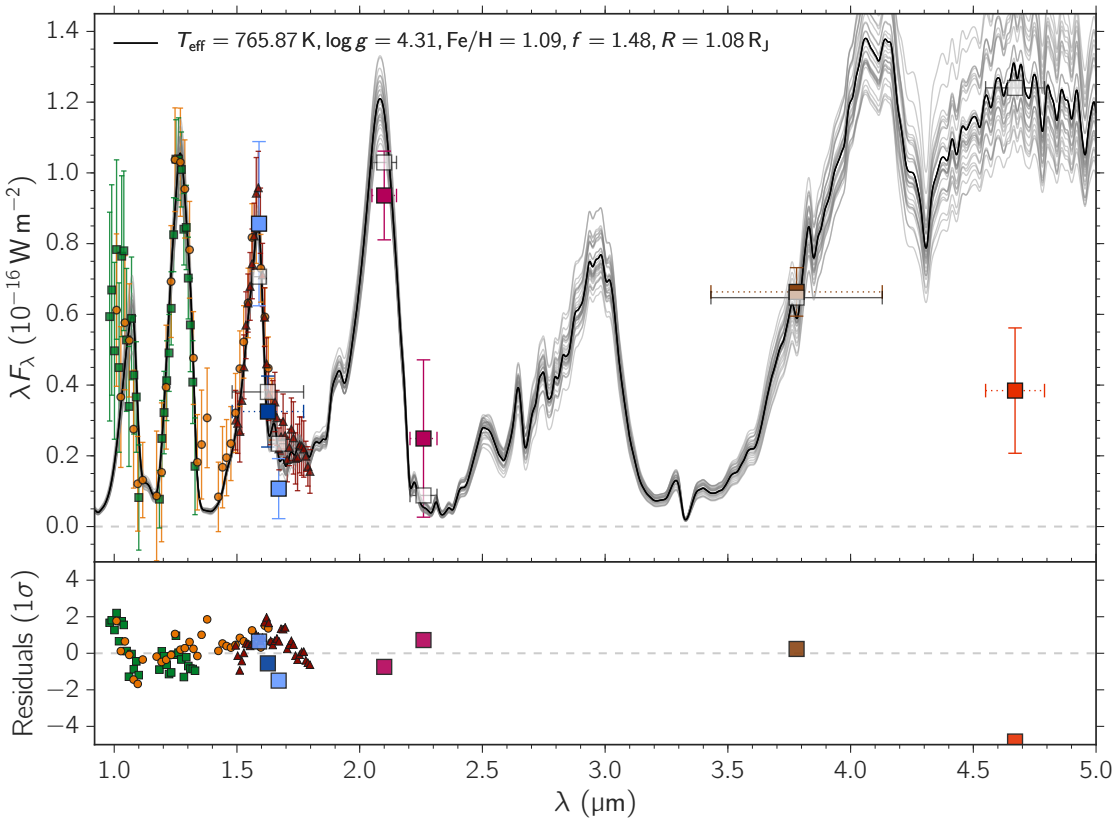


Figure 2.15: Same as Fig. 2.11, but including the updated L-band, and newly obtained M-band data from Rajan et al. (2017).

M-band observatinos obtained in Rajan et al. (2017) are the combined measurements of 4 half-nights of observations. To make sure that the measured flux is reliable, the data reduction should be repeated independently, as well as new data with another instrument should be taken. We have proposed for LBT/LMIRCam L- and M-band observations of 51 Eri b with the newly commissioned AGPM coronagraph, and hope to independently evaluate these result.

If the M-band data can be independently confirmed this would be extremely interesting, showing that our model understanding of 51 Eri b is incomplete. How do we remove flux from the M-band while still explaining the rest of the spectrum? The solution could be incorporating disequilibirum chemistry in the models. The M-band coincides with the CO bands of the atmosphere. Increasing the CO mass fraction in the atmosphere by a factor of ten, while keeping everything else the same in the atmosphere can be used to explain the spectrum including the low M-band flux (Paul Mollière, priv. communication). This would require an atmospheric quenching mechanism that prevents mixing of these species.

In a future study, in addition to the newer, higher quality unpublished SPHERE data, and using the novel data reductions approaches introduced in Chapter 4 and 5, I plan to use self-consistent disequilibrium models, or free-retrievals were the quenching pressure is included as a free-parameter, to revisit the properties and physics of 51 Eridani b.

2.8.1 Astrometry of 51 Eridani b

The original confirmation that 51 Eridani b is a bound object and not a background object was provided by De Rosa et al. (2015). However, the baseline was too short to draw any detailed conclusions about the orbit of the planet. Now, with a baseline of over three years of observations with SPHERE, Maire et al. (2019) published this data, which seems contains the first hint of orbital curvature. The detection of orbital curvature resolves degeneracies in the solutions and helps a lot in constraining the real orbit. This new analysis suggests an orbital period of 32^{+17}_{-9} yr (a semi-major axis of 12^{+4}_{-2} au), an inclination of 133^{+14}_{-7} deg, and eccentricity of $0.45^{+0.10}_{-0.15}$, and an argument of periastron passage of 87^{+34}_{-30} deg. It cannot clearly be said yet, whether the planet is currently accelerating or decelerating on its orbit, and further data is necessary. The orbital and stellar rotation axis are parallel to within 18°, and the planet orbit and star are either co- or counter-rotating.

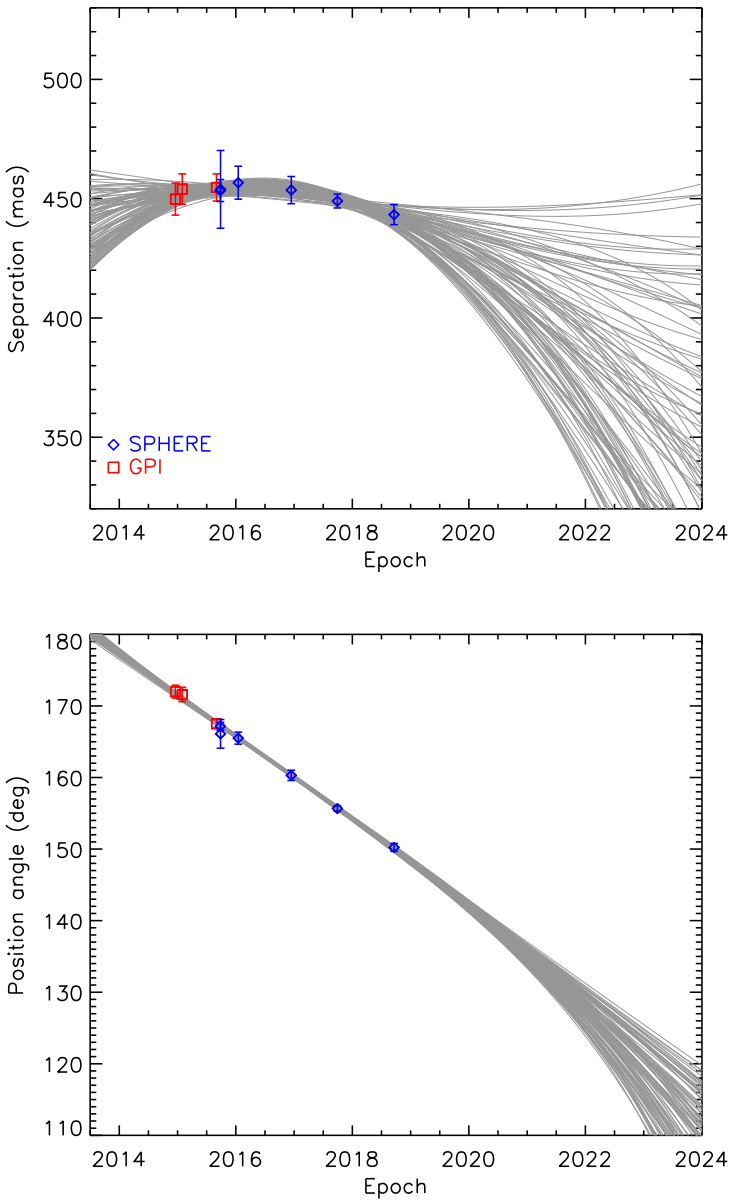


Figure 2.16: Predictions for the separation (*upper panel*) and position angle (*lower panel*) of 51 Eridani b for 100 randomly selected orbital solutions from the posterior distribution of orbital parameters. Source: Maire et al. (2019)

2.9 Appendix

2.9.1 Alternative reductions

Shown in Figure 2.17 are the extracted spectra with the different algorithms that were tested (see Sect. 2.3.1): ANDROMEDA (panel 1), PCA (panel 2), TLOCI (panel 3, Specal), and additionally PCA with simultaneous use of ADI and SDI reference images (panel 4). Additionally, panel 5 shows all reductions for the YJ data set and panel 6 all reductions for the YH data set. The ANDROMEDA, PCA, and TLOCI reductions all rely on the same cSDI prereduced data cube, whereas the simultaneous PCA ADI plus SDI reduction is completely independent based on the pipeline introduced in Mesa et al. (2015). The YH spectrum is fully consistent between all reduction methods and the only exception is the low quality of the H band extraction using PCA.
We notice more uncertainties in the absolute calibration of our YJ-spectral data, which changes depending on the exact algorithm used to reduce the data. The ANDROMEDA code yields compatible fluxes between the two observations, as does the independent reduction using spectral PCA with simultaneous ADI and SDI references. Whereas the PCA and TLOCI reductions using the same classical SDI prereduced frames, show higher peak fluxes in the J band. Although these reductions show comparable peak values to the GPI J-band spectrum, they do not follow the same spectral shape over the entirety of the GPI J-band spectrum.

2.9.2 Testing metallicity determination with benchmark brown dwarfs

The two T7.5 brown dwarfs Gl 570D and HD 3651B are considered to be benchmark objects, because they are on wide orbits around extensively studied K stars with known properties (Line et al., 2015, 2016). Having formed from the same cloud as their host stars, these brown dwarfs provide the opportunity to compare the derived parameters for the brown dwarfs, especially their composition, with their respective host star. Given the high metallicity inferred by our model for 51 Eri b, we want to make sure that our methodology and models are not biased toward obtaining high metallicity results. Below we compare the host star metallicities with the brown dwarf metallicities obtained with our self-consistent equilibrium *petitCODE* models. We further compare the parameters with the parameters derived by Line et al. (2015).
To make a comparison with Line et al. (2015) easier, we follow the same methodology, using every third pixel to avoid correlations between neighboring data points and the same additional free fit-parameter b in the likelihood function, which accounts for the underestimated uncertainties in the data by adding a constant 10^b term to the flux uncertainties. A flat prior is assumed for this parameter. All systematics in the absolute photometric calibration and distance are absorbed into the brown dwarf "radius"-parameter R because, with a flat prior, it allows the spectrum to freely float up and down. As also pointed out in Line et al. (2015), the absolute calibration is not necessarily reliable, so the radius should *not* be seen as a physical quantity, but rather as a data scaling parameter. On the positive side, this means that the determination of $T_{\rm eff}$, $\log g$, and [Fe/H] is independent of the absolute photometry and distance of the objects and is purely determined by the model shape and

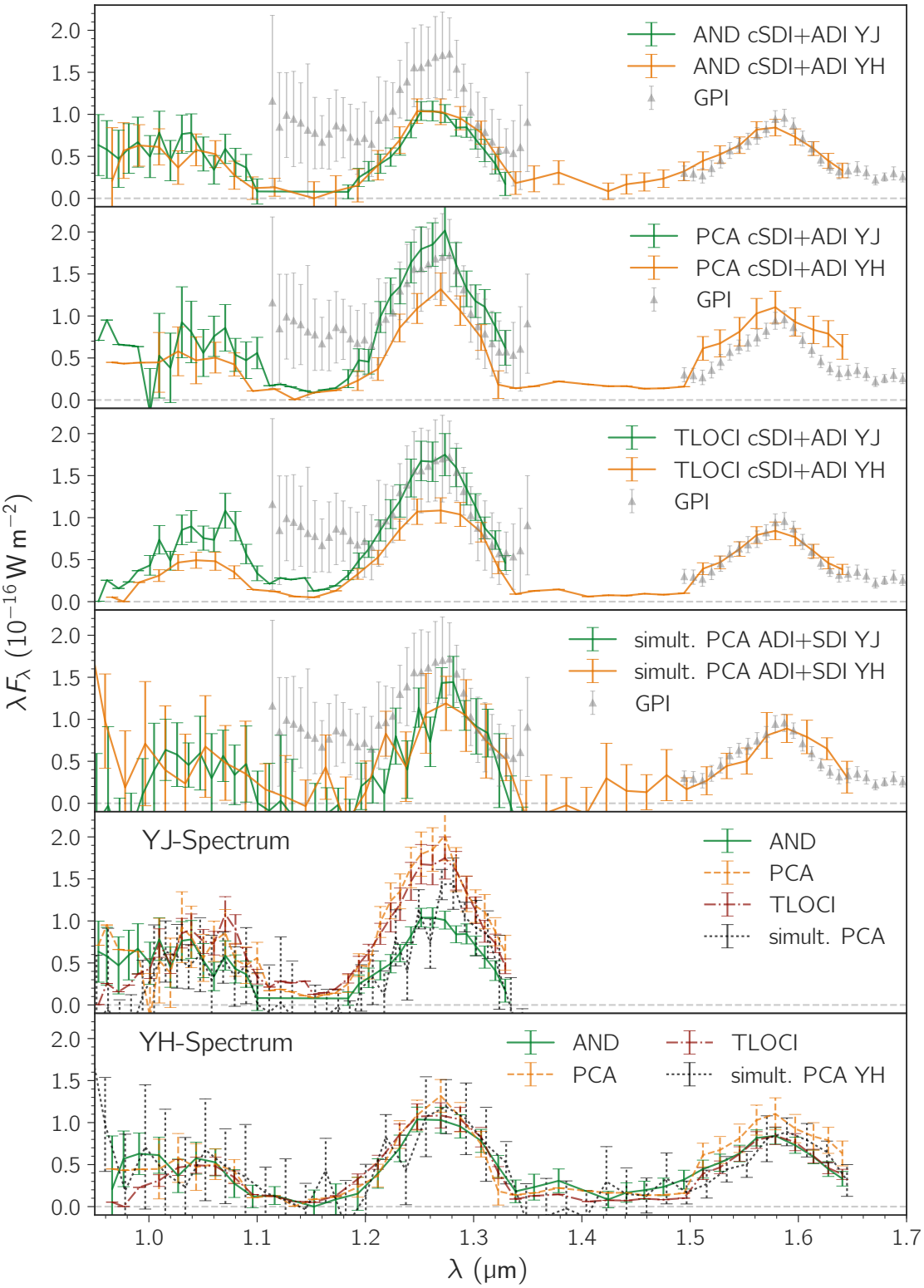


Figure 2.17: Shown in the panels are (from top to bottom) the extracted spectra with ANDROMEDA (panel 1), PCA (panel 2), TLOCI (panel 3), and simultaneous ADI+SDI with PCA (panel 4), all reductions for YJ (panel 5), and all reductions for YH (panel 6). The GPI spectra are plotted for comparison in the first 4 panels. The PCA and TLOCI pipeline used automatically exchange non-significant detections with upper limits, in which case no uncertainty is displayed on the data point.

relative strength of the features. A summary of the derived parameters for Gl 570D and HD 3651B for the Line et al. (2015) retrieval and the *petitCODE* clear models is shown in Table 2.6. The best fit *petitCODE* model spectra are shown in Fig. 2.18 and the respective posterior probability distributions in Fig. 2.19 and 2.20. Line et al. (2016) gives a summary of literature metallicity values for the host star Gl 570A and Line et al. (2015) derived the metallicity for HD 3651A. For Gl 570D, Saumon et al. (2006) gave an averaged metallicity data based on recent literature of $[\mathrm{Fe/H}] = 0.09 \pm 0.04$, and Casagrande et al. (2011) found a more recent value of 0.31 and -0.05 ± 0.17 from Line et al. (2015). The preponderance of evidence seems to suggest a slightly super-solar metallicity, whereas HD 3651A has a super-solar metallicity in the range of $[\mathrm{Fe/H}] = 0.18 \pm 0.07$ (Ramírez et al., 2013).
The metallicity we determined using the *petitCODE* models is within the given range of host star metallicities, showing that our model and fitting approach can be used to estimate metallicities reliably. Comparing the results to the free retrieval performed by Line et al. (2015), their metallicities fall on the lower end, whereas ours fall on the higher end of the metallicity range, which may reflect a difference in the free retrieval versus self-consistent model approach. For example, their retrieval requires an additional step to compute the chemical bulk metallicity from the retrieved abundances. On the other hand, our models assume a fixed solar C/O ratio. Both objects share very similar properties according to our fits, except that HD 3651B is more metal rich. This is consistent with our expectations as they are both classified as T7.5 dwarfs and the normalized spectra are virtually indistinguishable at the resolution of the SpeX instrument. Only the strength of the K-band flux, which is an indicator of metallicity mainly due to its sensitivity to collision-induced absorption (CIA) of H_2–H_2 and H_2–He pairs (see discussion in Sec. 2.4.2), is stronger in HD 3651B. Compared to Line et al. (2015) our models are about 50 K hotter in effective temperature. The derived surface gravity for Gl 570D is almost the same, whereas Line et al. (2015) arrive at a significantly higher surface gravity for HD 3651B.

2.9.3 Model spectra

Figure 2.21 shows a zoom in on the IFS spectra in the best fit cloudy *petitCODE* model. It can be seen that there is very good agreement between the model and the data in shape and amplitude, except for a systematic offset in the amplitude of the H band, which still exists and could not be modeled without negatively impacting the overall fit to the rest of the spectrum.
Figure 2.22 shows the the same plot as Fig. 2.11, but with the cloud-free *petitCODE* model. It is immediately apparent that the cloud-free model is incapable of explaining the long wavelengths of the spectrum (K1 and L' band), which results in unphysical parameters in temperature and radius and the lack of clouds in the model is compensated with extremely high metallicities.
Figure 2.23 of the Morley et al. (2012) shows that the lack of metallicity as a free-parameter (only solar metallicity was available) also skews the overall parameters, especially in order to fit the K1 peak. Since high metallicities are not allowed, the f_{sed} parameter increases in an attempt to compensate, again the resulting physical parameters are unreliable. All of this shows that a complex model that allows coverage of at least the basic physics (e.g., a cloud model with free parameters and non-solar metallicity), is a bare minimum to model

Table 2.6: Benchmark brown dwarfs

Object	$T_{\rm eff}$	$\log g$	Chemically derived bulk [Fe/H]	$[{\rm Fe/H}]_{\rm BD}$ - $[{\rm Fe/H}]_{\rm host}$ [a]
petitCODE clear model				
Gl 570D	769^{+14}_{-13}	4.67 ± 0.04	0.11 ± 0.04	$+0.16 \pm 0.18$
HD 3651B	783^{+13}_{-12}	4.64 ± 0.04	0.25 ± 0.04	$+0.07 \pm 0.08$
Line et al. 2015 retrieval				
Gl 570D	714^{+20}_{-23}	$4.76^{+0.27}_{-0.28}$	-0.15	-0.10 ± 0.17
HD 3651B	726^{+22}_{-21}	$5.12^{+0.09}_{-0.17}$	$+0.08$	-0.10 ± 0.07

Summary of modeling results for Gl 570D and HD 3651B using *petitCODE* clear models and comparison to atmospheric retrieval results by Line et al. (2015). [(a)] Difference of best fit model to middle of metallicity range of host star. Uncertainty includes the width of host star metallicity range in all cases, as well as uncertainty for brown dwarf metallicity in the *petitCODE* case. Host star metallicities used here: Gl 570D: -0.22 – 0.12 or -0.05 ± 0.17 (Line et al., 2015); HD 3651B: 0.11 – 0.25 or 0.18 ± 0.07 (Ramírez et al., 2013).

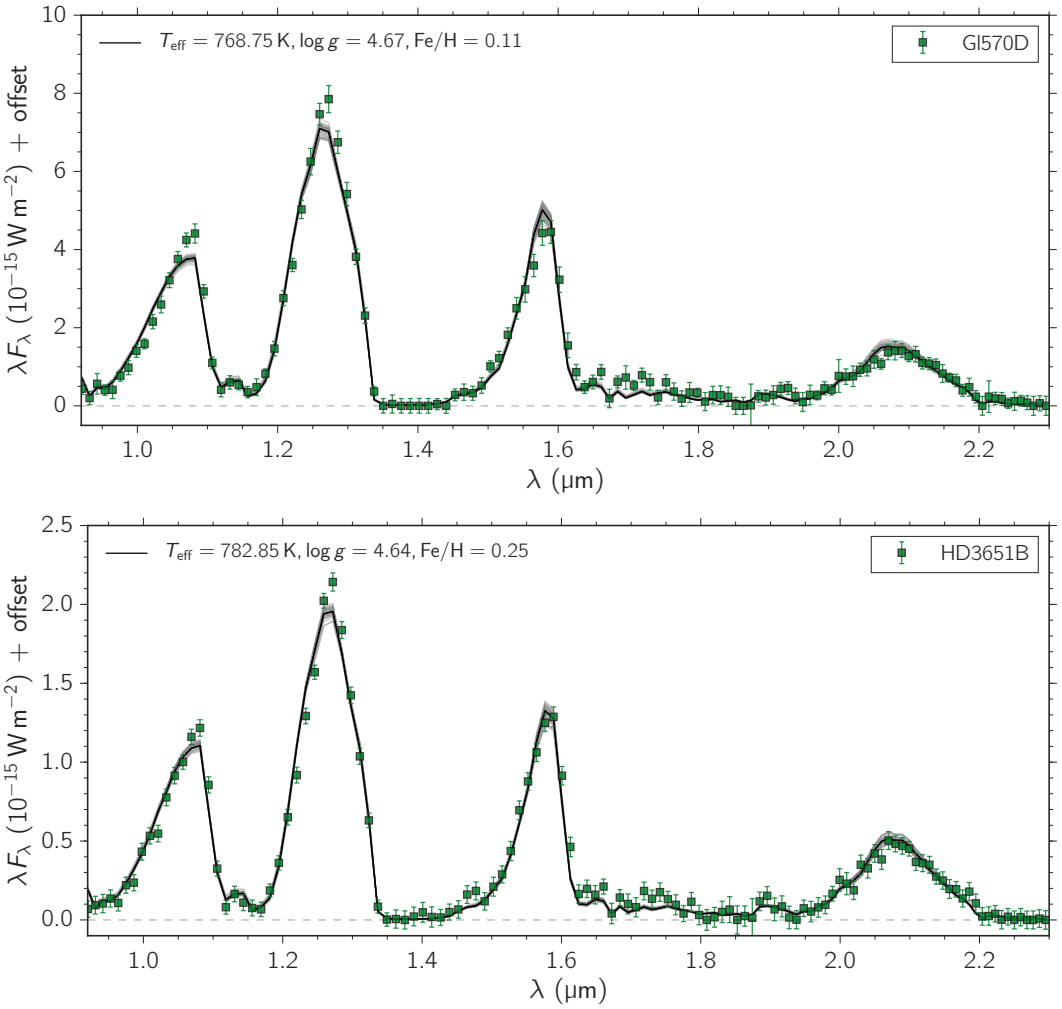


Figure 2.18: Upper panel shows best fit *petitCODE* clear spectrum for Gl 570D; the lower panel shows the same for HD 3651B. The overplotted gray lines represent the model scatter with spectra generated from randomly drawn samples of the posterior parameter distribution. Error bars plotted include the best fit value of the b parameter, correcting for the underestimated data uncertainty.

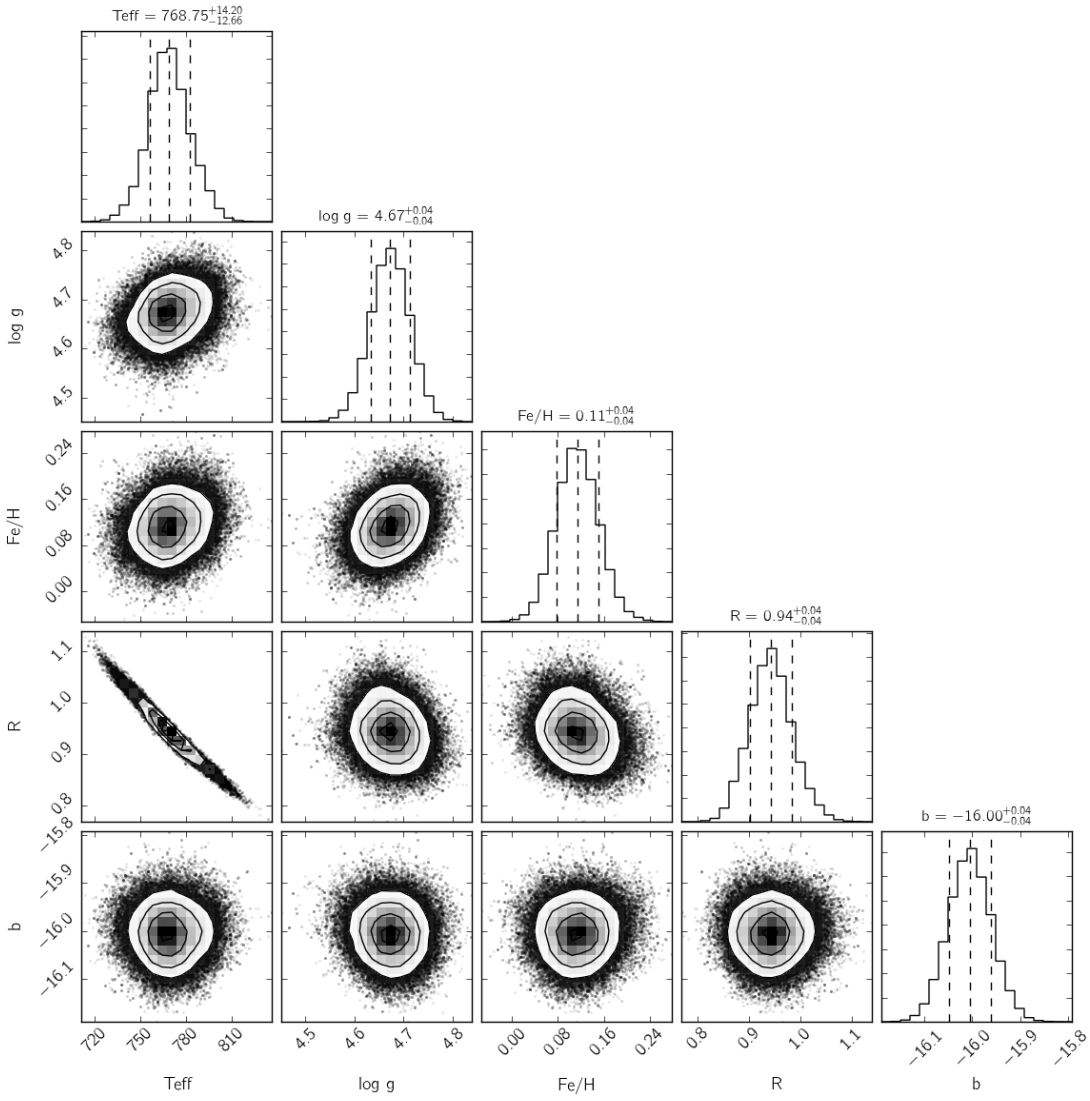


Figure 2.19: Corner plot showing the posterior probability distribution of the clear *petitCODE* fitted to Gl 570 D, including a further scale parameter b as an additive term to the flux uncertainty.

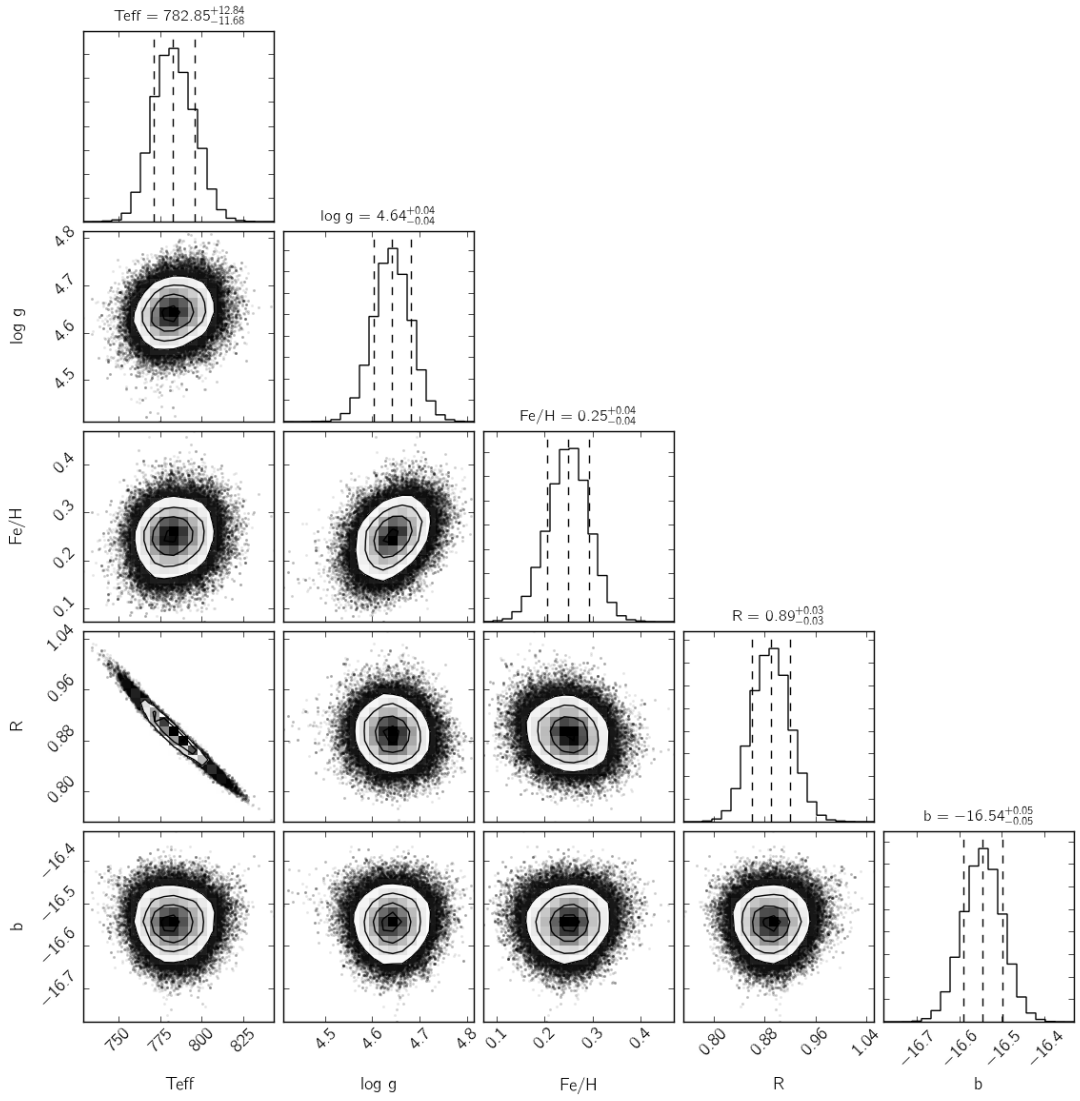


Figure 2.20: Corner plot showing the posterior probability distribution of the clear *petitCODE* fitted to HD 3651 B, including a further scale parameter b as an additive term to the flux uncertainty.

these cold giant planets.
Figure 2.24 shows the spectrum resulting from the patchy cloud model introduced in Sec. 2.4.2. The spectrum is almost indistinguishable from a pure cloudy model and does not improve the result significantly.

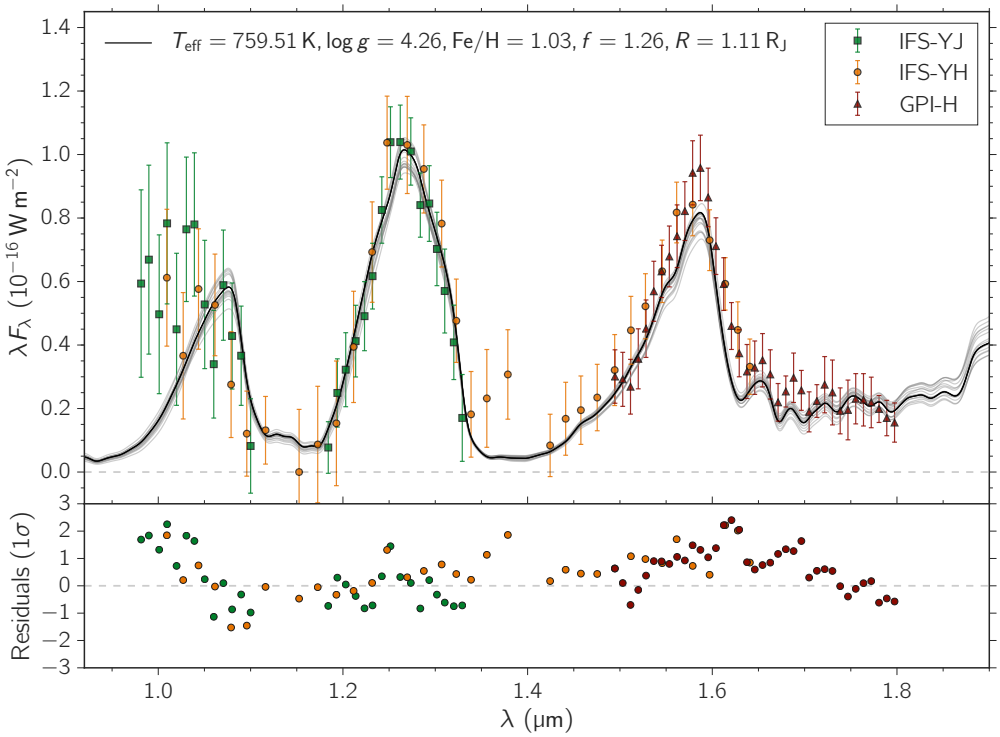


Figure 2.21: Same plot as Fig. 2.11, but zooming in on the wavelength range covered by spectral data.

2.9.4 Corner plots

The corner plots for additional models and data sets are shown. Figure 2.25 shows the parameter distribution for the cloudy *petitCODE* model in the case in which we exclude the Y-band data completely. We see that the Y band does not significantly constrain the models. Figure 2.26 shows the corner plot for the clear *petitCODE* model and Figure 2.27 for the Morley et al. (2012) model. As discussed above, both lead to skewed results, because important physics is missing. Figure 2.28 shows the corner plot for the patchy cloud model, a linear combination of cloudy and cloud-free models, which share the same parameters except for temperature and are linked by a cloud fraction parameter. Cloud fractions are very high and, as pointed out above, the resulting spectrum does not improve the fit significantly.
As an additional experiment, Figure 2.29 shows the posterior distribution for the cloudy model grid when only data from Macintosh et al. (2015) are used. This fit does not include the covariance matrices and should reduce roughly to a straightforward fit as the discovery paper described (except that the model can vary in metallicity). With the $\sim 40\%$ higher J-band flux and missing K band, we retrieve a very low surface gravity (same as the discovery paper), but significantly higher temperature outside of our model grid. In the a patchy cloud

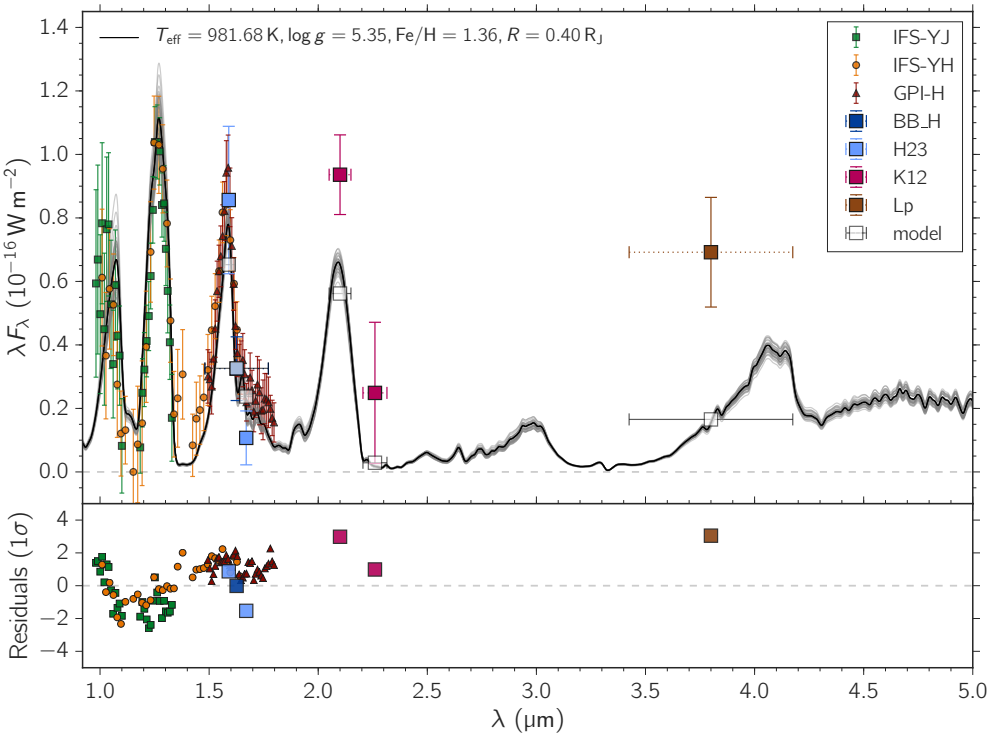


Figure 2.22: Same plot as Fig. 2.11, but using the clear model.

model of the original paper a higher J-band contribution can come from a clear model, but this is more difficult to explain in a pure cloudy model.

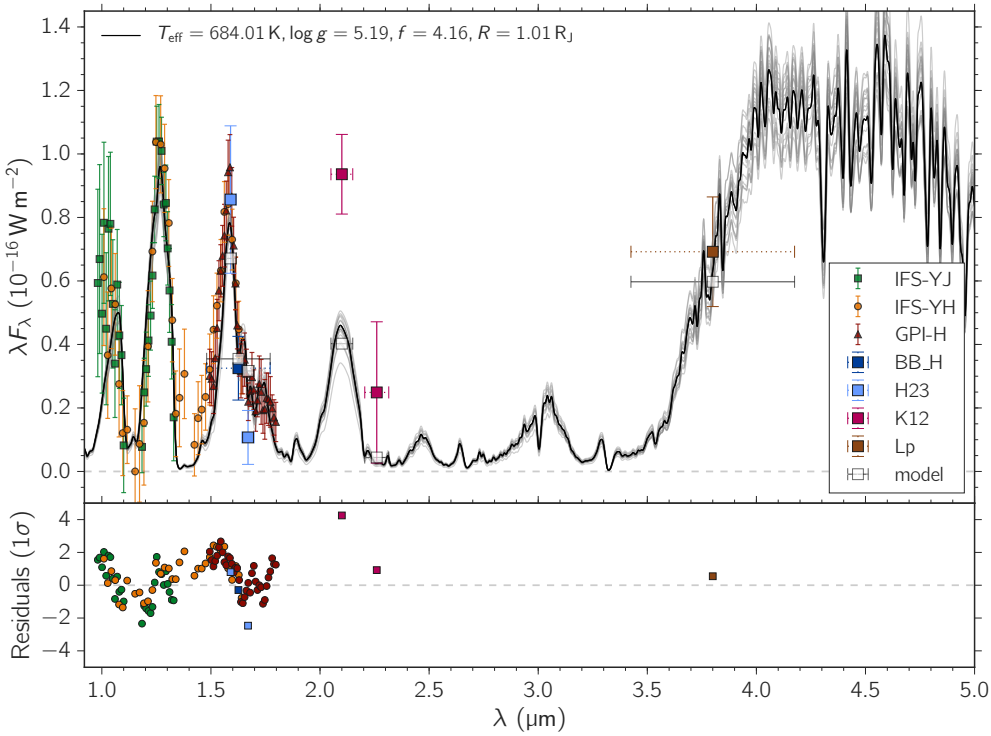


Figure 2.23: Same plot as Fig. 2.11, but using the Morley et al. (2012) model.

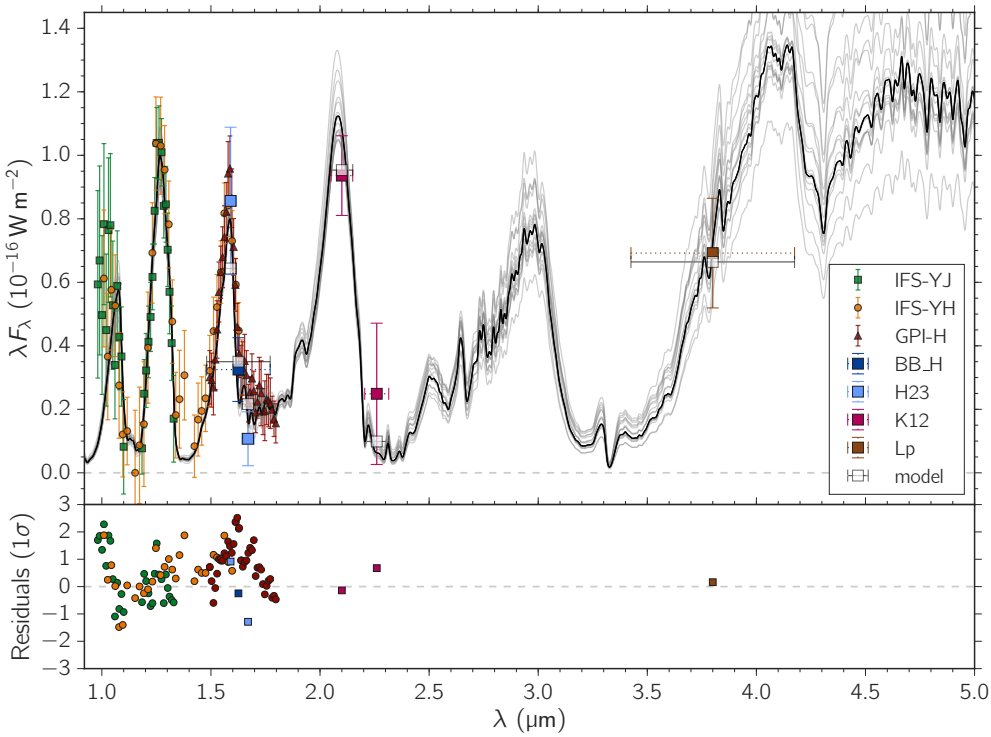


Figure 2.24: Same plot as Fig. 2.11, but using the patchy-cloud model described in Sec. 2.4.2. Best fitting parameters are: $T_{cloud} = 750 \pm 25$ K, $T_{clear} = 815^{+70}_{-40}$ K, $CF = 0.9 \pm 0.05$, $\log g = 4.5 \pm 0.3$, [Fe/H] $= 1.25$, $f_{\rm sed} = 1.10 \pm 0.15$, $R = 1.10 \pm 0.15$ R_J

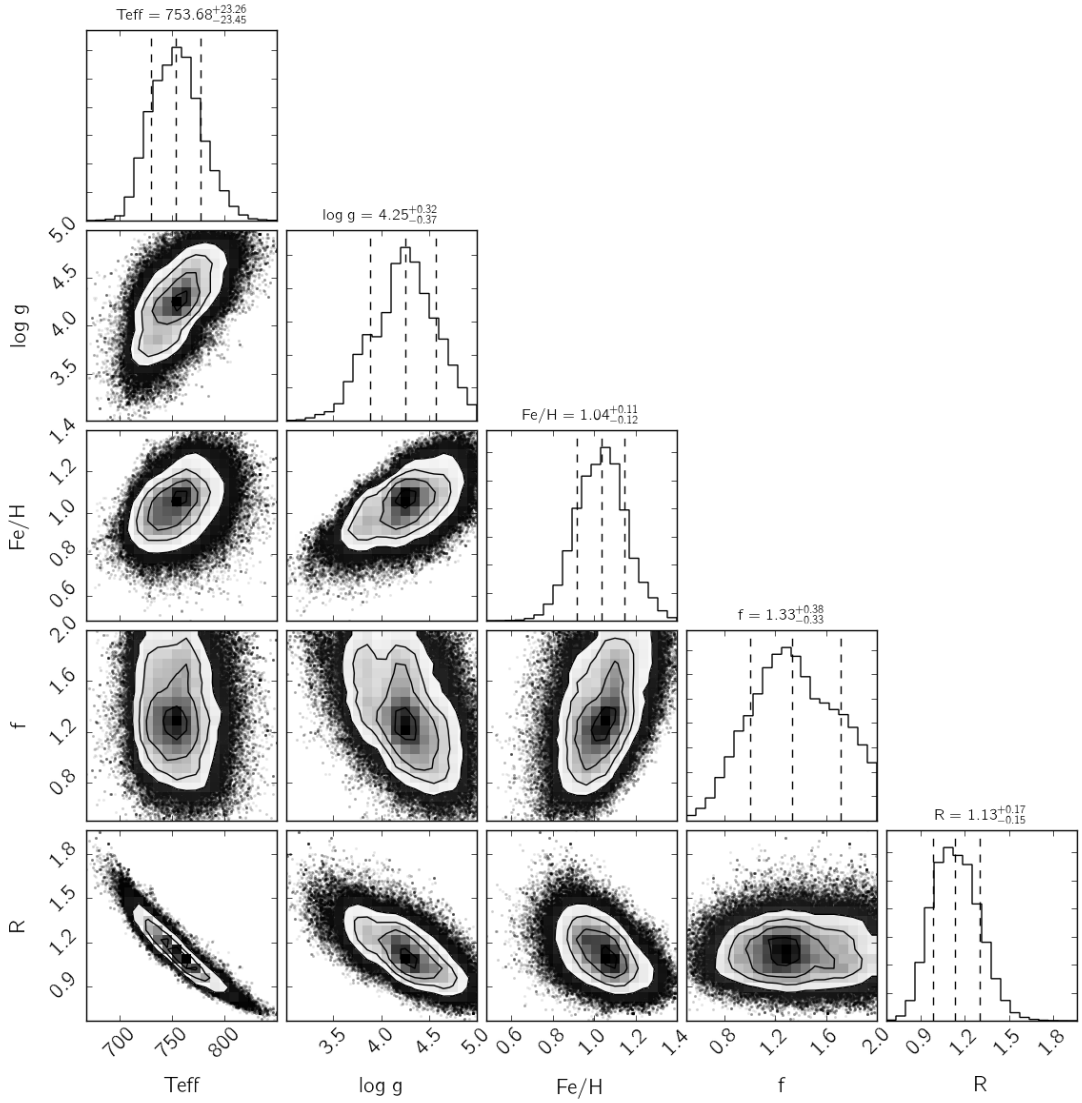


Figure 2.25: The same as Figure 2.12, but excluding the Y-band data. Corner plot showing the posterior probability distribution of the cloudy *petitCODE* grid with respect to each of its parameter pair as well as the marginalized distribution for each parameters. The uncertainties are given as 16% to 84% percentiles as commonly done for multivariate MCMC results.

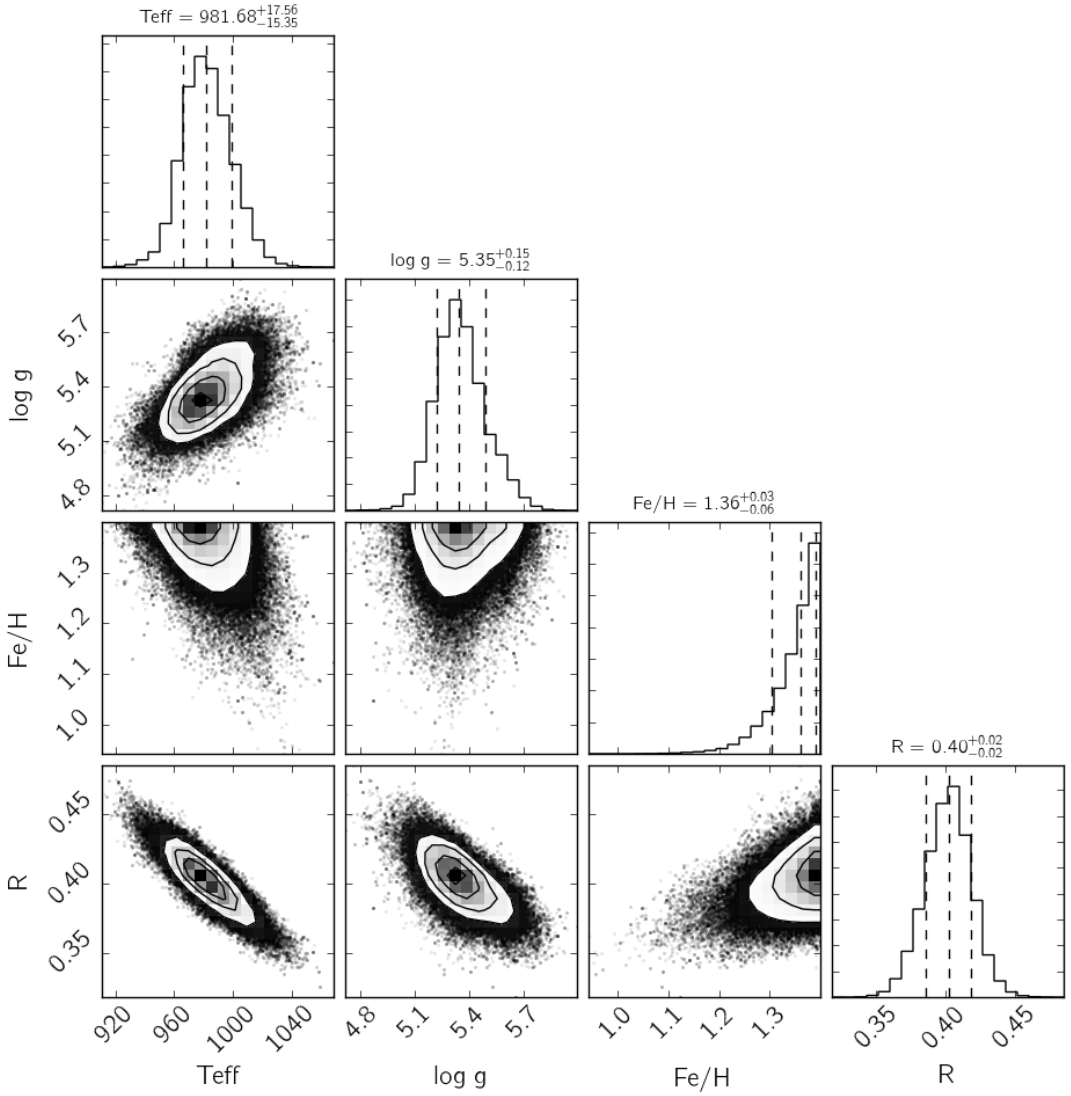


Figure 2.26: Corner plot showing the posterior probability distribution of the clear *petitCODE* grid, with respect to each of its parameter pair as well as the marginalized distribution for each parameters. The uncertainties are given as 16% to 84% percentiles as commonly done for multivariate MCMC results. Note that a clear model atmosphere requires a small radius, which speaks against this model.

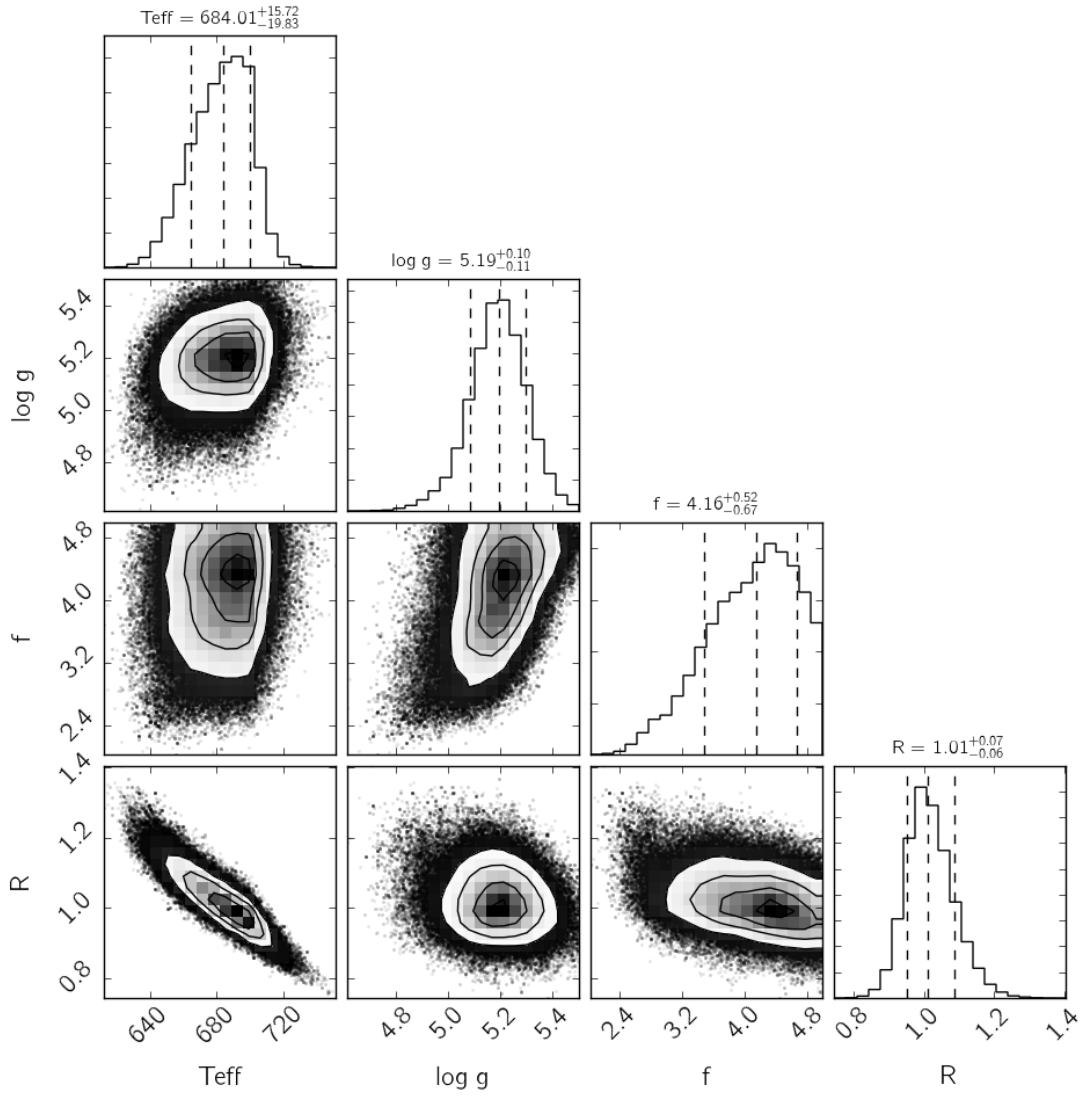


Figure 2.27: Corner plot showing the posterior probability distribution of the Morley et al. (2012) grid, with respect to each of its parameter pair as well as the marginalized distribution for each parameters. The uncertainties are given as 16% to 84% percentiles as commonly done for multivariate MCMC results.

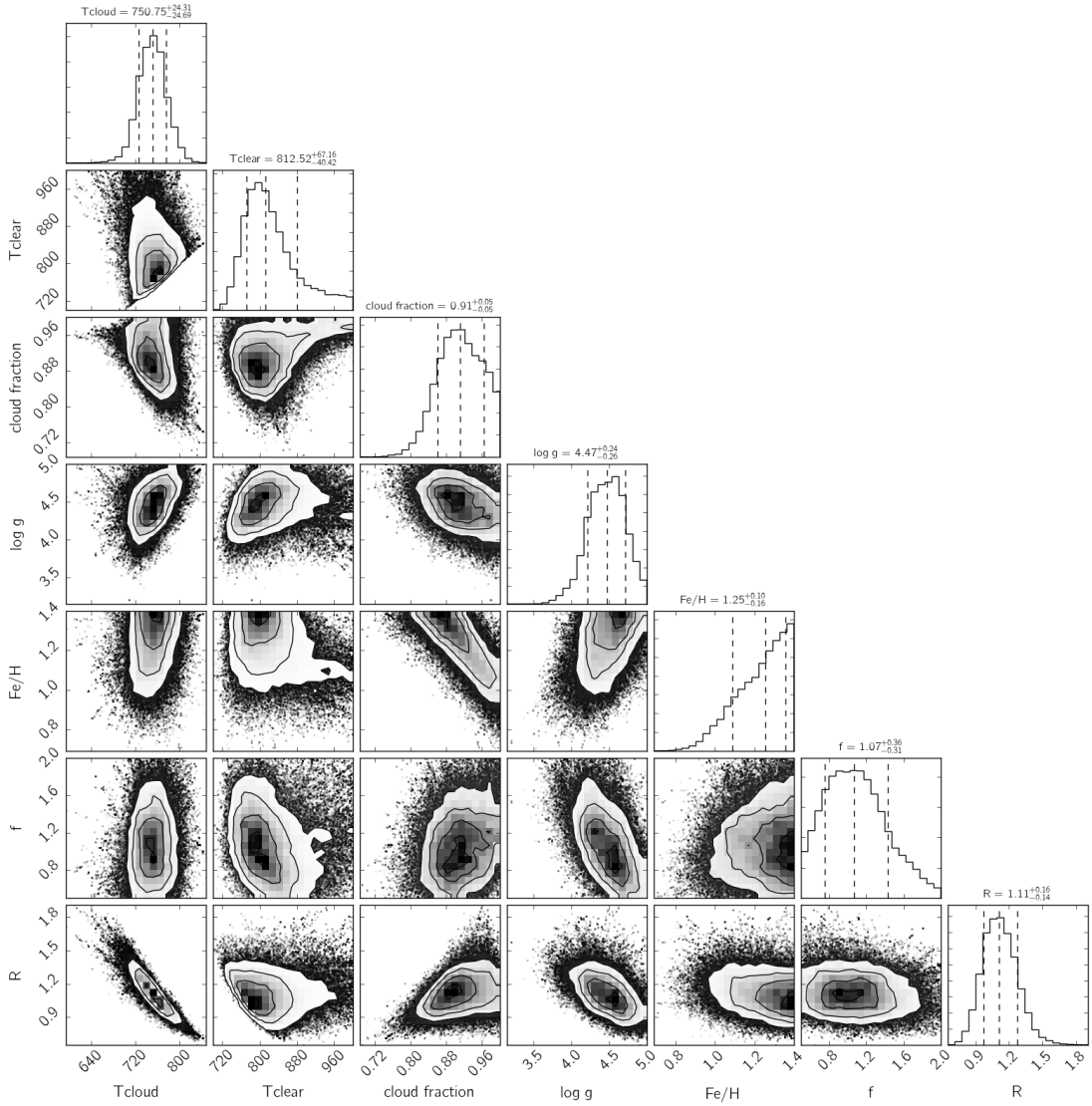


Figure 2.28: Corner plot showing the posterior probability distribution of the patchy cloud model described in Sec. 2.4.2, with respect to each of its parameter pair and the marginalized distribution for each parameters. The uncertainties are given as 16% to 84% percentiles as is common for multivariate MCMC results.

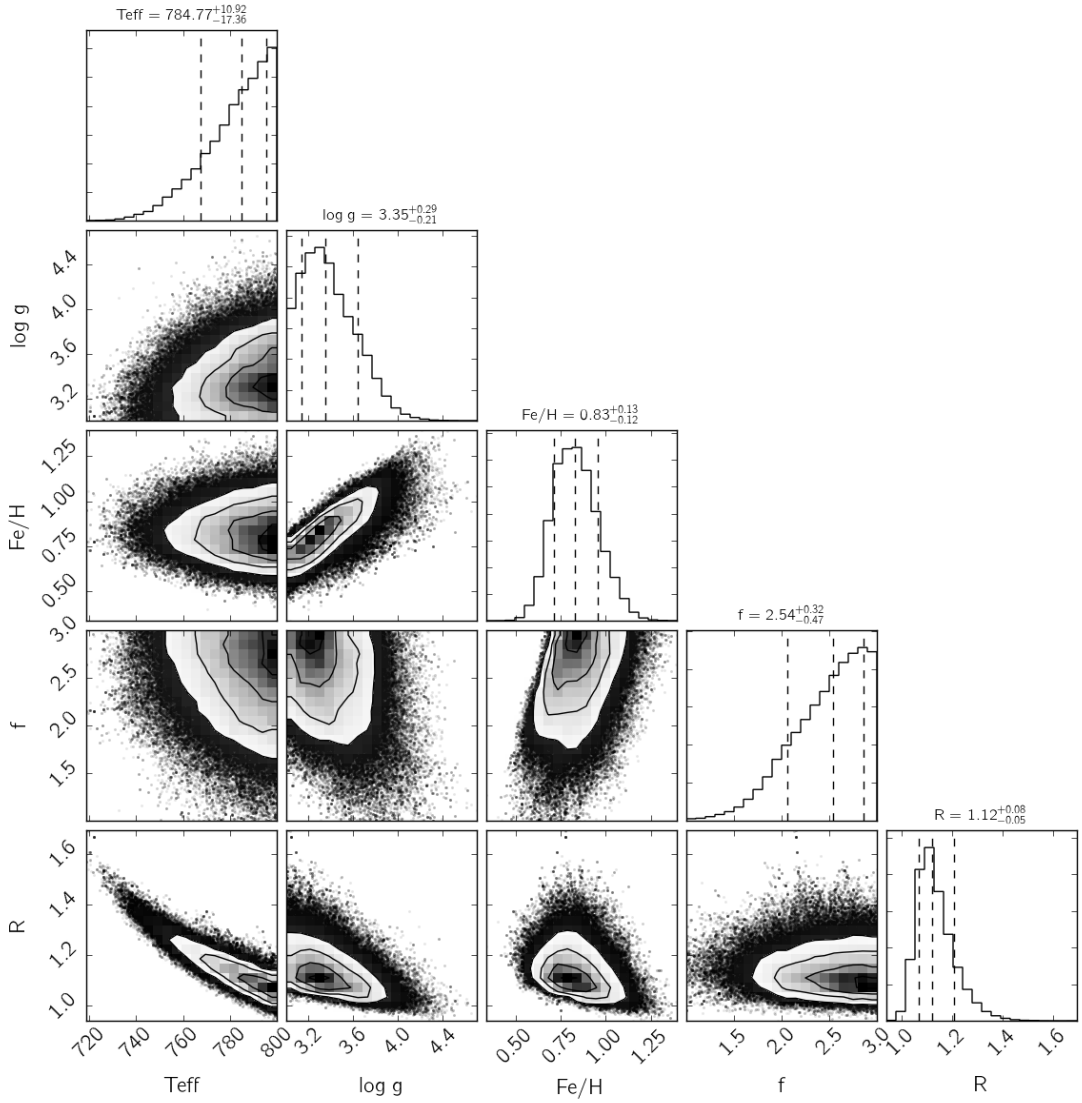


Figure 2.29: Corner plot showing the posterior probability distribution using only data published in Macintosh et al. (2015) (without taking the covariance into account) with respect to each of its parameter pairs as well as the marginalized distribution for each parameters. The uncertainties are given as 16% to 84% percentiles as is common for multivariate MCMC results.

3 Contributions to the atmospheric characterization of directly imaged planets

This chapter summarizes my contributions to the atmospheric characterization of directly imaged planets published in a number of refereed articles: "Orbital and atmospheric characterization of the planet within the gap of the PDS 70 transition disk" (Müller et al., 2018); "Spectral and orbital characterisation of the directly imaged giant planet HIP 65426 b" (Cheetham et al., 2019); "The GJ 504 system revisited: Combining interferometric, radial velocity, and high contrast imaging data" (Bonnefoy et al., 2018).

3.1 Introduction

The software package for statistical inference of atmospheric parameters I developed for the work published described in Chapter 2 Samland et al. (2017) was deliberately written in a flexible way, such that any atmospheric model grid and any spectrophotometric data could be used. I subsequently named this package BACON (Bayesian Atmospheric CharacterizatiON) and have used it to contribute to the atmospheric characterization of a number of direct imaging studies of other exoplanets: the first planet discovered inside the gap of a transition disk PDS 70 b, the first planet discovered with the SPHERE instrument HIP 65426 b, and a new analysis of GJ 504 b, which depending on its age could either be a planet or brown dwarf. Including 51 Eri b, these analyses cover the L-T spectral types of sub-stellar objects with two representatives each.

Both PDS 70 b and HIP 65426 b are hot young planets with mid L-type spectra, whereas 51 Eri b and GJ 504 b are cool objects of mid-to-late T-type spectra. The first two are dominated by dusty, probably silicate rich atmospheres, whereas the latter two are show atmospheres with strong methane and water absorption features, as well as clouds that are dominated by sulfide/salt species.

Table 3.1 gives an overview of the best-fitting physical parameter ranges derived using BACON. The mass quoted is the respective "hot-start" mass, derived from evolutionary models. This denotes the lowest possible mass range for these objects. The real masses are likely higher, but hot-start masses are the easiest to directly compare between different studies, as they only depend on the evolutionary model (in this case all studies used similar "dusty" atmospheres for the pure hot-start case), the age, and the bolometric luminosity of the planet. As such it is relatively safe to assume that 51 Eridani b has the lowest mass, followed by PDS 70 b, HIP 65426 b, and GJ 504 b (if the stellar age is old). PDS 70 b is

the youngest at 5.4 Myr, located in the gap of the transition disk around the star, followed by HIP 65426 b, 51 Eridani b, and GJ 504 b.

Table 3.1: Overview of best fitting physical parameter range.

Name	Temp. (K)	Surf. gravity ($\log_{10}$(cgs))	[M/H] (dex)	Radius (R_J)	Luminosity ($\log_{10}(L_{bol}/L_\odot)$)	Mass (M_J)
PDS 70 b	1000–1600	$\lesssim 3.5$	> 0	1.4–3.7	-3.86±0.2	5–9
HIP 65426 b	1620±10	~ 3.8	~ 0	~ 1.2	-4.05±0.03	8–11
51 Eridani b	760± 25	4.25±25	> 0	~ 1.1	-5.4±0.07	3±1
GJ 504 b	600±30	~ 3.7	≥ 0	~ 1	-6.13±0.03	3–30

Name	NIR SpT	Sep. (")	Sep. (au)	Primary Mass ($M_\odot$)	Age (Myr)
PDS 70 b	mid L	0.2	22	0.76	5.4±1
HIP 65426 b	mid L	0.83	120	1.96	14±4
51 Eridani b	mid T	0.45	12	1.75	23±3
GJ 504 b	T9	2.5	44	1.16	100–6500

3.2 PDS 70 b

PDS 70 b has been newly discovered by SPHERE (Keppler et al., 2018). It is the first bona-fide detection of a giant planet inside the gap of the transition disk. As such it presents an exciting view, as the first confirmed exoplanet that is still in the process of formation and showing active accretion (Wagner et al., 2018). On the other hand, such confounding factors as a possible envelope, circumplanetary disk (CPD), accretion shock, pose immense difficulties for the actual characterization of the object and its atmosphere.

3.2.1 Atmospheric modeling of PDS 70 b

In order to learn more about the atmosphere of PDS 70 b, I made use of new grids computed with the self-consistent 1D radiative-convective equilibrium tool petitCODE (Mollière et al., 2015a, 2017), by my collaborators. This resulted in three different grids of self-luminous cloudy planetary model atmospheres (see Table 3.2). For a detailed model description, please refer to Section 2.4.2. As the atmosphere of PDS 70 b is hotter than 51 Eridani b these grids mainly differ in their treatment of clouds: *petitCODE*(1) does not consider scattering and includes only Mg_2SiO_4 cloud opacities; *petitCODE*(2) adds scattering; *petitCODE*(3) contains four more cloud species including iron (Na_2S, KCl, Mg_2SiO_4, Fe). Additionally, I also used the same cloud-free *petitCODE* model grid as in Chapter 2 (Samland et al., 2017), here called *petitCODE*(0) and the public PHOENIX BT-Settl grid (Allard, 2014; Baraffe et al., 2015b).

In order to compare the data to the petitCODE models I used the same tools (BACON) as described in Chapter 2, that uses the python MCMC code *emcee* (Foreman-Mackey et al., 2013) on N-dimensional model grids linearly interpolated at each evaluation to map the

posterior probability distributions of the model parameters. Again, I assume a Gaussian likelihood function and take into account the spectral correlation of the IFS spectra (Greco & Brandt, 2016).

For an additional independent confirmation of the results obtained using petitCODE, I also used cloudy models from the Exo-REM code. The models and corresponding simulations are described in Charnay et al. (2018). Exo-REM assumes non-equilibrium chemistry, and silicate and iron clouds. For the model grid Exo-REM(1) the cloud particles are fixed at $20\,\mu$m and the distribution takes into account vertical mixing (with a parametrized K_{zz}) and sedimentation. The Exo-REM(2) model uses a cloud distribution with a fixed sedimentation parameter $f_{sed} = 1$ as in the model by Ackerman & Marley (2001) and petitCODE. Table 3.3 provides a compilation of the best-fit values and Fig. 3.1 shows the respective spectra. The values quoted correspond to the peak of the respective marginalized posterior probability distribution. The cloud-free models fail to represent the data and result in unphysical parameters. In contrast, the cloudy models provide a much better representation of the data. The results obtained by the petitCODE and Exo-REM models are consistent with each other. However, because of higher cloud opacities in the Exo-REM(2) models the $\log g$ values are less constrained and the water feature at $1.4\,\mu$m is less pronounced. Therefore, the resulting spectrum is closer to a blackbody and the resulting mass is less constrained. All these models indicate a relatively low temperature and surface gravity, but in some cases unrealistically large radii.

Evolutionary models predict radii smaller than $2\,R_{\rm J}$ for planetary-mass objects (Mordasini et al., 2017). The radius can be pushed toward lower values if cloud opacities are removed, for example by removing iron (*petitCODE*(2)). However, a direct comparison for the same model parameters shows that this effect is very small. In *petitCODE*(1) this is shown in an exaggerated way by artificially removing scattering from the models, which leads to a significant reduction in radius. In general, a wide range of models are compatible with the current data. The parts of the spectrum most suitable for ruling out models are the possible water absorption feature at $1.4\,\mu$m, and the spectral behavior at longer wavelengths (K to L' band). Given the low signal-to-noise ratio in the water absorption feature and the large uncertainties in the L' flux, it is very challenging to draw detailed physical conclusions about the nature of the object. It has to be emphasized that other possible explanations for the larger than expected radii from evolutionary models include the recent accretion of material, additional reddening by circumplanetary material, and significant flux contributions from a potential circumplanetary disk. The third possibility is especially interesting in the light of possible features in the reduced images that could present spiral arm structures close to the planet, as well as the detection of H_α emission at the location of the planet (Wagner et al., 2018).

Table 3.2: Model grids used as input for MCMC exploration. The radius of the planet was included as an additional analytic fit-parameter regardless of the model, ranging from $0.1\,R_{\rm J}$ to $5\,R_{\rm J}$.

Model	$T_{\rm eff}$ (K)	ΔT (K)	$\log g$ ($\log_{10}$(cgs))	$\Delta\log g$ ($\log_{10}$(cgs))	[M/H] (dex)	Δ[M/H] (dex)	$f_{\rm sed}$	$\Delta f_{\rm sed}$	Remarks
BT-Settl	1200 – 3000	100	3.0 – 5.5	0.5	0.0	–	–	–	–
petitCODE(0)	500 – 1700	50	3.0 – 6.0	0.5	-1.0 – 1.4	0.2	–	–	cloud-free
petitCODE(1)	1000 – 1500	100	2.0 – 5.0	1.0	-1.0 – 1.0	1.0	1.5	–	w/o scattering, w/o Fe clouds
petitCODE(2)	1000 – 1500	100	2.0 – 5.0	0.5	0.0 – 1.5	0.5	0.5 – 6.0	1.0[a]	with scattering, w/o Fe clouds
petitCODE(3)	1000 – 2000	200	3.5 – 5.0	0.5	-0.3 – 0.3	0.3	1.5	–	with scattering, with Fe clouds
Exo-REM(1)	400 – 2000	100	3.0 – 5.0	0.1	0.32, 1.0, 3.32	–	–	–	cloud particle size fixed to 20µm
Exo-REM(2)	400 – 2000	100	3.0 – 5.0	0.1	0.32, 1.0, 3.32	–	1.0	–	–

[a] Except additional grid point at 0.5

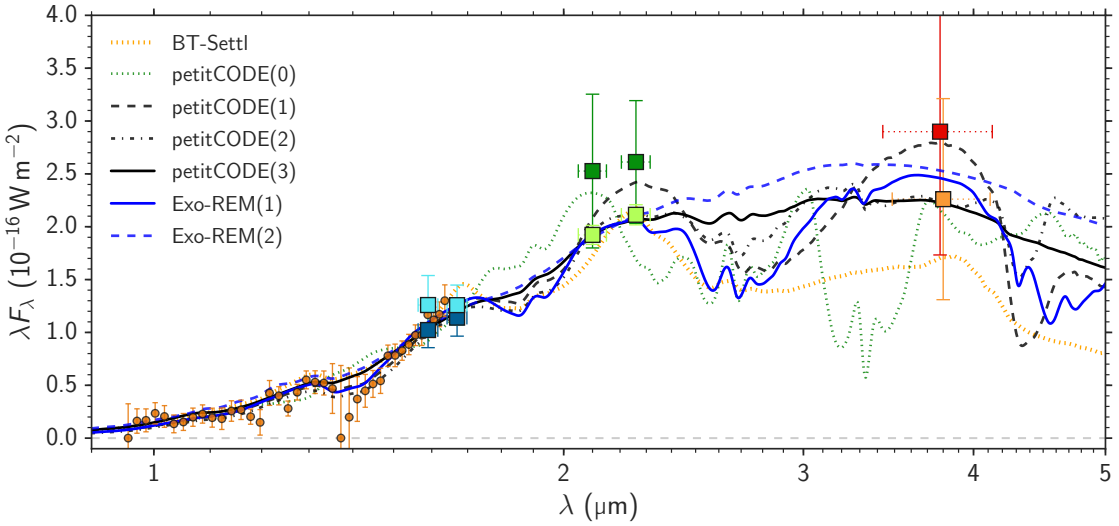


Figure 3.1: Spectral energy distribution of PDS 70 b as a function of wavelength constructed from Y- to H-band IFS spectra (orange dots), IRDIS H_2H_3 (first epoch in dark blue squares, second epoch in light blue squares), and K_1K_2 (first epoch in dark green, second epoch in light green), NaCo (red), and NICI (orange) L'-band images. Error bars show 1-σ uncertainties. The horizontal error bars show the width of the SPHERE and NACO filters, and the grey squares show the predicted flux calculated using the filter transmission curves. Plotted are the best fits for the seven model grids smoothed to the resolution of IFS.

Table 3.3: Parameters of best-fit models based on the grids listed in Table 3.2. The last two columns indicate qualitatively whether or not the corresponding model is compatible with the photometric points in K and L' band, whereas all models describe the Y- to H-band data well.

Model	$T_{\rm eff}$ (K)	$\log g$ ($\log_{10}$(cgs))	[M/H] (dex)	$f_{\rm sed}$	Radius $R_{\rm J}$	Mass[b] $M_{\rm J}$	K flux	L' flux
BT-Settl	1590	3.5	–	–	1.4	2.4	yes	yes
petitCODE(0)	1155	5.5	-1.0	–	2.7	890.0	no	(yes)
petitCODE(1)	1050	≤ 2.0	≥ 1.0	1.5[a]	2.0	0.2	yes	yes
petitCODE(2)	1100	2.65	1.0	1.24	3.3	1.9	yes	(no)
petitCODE(3)	1190	≤ 3.5	0.0	≤ 1.5	2.7	8.9	yes	yes
Exo-REM(1)	1000	3.5	1.0	–	3.7	17	yes	yes
Exo-REM(2)	1100	4.1	1.0	1	3.3	55	yes	yes

[a] Only grid value. [b] As derived from $\log g$ and radius.

3.3 HIP 65426 b

The planet HIP 65426 b is the first newly discovered planet with the SPHERE instrument. It was announced in Chauvin et al. (2017) and later further characterized in Cheetham et al. (2019) of which I am second author, and worked on the atmospheric characterization part.

As shown in Table 3.1, at first glance it is very similar to PDS 70 b. Both are hot, dusty, L-type planets. HIP 65426 b differs by being about 10 Myr older than PDS 70 b, which is significant, as young planets experience rapid cooling after formation and only later "lose" information on their formation process, as the cooling curves for hot-start and cold-start models converge and the bolometric luminosity is dominated by mass rather than age. The next important difference is that HIP 65426 b is further away from the star (∼120 au) compared to PDS 70 b (∼22 au), which could also imply a strongly different formation history. The last potentially important difference is that PDS 70 is a K-type star ($M \sim 0.76 M_{\odot}$), whereas HIP 65426 is an A-type star ($M \sim 1.96\,M_{\odot}$), which also suggests that the size and mass of the progenitor planetforming disks would have been different (Pascucci et al., 2016; Andrews et al., 2018a).

3.3.1 Atmospheric modeling of HIP 65426 b

I used BACON to probe the parameters for HIP 65426 b, in the same way as described above. However, it turned out that the models with an Ackerman-Marley cloud model (i.e. $f_{\rm sed}$ prescription) (Ackerman & Marley, 2001), we had available were not a good match to the data and were unable to explain the spectrum as a whole at all wavelength. This could have numerical reasons: the hot iron/silicate clouds in petitCODE for certain parameter combination become numerically unstable, i.e. they become so thick that the model locally heats up, evaporates the clouds, cools down again, and repeats this cycle. Thus, models can be stuck in such an oscillation without converging to a stable solution. It can also be that for the atmosphere of HIP 65426 b, some combination of mixing parameters and settling parameters could work that was not explored by the available grid.

In the end, the analysis focused on publicly available PHOENIX BT-Settl grid (Allard, 2014; Baraffe et al., 2015b). These models follow a different approach to cloud opacities, in which opacities are based on a settling time scale argument (Rossow models, e.g. as described in Helling et al., 2008), that seems to work relatively well for hot planetary objects. This is also confirmed by my model fits using the BT-Settl grid, which provide an excellent fit to the data, and even predicted the correct K-band flux before this data was available. The main drawback of BT-Settl models is that they are based on outdated opacities and line lists; fortunately this is not such a big problem for relatively hot objects, and more of a problem for cool, methane and water dominated atmospheres, where molecular line lists and opacities have seen major revisions (e.g., Yurchenko & Tennyson, 2014).

The observational data with the best-fitting models are shown in Fig. 3.2 and the corresponding corner plot of the Bayesian analysis is shown in Fig. 3.3. The fit is excellent over the whole wavelength range. The only small deviation is the M-band data, which is still consistent with the models within 1-σ. The BT-Settl model unfortunately does not explore metallicity space, so we cannot draw any conclusion on the metallicity of the planet, but considering the good fit, we can say at the least, that the solar metallicity model is consistent

with the data.

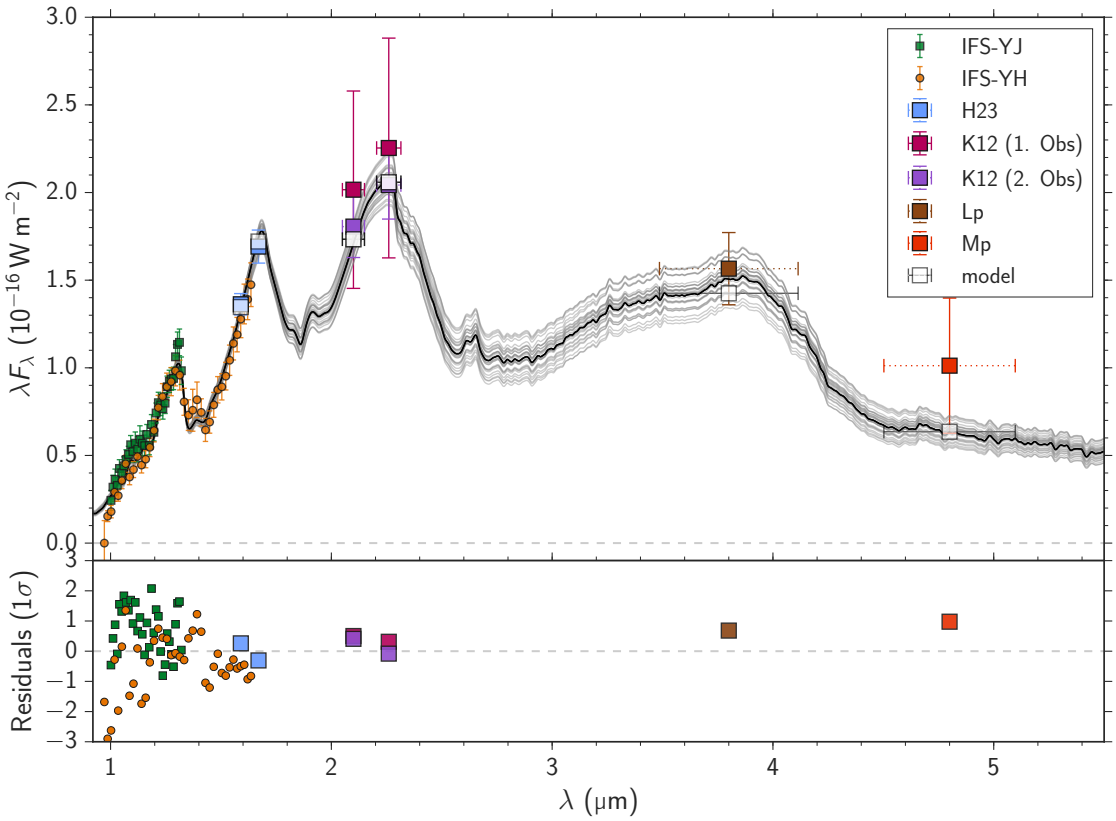


Figure 3.2: The observed spectrum of HIP 65426 b (colored squares) compared to the best-fit spectral model from the BT-Settl grid, with $T\mathrm{eff} = 1618\,\mathrm{K}$, $\log g = 3.78$ and $R = 1.17\,\mathrm{R_J}$. The data from Y- to K-band were taken with SPHERE. L- and M-band observations are from NACO. Error bars show 1-σ uncertainties. The horizontal error bars show the width of the SPHERE and NACO filters, and the grey squares show the predicted flux calculated using the filter transmission curves. Several models drawn from the MCMC posterior distribution are shown in grey.

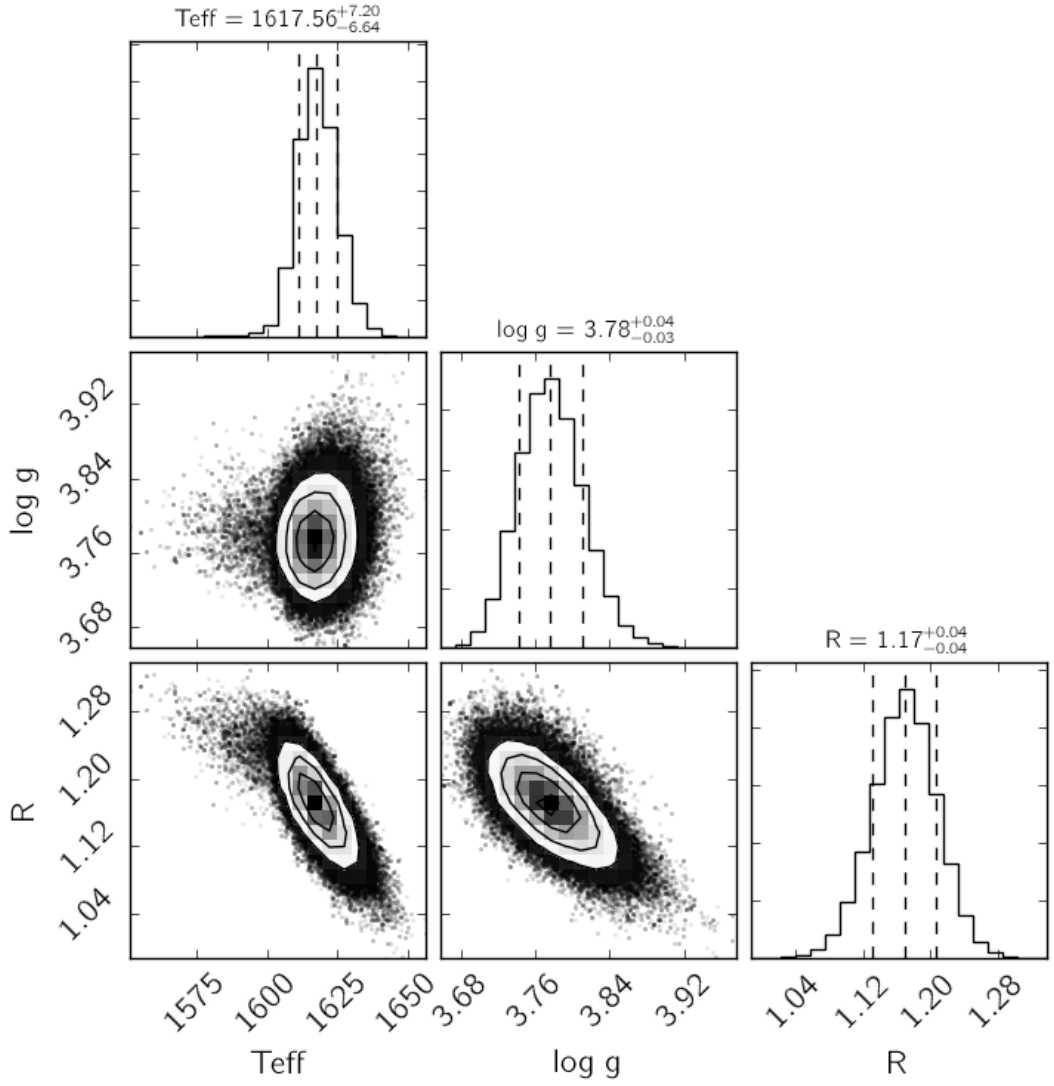


Figure 3.3: Corner plot showing the posterior probability distribution for HIP 65426 b for the BT-Settl grid with respect to each of its parameter pairs as well as the marginalized distribution for each parameter. R is given in units of $\mathrm{R_J}$. The uncertainties are given as 16% to 84% percentiles as is common for multivariate MCMC results.

It is an interesting exercise to compare the spectrum of the seemingly similar objects HIP 65426 b and PDS 70 b, with each other. Both are at almost the same distance from the Earth (113 pc vs 109 pc), such that we simply plot both SEDs in the same units, shown in Fig. 3.4.

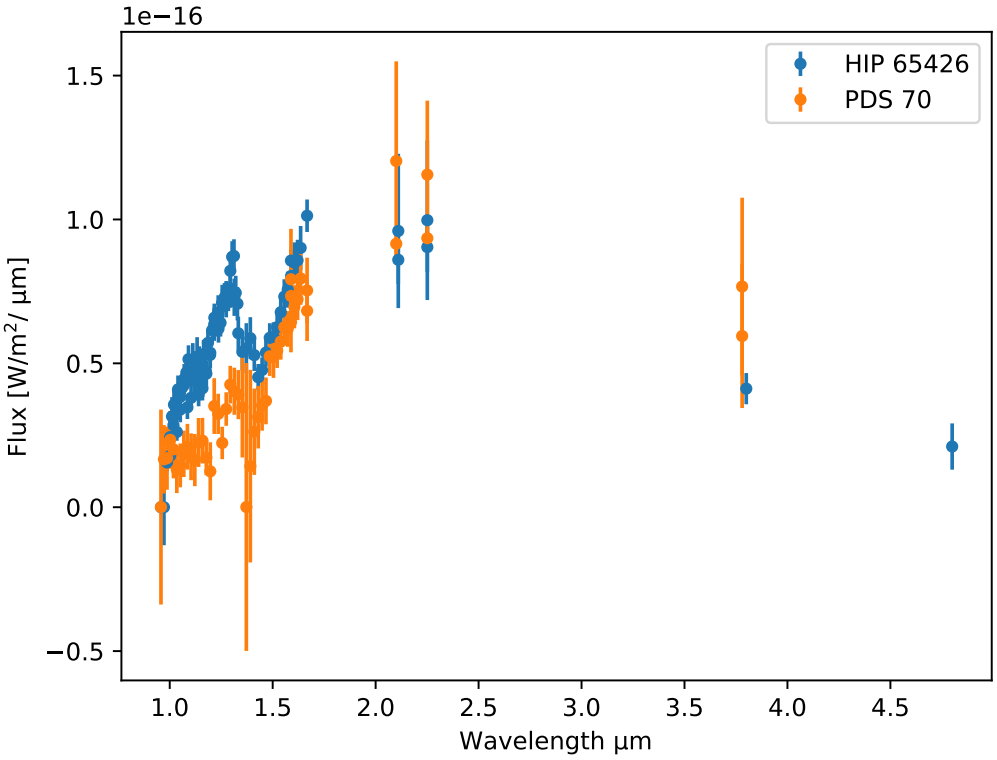


Figure 3.4: Direct SED comparison for HIP 65426 b and PDS 70 b.

We can see that if we only had data in the H-, and perhaps K- and L-band, these two objects would look relatively similar, and their similar spectral classification derives from this fact. This local similarity between two objects at specific wavelengths is a dangerous pitfall for modelers and empirical comparisons in general.

We see a strong difference in the J-band. The dampening of the J-band in PDS 70 b by strong opacity sources in the atmosphere makes the overall spectral very compatible with a blackbody. On the other hand, HIP 65426 b has a more pronounced absorption structure from the photosphere of the planet. We also see a potentially higher flux of PDS 70 b at longer wavelengths. The dampened J-band and higher L-band flux can be explained by opacities "shifting" emission to longer wavelengths, as well as an overall lower temperature of the atmospheres of PDS 70 b. This qualitative view of the two objects is compatible with the quantitative modeling results we obtained in the above studies.

3.4 GJ 504 b

The companion GJ 504 b (e.g., Kuzuhara et al., 2013; Bonnefoy et al., 2018) is a bit of an oddball, as there has been a long debate about the age of the system which has two solutions, one of several Myr and one of several Gyr. This debate is still not decisively concluded one way or the other. Depending on the age of the system GJ 504 b is either a planet or a brown dwarf in terms of mass.

Here I will focus on the results obtained with BACON and petitCODE for GJ 504 b in the context of Bonnefoy et al. (2018). The data is based on SPHERE dual-band photometry as well as four narrow band photometric points obtained with LBT/LMIRCam in the L-band.

3.4.1 Atmospheric modeling of GJ 504 b

As the temperature of GJ 504 b is in a similar range as 51 Eri b, we can make use of the same cloudy grid already used in Chapter 2 (Samland et al., 2017). We extended the models with grid points using higher and lower K_{zz} (mixing strength) and included this as a free-parameter.

Figure 3.5 shows the data and best-fit models for GJ 504 b, and Fig. 3.6 the associated corner plot. The data is fit well by the models, and the main physical parameters (temperature, gravity, radius), are well constrained by the data. The corner plot suggests that a bigger grid in the cloud and vertical mixing parameters would ultimately be useful. The data is compatible with petitCODE models of solar metallicity, relatively low surface gravity, and radius.

It should be pointed out that unless the true surface gravity of GJ 504 b is more than a factor 10 higher than we estimate ($\log g \sim 5$ dex), the mass derived from surface gravity and radius (even fixing the radius to $1 \mathrm{R_J}$) is compatible with that of a planet, which would suggest a young age. Of the models tested in Bonnefoy et al. (2018), all with the exeption of Exo-REM, suggest a planetary $\log g$. From the corner plot shown in Fig. 3.6, we see that neither K_{zz} nor f_{sed} is strongly positively correlated with the surface gravity. In fact K_{zz} and $\log g$ are slightly anti-correlated from which I conclude that extending the grid to higher K_{zz} would not significantly lower the best-fit surface gravity.

As always, these results and parameters are obtained under the assumption that the physical model is a good approximation of reality.

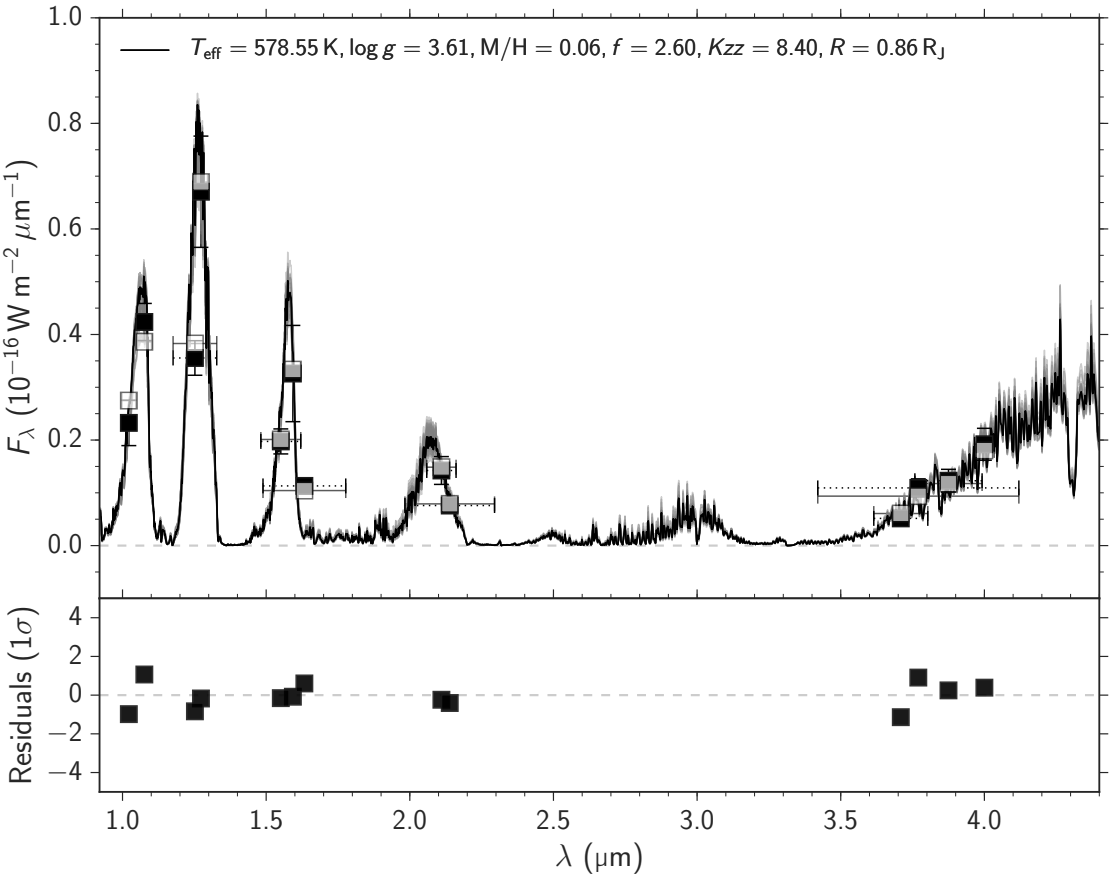


Figure 3.5: The observed spectrum of GJ 504 b (squares) compared to the best-fit spectral model from the cloudy *petitCODE* grid. The data from Y- to K-band were taken with SPHERE; L-band photometry was obtained with LBT/LMIRCam. Error bars show 1-σ uncertainties. The horizontal error bars show the width of the SPHERE and LBT/LMIRCam filters, and the grey squares show the predicted flux calculated using the filter transmission curves. Several models drawn from the MCMC posterior distribution are shown in grey.

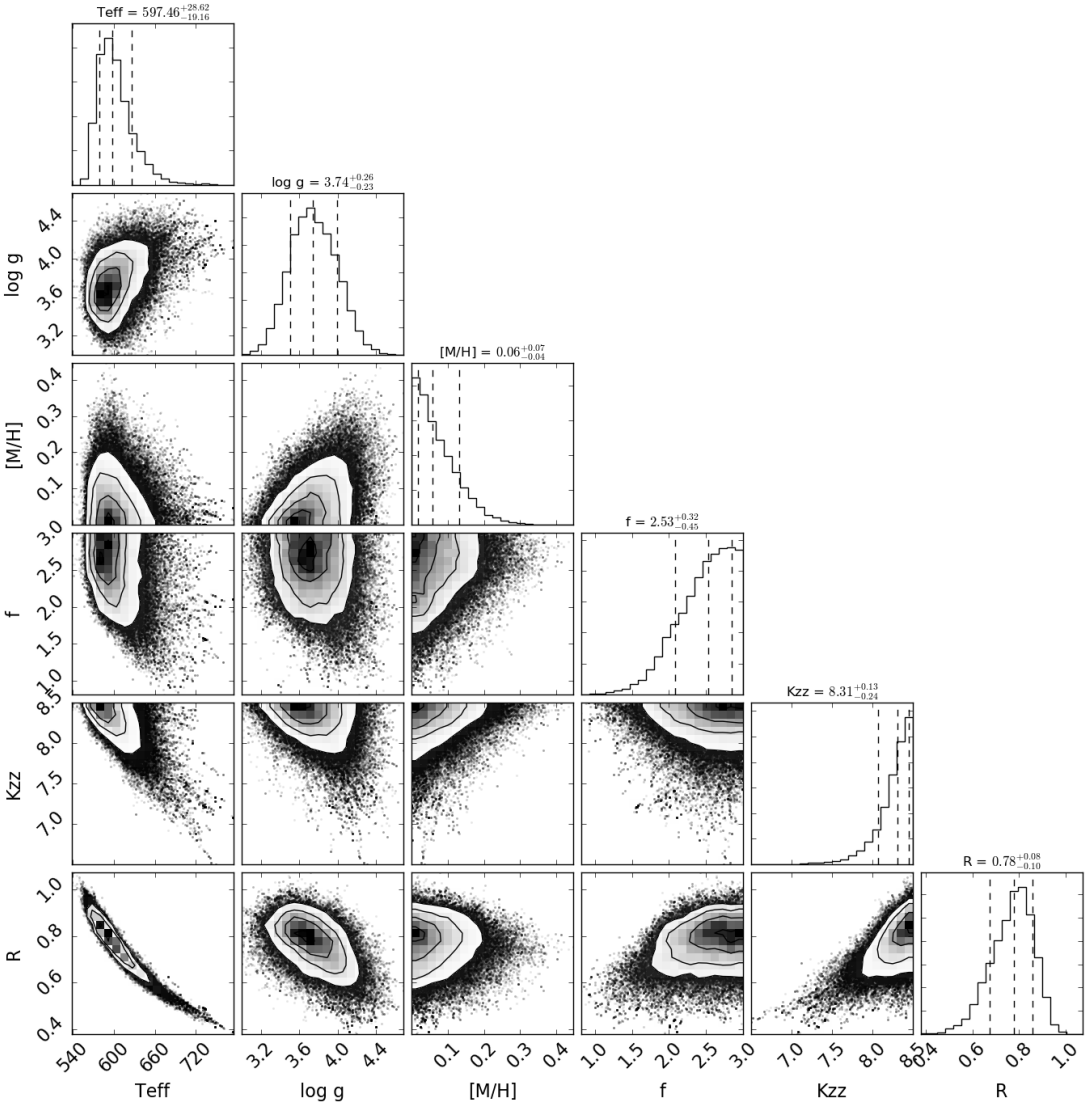


Figure 3.6: Corner plot showing the posterior probability distribution for GJ 504 b for cloudy *petitCODE* models with respect to each of its parameter pairs as well as the marginalized distribution for each parameter. The uncertainties are given as 16% to 84% percentiles as is common for multivariate MCMC results.

3.5 Summary and conclusions

In this Chapter we have seen that objects that have superficially similar properties, i.e. similar spectral type, or temperature, can still exhibit a wide range of physical characteristics. To really explore these differences, and put them on a solid scientific footing, we have to study objects in a more homogeneous way, both in terms of the methodology of data reduction, as well as statistical inference, and underlying atmospheric models used. It is common to focus on one particular object in one study, following one particular methodology.

I conclude that in order to advance *comparative exoplanetology* in the high-contrast imag-

ing field, we will have to study the objects discovered so far in a standardized framework, following the same data reduction procedures, as well as use one set of generalizable atmospheric models. This could be achieved in the future with a homogeneous retrieval analysis of all of these objects.

Chapter 4

"*Binning is Sinning.*"

by Unknown scientist

"*It's a trap!*"

by Admiral Ackbar

4 A temporal, non-local systematics model for direct detection of exoplanets at very small angular separations

The work presented in this chapter will be the basis for a refereed article of the same title for which I am the lead author.

Overview: High-contrast imaging surveys have shown giant planets at large separations to be rare and that it is of paramount importance to push towards detections at smaller separations. Traditional methods for post-processing of pupil tracking observations perform badly at smaller separations, due to the larger field-rotation required to displace a source on the detector and the associated temporal exclusion criteria for the data. My goal is to develop a method of extracting planet signals which is not strongly negatively impacted by such exclusion criteria at small angular separations. To achieve this I create a data-driven model of the temporal behavior of the systematics (temporal trends) using reference pixels. This model is valid under the condition that the underlying causes of the systematics affect multiple pixels, which is the case for the speckle pattern in high-contrast imaging. I simultaneously fit a forward model of a planet signal "transiting" over detector pixels and reference light curves describing the temporal trends of the speckle pattern to find the best fitting model describing the signal.

With my implementation of a non-local, temporal systematics model, called TRAP, I show that it is possible to gain up to a factor of six in contrast at close separations ($< 3\lambda/D$) compared to a spatial model with temporal exclusion criterion. I demonstrate that the temporal sampling has a big impact on the achievable contrast, with better temporal sampling resulting in significantly better contrasts. At short integration times (4 seconds) for β Pic data, I increase the SNR of the planet by a factor of four compared to the spatial systematics model. Finally, I show that the temporal model can be used on unaligned data which has only been dark and flat corrected, without the need for further pre-processing.

4.1 Introduction

The field of direct observation of extrasolar planets has seen tremendous progress over the last ten years, both in terms of observational capabilities, as well as calibration techniques. Especially the recent advent of instruments dedicated to high-contrast observations like SPHERE (Spectro-Polarimetric High-contrast Exoplanet REsearch; Beuzit et al., 2019), GPI (Gemini Planet Imager; Macintosh et al., 2014), and CHARIS (Groff et al., 2015), as well as the development of more sophisticated observational strategies and post-processing algorithms have paved the way to make this a unique technique to study atmospheres and orbital characteristics of exoplanets directly. However, giant planets and substellar companions at large orbital separations ($\gtrapprox 5-10$ au) have proven to be rare by multiple large direct imaging surveys (e.g., Brandt et al., 2014; Vigan et al., 2017). Pushing our detection capabilities towards smaller inner-working angles is therefore one of our primary goals to tap into a parameter space regime in which planets are more abundant and which overlaps with other indirect detection techniques.

Detecting planets at small separation from their host star is intrinsically more difficult the smaller the separation becomes, because the speckle background will be higher, and less independent spatial elements are available for comparison and calibration. However, this problem is further exacerbated because most algorithms used to model the stellar contamination in the data, require an exclusion criteria for what data can be used to construct the systematics model that prevents incorporating planet signal into the model.

In this work, we present a novel algorithm designed to significantly improve performance at small angular separations compared to conventional algorithms for pupil tracking observations. Especially as coronagraphs become more powerful and allow us to probe smaller inner working angles (IWA[9]), this will become more and more important. We achieve this by building a *data-driven, temporal, and non-local systematics forward model*, that does not require a temporal exclusion criterion and allows using all frames in the observation sequence regardless of separation. This classification scheme and how the algorithm differs from traditional *angular differential imaging* (ADI) algorithms will be explained in detail in the next sections.

In Section 4.2 we will give an overview of general approaches high-contrast imaging and show how our new temporal systematics model approach relates to the spatial systematics model approach traditionally used in pupil tracking observations. In Section 4.3 we motivate the non-local, temporal systematics approach, based on the shared underlying causes of systematics in the data. Section 4.4 then goes on to explain the application to pupil tracking high-contrast imaging data. Section 4.5 discussed the data used to demonstrate the performance of our algorithm TRAP and section 4.6 shows the results obtained. We end in Section 4.9 with our conclusions and an outlook on how the algorithm can further be improved in the future.

[9]The IWA is defined as the angular separation at which the coronagraphic transmission crosses 50%.

4.2 Model generalities: classes of data and algorithms

The main challenge in high-contrast imaging is distinguishing a real astrophysical signal from light of the central star, the so-called speckle halo. This systematic noise[10] is generally orders of magnitude brighter than a planetary companion object in raw data, and furthermore can locally look like a genuine point source. In our classification we will distinguish algorithms based on three questions:

1. Is the systematics model physics driven or data driven?
2. Is the systematic noise corrected first before measuring the planet signal or are both fitted for simultaneously?
3. What is the distinguishing property between the planet and star light?

1. Physics vs. data driven models There are in general two approaches to modeling systematics. One is to build up a complete *physics driven model* (i.e. a simulation) of the data that describes its systematic behavior and various sources of uncertainty. While detector related uncertainties are comparatively easy to understand, subtle changes in the instrument and atmospheric variability have consequences that are hard predict from first principles. This approach is generally not followed, due to the sheer complexity of the problem. This is not to say that the choice of data driven model cannot be informed by the physics of the instrument and observation.

The second approach is to use a *data-driven model*, in which each data point or data source is *predicted* using a parametrized mathematical function of *other data*. This is usually the preferred way for high-contrast imaging, as we are limited by unknown or untracked systematics induced by the instrument and atmosphere. This is also the approach we follow in this work.

2. Subtracting systematics model vs. explicit forward model Another decision is between, a) using a two-step approach in which we build a systematics model that we then subtract in order to measure the signal of interest, and b) fitting an explicit model of the signal of interest together with the systematics.

As long as we have an understanding of the type of signal we want to recover, forward models are generally preferable to the subtraction of a noise model and subsequent measurement of the signal. The systematics model is never perfectly orthogonal to the signal, consequently some distortion of the signal shape and amplitude is unavoidable. Fitting the signal simultaneously with the noise model reduces over-fitting, and also allows using a more flexible systematics model.

3. Distinguishing property between stellar and planetary light (diversity) In order to differentiate between an observation that contains only a star, and one that includes an additional faint plant signal, the two signals have to differ in some way. Over the years,

[10] We may refer to noise in this work as *an unwanted signal or a disturbance*, and not strictly a stochastic entity. Mostly, we will refer simply to "systematics".

many approaches to distinguish light from a star and from planetary light have been been developed. In general there are multiple such properties of light that can be exploited. For example: temporal, spectral, spatial, and polarimetric differences. A detailed overview of diversities is given in Chapter 1.2.2 of this thesis.

4.2.1 Pupil tracking mode: inducing temporal variation from spatial separation

In this work we will focus primarily on one of the most commonly used observational technique used for high-contrast exoplanet imaging observations: *pupil tracking mode* observation. A general overview about this observational technique can be found in Chapter 1.4.3.

Pupil tracking observations, due to the field-of-view (FoV) rotation, induce temporal variation in the planetary signal on the detector, thus creating a signal that varies in both space and time. Because there are two ways in which the planet signal differs from the stellar systematics, we have the freedom to build our systematics model also either on correlations in space or correlations in time. Figure 4.1 shows the two main paths we can take from there, which correspond to these two paradigms. It is also possible to build spatio-temporal hybrid models, that have not yet been explored, and will be part of future work.

The difference in the two approaches can be understood based on whether we treat the data as being either: a) a set of images, to be explained in a basis constructed from images taken at other times (spatial systematics model), or b) a set of time series, to be explained in a basis constructed from other time series taken at other locations (temporal systematics model).

Let us a assume we have a set of data vector $\mathbf{Y}$ that contain planet signal $\mathbf{S}$, and are afflicted by some functional form of stellar systematics N and random noise. The data has a spatial dimension x and a temporal dimension t.

$$Y(x_i,t_j) := \{S(x_i,t_j) + f(N,x_i,t_j) + \epsilon\}, \quad \forall i,j : S(x_i,t_j) > 0. \tag{4.1}$$

Let us further assume there are other vectors of data that do not contain planetary signal, but are affected by the same underlying causes of systematics

$$X(x_i,t_j) := \{f(N_s,x_i,t_j) + \epsilon\}, \quad \forall i,j : S(x_i,t_j) = 0. \tag{4.2}$$

It has been shown, that the noise functions f can be reconstructed via a regression of other data, for example using a linear model of other data vectors unaffected by the signal we are looking for, if the underlying systematic causes in the data are shared. The spatial approach can then be formulated as follows,

$$\mathbf{y}\,(\mathbf{x}, t = t_k) \sim \mathbf{S}(\mathbf{x}, t = t_j) + \sum_{j \neq k} \alpha_j \mathbf{X}_j(\mathbf{x}), \quad \mathbf{y} \in Y, \quad j = \text{const}, \tag{4.3}$$

where the data to be fitted takes the form of an image at one point in time t_k, and the systematics are reconstructed, using a linear combination of base images (vectors) $\{\mathbf{X}\}$ that correspond, for example, directly to other reference images (as in LOCI), excluding (at least) the time $t_j = t_k$ the image $\mathbf{y}(\mathbf{x}, t_k)$ was taken.

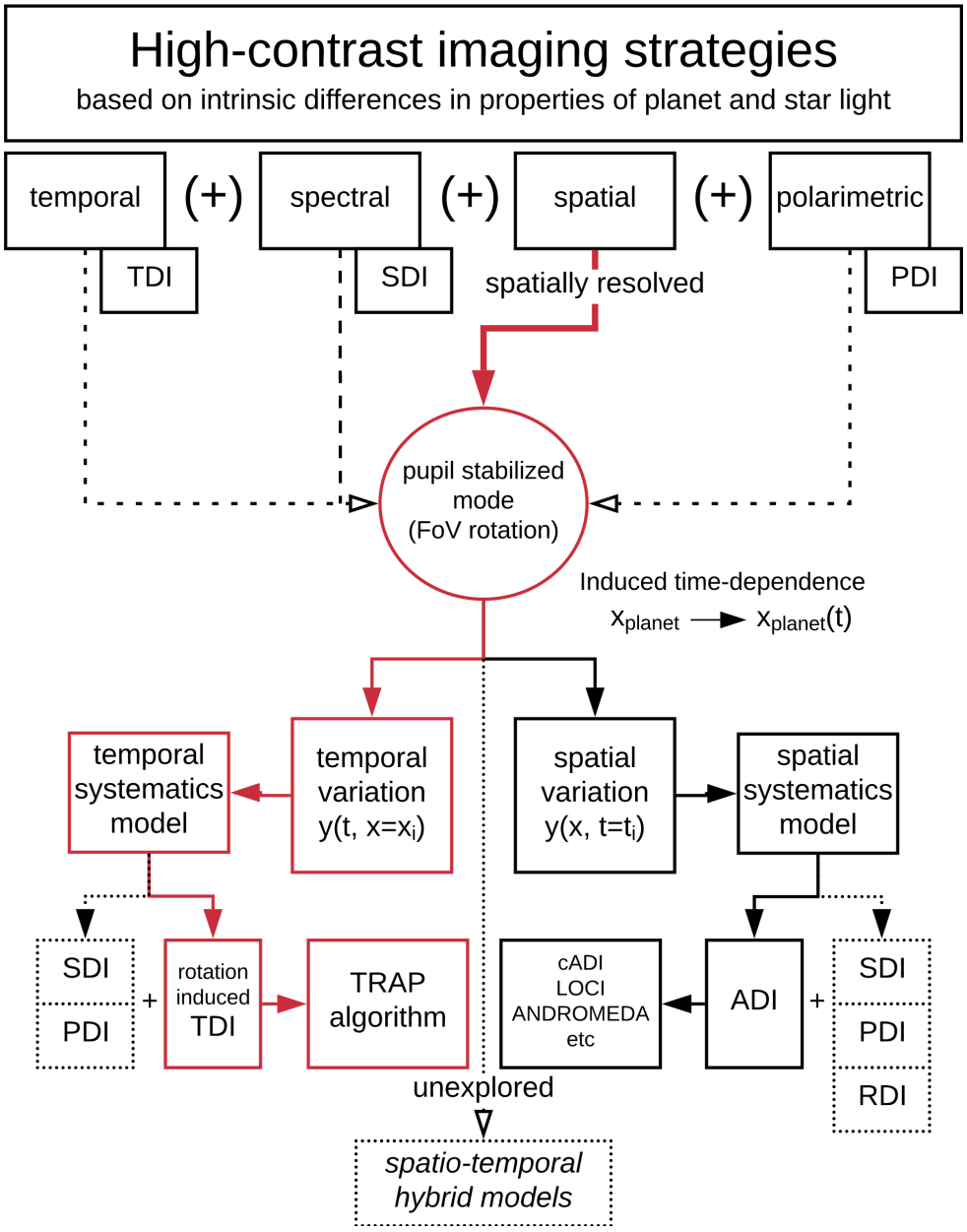


Figure 4.1: Schematic overview of high-contrast imaging strategies. The upper part of the diagram shows the diversities: differences in properties of the detected planetary and stellar light. They are not mutually exclusive and can be combined. Pupil tracking observations can be used in conjunction with any type of imaging observation, and are special in that rotation of a spatially resolved system induces a predictable temporal change in the *planet signal only*. This allows two approaches. The red color shows the path followed in this work (left branch). This branch shows the temporal systematics approach, where data y is viewed as a time-series at a specific location, that is explained in terms of other data that are time-series at other locations. The right branch corresponds to the spatial systematics approach, where data y is viewed as an images at one specific time, that is compared to other images at other times. Beside these two classes there are spatio-temporal hybrid models, that have not yet been explored.

The temporal approach can be formulated as

$$\mathbf{y}\,(x = x_k, \mathbf{t}) \sim \mathbf{S}(x = x_i, \mathbf{t}) + \sum_{i \neq k} \alpha_i \mathbf{X}_i(\mathbf{t}), \quad \mathbf{y} \in Y, \quad i = \text{const}, \tag{4.4}$$

where the data to be fitted takes the form of the time-series (lightcurve) of one pixel x_k and is reconstructed in a basis set of reference time-series $\{\mathbf{X}\}$, excluding (at least) the time-series $\mathbf{y}(x_k, \mathbf{t})$.

The α-coefficients are determined through a regression with the data to find the optimal combination. This means in terms of correlations, that the spatial systematics approach determines the α-coefficients by optimizing based on spatial correlations ("similarity in space") between the images, whereas the temporal systematics approach optimizes for temporal correlations ("similarity in time").

In the light of the discussion above, spatio-temporal hybrid models can be thought of extending the data vectors from being either time series of individual pixels or images at specific times, to time-series of patches of pixels. In such models, the time-series of one patch would be re-constructed in a basis set of vectors each containing the time-series of another patch. Such an approach would then be optimized based on both, spatial and temporal similarity between multiple such patches.

Traditionally, pupil tracking observations have been viewed as an intrinsically spatial problem, i.e. from the point of view of spatial diversity of the planet signal and a corresponding spatial systematics model. Following the invention of the *roll deconvolution* (or *roll angle subtraction*) method for space-based observatories (Müller & Weigelt, 1985), classical ADI (Marois et al., 2006) for ground-based observatories, and its evolution *Locally Optimized Combination of Images* (LOCI Lafrenière et al., 2007), virtually all subsequent papers and algorithms focused on using the local/spatial similarity between systematics training set and the data to remove the quasi-static speckle pattern covering the signal of interest. This is the case for most commonly used algorithms, regardless of the implementation of model construction, e.g., PynPoint (PCA, Amara & Quanz, 2012a; Stolker et al., 2019), KLIP (Soummer et al., 2012; Ruffio et al., 2017), VIP (Gomez Gonzalez et al., 2017), and ANDROMEDA (Cantalloube et al., 2015). One exception to this is the wavelet-based temporal de-noising approach (Bonse et al., 2018), which, however, does not attempt to model the systematics or planet signal, but rather temporally filter/pre-conditions the data before applying the spatial systematics approach. In this work we will make the case for the temporal paradigm and why it may be preferable.

4.3 Causal structure of systematics: what are the requirements for each approach to work?

In order to decide which approach to take, we first have to think about the causes of the systematic noise. Why does the spatial approach work, and why does the temporal approach work? Firstly, for the spatial approach to work either of two conditions have to be met: a) the speckle pattern is locally stable (strongly correlated in time) over the time scales that are required to create displace the signal on the detector, or b) the pattern is

statistically recurrent (e.g., finite probability of pointing jitter to revisit the same position). Both cases have a high requirement on the overall stability of the instrument and observing conditions, but in modern instruments a large part of the systematics is indeed correlated on timescales of minutes to hours ("quasi-static") (e.g., Milli et al., 2016; Goebel et al., 2018).

Why does the temporal approach work? At first glance this is not obvious, until we think about the causal structure of the systematic noise in the data, or what statisticians also call "confounding" of the signal. If you take the time series of one pixel and the time series of another pixel at a different location, what do they have in common? They do not directly "talk to each other" and are, at first order, independent measuring devices. However, they *are* influenced by same underlying causal influence that generates the systematics in the first place.

The mathematics of statistically modeling *confounding* in time-series data that shares systematics generated by the same underlying causes has been described in detail in Schölkopf et al. (2016). Figure 1 of Schölkopf et al. (2016) shows the general structure of the causal inference as directed graphical model (or directed acyclic graph, DAG).

In the above Equation 4.1 and Equation 4.2, we see that the pixels of set Y are independent from those in X in terms of whether the planet signal is seen, and they do have different expressions of noise functions $f(\mathbf{x}, \mathbf{t})$, that differ spatially and in time. However, their systematic noise terms are functions of the same underlying causes N. Schölkopf et al. (2016) showed that by using a regression model of a sufficiently number of such "half-siblings", we can reconstruct a specific instance of the noise function. Further, we can include a model of the planet signal itself in the regression, such that the overall regression explains the data in Y up to a stochastic noise term, which remains the fundamental limit of the data. This fundamental limit is the combination of photon noise and detector read-out noise.

4.3.1 The problems with spatial systematic models

The main drawback of the spatial modeling approach is the need to prevent contamination of the systematics training set with real signal by using an exclusion criterion (protection angle). Its selection will always be a balancing act between excluding more frames to prevent self-subtraction, and losing information on the speckle evolution due to exclusion of frames close in time. A more detailed discussion of speckle de-correlation time scales can be found in Chapter 1.4.2. This trade-off gets gradually worse the closer one approaches the central star in angular separation, because the same field-of-view rotation corresponds to smaller and smaller physical displacement on the detector between subsequent frames. Figure 4.2 shows the effect of the protection angle in terms of loss of available training data (left panel) and the average separation in time to the closest exposure that can be used for modeling the systematics (right panel). For example, for a typical target (meridian passage 60° over the horizon), at a separation $3\,\lambda/D$, it takes 25 minutes to displace a point source by half its FWHM, which can be a significant fraction of the total on-sky exposure time. This effect gets significantly worse for separations $< 2\lambda/D$ and targets of lower elevation or offset from meridian passage. These long exclusion times at small separations are on the order of the linear decorrelation timescale for relatively stable instrumental speckles (Milli et al., 2016; Goebel et al., 2018). Even at large separations the exclusion will still be on the order

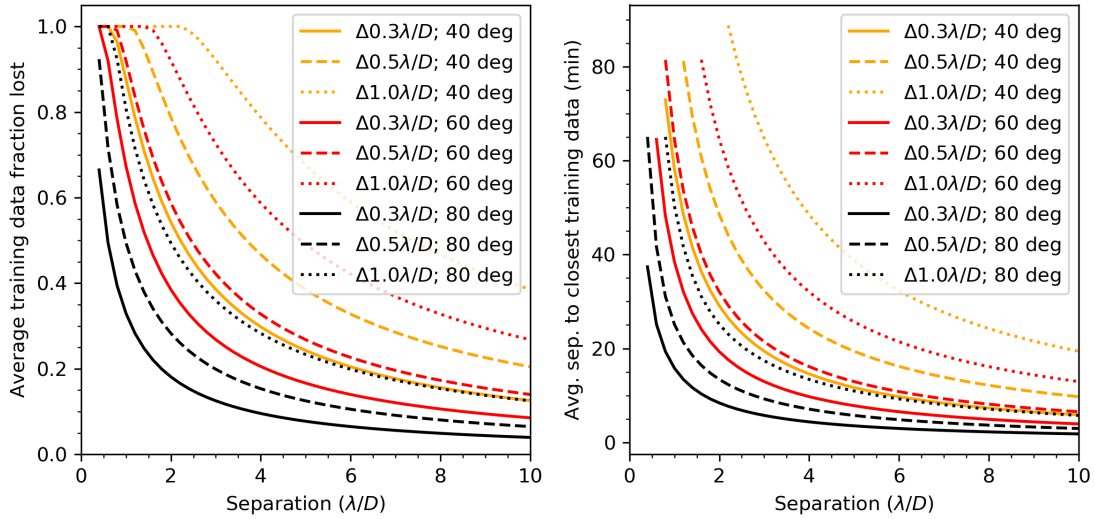


Figure 4.2: Shown is the effect of a different protection angle (0.3, 0.5, 1.0 λ/D) over separation from the central star. The line style gives the protection angles, where as the color of the line encodes different elevations of the target over the horizon at meridian passage. All values are computed for an observation sequence of 90 minutes. The left panel shows the total fraction of training data lost due to the protection angle averaged over the whole sequence. The right panel shows the average temporal separation to the nearest viable frame outside the protection zone.

of minutes, which means that using a spatial approach, modeling the turbulence-induced short-lived speckles with an exponential decay timescale of a few seconds (τ=3.5s, Milli et al., 2016) is intrinsically impossible.

Excluding frames to construct a local speckle model is therefore equivalent to strongly reducing both the size and quality of the systematics training set. On the other hand, not excluding frames leads to strong attenuation of the signal of interest due to incorporating signal into the training set. Traditionally, resolving this issue requires *co-temporal* data and a systematics model that makes use of data taken at the same time, e.g. in the form of SDI or PDI. However, in this work we demonstrate how it is possible to apply a temporal systematics model to monochromatic pupil tracking observations, using non-local reference data.

4.4 Causal data-driven modeling of temporal systematics for high-contrast imaging

We will construct a non-local, temporal systematics model which allows us to avoid the harsh exclusion criterion for preventing self-subtraction, while still capturing the systematic noise. The mathematical idea of using time-series regression models, as explained in Section 4.3, has been used very successfully for transit observation using the Kepler spacecraft (e.g., Wang et al., 2016). This is a form of temporal differential imaging (TDI, Fig. 4.1, top row), in which temporal differences in the signal are inherent to the data (not induced by the method

of observation), and the systematic trends are caused, for example, by pointing drifts, and other instrumental effects, and are shared by multiple pixels[11] on the detector.

The inclusion of many *regressors* (pixels in training-set or derived quantities thereof) make data-driven models very *flexible*, such that they can explain even complex systematic behavior. At the same time we have to worry about *overfitting* the signal.

Another important aspect of data-driven modeling is the selection of the data that should be used as input for constructing the model (the regressors). The scientist's understanding of the underlying causal structure in the data plays an important role in the selection of the training set. In the Kepler spacecraft example, this could mean, for example, selecting stars of similar magnitude to the target star.

In this work we show how a similar approach can be implemented for pupil tracking high-contrast imaging data.

4.4.1 Implementation and application to high-contrast imaging data

The situation is very similar for high-contrast imaging data. Causes of systematics, with the exception of detector artifacts (e.g., bad/warm/dead pixels, flat field), usually influence either the image globally (e.g., most atmospheric effects, Strehl ratio), or a significant region of the detector (e.g., wind-driven halo, (mis-)alignment of the coronagraph). Even slowly evolving changes in the quasi-static speckle pattern caused by the instrument usually are not strictly confined to one region of the detector (e.g., speckle patterns can display symmetries).

A detailed study about an objective optimal choice of regressors is beyond the scope of this work, but multiple heuristic choices can be made: 1) preference to pixels at same separation (brightness), 2) inclusion of pixels located on the exact opposite side of the star, 3) pixels at similar position angle. Figure 4.3 shows an example for the reference pixel (regressor) selection geometry chosen in this work, for an assumed companion that passes 20 pixels north of the host star at the midpoint of the observation.

There are many possible choices for the exact implementation of our algorithm. As the main objective of this work is to demonstrate the principle, we try to achieve a balance between simplicity and effectiveness. This means: a) we use a linear model with a quadratic objective function (here, χ^2 statistics), b) we use principal components of the regressor pixel light curves, not the pixels themselves (principal component regression). This reduces co-linearity in the systematics model and transforms it into an orthogonal basis. It also allows us to forgo regularization in favor of truncating the principle components after a certain number, c) we fit each pixel individually instead of all affected pixels at the same time, d) we fit the planet model and the systematics model simultaneously in order to prevent over-fitting, and thus, e) we employ a complete forward modeling approach, i.e. fit the contrast of the planet for any assumed position, similar to ANDROMEDA (Cantalloube et al., 2015) for direct imaging, or Foreman-Mackey et al. (2015) for Kepler. As such, algorithmic throughput corrections as commonly used in direct imaging pipelines to correct for over-fitting, are not necessary. Also, as we do not use the pixels affected by the planet signal, i.e. self-subtraction (a specific form of over-fitting), is avoided. Unless we perform a blind-search in which we

[11] for simplicity we use the term "pixels" here, but the same can be applied to spaxels, microspectra, or any more general quantized "measurements". More generally, one can also use quantities derived from the pixels such as principal components, instead of the pixels themselves, as will be discussed later.

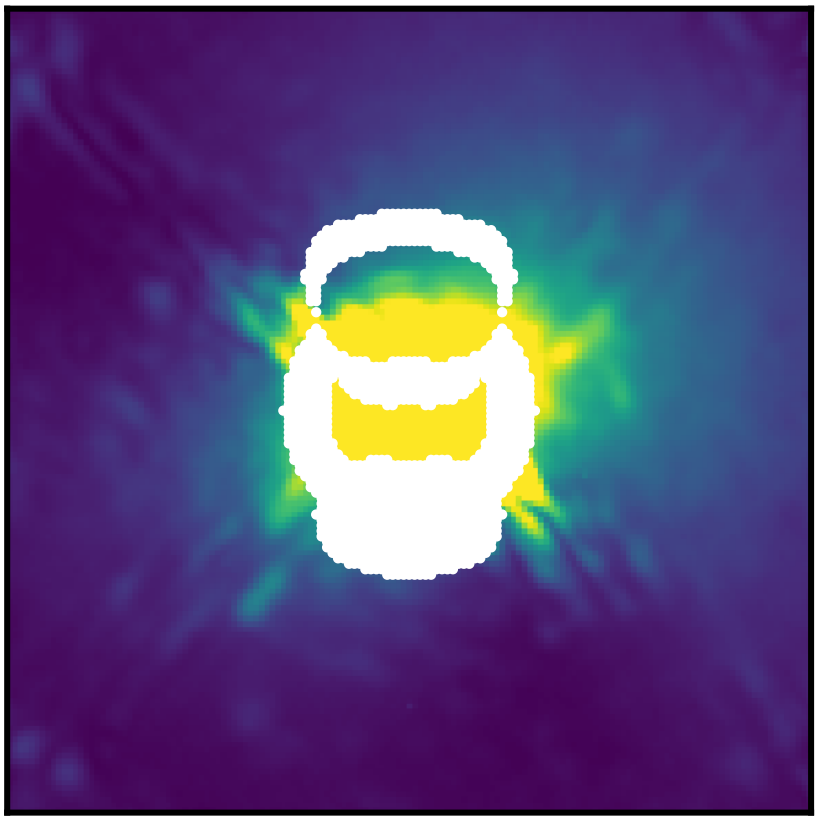


Figure 4.3: Example of reference pixel selection for one assumed planet position. The signal here is assumed to move through the position $(\Delta x, \Delta y) = (0, 20)$ pixel above the star, at the midpoint of the observation sequence. The reference pixels are shown in white. This example is based on the 51 Eri observation's parallactic angles.

do not a-priory know the model of what we are looking for, or the model of the signal is too complicated to describe analytically, a forward model is always preferable to speckle subtraction. An example where forward models are difficult to implement are protoplanetary disks, that can have multiple rings, spirals and asymmetries, and generally require detailed hydrodynamical simulations to model. On the other hand, most debris disks can be well modeled using analytic descriptions and radiative transfer codes (Olofsson et al., 2016).

The resulting algorithm looks as follows for *one* assumed relative position $(\Delta\text{ra}, \Delta\text{dec})$ of a point source on-sky:

1. Generate forward model of signal at assumed position for the time-series, i.e. embed model PSF[12] at appropriate parallactic angle for each exposure to obtain the set of light curves S for all affected pixels. The set of affected pixels for this (on-sky) position will be called Y. Optional: exclude all bad pixels from Y and S.

2. Construct training set for systematic trends T of non-signal pixel with desired constraints (heuristically chosen here: similar distance, mirrored area on other side of star, strips of pixels at same position angles as signal just inwards and outwards of signal). Recommended: exclude all bad pixels and/or pixels affected by a known signal.

3. Compute temporal noise model from T: singular value decomposition (SVD) of scaled training pixels, obtain orthogonal base vector matrix B. This matrix contains our *systematics light curves* or in transit terminology, the *temporal trends*. We add another constant column, to account for constant offsets in our data. This constitute our regressor matrix for the noise systematics and is the same for all pixels in Y, although the coefficients will vary.

4. For each pixel p in Y, i.e. affected by signal:
   a) Add additional column to B containing the light curve shape for the signal (e.g., planet) for this specific pixel from S, we call this complete systematics+signal matrix $\widetilde{B}$
   b) System of equations $\widetilde{B}\,w = d_p$, where w is the vector of weight coefficients and d_p is the time-series of pixel p.
   c) Solve above system using a χ^2-fit of $\widetilde{B}$ to p to obtain optimized coefficients: $w = [\widetilde{B}^T C^{-1}\,\widetilde{B}]^{-1}\,[\widetilde{B}^T C^{-1} d_p]$, where C is assumed to be the unity matrix. The variance associated to the coefficients are then $\sigma_w = [\widetilde{B}^T C^{-1}\,\widetilde{B}]^{-1}$. The last coefficient of w_p and σ_w correspond to the planet model.
   d) Alternative to 4c: marginalize over systematics model

5. Remove significant outliers (e.g., remaining bad pixels that could not be fit properly)

6. Perform average of planet coefficient weighted with their uncertainties over all pixel in S.

Iterating over a grid of possible positions allows us to construct a conditional contrast map, i.e. the contrast and its uncertainty given the position of an assumed object relative to the central star. From these we can construct a signal-to-noise (SNR) map that can be used for detection. However, the uncertainties have been computed under the simplified assumption of independent and Gaussian noise, which does not accurately reflect the reality

[12]We use an unsaturated PSF obtained without the coronagraph directly before or after the sequence, but artificially induced satellite spots could be used. Other ways of reconstructing the PSF such as especially designed coronagraphs that act as focal plane wave-front sensors (Wilby et al., 2017) could in principle be used.

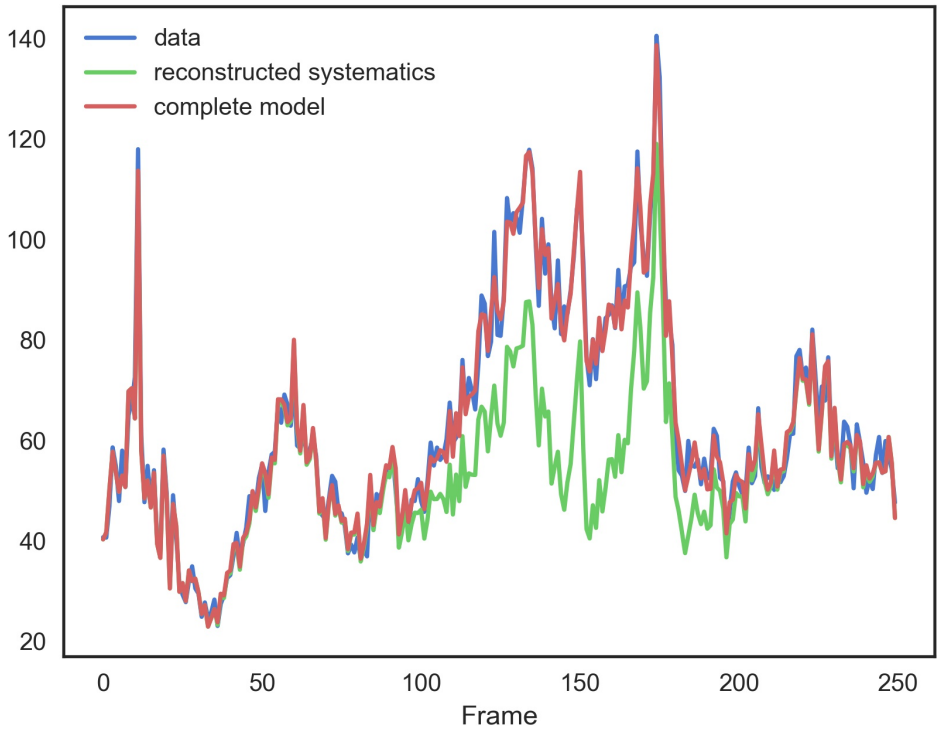


Figure 4.4: Comparison of fitted lightcurve systematics model (with and without planet) with data.

and complexity of high-contrast imaging data. Multiple studies have shown residuals after post-processing with high-contrast imaging pipelines, although significantly whitened, to not be strictly Gaussian and independent (e.g. Marois et al., 2008a). The most direct solution would be to account for these effects in the likelihood function used, but in practice this proves to be challenging since the uncertainty distributions depend both on the observing conditions and instrument. Therefore, we apply the same solution to this problem as in ANDROMEDA Cantalloube et al. (2015) and subsequently pyKLIP (Ruffio et al., 2017) and empirically calibrate the SNR map using the azimuthal standard deviation[13] of the SNR map itself as a function of separation to normalize the map. If not specified otherwise, any reference to S/N in this work refers to the calibrated S/N. This calibrated SNR map will also be referred to as the (normalized) *detection map*. Any known sources in the field should be masked out when deriving the empirical calibration. At the smallest separations ($< 3\,\lambda/D$) small-sample statistics may become important for the empirical normalization aspect (Mawet et al., 2014). As this work focuses on demonstrating the performance of this method in direct comparison to another method, all contrasts shown in this work are without small sample statistics or coronagraphic transmission correction.

[13]In our case we use the robust standard deviation based on the median absolute deviation for the detection limits to be less affected by remaining outliers and real signals.

4.5 Data sets used for demonstration

We demonstrate the above described algorithm on two real data sets obtained with the extreme-AO (SAXO; Fusco et al., 2014) fed infrared dual-band imager and spectrograph (IRDIS; Dohlen et al., 2008) mounted on SPHERE (Beuzit et al., 2019) at the VLT. Both data sets were obtained as part of the SHINE (SpHere INfrared survey for Exoplanets) guaranteed time observation in pupil tracking mode and use the apodized Lyot coronagraph (Soummer, 2005; Boccaletti et al., 2008) with focal-mask diameter of 185 milli-arcsec and an inner-working angle (IWA) of $\sim 2\,\lambda/D$. The two data sets are described in Table 4.1 and Fig. 4.5 and 4.6 show the pre-processed (bad pixel/background/flat corrected and aligned) temporal median image for the 51 Eridani and β Pic observations, respectively.

Both data sets were pre-reduced using the newest version of the SPHERE Data Center pipeline (Delorme et al., 2017b), which uses the Data Reduction and Handling software (v0.15.0, Pavlov et al., 2008) and additional custom routines. It corrects for detector effects (dark, flat, bad pixels), instrument distortion, registers the frames, and calibrates the flux.

1) The first test case is the directly imaged planet in the 51 Eridani system (Macintosh et al., 2015) published in Samland et al. (2017). This data set was obtained in the K1 and K2 bands simultaneously, but we focus our discussion on the K1 channel. The sequence was not taken under ideal conditions. It exhibits many common problems encountered in high-contrast imaging, such as a strong wind-driven halo as seen in Fig. 4.5, and tilt (coronagraph misalignment) causing spider features to appear. These effects make for a realistic test case, and allow us to demonstrate the algorithm performs well under adverse and changing conditions.

2) The second test case uses IRDIS H2 data of the β Pic system (Lagrange et al., 2009, 2019) taken in continuous waffle mode, i.e. using sine-wave modulations on the deformable mirror to induce four satellite spots in all frames of the sequence that can be used to determine the center accurately. This also allows to measure the satellite spot amplitude variations as proxy for the PSF model flux modulation over time. The purpose behind including this data set is two-fold. Firstly, the exposure time is short for an IRDIS sequence (4s), allowing us to test our time-domain based algorithm on a data set with better temporal sampling, approaching the half-life time of fast-decaying speckles (τ=3.5s), so that we do not average over multiple correlation timescales. Secondly, data sets taken in continuous waffle mode do not use detector-stage dithering, which is the default for IRDIS unless switched off. We will use this non-dithered data set to test our algorithm directly on the non-aligned data cube (only dark and flat corrected). This is only possible in a non-local modelling framework. We demonstrate that we can skip bad pixel interpolation and shifting steps by including the shift in the forward model of the planet signal, and exclude the stationary bad pixels from our training and test sets. We also account for SPHERE's anamorphism in the position of the model. Thirdly, we demonstrate the possibility of including the satellite spot modulations in the forward model to reduce a systematic bias of the photometric calibration.

Table 4.1: Observing log

Target	UT date	Waffle mode	IRDIS Filter	IRDIS DIT[a] (sec, #)	T_{exp} (min)	Field Rot.[c] (deg)	Sr[d]
51 Eri	2015-09-25	no	K12	16×256	68	42	$0.80 - 0.90$
β Pictoris	2015-11-30	yes	H23	4×200	53	40	N/A

Observational data used for the analyses. [a] Detector integration time [b] All observation were centered on the meridian passage of the target with an airmass between 1.08 and 1.13. [c] Strehl ratio computed by AO system's Real Time Computer, and scaled to a wavelength of 1.6 µm.

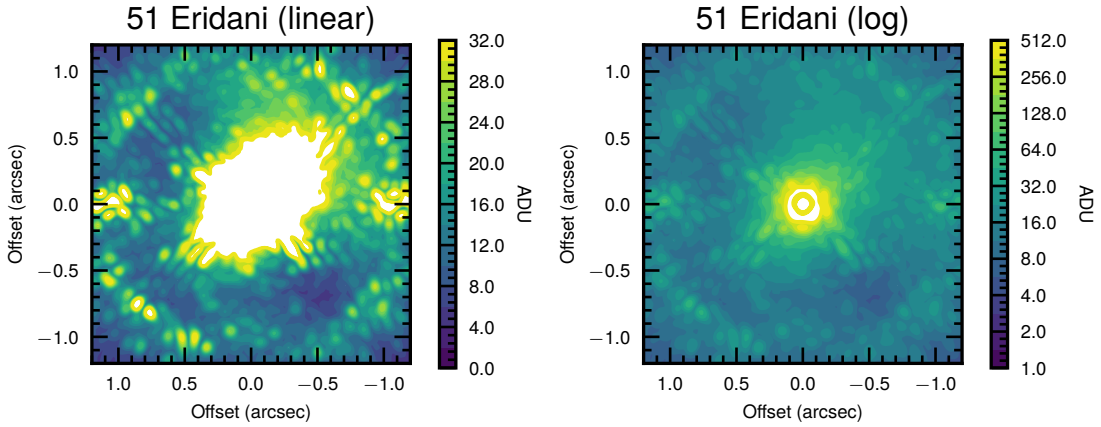


Figure 4.5: Contour plot of median combined image of pre-processed 51 Eridani data cube. The left panel shows data in linear scaled brightness bins, the right panel in logarithmic scaling.

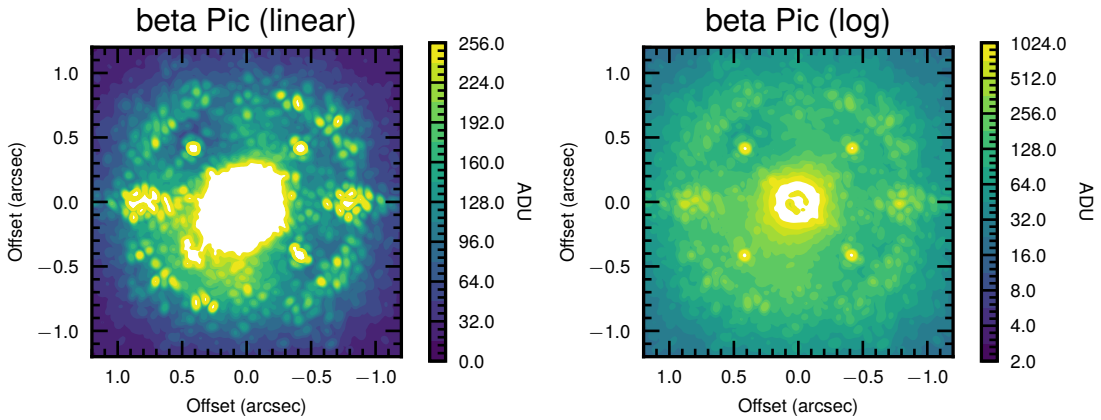


Figure 4.6: Contour plot of median combined image of pre-processed beta Pic data cube. The left panel shows data in linear scaled brightness bins, the right panel in logarithmic scaling.

4.6 Results

We perform a direct comparison of results between our implementation, called TRAP (Temporal Reference for the Analysis for Exoplanets), and ANDROMEDA. TRAP and ANDROMEDA are both based on an inverse problem solving approach using likelihoods instead of speckle subtraction. Therefore the output data products can be directly compared regardless of implementation details of the pipelines. This would not be the case if the comparison pipeline used its own method of computing the algorithmic throughput corrections for the contrast curve, and not an empirical normalized of the pseudo-likelihood (detection) map. For our tests we use the exact same normalization method of the detection map and contrast curve computation for both TRAP and ANDROMEDA outputs. The implementation is included in the public TRAP repository and computes the normalization factor in annuli of 3

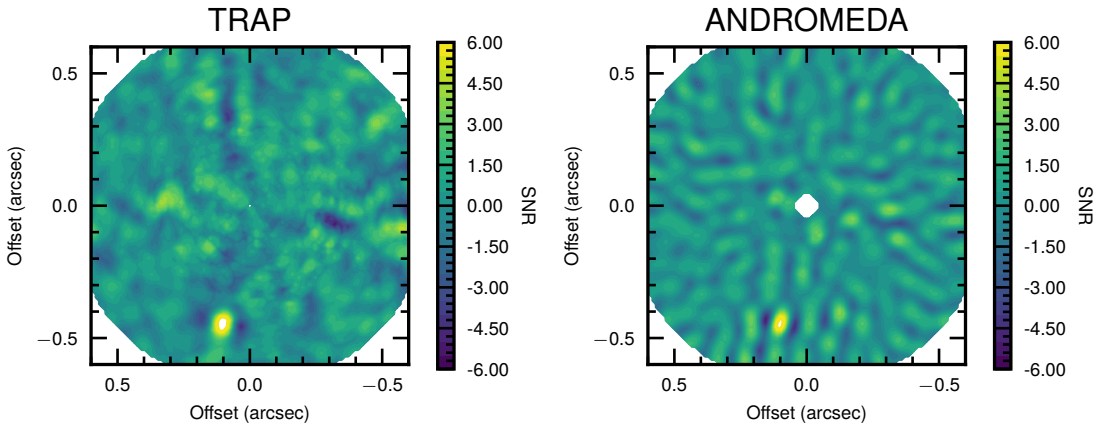


Figure 4.7: Contour plot of normalized detection maps obtained with TRAP ($f = 0.3$) and ANDROMEDA ($\delta = 0.5\lambda/D$) for 51 Eridani b. These maps must not be confused with derotated and stacked image. They represent the forward model result for a given relative planet position on the sky (ΔRA, ΔDEC), i.e. the conditional flux of a point-source predicted by the forward model given a fixed position, corresponding to trajectory over the detector (all pixels affected during the observation sequence).

pixel width using a robust azimuthal standard deviation. Additionally we mask the position of the planet with a mask size of $r_{\mathrm{mask}} = 15$ pixel.

4.6.1 51 Eridani b: centrally aligned data cube

For the first direct comparison between TRAP and ANDROMEDA on exactly the same pre-reduced and centered data cube, we use data of 51 Eridani.

Comparison to ANDROMEDA

The normalized detection map for both TRAP and ANDROMEDA are shown in Fig. 4.7. The first thing we notice is that the remaining structures in the detection map using TRAP are less point-source like and therefore less prone to be confused with a real signal. Besides 51 Eri b, we do not detect any other signal with $> 5\sigma$. Table 4.2 shows a summary of the obtained photometry for 51 Eri b using TRAP and ANDROMEDA.

Figure 4.8 shows the contrast curve for both reductions (top panel) and the factor gained in contrast by using TRAP compared to ANDROMEDA (bottom panel). The ANDROMEDA results have been obtained using two different values for the protection angle $\delta = 0.5$ FWHM and $\delta = 1.0$ FWHM. As we obtain a 2D detection map from our forward model, not only can we determine the median detection limit at a given separation, but also the azimuthal distribution of contrasts for a given radius. Thus, in addition to the median, we plot the 68% and 95% percentile ranges (corresponding to 1σ and 2σ) as shaded regions reflecting the variability of the detection contrast along the azimuth. This is another powerful quantity to evaluate the performance of algorithms.

The detection limit obtained with TRAP is consistently lower than ANDROMEDA for separations up to about $10\,\lambda/D$, at which point the results between TRAP and ANDROMEDA

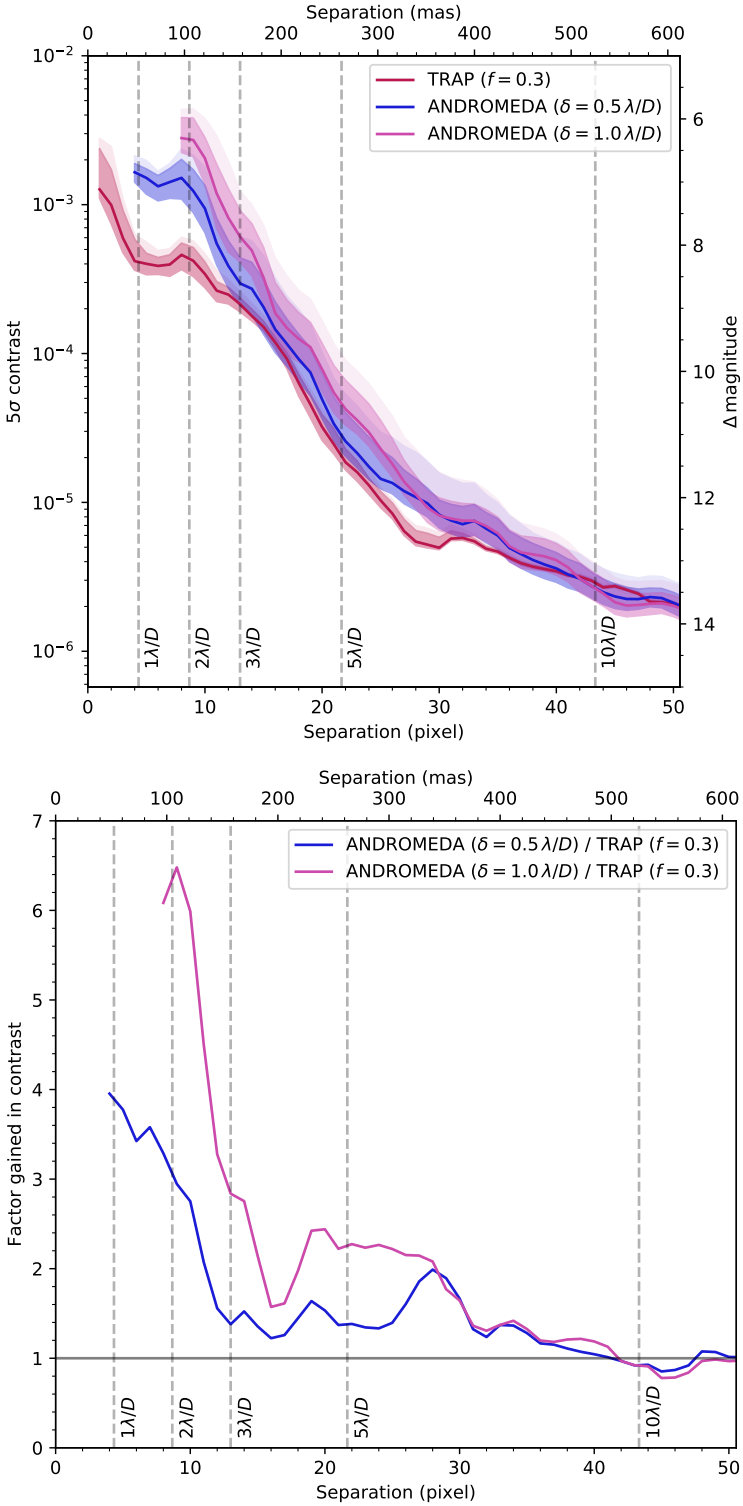


Figure 4.8: Comparison between the contrast obtained with TRAP and two ANDROMEDA reductions for 51 Eridani using the same input data for the K1 band. TRAP has been run with 30% of available principal components, whereas the two ANDROMEDA reductions correspond to a protection angle of $\delta = 0.5$ FWHM and $\delta = 1.0$ FWHM. (top) The shaded areas correspond to the 16%-84% percentile interval of contrast values at a given separation. (bottom) Factor in contrast gained using TRAP.

are consistent. Especially at small separation we clearly see a gain in sensitivity due to the absence of a protection angle in our reduction. The ANDROMEDA curves cut off at about $1\lambda/D$ and $2\lambda/D$ respectively, because no reference data that fulfills the exclusion criterion exists (see Fig. 4.2, $\sim 40°$ elevation). The relative gain in contrast with respect to the spatial model (ANDROMEDA) as shown in Fig. 4.8 is highly significant, with ramifications for the detectability of close in planets. It can be as high as a factor of six at $2\lambda/D$ for a protection angle of $1\lambda/D$ and four at a separation of $1\lambda/D$ for a protection angle of $0.5\lambda/D$.

These results of a diminishing gain from our model with separation and subsequent agreement with ANDROMEDA at about $10\lambda/D$, is consistent with our expectation that the limiting factor of field rotation for the spatial model ceases to be important with larger separation. We note that in general the azimuthal variation of the contrast at each specific radius is consistently and significantly smaller for our model. This could be explained by weaker spatial correlations of small scale structures in the detection map. This difference in azimuthal variation is especially noticeable at separations > 350 mas, where the influence of the wind-driven halo declines (Fig. 4.5).

Changing the number of principal components

To confirm the reliability of our results we perform the same reduction of the data with different complexity of the systematics model. We test the impact of the fraction f of the maximum number of components $N_{\max} = N_{\text{frames}}$ used. Figure 4.9 shows the contrast curve for the same data as used in Fig. 4.8, with $f = 0.1, 0.3, 0.5, 0.7$. Firstly, we note the absence of a trend towards better detection limits with increasing number of principal components. In fact the contrast gets worse. Counter to intuition, this is to be expected and shows that fitting a forward model of the planet signal simultaneous to the systematics model works to counteract over-fitting. By increasing the complexity of the systematics model, the planet model component gets less constrained, resulting in a larger scatter. Secondly, the value of $f = 0.3$ provides good results irrespective of separation, meaning that we do not have to make a significant distinction in the complexity of our temporal model depending on separation, although models that are not sufficiently complex ($f = 0.1$) are performing slightly worse at close separation, which could be related to the presence of the strong wind-driven halo and effects from the (mis)alignment of the coronagraph with the star in addition to the quasi-static speckle pattern. In spatial models one may have to chose a different model complexity based on the separation, trying to compensate self-subtraction effects from using smaller protection angles by using a less complex model, but also having fewer spatial modes to reconstruct.

However, for our temporal approach the existence of a model complexity that fits well globally is in agreement with our expectation and past experiments. Because we do not have a temporal exclusion criterion, the number of frames of our data does not depend on separation. Additionally, a speckle lifetime analysis performed by Milli et al. (2016) for SPHERE did not show a separation dependence of the linearly decreasing speckle correlations on timescales of tens of minutes. The fast evolving and exponentially decaying correlations ($\tau \sim 3.5$ seconds), that are likely caused by turbulence internal to the instrument, did show a slight separation-dependence, but our integration times are too long and the effect too small to be expected to be visible.

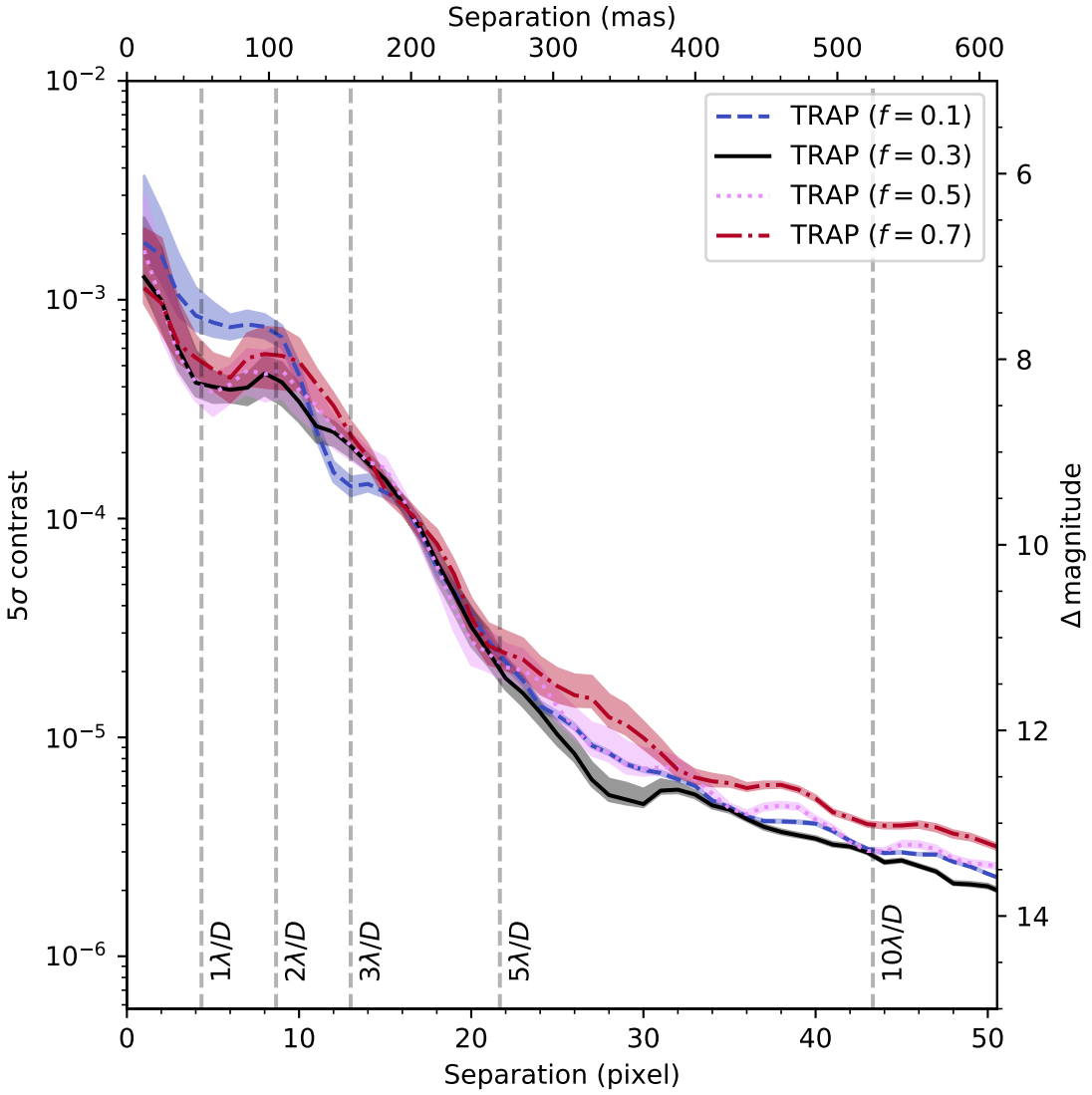


Figure 4.9: Contrast curves obtained with TRAP for 51 Eridani when different fractions of the maximum number of principal components are used.

Changing the temporal sampling

For a temporal systematics model, it is highly probable that the performance of the algorithm scales with temporal sampling. In Fig. 4.10 we repeat our reduction with the same data set binned down by a factor of four (16s to 64s exposures). We again plot the contrast of TRAP reductions with varying principal component fractions f and ANDROMEDA with the standard setting of $\delta = 0.5$ FWHM (top panel) for the same binned data, as well as the factor of gain in contrast we obtain compared to ANDROMEDA (bottom panel). We can still see a gain at very small separations, but the advantage of using a temporal model drops of faster, and may even perform worse at longer separations. From $\sim 5\lambda/D$ we do not see any improvement and the improvement at small separations is smaller. This is consistent with our expectation of temporal models being able to capture more systematic variations with a denser time sampling.

Table 4.2: Photometry and SNR for 51 Eridani b

Method	Bin 4x	Contrast (10^{-6})	SNR
TRAP	no	6.690 ± 0.787	9
ANDROMEDA	no	6.323 ± 0.801	8

Figure 4.11 shows the advantage in contrast when using the unbinned (16s exposure) compared to the binned (64s exposure) data. We see that temporally binning the data has an adverse effect on the contrast over almost the whole range of separations. Averaged over separations between 0.5 and 10 λ/D, we see a 40% gain in contrast by using the faster temporal sampling. We note that both of the sampling rates shown here are too coarse to model the short-lived speckle regime, and expect a further improvement by reducing the exposure time by another factor of four or more ($\leq$4s).

4.6.2 Reconstructing the companion PSF

We further demonstrate the fidelity of the algorithm by extracting an image of the companion's PSF from the data. This image is shown in the left panel of Fig. 4.12 and the temporal mean of the residuals are shown in the right panel.

This image was obtained by, 1) subtracting the best-fit model of the planet from the data, 2) fitting the coefficients for the temporal systematics for the planet position without including the planet forward model, 3) subtracting the best-fit systematics model from the data including the planet, 4) cutting out 2D stamps at the position of the planet in each frame and align them at sub-pixel precision, 5) weighted average along time-axis using the residuals of the planet and systematics subtracted stamps as weights.

With this test we demonstrate that even for data taken in challenging conditions, we can extract the PSF of the planet with enough fidelity to clearly detect the first airy ring. It can be seen that our algorithm does not suffer from any self-subtraction effects as traditional spatial models do. Additionally, as we cut out 2D stamps instead of derotating the frames we do not distort the shape of the PSF.

4.6.3 β Pic: continuous satellite spot data with short integrations

A comparison between the detection maps obtained with TRAP and ANDROMEDA is shown in Fig. 4.13. For both reductions we used the same pre-reduced and aligned input data cubes, and the reduction parameters were the same as for the above 51 Eridani data ($f = 0.3$ for TRAP and $\delta = 0.5$ FWHM for ANDROMEDA). We focus our discussion on one of the two channels (H2). The color scaling differs to account for the difference in SNR of the detection. With an SNR of 40, the TRAP detection is about four times higher than in ANDROMEDA. The SNR in H3 is even slightly higher, as the contrast for β Pic is favorable at these wavelengths. Table 4.3 shows a summary of the photometry results obtained for all reductions of β Pic b discussed in this section.

Figure 4.14 shows the obtained contrast analogous to Fig. 4.8. Qualitatively, we see the same effects as for 51 Eridani, an increasingly significant gain at small separations by using

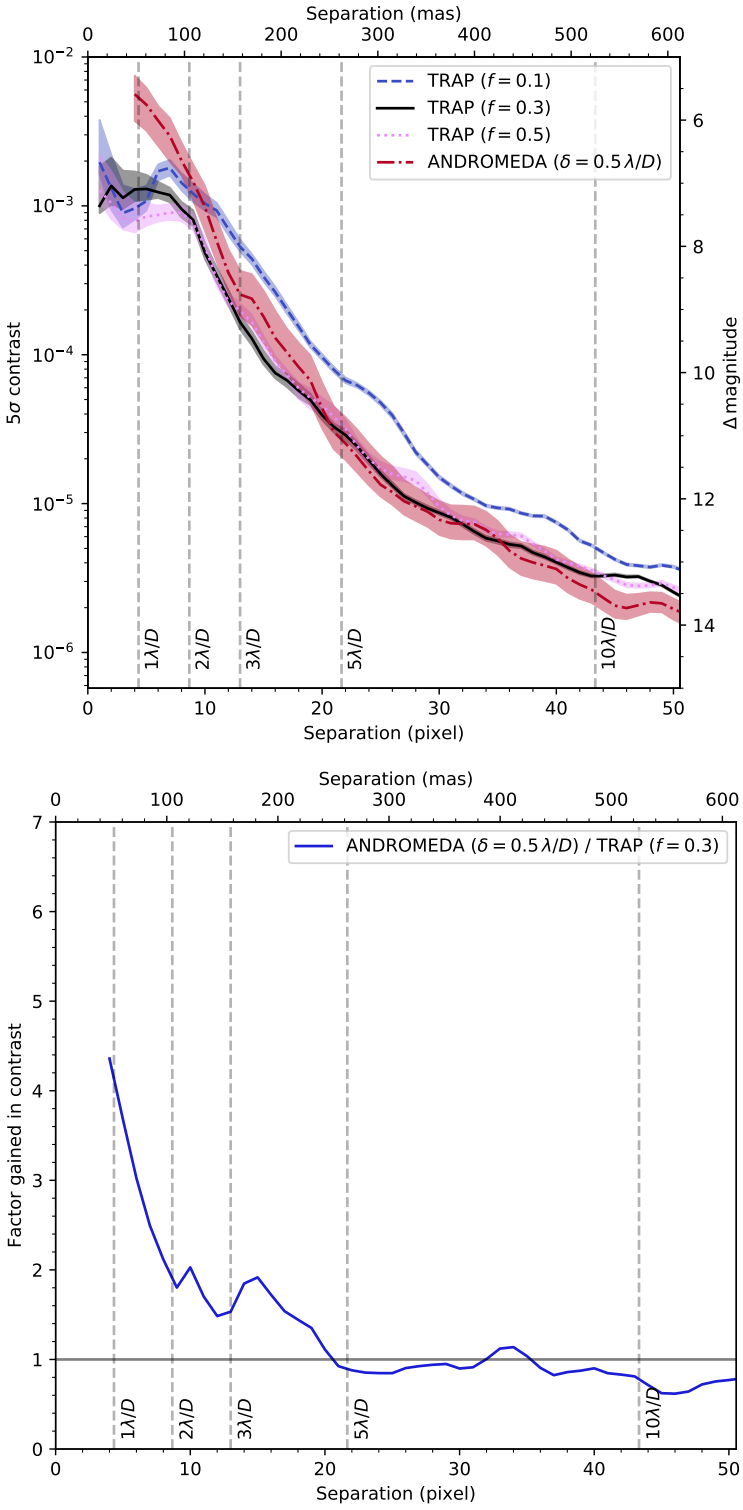


Figure 4.10: Same as Figure 4.8, but each four frames of the 51 Eri b are temporally binned to achieve a DIT of 64s.

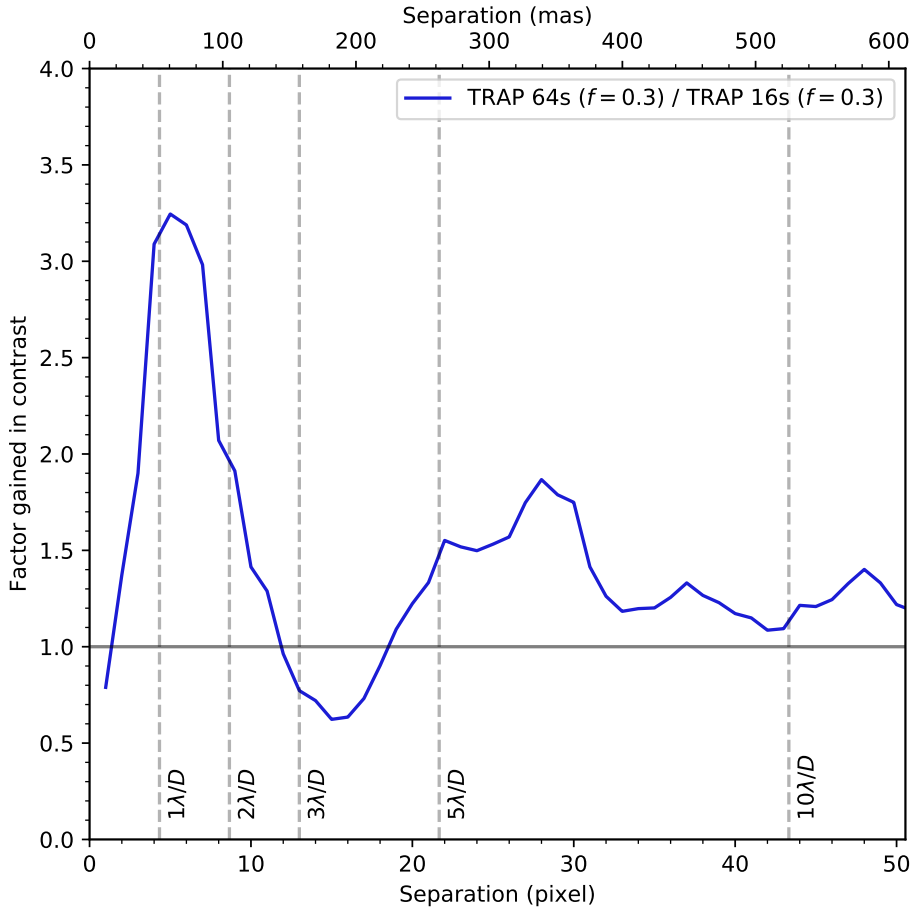


Figure 4.11: Effect of using faster temporal sampling DIT 16s (unbinned) vs slower sampling with DIT 64s (binned) on contrast obtained with TRAP.

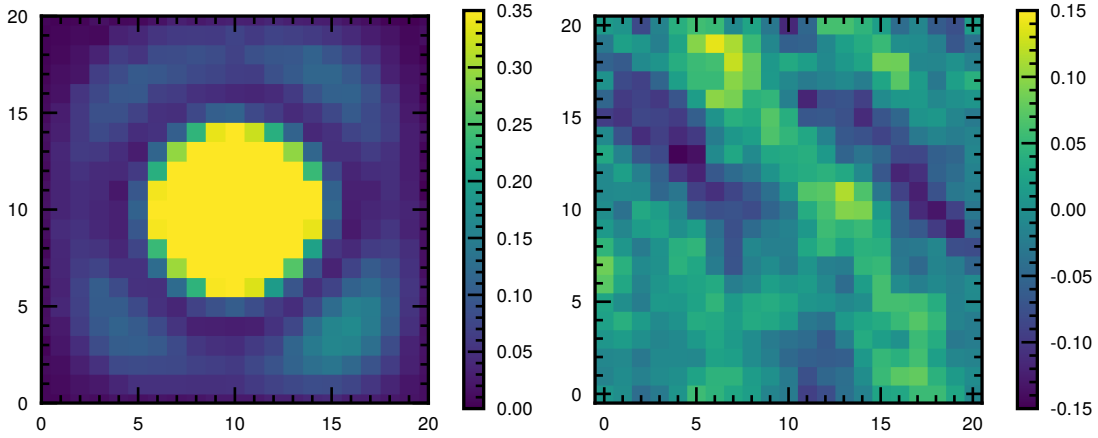


Figure 4.12: (left) Extracted image of 51 Eri b with temporal systematics removed. (right) Temporal mean of residual. (x,y)-axes in pixel and flux in arbitrary units.

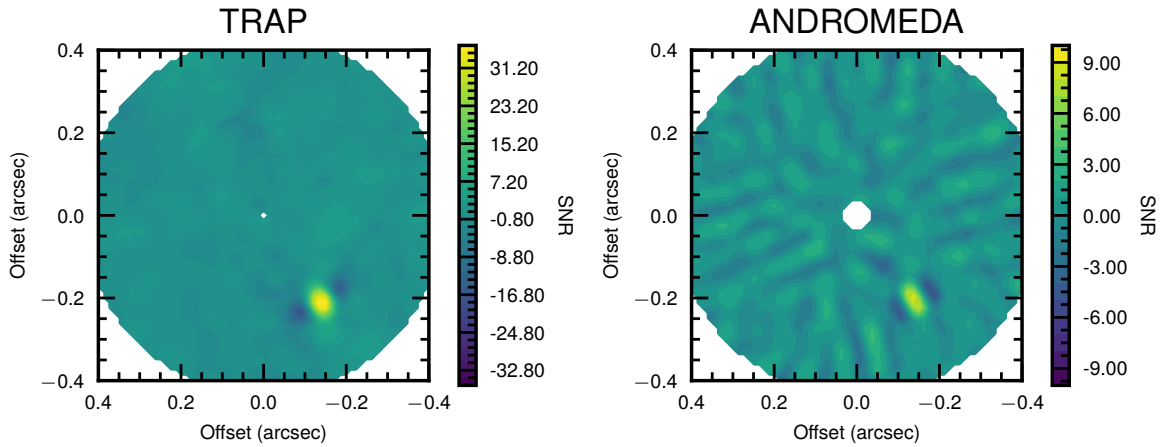


Figure 4.13: Contour plot of normalized detection maps obtained with TRAP ($f = 0.3$) and ANDROMEDA ($\delta = 0.5\lambda/D$) for β Pic. These maps must not be confused with derotated and stacked image. They represent the forward model result for a given relative planet position on the sky (ΔRA, ΔDEC), i.e. the conditional flux of a point-source predicted by the forward model given a fixed position, corresponding to trajectory over the detector (all pixels affected during the observation sequence).

a temporal model. We see an even more pronounced reduction in azimuthal variation in contrast, i.e. the "width" of the contrast curve. For this data, we additionally see a baseline increase in performance between $50-200\%$ at larger separations ($> 3\,\lambda/D$), that we attribute to the better temporal sampling. To confirm this hypothesis, we have binned down the data by a factor of 16 to obtain 1 minute exposures. The normalized detection map obtained with TRAP is shown in Fig. 4.15 and the contrast compared to the unbinned ANDROMEDA reduction is shown in Fig. 4.16. The planet signal is detected with a SNR of about 13, only slightly better than the SNR obtained with ANDROMEDA on the unbinned data. The obtained contrast curve is also comparable to ANDROMEDA at this separation. We therefore attribute the SNR improvement by a factor of four to the fact that our algorithm is capable of taking into account the bulk of short-timescale variations. We note that at the separation of β Pic b for the epoch of the observation ($\sim 6\,\lambda/D$), the exclusion time for spatial models is on the order of minutes even for very small protection angles ($\delta = 0.3$ FWHM) and above 10 minutes for the standard setting of $\delta = 0.5$ FWHM (see Fig. 4.2).

Applying the algorithm to unaligned data

As we use a non-local, temporal systematics model and do not attempt to reconstruct a spatial model of how the speckles "look", we can in principle forego aligning the data and run the algorithm on minimally pre-reduced (background and flat corrected) unaligned data. To demonstrate this, we measure the center based on the satellite spots for each frame, and use this varying center position to construct the lightcurve model for the planet, i.e. we do not shift the frames but the forward model for our planet. We also apply the anamorphism correction for SPHERE to the relative position of the planet by reducing the relative separation of the model by 0.6% in y-direction (Maire et al., 2016b), instead of stretching the images. We also modulate the contrast of the planet model with the satellite spot amplitude

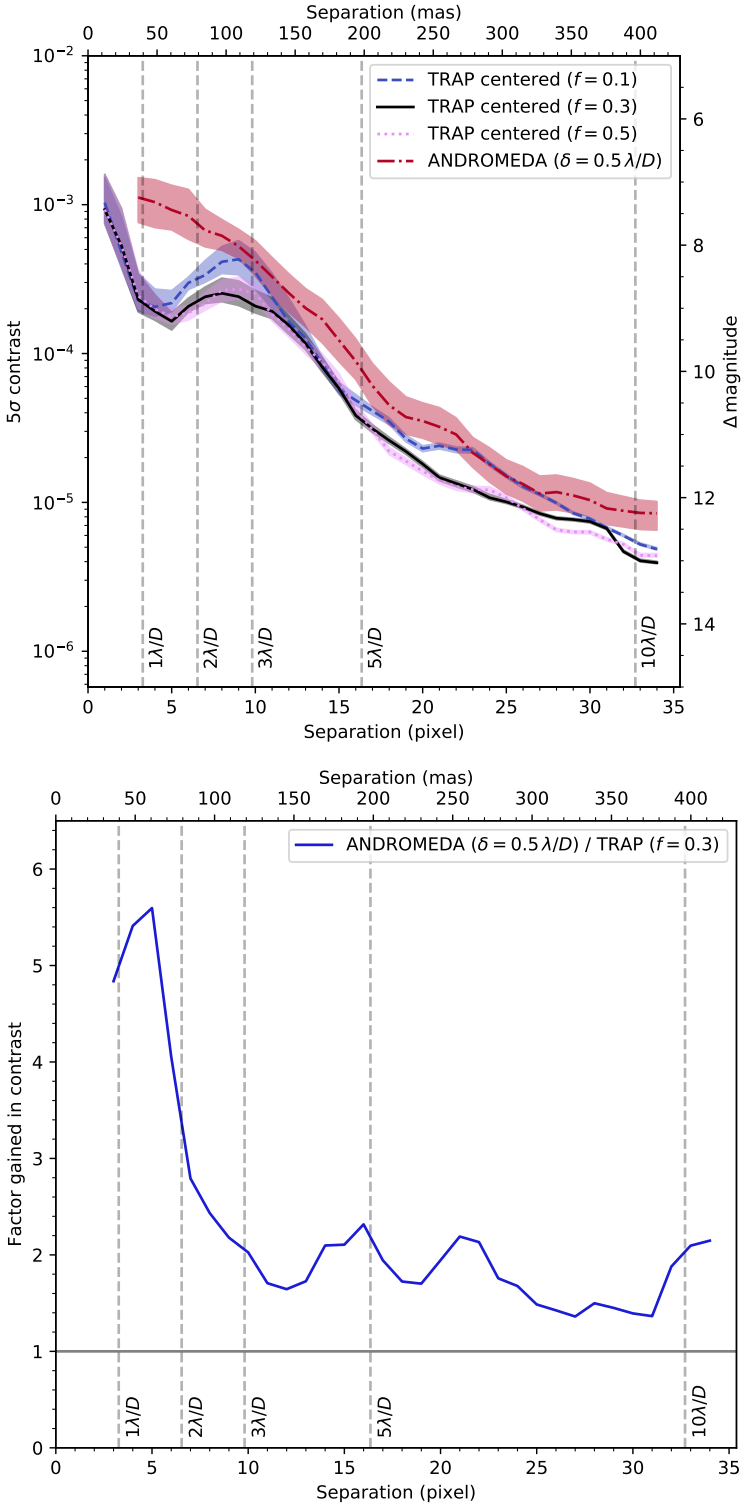


Figure 4.14: Comparison between the contrast obtained with TRAP and ANDROMEDA reductions for β Pic using the same input data for the H2 band. TRAP has been run with 30% of available principal components, whereas the ANDROMEDA reduction correspond to a protection angle of $\delta = 0.5$ FWHM. (top) The shaded areas correspond to the 14%-84% percentile interval of contrast values at a given separation. (bottom) Contrast gained by using TRAP.

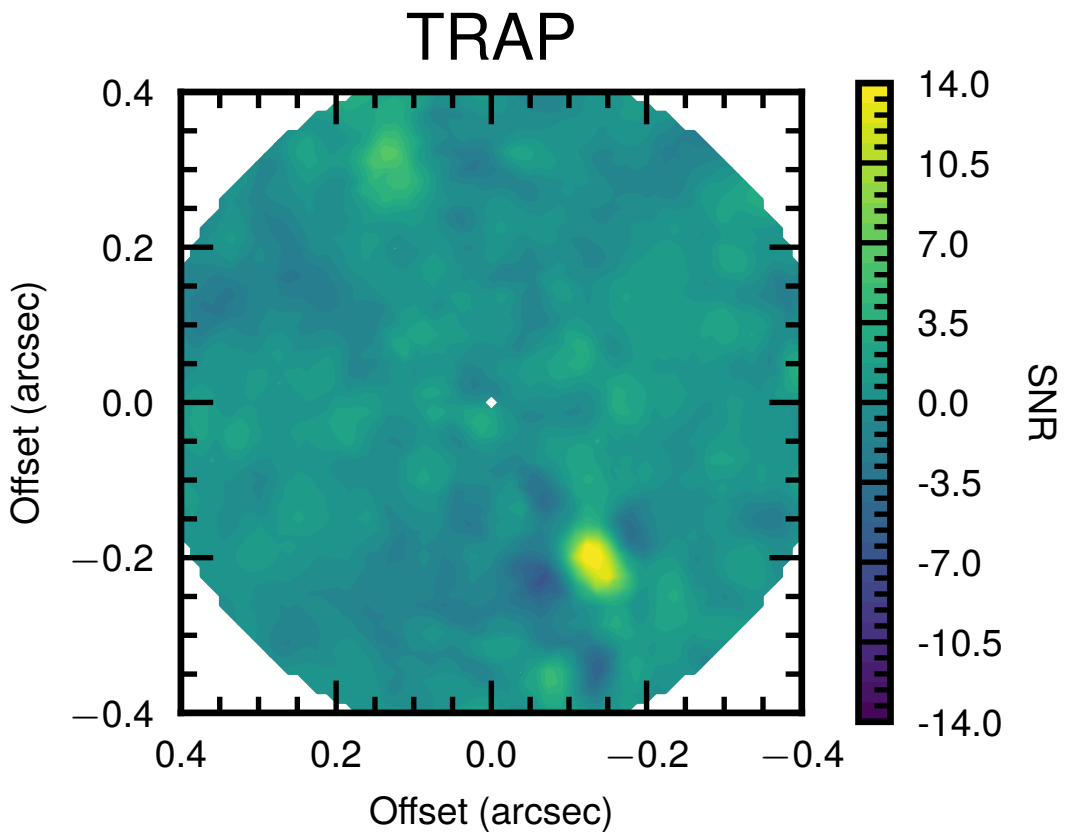


Figure 4.15: Contour plot of normalized detection maps obtained with TRAP ($f = 0.3$) binned data of beta Pic (16x binning, 64s exposures).

variation of each frame. The result is shown in Fig. 4.17, with the aligned data on the left and the unaligned data on the right. We note that the SNR of the detection is virtually the same. There are only slight difference in the residuals structures. We do see a blob above the position of β Pic b that edges above 5σ in the reduction on unaligned data. It is difficult to evaluate the veracity of these structures due to the presence of disk structure in the β Pic system.

The contrast curves for 1) aligned data without planet brightness modulation, 2) including planet brightness modulation, and 3) unaligned data with planet brightness modulation are shown in Fig. 4.18. Taking into account the brightness modulation derived from the satellite spots does not have a noticeable impact on the contrast limits. It does have a minimal impact on the derived contrast of β Pic b, as it reduces the flux calibration bias incurred by assuming an average or median contrast for the planet flux in the planet model, instead of the real distribution. In the case of this observation the scatter of brightness variation is roughly Gaussian with a variability of $\sim$6% centered on the mean of the satellite spot brightness, without a large systematic trend. Taking into account this variation becomes more important the more unstable the conditions are and when a correlated trend is present.

A bigger difference can be seen in the reduction of the uncentered data. The contrast appears to be worse in the inner-most region covered by the coronagraphic mask, but slightly

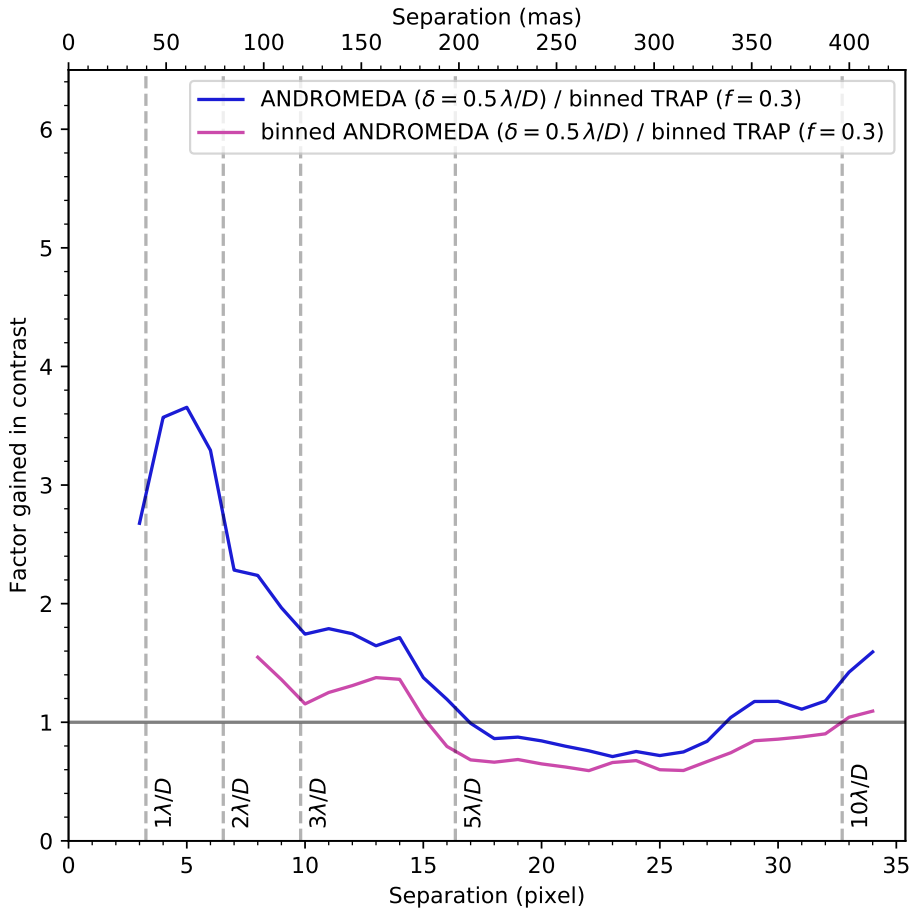


Figure 4.16: Contrast ratio between TRAP ($f = 0.3$) reduction on temporally binned data (16x binning, 64s exposures) and ANDROMEDA reduction of the same binned data, as well as unbinned data, of β Pic.

better outside of $3\lambda/D$. The SNR of β Pic b, again, is virtually the same as in the centered case, but astrometry and photometry change slightly.

Table 4.3: Photometry and SNR for β Pic b

Method	Mod.	Align	Bin 16x	Contrast (10^{-4})	SNR
TRAP	no	yes	no	1.330 ± 0.035	38
TRAP	yes	yes	no	1.318 ± 0.035	38
TRAP	yes	no	no	1.244 ± 0.033	38
TRAP	no	yes	yes	1.290 ± 0.085	15
ANDROMEDA	no	yes	no	1.196 ± 0.128	9
ANDROMEDA	no	yes	yes	0.963 ± 0.121	8

Overview of reduction results depending on whether waffle amplitude modulate, pre-aligned (centered) data, and temporal binning is used. All TRAP reductions used $f = 0.3$ and ANDROMEDA reductions $\delta = 0.5$ FHWM.

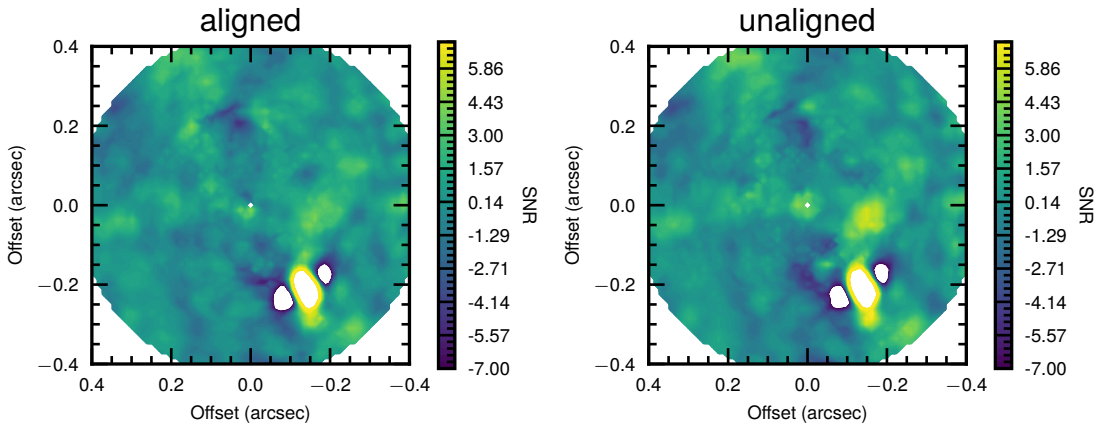


Figure 4.17: Contour plot of normalized detection maps obtained with TRAP ($f = 0.3$) on aligned and unaligned data for beta Pic. These maps must not be confused with derotated and stacked image. They represent the forward model result for a given relative planet position on the sky (ΔRA, ΔDEC), i.e. the conditional flux of a point-source predicted by the forward model given a fixed position, corresponding to trajectory over the detector (all pixels affected during the observation sequence).

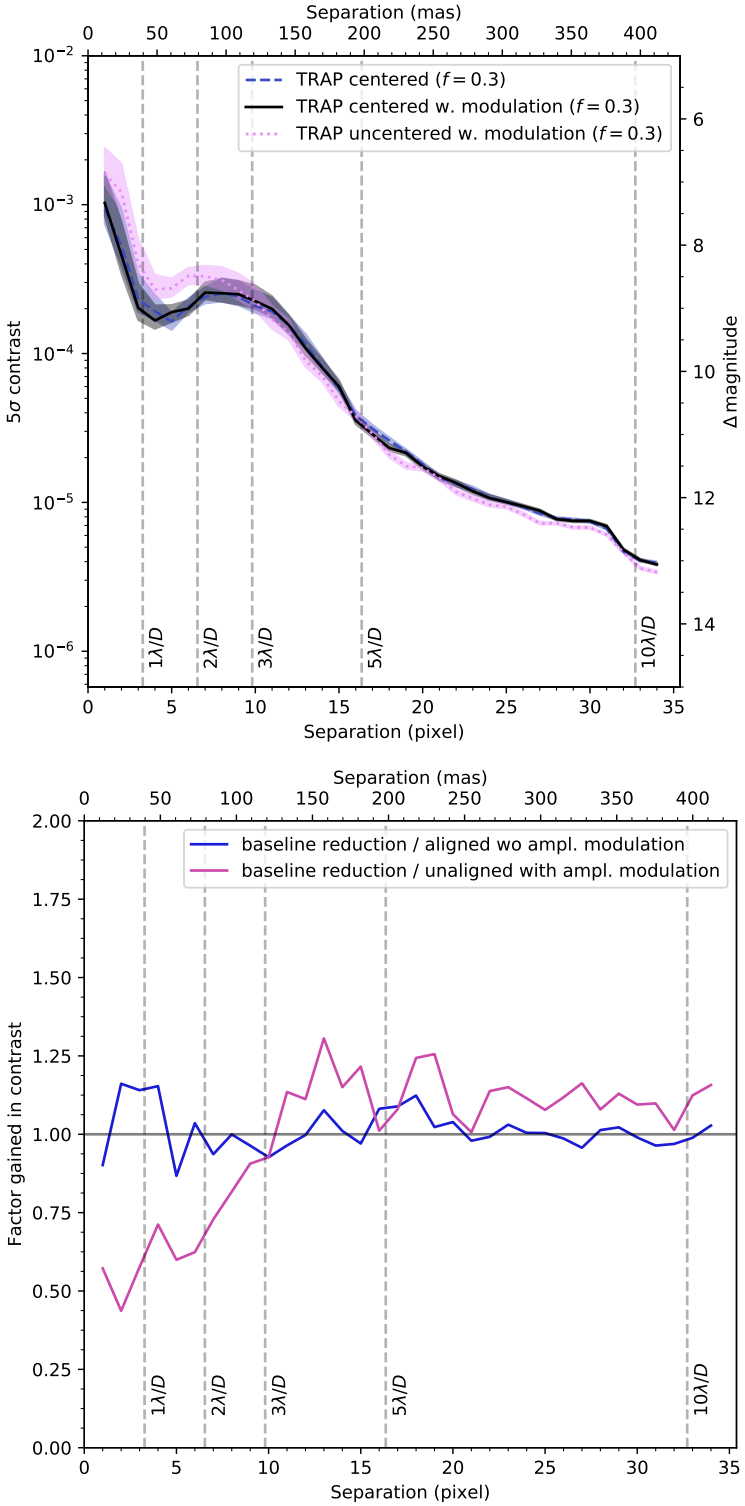


Figure 4.18: (top) Comparison between the contrast for β Pic obtained with TRAP on 1) centered data not taking into account brightness modulation, 2) taking it into account, and 3) on uncentered data with brightness modulation. TRAP has been run with 30% of available principal components. The shaded areas correspond to the 14%-84% percentile interval of contrast values at a given separation. (bottom) Contrast gain compared to centered data without including amplitude variations (baseline reduction).

4.7 Computational performance

The computational time needed to reduce the 51 Eridani data set (256 frames) at one wavelength up to a separation of ∼45 pixel (∼550 mas) for standard parameters ($f = 0.3$, PSF stamp size 21x21 pixel) is about 170 minutes on one core on a relatively new laptop (intel CORE i7, 7th Gen, 16 GB memory). The algorithm is parallelized on the level of fitting the model contrast for a given position, such that the grid of positions to explore is divided among the available cores. At the current version of the code, using four cores roughly halves the time to about 90 minutes, because of inefficiencies in memory sharing. We expect a close to linear relation with the number of cores after improvements to the code's parallelization architecture.

4.7.1 Scaling with number of frames

It is noteworthy that this algorithm computational speed scales better with the number of frames in the observation sequence than traditional spatial-based approaches that include a protection angle. The absence of an exclusion criterion means that the principal component decomposition has to be performed only once for one assumed companion position, instead of having a separate training set for each frame in the sequence. The only increase in computational time stems from the need of decomposing a larger dimensional matrix once per model position and subsequently inverting a larger system of linear equations for each pixel. Testing the computational time for different binning factors, we note that the computational time, scales roughly linearly with the number of frames $t \propto N_{\mathrm{frames}}$, i.e. doubling the number of frames, roughly doubles the computation time.

4.7.2 Scaling with outer-working angle

We select a new training set depending on the position, since we have to exclude the reduction area from the training set. As such, the time spend on constructing the model is linearly proportional to the number of positions tested, which, when exploring a linear parameter space in ΔRA and ΔDEC means that the number of PCAs performed is proportional to the search area, such that $N_{\mathrm{training}} \propto r^2$, where r is the separation from the central star. At the same time the number of pixel affected by a potential target also increases with separation $N_{\mathrm{pix,\ affected}} \propto r$. In terms of computational time, however, the time spend on PCA is relatively minor (once per tested position), and the scaling with number of affected pixels outweighs. Testing the algorithm with increasing outer-working angle (OWA), we derive a power-law index of $t \propto \mathrm{OWA}^{2.3}$. If computation time is an issue, one could easily use our algorithm exclusively for the inner-most region and combine the results with an algorithm that scales better with separation further out, as we do not expect big performance improvements at large separations compared to other algorithms.

4.8 Discussion

It is always challenging to directly compare a new approach with existing pipelines, due to numerous differences in implementation, a wide range of possible reduction parameters, test

data taken in various conditions and observing modes, and differences in statistical evaluation of the outputs. No comparison will ever be complete. For our work, we have chosen ANDROMEDA as the most viable representation of the spatial systematics modeling class of algorithms, because its likelihood-based forward modeling approach is analogues to our own and allows us to directly evaluate and compare the outputs in a fair way. Additionally, ANDROMEDA has been shown to compare well with other established PCA and LOCI-based pipelines (Cantalloube et al., 2015; Samland et al., 2017).

4.8.1 Applicability to extended objects

One aspect of non-local models and regressor selection that has not been discussed in detail in this work and will require future research, is the topic of disks. The extracted image of 51 Eri b and its airy ring in Fig. 4.5 demonstrates the capability of our temporal algorithm to extract the undistorted shape of an object, unaffected by self-subtraction. This could be an invaluable property for disk imaging, where preserving the morphology of the object is paramount.

It will be relatively straightforward to mask out edge-on disks from the pool of reference pixels (regressors), whereas the optimal choice of regressors for face-on disks will have to be investigated in the future. Given typical field rotations, a significant fraction of relevant pixels on the detector will be affected by disk signal at some point during the sequence. Although it may be as simple as choosing the training set from a different separation range instead, one should still make sure that the result is unbiased. It should also be pointed out that a completely azimuthally homogeneous structure will not be picked up by the algorithm in its current form, because we include a constant term in our fit. This degeneracy has to be broken for example by including a self-consistent 2D disk model. Also, as soon as a pixel registers times with disk flux and times without, this degeneracy is immediately broken.

4.9 Summary and conclusions

We have presented a new paradigm of using a temporal, non-local systematic models to more effectively search for point sources at the scientifically most interesting small separations where traditional ADI algorithms have issues by design. Our implementation, called TRAP, is open-source and will be publicly available.

We have shown on two test data sets that the TRAP pipeline performs as good or better than the spatial approach with a strong improvement in contrast by a factor between 1.5 and 6 at angular separations $< 3\lambda/D$. Beyond this separation, the improvement strongly depends on the temporal sampling of the observation sequence. Increasing the integration time from 16s to 64s for the 51 Eridani b data set led to a decrease in gained average contrast by $\sim$40%. Regardless of integration time, the azimuthal variance of the achieved contrast at any specific separation is strongly reduced using the temporal systematics model compared to a spatial systematics model.

We have shown that for short integration times (4s, β Pic), we can achieve a significant overall improvement of the contrast by a factor of two, even at separation between 3 – 10 λ/D. The SNR measured for β Pic b significantly increased from about $\sim$10 with the spatial model to about $\sim$40, making full use of the systematic information on short time

scales. We conclude that the effect of exposure time on the achievable contrast is under explored in the literature, because spatial algorithms used are not able to make optimal use of the information contained in short time-scale variations, as typical exclusion time-scales are much larger than exposure times at short separations.

We have shown that fitting the planet and noise model simultaneously constitutes a self-regulating process on the achieved contrast when we increase model complexity: increasing the systematics model complexity (i.e. fraction of maximum number of principle components used) does not automatically lead to "better" contrasts. On the contrary, the constraints on a potential planet's contrast get worse.

Furthermore, we demonstrated that we can apply our temporal approach to minimally pre-reduced (dark and flat corrected) data without aligning the frames. This is achieved by adjusting the forward model position that generates the companion lightcurve according to the star center position, the anamorphism, and excluding all bad pixels from the training and reduction sets instead of interpolating them. This reduction achieves very similar results to elaborate pre-reductions and reduces the need for data manipulation steps (shifting, resampling, interpolation). This also leads to a strong reduction in processing time needed for alignment and bad pixel interpolation, which can take more than an hour for data sets with many exposures. Like ANDROMEDA, our algorithm does not require derotation of frames. Spatial filtering, which improves ANDROMEDA performance and is also used in pyKLIP, is not needed for our algorithm to perform well. The ability to post-process data without re-sampling could prove beneficial for SPHERE-IFS in the future, as IFS uses a hexagonal lenslet geometry. The output images have to be resampled to a rectilinear grid for traditional post-processing pipelines. With TRAP we can perform the analysis on the native image geometry.

We do not recommend dithering for pupil tracking data as it is not necessary and can interfere with the performance of temporal models. We further recommend the use of continuous waffle mode to improve the forward model performance with accurate center and amplitude variations. We strongly recommended exploring shorter integration times for observation sequences. Decreasing the integration time of IRDIS from 64s to 4s increases the observation overheads by about 13 percent points. If our scientific interest is companions at small separations, we are in the speckle limited regime and the increase in read-noise is likely negligible, easily balanced by the manyfold increase in algorithmic performance.

This work demonstrates the potential of the non-local, temporal systematics modeling approach. However, there are still many possible avenues for future improvements. The next step will be extending the algorithm to take into account spectral information. This can easily be achieved by adding the lightcurves of pixels at other wavelengths to the reference set. Currently, the area of the detector affected by the signal of interest is excluded from the training data. By shifting this area inward/outward proportional with wavelength and adding those signal-free pixels to the training set, it is possible to add an "autoregressive component" to the model that traces the temporal behavior of the exact speckles at the position of interest. Again, this can be achieved without rescaling the raw data or prescribing a detailed chromatic behavior other than the rough wavelength scaling.

Another important step will be improving the fidelity of the forward model, e.g. directly

including the coronagraphic transmission model in the forward model of the signal. This is a more consistent approach than post-hoc adjusting the contrast curve by the coronagraphic transmission. As we push to lower and lower inner-working angles, a good understanding of the coronagraph at all wavelengths will become important and should be modeled as well as measured on-sky as part of publicly available instrument calibrations. There is a multitude of effects that, in principle, can be included in the forward model, like distortion of the PSF shape (e.g., variations in Strehl, low-wind effect), smearing caused by integration time, geometric distortion.

The temporal modeling approach may also prove beneficial for instruments with large pixel scales, or undersampled PSFs, as is sometimes the case for IFUs. Large pixels/spaxels can exacerbate the problems caused by insufficient field-of-view rotation. An interesting use-case, that has not been explored in this work, is the application to space-based observatories. For observations taken at different roll angles we can also build a temporal forward model that would then take the form of a step-function for affected pixels. This, again, would allow us to take into account systematics on smaller time scales, e.g. from instrument jitter.

The algorithm as introduced in this work is optimized for companion searches over a grid of possible positions. If our goal is the detailed characterization of a planet of known position – or even a disk – this approach may not be optimal. Future development should include exploration of specific models using samplers, such as the nested-sampling approach. This would allow, both, the exploration of more complicated physical models such as debris disks or binary objects, as well as more sophisticated systematic models. Currently, our algorithm implicitly assumes that all necessary information on the systematic temporal trends is included in other pixels. It has been shown for transit photometry that "missing" information on systematic trends can be included in a Gaussian process trained on auxiliary data (Gibson, 2014). It is possible that including additional external information (e.g., on the wind, state of the AO, position of derotator, temperature, focal-plane wavefront sensing), could further improve the algorithmic performance significantly. Using nested-sampling, combined with a negative model injection approach, would provide an easy way to include all available information and perform direct model comparison based on Bayesian evidence.

Lastly, future and current development and deployment of coronagraphs with smaller inner-working angles of $1\lambda/D$ will further increase the importance of our algorithm.

5 Adaptation of the CHARIS Pipeline for SPHERE-IFS

The work presented in this chapter will be the basis for a refereed article of the same title for which I am the lead author.

Overview: In this chapter I present the current state of my work to improve the quality and reliability of the spectral extraction of high-contrast imaging IFS data obtained with SPHERE. In order to accomplish this goal, I am adapting the open-source, python-based, CHARIS instrument pipeline for SPHERE-IFS. CHARIS is an integral field spectrograph designed for high-contrast imaging and is similar in design to SPHERE-IFS, which allows us to create a common code base. Extracting image cubes from raw IFS data is a complicated process, leaving many areas in which improvements can be made. The CHARIS pipeline already implements many ideas that will be a major upgrade compared to the official ESO pipeline, and solves a number of known issues. It is well documented and open to the public, which will help to create community involvement, transparency, and future support. The goal is to have a completely independent confirmation of the spectra obtained with the IFS instrument, and rule out any potential systematic biases that could skew the interpretation of atmospheric models of exoplanets. I have used the CHARIS pipeline at its current stage of implementation to convert the raw microspectra data into a fully calibrated image cube, and extracted the spectrum of 51 Eridani b using the data and post-processing methods described in (Samland et al., 2017). Using this independent calibration of the raw data, I confirm my previously published J-band flux measurement.

At the end of this chapter I will give an overview about the potential synergies between the CHARIS pipeline for spectral extraction and the TRAP algorithm presented in the previous chapter, that will turn SPHERE-IFS into the true workhorse for exoplanet detection at small angular separations among the SPHERE instruments.

5.1 Introduction

In recent years a new generation of integral-field spectrographs (IFSs) on 8-meter class telescopes have been developed especially with the purpose of aiding the direct detection and characterization of extrasolar planets. Integral-field spectrographs allow us to obtain spectral information across the field-of-view (FoV) of the instrument, hence providing spatially resolved images at a wide range of wavelengths simultaneous. IFSs have proven to be a pow-

erful asset in the tool-box of high-contrast imaging; not only can the spectral information be used to characterize the properties of any astrophysical object detected, it can also be used to exploit the different chromatic behavior of diffraction speckles from a real astrophysical source to suppress noise and improve detection contrast.
These recent instruments combine the large angular resolution of the telescope with extreme adaptive optics (XAO) to provide near diffraction limited images of stars and their surrounding. The difference in brightness between the star and potential planets necessitates the use of further technologies like coronagraphs and dedicated observation strategies to suppress the central star's light, and help distinguish between the unavoidable residual starlight halo and a genuine astrophysical signal in its direct vicinity.

There are currently four such dedicated (XAO) high-contrast imaging instrument equipped with an IFS: GPI on Gemini South (soon Gemini North) (Macintosh et al., 2008), ALES at the LBT (Skemer et al., 2015), SPHERE at the VLT (Beuzit et al., 2019), and CHARIS (Peters et al., 2012; Groff et al., 2015) on the Subaru telescope. All of these instruments use a similar technology, lenslet arrays, to map spatial elements on the sky onto small spatially separated spots that will be dispersed onto the detector as distinct, mostly non-overlapping spectra. Although these instruments differ in implementation details and design (e.g., wavelength range, resolution, FoV-size), they are sufficiently similar for a suitably flexible software package to be adapted to the peculiarities of each individual instrument. Such a software framework will encourage co-operation between instrument teams, improve maintainability and reliability, as well as comparability of results obtained with different instrument, and by necessity, be open-source and well documented. Given the complexity inherent in the reduction and processing of IFS data, individual teams, after investing a large part of their budgets in the initial development and building of an instrument, often struggle to provide the manpower for continuous support and improvement of the software used for data reduction. This can be remedied by providing the scientific community access to source-codes and encourage community involvement in further development. This will encourage active improvement of the reliability of code, and remove "black box" elements that the user has to take for granted.

In this work we present our effort to take the first step in this direction by adapting the CHARIS data reduction pipeline (Brandt et al., 2017), an open-source package written in Python and Cython available on Github[14], to be able to process SPHERE-IFS data into scientifically usable products. The biggest difference of the SPHERE-IFS compared to the other IFS instruments is its hexagonal lenslet geometry and will receive special attention in this work, since this fact is not widely known in the community and has implications for future improvements of data reduction algorithms.
This chapter is structured as follows. Section 5.2 gives an overview of the design properties and peculiarities of the SPHERE-IFS, and how it differs from CHARIS. Section 5.3 summarizes the capabilities of the CHARIS pipeline. It also gives an overview about the currently available SPHERE pipeline and potential short-coming. Section 5.4 shows the detailed steps performed to obtain the necessary calibration data, and the static calibration products incorporated into the pipeline. Section 5.6 shows results obtained using the new pipeline on SPHERE-IFS data and compares the result to the official pipeline. Finally, we conclude and provide an outlook for future improvements in Section 5.7.

[14]https://github.com/PrincetonUniversity/charis-dep

Table 5.1: Comparison of CHARIS and SPHERE-IFS characteristics

Parameter	CHARIS	SPHERE-IFS
Detector	2048x2048 H2RG	2048x2048 H2RG
No. of lenslets	135x135	152x152
Lenslet lattice configuration	square	hexagonal
Lenslet area	16.4 mas^2	19.2 mas^{2,a}
Field-of-view	2.2" x 2.2"	1.73" x 1.73"
Wavelength coverage	1.15 – 2.38 µm	0.95 – 1.66 µm
Microspectrum length	∼30 pixel	∼35 pixel
Separation between spectra	∼7 pixel	∼5 pixel
R = $\lambda/\delta\lambda$	∼ 75 (high-res); ∼20 (low res)	∼55; ∼35
Available modes	J, H, K; J–K	Y–J; Y–H

Basic characteristics of both IFU instruments. $^{(a)}$ Area of hexagonal lenslet on sky. Resampled by pipeline to ∼7.4 mas^2 / pixel.

5.2 SPHERE Integral-Field Spectrograph

The SPHERE IFS is part of the SPHERE instrument (Beuzit et al., 2008) located on the stable Nasmyth platform of the ESO Very Large Telescope (VLT), as part of a system of three dedicated high-contrast imaging instrument: ZIMPOL (visible light polarimetry, Thalmann et al. 2008), IRDIS (dual-band imager, Dohlen et al., 2008), and IFS (integral field spectrograph, Claudi et al., 2008). These instruments are behind a Common Path and Infrastructure (CPI) module which feeds the subsystems with a highly stabilized, AO corrected beam. The SAXO XAO system for SPHERE achieves Strehl ratios better than 90% (Petit et al., 2012) in the near-infrared (NIR). The deformable mirror (DM) is described in Hugot et al. (2012).
In this paper we will focus exclusively on the IFS subsystem and its similarities and differences to CHARIS. Table 5.1 shows a comparison of basic design characteristics for SPHERE-IFS and CHARIS.

In the context of this paper there are several noteworthy aspects and differences between these two instruments, although the overall design principles are the same. Both instruments are fed by extreme adaptive optics corrected beams, use H2RG detectors of the same dimension, lenslet arrays as the method of choice to keep neighboring spectra distinct on the detector, and prisms as dispersive element. Let us discuss the difference that do not directly impact the data reduction first. CHARIS has a wavelength coverage that extents further into the near-infrared, including the K-band. As such it has to be cryogenically cooled, whereas SPHERE-IFS is operating at warm temperatures. Both instruments have a lower and higher resolution mode: in the case of CHARIS, higher-resolution can be used to study individual NIR bands, whereas the low-resolution mode covers the whole spectral range from J- to K-band. SPHERE-IFS has two modes covering YJ or YH, with different spectral resolutions. In YJ-mode, the H-band is simultaneously observed with the IRDIS dual-band imager, whereas YH-mode is usually combined with simultaneous K-band observation. These differences in wavelength coverage, resolution, and also the difference in field-of-view size, do not fundamentally change the way the data reduction is done. Except

that for each mode instrument specific calibration data has to be taken.

The main difference between the two instruments are the lenslet array design, the calibration unit, and the readout mode of the detector. While CHARIS (and GPI and ALES) use the more traditional TIGER lenslet array design with a square grid of lenses, SPHERE uses a novel hexagonal BIGRE design (Antichi et al., 2009) which is especially optimized to reduce cross-talk (cross contamination) between both adjacent lenslets and adjacent spectra for a given packing density of spectra on the detector. As a result the image is spatially sampled on a honeycomb grid, and either has to be resampled onto a regular "pixel" grid, or especially taken into account in all post-processing. The ESO pipeline resamples the extracted image cubes to allow them to be handled as normal images, at the cost of a relatively aggressive processing (resampling) step that spatially correlates signals and noise properties.
Another difference is that SPHERE has an internal calibration unit that takes flat fields, dark and background, and wavelength calibration data during the day for calibration purposes and monitoring of the health of the instrument. This saves observation time during the night for actual science data. The calibration unit consists of: a) a series of lamps, both narrow and broad, for taking both detector flat fields and flat fields taken through the IFU (integral field unit), and b) a series of monochromatic lasers for wavelength calibration. While the distinction between detector flat and lenslet flat is the same for both instruments, there is a difference in the way the wavelength calibration is performed. For SPHERE the 'wavecal' calibration template that is routinely executed takes images with three (YJ-mode) or four (YH-mode) lasers simultaneously (see left panel, Fig. 5.2). CHARIS on the other hand has a supercontinuum source with tunable narrow-band filter of 5-nm width to take a series of nearly monochromatic spot images for various wavelengths to perform the wavelength calibration.
Lastly, SPHERE employs correlated double sampling for reading out the detector, while CHARIS saves all the reads of each exposure.

5.3 CHARIS Pipeline

The CHARIS pipeline (Brandt et al., 2017) is an open-source, python based software package available on Github. Some routines have been implemented in Cython to optimize performance. It has been – and is being – developed to reduce data obtained with the CHARIS instrument (Peters et al., 2012; Groff et al., 2015) attached to the SCExAO extreme-adaptive optics system (Jovanovic et al., 2015b) at the Subaru telescope in Hawaii. It provides all necessary tools to go from raw detector reads (sampling up-the-ramp) of the dispersed microspectra images to extracted (λ, y, x) image cubes with associated uncertainties. Aside from its accessible open-source nature, and its ability to fit detector read-out ramps directly, which is not strictly of interest only to IFS instruments, there are some notable improvements that set it apart from other lenslet-based IFS pipelines that are currently in use. These improvements are primarily in the algorithms used to extract the microspectra. The official SPHERE and GPI pipelines, for example, sum up the flux measured along the orthogonal to the dispersion direction in an aperture photometric approach. The CHARIS pipeline in addition this, implements two more sophisticated approaches: an instrument profile weighted sum (optimal extraction) and an entirely lenslet PSF-model based least-square approach.

Here we provide an overview of these pipeline capabilities, for an even more detailed

description of the algorithms we refer the reader to Brandt et al. (2017). The summary is necessary, however, to highlight the differences compared to the official SPHERE pipeline and what has to be done to make the CHARIS pipeline work for SPHERE. We will discuss the capabilities in chronological order of application: (1) measure count rates from individual reads up-the-ramp (if available), (2) calibrate the wavelength solution using monochromatic or narrowband flatfields, (3) extract the microspectra to make the final image cube.

5.3.1 Sampling up-the-ramp (UTR)

The CHARIS instrument, as well as SPHERE-IFS and many other near-infrared instruments, uses a Hawaii2-RG (H2RG) detector in which each pixel has its own amplifier and is read out by comparing its voltage to a reference voltage. This means that pixels can be read out multiple times without resetting the detector, which also happens pixel-by-pixel. Such a detector therefore allows to save a large number of reads for each exposure to disk. The slope in counts over time allows the measurement of the electron count rate and therefore the photon count rate given the gain of the instrument. As CHARIS saves all reads, the CHARIS pipeline implements performance optimized routines to fit the count rate for each pixel. For N reads this yields an improvement of signal-to-noise ratio (SNR) of

$$\frac{SNR_{UTR}}{SNR_{CDS}} = \sqrt{\frac{N\,(N+1)}{6\,(N-1)}}, \tag{5.1}$$

compared to correlated double sampling ($N = 2$) in the limit of uncorrelated read noise.

By introducing additional terms in the χ^2-fit of the ramp other common problems can be addressed. For some H2RG the first read in a ramp is contaminated by an exponential decay of the reference voltage. This effect can be fitted for and corrected using the individual reads. Another highly relevant advantage of individual reads is that they allow to correct for nonlinearity and saturation of pixels, and account for leakage of electrons into neighboring pixels upon saturation. This allows to compensate non-linearity effect, e.g. when seeing improves dramatically during an observation sequence, or generally integrate longer to up efficiency and save overheads.

The read-out electronics in SPHERE-IFS use correlated double sampling (CDS), i.e. using differences in counts of a restricted number of samples at the beginning and end of the exposure to determine the count rate. Only the resulting count rate is saved. While this has benefits in terms of data volume and data rates, it constitutes a loss of information and prevents us profiting from the above stated advantages. The contamination of the first read of an exposure as been confirmed for SPHERE-IFS as well during commissioning and is currently being mitigated by introducing a 0.2 second delay for the first read, increasing the overhead especially for short exposure time observations. My research in post-processing shown in Chapter 4 suggests that short exposure times can have big advantages for the modeling of systematic noise in the data (speckles), significantly improving the achievable contrast. Implementing an UTR mode for SPHERE would certainly be a cost efficient option for a future instrument upgrade.

5.3.2 Calibration using monochromatic flatfields

The raw images consist of microspectra arranged in the geometry of the lenslet array. One of the most important steps in the data reduction is therefore assigning each pixel membership to a specific spectrum (lenslet) and the wavelength corresponding to the position in the dispersed spectrum.
For this purpose, the CHARIS pipeline comes shipped with a set of important calibration data. This calibration data contains a library of (near-)monochromatic images of the point-spread function (PSF) for each individual lenslet (henceforth called PSFlets) taken at various wavelengths. This data is taken for each wavelength individually to prevent cross-talk and allow the characterization of the PSFlets both spatially and over wavelength. This PSF library can later be used to extract the spectrum in a χ^2-based model-fitting approach (as described below, Sec. 5.3.3). It also allows to precisely determine the wavelength solution: the assignment of a position on the detector to each wavelength of each spectrum. The precisely calibrated wavelength solution using the monochromatic images is provided in the pipeline as well. However, the wavelength solution can and does vary, for example, due to the ambient temperature. The approach of the CHARIS pipeline is therefore to use this one-time, precise calibration as a template and use a monochromatic flat taken close in time to the actual observation in order to refine the wavelength solution by fitting a linear shift in (x, y)-direction and a rotation term. The overall dispersive properties of the optical system are observed, and also expected, to be stable, such that a linear shift and rotation is sufficient to capture the temporal change in the system (Brandt et al., 2017). We can therefore use the usual wavelength calibration data routinely taken to fit this offset, whilst enjoying the benefit of the more precise wavelength solution, as long as the one-time calibration data is provided.

The pipeline adaptation described in this paper delivers the necessary PSFlet library and wavelength solution such that the same approach is possible (see Sec. 5.4).

5.3.3 Microspectra extraction

An example of how the raw data for SPHERE-IFS looks like is shown in Fig. 5.1. The DRH (ESO) pipeline, as well as other pipelines, like the GPI pipeline rely on a simple method of extracting the microspectra from the raw data. The method used is **aperture photometry**, were pixels orthogonal to the dispersion direction are summed up to obtain the flux at one wavelength, which is known from the wavelength calibration. However, this has multiple serious drawbacks: a) it is not straightforward to exclude bad pixels, b) each spectrum is slightly displaced with respect to the static pixel grid; this means that the wavelength sampling for each spectrum is different. In order to obtain one image at one specific wavelength, it is therefore necessarily to interpolate the data to the same wavelength grid. c) we do not make efficient use of our knowledge of the instrument.

To address these short-coming, the CHARIS pipeline implements two more sophisticated approaches to extracting the microspectra. The first of these is called **optimal extraction** (Horne, 1986), which solves a) and c) by not simply summing up the pixels orthogonal to the dispersion direction, but weighing them with the known PSFlet profile along the direction orthogonal to the dispersion direction. This way, higher SNR can be obtained by putting

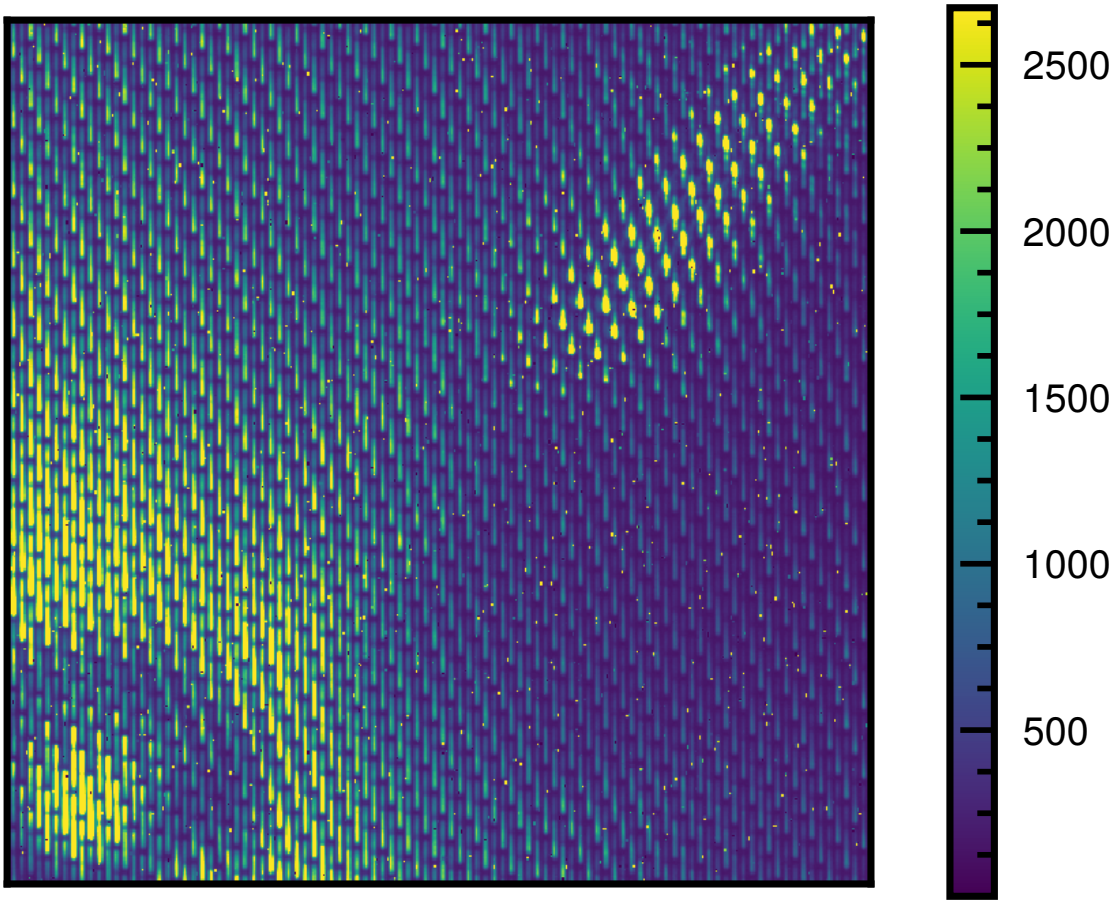


Figure 5.1: Cutout of unprocessed IFS raw data of 51 Eri b in YH-mode with satellite spots induced by the deformable mirror. The dispersion is in vertical direction from longer wavelengths at the top, to shorter wavelengths at the bottom.

more emphasis on the central pixels and less on the low SNR pixels, which are also more likely to be contaminated by the neighboring spectra. As we apply a weighing in the sum, that also provides an easy mechanism to ignore pixels flagged as bad entirely. This still leaves us with problem b): the need to perform interpolation in the wavelength direction.

The last, and preferred way of extracting the spectrum, is called χ^2-extraction. In this approach, a forward model of the spectrum is fit to the data. We use the actual library of PSFlets for each position and wavelength to construct a linear model of the spectrum from monochromatic PSFs, and optimize the coefficients for each wavelength to extract the spectrum. In this way, we can ignore bad pixels in the fit, we do not have to interpolate, because we fit a forward model and the wavelengths at which the flux is extracted can be chosen freely, and we make full use of our knowledge how the IFS works, because we have a detailed model.

This actually comes with another big advantage. Because we fit a model, we have an idea how much flux is in the wings of PSFs of each lenslet. Therefore, running the algorithm a second time allows us to account for cross-talk and contamination between neighboring spec-

tra on the detector self-consistently. Lastly, it should be pointed out that the simultaneous fitting of different wavelengths with PSFlets, can be interpreted as a *deconvolution with the linespread function*. This will make the images look less "blury", because a monochromatic image is extracted. This may also make them look more noisy to the eye. Usually uncertainties of neighboring wavelength are strongly positively correlated, because of the width of the lenslet PSF in dispersion direction (i.e. the linespread function), using this approach the uncertainties will be anti-correlated between neighboring wavelengths. This should be kept in mind, but will probably not have any strong implications as long as we include the correlation matrix in atmospheric model fits derived from the extracted spectra, as is done in Chapter 2.

The wavelength grid on which the images are extracted is defined over the spectral resolution R of the respective mode as

$$R = \lambda/\delta\lambda = \left(2\frac{\mathrm{d}\ln\lambda}{\mathrm{d}x}\right)^{-1}, \tag{5.2}$$

where x is in pixels, and the dispersion is defined as wavelength shift per two pixels. The extracted cube is defined on a logarithmic wavelength array $\ln\lambda$ as opposed to the output of the SPHERE pipeline which is given linearly with wavelength.

5.3.4 Official SPHERE pipeline

The ESO pipeline (DRH, Data Reduction and Handling) (Pavlov et al., 2008) has some known short-comings, which we hope to address with the adaptation of the CHARIS pipeline to SPHERE-IFS. One is that properly reducing IFS data currently requires the combination of DRH and addition custom routines designed as described for example in Chapter 2.3 (Samland et al., 2017). Cross-talk and bad-pixel corrections are done using additional consortium routines, the wavelength solution is not entirely correct and has to be corrected in an additional step by measuring the satellite spot positions in the data, that should scale proportionally to the wavelength (see, Vigan et al., 2017). It is also a bit alarming how many time the data is interpolated in the process of extracting the data: 1) interpolation in wavelength direction; 2) an interpolation in spatial direction to align the spectrum before summing up the flux along the orthogonal direction; 3) resampling the hexagonal geometry to a rectilinear grid (probably the most invasive step); 4) rescaling the images radially to correct for the initially wrong wavelength solution; 5) aligning the data such that the star is in the center of the image; 6) rescaling the y-direction by 6% to correct for the anamorphism of the instrument (geometric distortion)

All of these steps are done before any post-processing is done. While these steps should not strongly bias the results the lead to correlations in the signal, both spatial and spectral. With the CHARIS pipeline, we can avoid interpolation steps 1 to 4, while steps 5 and 6 can be made superfluous by using the TRAP algorithm described in previous Chapter 4.

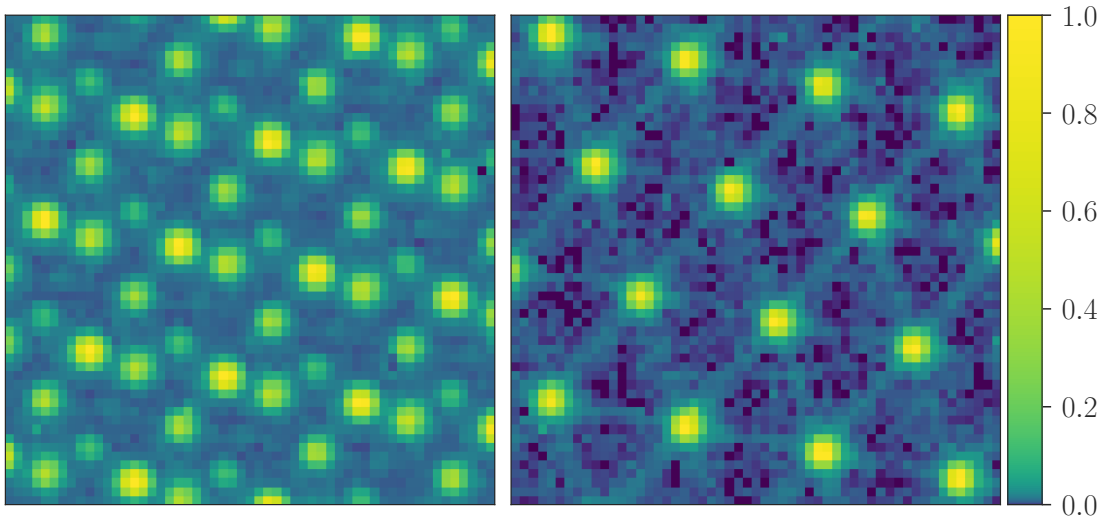


Figure 5.2: Shown are normalized monochromatic laser flat fields in logarithmic scale. The dispersion direction is from bottom to top. (left) Standard wavelength calibration image for the extended (YH) mode using all four lasers simultaneously: provided by ESO as part of the standard calibration sequence and used by the ESO pipeline. (right) Only 1123.7 nm laser: data at this and three other wavelengths was taken and used to construct our lenslet PSF library.

5.4 New calibration of the SPHERE-IFS

5.4.1 Calibration data

In order to construct the PSflet library we acquired data of all four available lasers. For YJ-mode this corresponds to 987.7 nm, 1123.7 nm, and 1309.4 nm. In YH-mode we have one additional laser at 1545.1 nm. Figure 5.2 visualized the difference between using all four lasers at the same time and using an individual laser. These calibration data were then dark and flat corrected. As the individual lenslet PSFs are sampled on an integer pixel grid and each PSFlet is slightly shifted, we locally combine nearby PSFlets to obtain artifact free, super-sampled versions for our library. Each lenslet PSF is then normalized. Figure 5.5 shows the oversampled PSFlets at 1123.7 nm over 25 subregions of the detector. Similar data is obtained for the other three wavelengths.

5.4.2 Masking and flat fielding

The flat fielding is still a work in progress. However, our strategy is to build a master detector flat field using the past four years of calibration data and included in the pipeline. We also include a master bad pixel mask, which combines the warm pixels from dark frames with dead and unresponsive pixels from flat field observations. The master flat field created by combining hundreds of individual flat fields is shown in the left panel of Fig. 5.3. The right panel of Fig. 5.3 shows the result of decomposing the temporal variance of these hundreds of flat fields taken over the span of multiple years into principal components and checking visualizing the component with the smallest overall contribution to the variance. This is

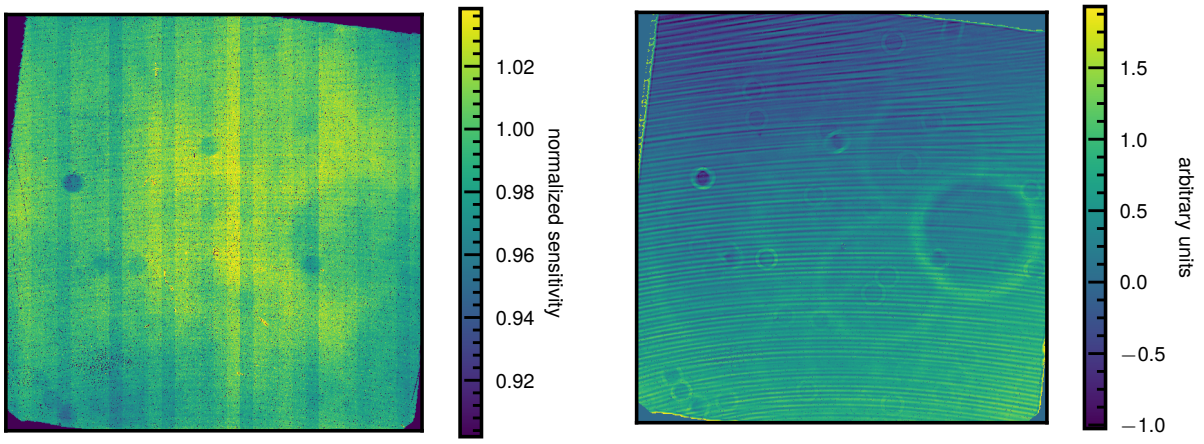


Figure 5.3: (left) Master flat field generated by averaging hundreds of individual flat field frames. (right) The smallest component of a principal component decomposition of the temporal variance of the flat fields. This image highlights minute details of the flat field variability. Details normally not clearly visible become apparent due to the changes in illumination by the internal calibration lamp affecting diffraction patterns over time. This includes the fringe-like horizontal stripes, as well as dust diffraction by various optical surfaces.

an interesting experiment that is only made possible by the wealth of daytime calibration data, taken in each run over the years. It highlights the aspect of the flat field that are variable over time (e.g., because of changes in illumination by the internal lamp). We can see diffraction rings by dust sitting on various optical surfaces, as well as a fringe-like pattern of stripes that may be caused by the surface of the detector itself.

In the final flat field distributed with the pipeline, the vertical stripes caused by the difference in reference voltage of detector columns should be filtered out from the flat field. Also the most large scale patterns should be removed, as they are likely caused – at least partially – by inhomogeneous illumination of the detector and not by intrinsic differences of the pixel sensitivity.

Additionally we need an IFU flat field, that tells us the relative throughput of each lenslet. CHARIS uses the wavelength-averaged (near-)monochromatic spot amplitude patterns obtained with the supercontinuum source and a tunable narrow-band filter directly as an indicator for the lenslet transmission. Unfortunately this is not possible for SPHERE as the lasers are coherent light sources and susceptible to fringing and spurious amplitude variations across the lenslet array. This is not an issue when we make the PSFlet library as the models are normalized, but they are not reliable indicators for absolute transmission. We therefore plan to use a similar approach as used in the DRH pipeline. We will extract the spectra for a white lamp illuminating the IFS using the optimal extraction algorithm, and take the wavelength averaged transmission, and normalize by the average of all lenslets. These values will be included as the IFU flat field.

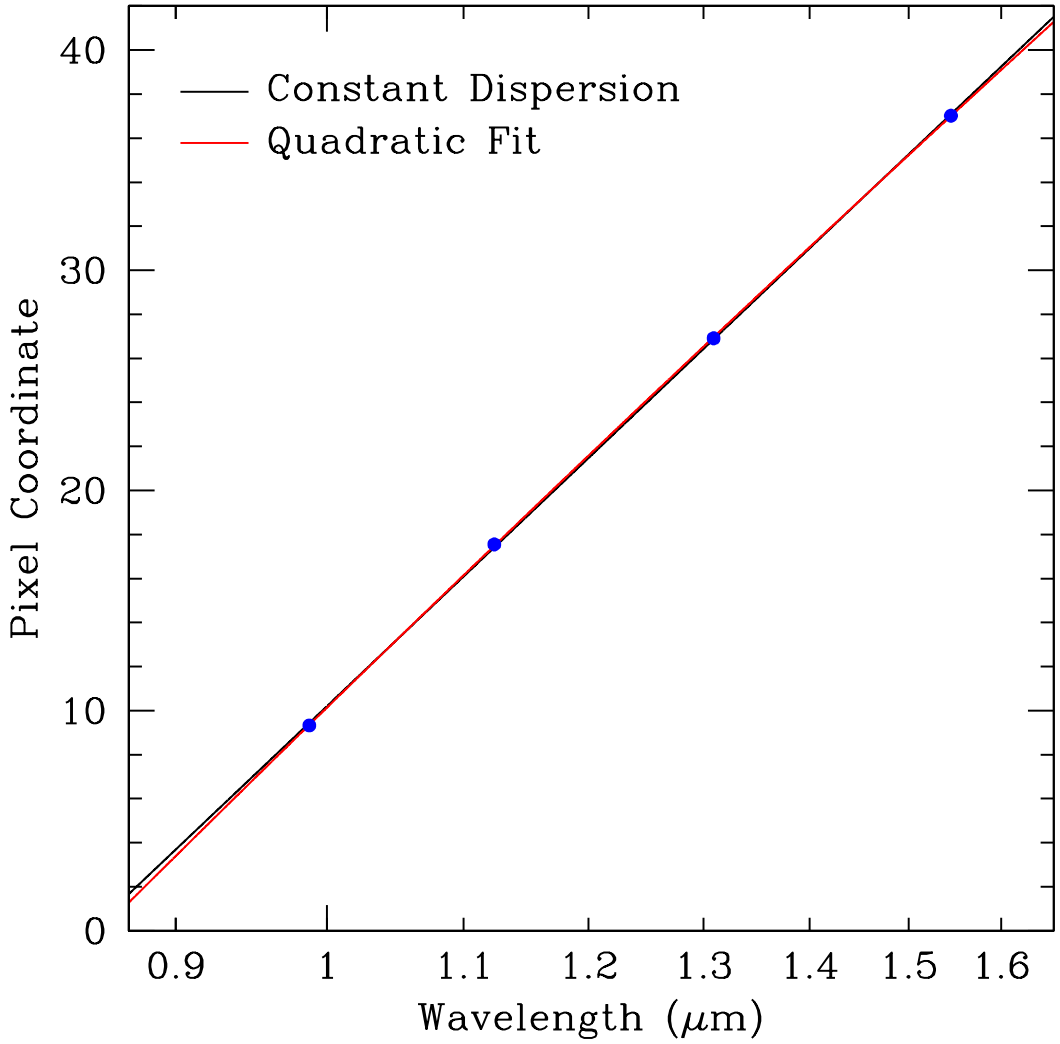


Figure 5.4: Wavelength solution for the extended YJH mode measured using the four individual calibration laser at 987.7, 1123.7, 1309.4, and 1545.1 nm respectively. Shown are fits for both linear and quadratic dispersion.

5.4.3 Wavelength solution

We follow the same method as Brandt et al. (2017) to fit the wavelength solution using the monochromatic flat fields. The wavelength solution for the YH-mode is shown in Fig. 5.4 and shows an almost linear dispersion relation that fits our data very well. Unlike for CHARIS which uses a 3rd-order polynomial, a quadratic solution for SPHERE is sufficient.

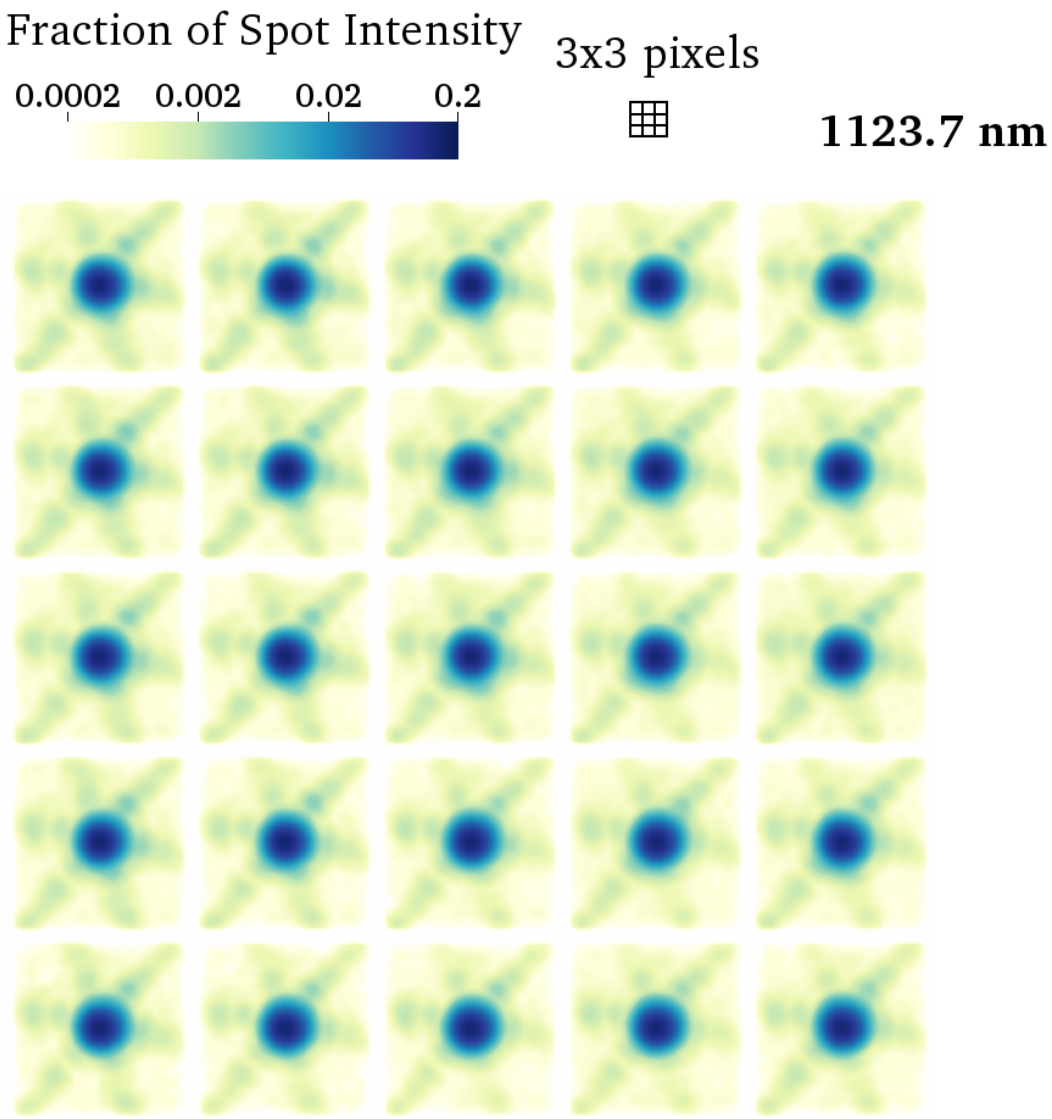


Figure 5.5: Oversampled PSFlets at 1123.7 nm over 25 regions of the detector. Constructed by locally combining the normalized PSFlets in an area. A 3x3 pixel grid is shown as size comparison. Similar libraries are created at the other three calibration wavelengths. The oversampled PSFlets can be used to construct a model of the pixellated microspectra corresponding to monochromatic or broadband light imaged by the lenslet array. Contrary to CHARIS we see six and not four diffraction spikes due to the difference in lenslet geometry. The shapes are extremely homogeneous across the field of view. There are no shape variations that would indicate a changing focus.

5.5 Adaptations for SPHERE

In addition to the necessary new calibration data, we had to adapt the pipeline itself to be more flexible. An instrument class has been added to the software package and instrument specific constants and settings were moved to this class to make the package easier to maintain and adaptable to other lenslet-based IFS instruments in the future. This class includes the various observing modes, their corresponding wavelength ranges and resolution. It also contains the geometry of the lenslet lattice that is used to index the spectra (rectilinear or hexagonal). Additionally the following points needed to be addressed:

Visualization routine for spatial hexagon sampling We implemented routines that allow to visualize the extracted cubes in the native hexagonal geometry without resampling, as most normal image viewers like DS9 can only view pixel-based images. A command-line script allows quick access to viewing an extracted data cube, similar to the other routines of the package. It allows the interactive viewing of an extracted image cube with sliders for wavelength and adjusting the image scale.

Resampling to rectilinear grid We included additional routines that allow for computing the overlap of arbitrary polygons (and grids thereof) using the Sutherland-Hodgman algorithm. For our case we construct two grids, the hexagonal grid in pointy-top configuration, and a square pixel grid with squares of the same side-length as the hexagons. This choice of size is natural in that the centers of hexagons and squares align every 4 pixels in y-, and every 6 pixel in x-direction. This also corresponds to pixel scale chosen in the DRH pipeline.

Parallactic angle computation We adjusted the parallactic angle (PA) computation from the python based VLTPF pipeline[15] (Vigan et al., 2015) to add the correct PA to the header of the output files. The VLTPF pipeline provides a set of tools used to execute DRH pipeline recipes and adds additional functionality like finding the centers of images, the PA computation, and fixing the wavelength offset from the wrong DRH IFS wavelength calibration. The PA computation addresses various issues found during the guarantied time observations of SPHERE, such as a steady shift in the internal instrument clock, which was fixed after the problem was identified, but has to be taken into account for older data.

Alternative cross-talk correction While the χ^2-extraction method provides the option to model and correct cross-talk by iteratively fitting the model, cross-talk is not corrected for in the optimal extraction and aperture photometry approach. I will, in the future, also included the cross-talk correction routine as described in the Appendix of Vigan et al. (2015), which corresponds to a spatial filter that removes flux at the spatial scales at which the cross-talk occurs, as an optional step. Naturally, this correction should not be used together with the χ^2-method's cross-talk correction.

[15] https://github.com/avigan/VLTPF

5.5.1 The virtues of the hexagon

At the current point in time most astronomical instruments make use of rectilinear grids of pixels, because they are easy to use and manufacture. However, there are some general properties of hexagonal grids that make them suited for astronomical tasks.

1. Hexagons have a low perimeter-to-area ratio and therefore reduce effects due to sharp edges. Circles have the highest ratio, but hexagons are the closest polygon to circles that can still tessellate to form an evenly spaced grid.

2. This property becomes especially important at the edges of often used masks in astronomical post-processing, like circular apertures or annuli, e.g, most ADI post-processing uses annuli to divide the image into sub-regions. Squares are much less suited for these geometries with curves (alas easier to work with).

3. As opposed to square grids, where each diagonal neighbor is a factor $\sqrt{2}$ further away than the four closest neighbors; all six neighbors of a hexagon are equally far away. This makes correcting outliers by local averaging or median filters more accurate

4. Hexagons are preferable when dealing with problems that relate to connectivity or movement paths. In pupil-stabilized observations astrophysical signals following an arc-like movement path, which is easier to represent on a hexagonal grid.

5.5.2 Bad lenslet correction

The badpixel correction which is used for cosmetically replacing bad pixels and setting the inverse-variance of the outlying pixel to zero in the CHARIS pipeline cannot be used one-to-one for SPHERE. The reason is that any outlier correction should act on the hexagonal grid before the image is re-sampled onto the rectilinear grid. We have therefore included an additional calibration file that stores the indices of all neighboring hexagons to each hexagon. This allows the quick application of median and robust standard deviation to all neighbors to both identify and replace outliers.

5.6 Results

The stable pipeline version that will be released to be public is still work in progress. Work still has to be performed to optimize the lower and upper limit of the wavelength ranges in the individual modes, and the spectral sampling at which the images are extracted in the χ^2 approach has be explored. We have yet to include the IFU flat field to correct for the lenslet transmission efficiency. There is still some work to be done to make the χ^2-extraction more robust against remaining and unidentified bad pixels. This last part is necessary, because for CHARIS the reading-up-the-ramp is utilized to get a reliable uncertainty on the counts for each pixel that can be passed to the extraction pipeline, information that is lacking for SPHERE as currently not all frames are saved.

Figure 5.6 shows an image extracted with the current version of the pipeline, and a comparison to the same frame extracted with the DRH pipeline and additional routines for cross-talk and wavelength correction used in the SPHERE GTO consortium. The data shows the star

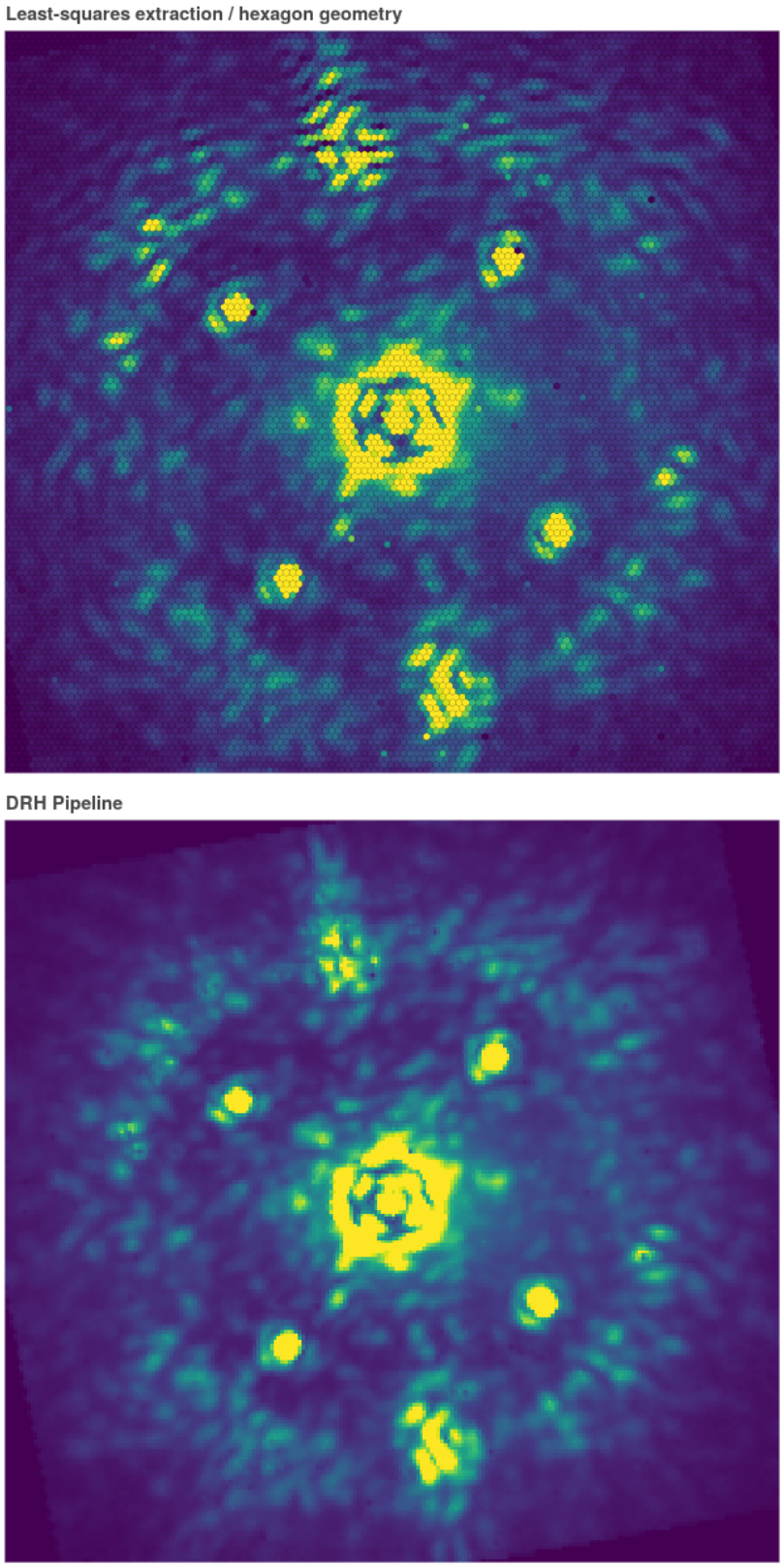


Figure 5.6: (top) least-square extracted frame (YH-mode) in hexagon geometry; (bottom) same frame reduced with DRH pipeline and additional routines for cross-talk correction and wavelength correction.

51 Eridani observed with SPHERE-IFS in YH-mode. This is a newer observation than the one used in Chapter 2 and 4, in which no wind-driven halo is visible and the speckle pattern is less obscured.

I show that I can already extract image cubes from the raw microspectra data, and that the pipeline is nearly operational. While detailed tests and comparisons still need to be performed, we can already note that the speckle pattern is less distorted than the DRH pipeline reduction for the same frame, which I attribute to less interpolations, the deconvolution with the linespread function in the χ^2-approach, and keeping the hexagonal geometry.

5.6.1 Extracting the spectrum of 51 Eridani b

One of the motivations driving the adaptation of the CHARIS pipeline for SPHERE was also to have another independent way to confirm the spectra of detected objects, such as the spectrum of 51 Eridani b shown discussed in Chapter 2 of this thesis and shown in Figure 2.4, and the discrepancy in the J-band where we detect flux 40% lower than GPI. It is common practice to extract the spectrum of an object with many different post-processing tools to judge the robustness of the detection with independent tools. However, no amount of post-processing comparison can assure us that there may not be something awry on a deeper level of the data handling.

With the CHARIS pipeline we now have a completely independent tool in hand to confirm our findings. To do this, we extracted the same YH-mode data of 51 Eridani b as used in Samland et al. (2017) with the CHARIS pipeline in its current state. We used the "optimal extraction" method, as described above, because it is closest to the approach taken in the DRH pipeline, but still presents an improvement. The χ^2-approach, on the other hand, at this point still needs further testing and optimization, although I am confident that ultimately the χ^2 method will be the approach generally recommended and used.

Furthermore, for this test, we did not employ the cross-talk correction from the DRH pipeline, which means no cross-talk correction is performed. As pointed out above, we have not included the lenslet flat field yet. Lastly, we extracted the images at 39 equally spaced wavelengths, the same number of channels in the DRH, the first and last of which correspond to the same wavelengths as the DRH extracted and re-calibrated wavelengths.

Figure 5.7 shows a comparison between the shortest wavelength extracted with the CHARIS pipeline using optimal extraction (left panel), and the same wavelength extracted with the DRH pipeline and the additional routines (right panel). One difference that can be seen clearly, is that the CHARIS pipeline reduction does not show the "checkerboard" like structure in the background, that is likely an artifact from the extraction process, and especially visible at shorter wavelengths. It gradually reduces at longer wavelengths, but re-surfaces at wavelengths with lower signal, e.g. in the telluric absorption band.

Figure 5.8 shows the extracted spectra of 51 Eridani b from: a) the DRH extracted and re-calibrated data cube, b) the CHARIS pipeline extracted cube using optimal extraction. The conversion from contrast to flux was done using the same methodology and flux calibrated spectrum as in Chapter 2.

We can see that the overall performance is very good. The uncertainties obtained from post-processing are significantly smaller, which is probably due to two main reasons: using optimal extraction instead of aperture photometry, and not temporally binning the data.

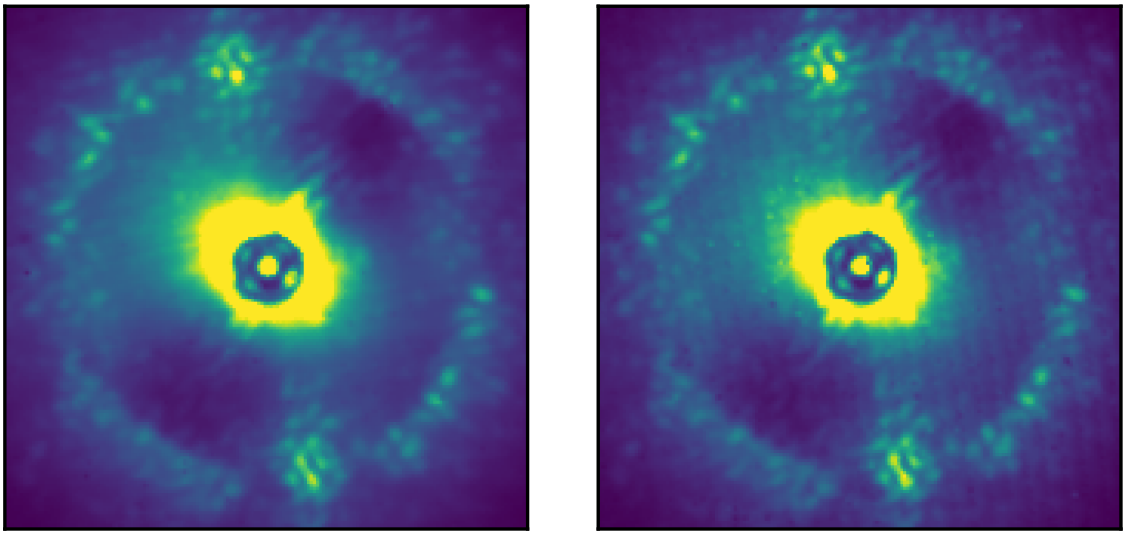


Figure 5.7: (left) Extracted with the CHARIS pipeline using optimal extraction. (right) Extracted using DRH and additional routine (Samland et al., 2017). Both panels show a median image of all frames extracted at the same wavelength (shortest wavelength bin) and are normalized to the same scale in arbitrary units.

The DRH pipeline is relatively slow, to save processing time by default the Data Center (Delorme et al., 2017b), usually bins down the IFS data which may hinder post-processing performance.

The CHARIS pipeline extracted spectrum does not significantly degrade in quality towards shorter wavelengths. This could prove invaluable for more reliable Y-band characterizations of bright exoplanets in the future.

Regarding the spectral features, the overall strength of the J- and H-band peak are similar to the published results. We can definitely rule out the high J-band flux reported by GPI (Macintosh et al., 2015; Rajan et al., 2017); if anything the flux is lower. There is, however, a strong deviation in the spectral shape of the H-band. The most probably explanation is that this is caused by the lack of cross-talk correction in this analysis. The reason is simple, the star 51 Eridani, and therefore the speckle pattern, is brighter at shorter wavelength than longer wavelengths. As can be seen in Fig. 5.1, the way the spectra are organized on the detector, by necessity, the tail of one spectrum will be neighbor to the opposite end of the adjacent spectrum. If the overall speckle background is brighter at short wavelengths, the bright part of the spectrum will be next to the faint part of the adjacent spectrum. For this reason, shorter wavelengths are much more likely to contaminate longer wavelengths than vice-versa. Optimal extraction helps mitigate this problem by putting less emphasis on pixels further from the center of the spectrum, but does not prevent it completely.

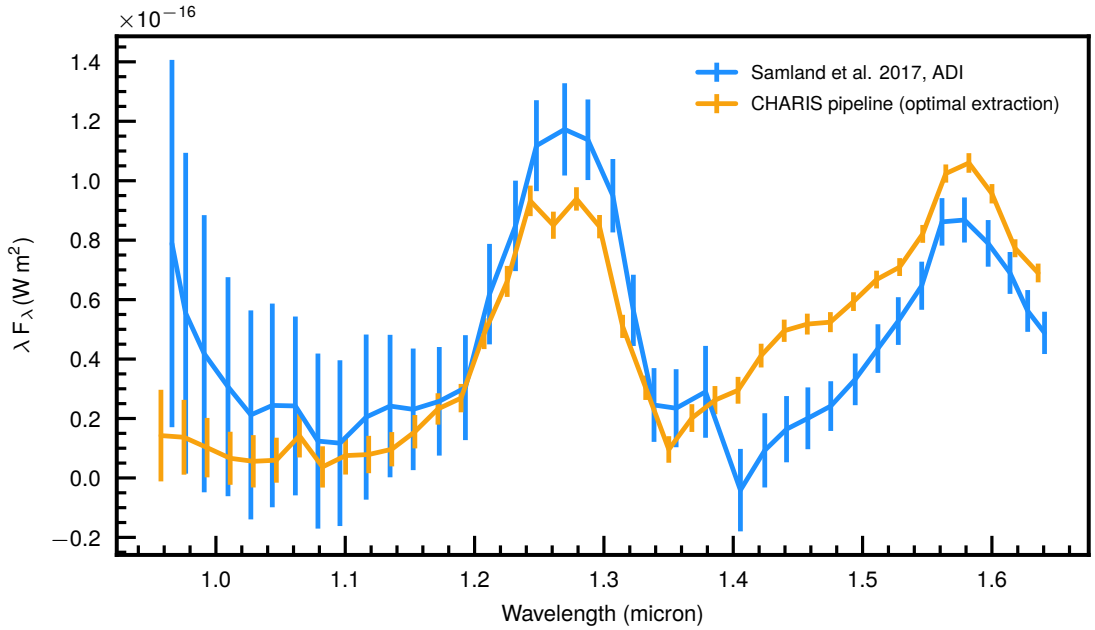


Figure 5.8: Comparison of 51 Eri b spectra extracted with ANDROMEDA (ADI-only) from: a) the DRH extracted and recalibrated data cube (Samland et al., 2017) in blue; b) the CHARIS pipeline extracted cube using optimal extraction and no cross-talk correction. The error bars denote 1-σ uncertainties.

5.7 Summary and conclusions

With my work of adapting the CHARIS pipeline to work with SPHERE-IFS, I took an important step to: 1) fix known short-comings of the official IFS pipeline and increase its reliability, 2) unlock the full potential of the IFS instrument by making better use of calibration files and our knowledge of the instrument, 3) enable the community to participate in the development and future improvement of SPHERE data reduction with open-source, well-documented reduction tools, 4) make CHARIS and SPHERE-IFS results directly comparable, 5) let both CHARIS and SPHERE teams profit by improvements made to the pipeline, 6) provide a tool that is easy to use correctly. There are many steps involved in reducing IFS data with DRH, some reliant on unofficial tools. Users who are not part of the SPHERE consortium will find it difficult to navigate the pitfalls.

I used the current version of the adapted pipeline to obtain a calibrated data cube for 51 Eridani b and directly compared the result to our spectrum published in Samland et al. (2017). I independently confirmed the J-band flux to be lower than the published GPI results Macintosh et al. (2015); Rajan et al. (2017). While work still needs to be done on the pipeline, its potential can already be seen in that it increases the flux accuracy at shorter wavelengths, results in overall higher signal-to-noise in post-processing, and does not show the checkerboard artifacts in the extracted images that are visible in the DRH reduction.

There is still a lot of untapped potential in the instrument that could be made use of in the CHARIS pipeline. Firstly, the IFS instrument has a detector dithering stage which is not being used. Originally it was intended to use dithering during the scientific observations, but

it turned out that extracting the shifting microspectra was difficult. However, I think the dithering stage would be the perfect tool to create high signal-to-noise sets of supersampled lenslet PSF models. At the moment several nearby PSFlets have to be averaged locally to create the lenslet PSF models, which degrades the fidelity of the model. This is especially true for individual out-of-focus lenslets. With dithering this averaging can be avoided. Secondly, the SPHERE detectors do not save all reads and only the extracted count rate is saved. The CHARIS pipeline implements tools that can make use of all individual reads, which allows to correct for non-linearity and also provides uncertainties for all count rate measurements that can be used for the χ^2-extraction.

Going forward, I would like to emphasize the synergistic potential between the CHARIS pipeline and my TRAP algorithm introduced Chapter 4. When using a post-processing algorithm that creates a non-local, temporal systematics model, the images need not be aligned. For SPHERE-IFS this further means that the post-processing can be performed directly on the hexagonal grid, without the need to resample the images onto a rectilinear pixel grid. As mentioned in the conclusions of Chapter 4, "SDI" can easily be implemented in TRAP by adding regressors from other wavelengths to the temporal model. In addition to this capability, I will also implement the regressor selection in a way that will allow the data to be hexagons. The only thing that has to be shifted/interpolated will be the forward model of the companion signal.

Finally, the combination of the CHARIS pipeline with TRAP will give us a new state-of-the-art view of exoplanet spectra obtained with SPHERE-IFS.

6 Conclusions and Future Perspective

6.1 Summary

The field of directly detecting exoplanet atmospheres is still very young in the larger context of astronomy. This means that many techniques are being developed, in terms of observational strategies, technology, algorithms, as well as the theoretical atmospheric models themselves. With this thesis I have contributed to both the understanding of the atmospheres of directly imaged exoplanets, as well as the development of new algorithms which enable the detection of more challenging objects very close to the host star, and the extraction of more reliable spectral information. All of these aspects together determine what we can learn about exoplanets from direct imaging and how reliable the conclusions are.

The first part of this thesis focused on the characterization of the atmospheres of directly imaged exoplanets, the second part focused on ways to improve the reduction of high-contrast imaging data.

6.1.1 Characterization of Exoplanet Atmospheres

So far, only a handful of planets have been discovered with direct imaging. Obtaining spectra that are of sufficient quality to characterize their properties is even more challenging. However, thanks to a new generation of instruments on 8m-class telescopes, we are now able to study such planets in detail with low-resolution spectroscopy and narrow-band photometry. In the future, medium and high-resolution spectroscopic observations (e.g. Snellen et al., 2014; Konopacky et al., 2013; Hoeijmakers et al., 2018) of non-transiting planets will become increasingly important.

As part of the SPHERE consortium, I have contributed to a large number of studies on directly imaged planets, both in terms of data reduction and characterization. I led a detailed study on the cold, methane-rich planet 51 Eridani b (Chapter 2) for which I developed an MCMC-based statistical inference code called BACON (Bayesian Atmospheric CharacterizatiON) to determine atmospheric parameters, which has been used for my subsequent characterization work, an overview of which was given in Chapter 3. I described my atmospheric characterization work on the first planet discovered inside the gap of a transition disk (PDS 70 b, Müller et al., 2018), the first planet discovered with SPHERE (HIP 65426 b, Cheetham et al., 2019), and my contributions to the recent study of GJ 504 b (Bonnefoy et al., 2018). Table 3.1 summarized the main characteristics derived for these planets.

Chapter 2: 51 Eridani b

For 51 Eridani b, the bulk of the analysis focused on the physical interpretation of the spectrum of the planet. In order to infer the atmospheric parameters, I developed the BACON (Bayesian Atmospheric CharacterizatiON) code to use affine-invariant ensemble MCMC sampling (Foreman-Mackey et al., 2013) of self-consistent atmospheric forward models of arbitrary dimensions, while taking into account spectral correlations in the data. I closely worked with Paul Mollière for this project to obtain a model grid that can adequately explain the data, including the dimensions of metallicity and the settling strength of clouds.

Only models with high metallicity were able to explain the high flux in the K-band while simultaneously fitting the spectrum on both the shorter and longer wavelength regimes. The collision-induced absorption (CIA) cross-section of H_2–H_2 and H_2–He pairs peaks in the K-band (Borysow, 1991; Allard et al., 2001), making it pressure sensitive. Higher metallicity means shifting the photosphere to lower pressures reducing the contribution of CIA to the spectrum. This result shows the importance of super solar metallicity models as well as the significance of K-band data for constraining the properties of atmospheres.

Additionally, I used BACON to fit linear combinations of cloud-free and cloudy atmospheres to simulate a patchy cloud layer, similar to Macintosh et al. (2015). These fits showed that a uniform cloud layer for this model is sufficient to explain the data.

The empirical comparison with Mickaël Bonnefoy's database of substellar objects confirmed the peculiarity of 51 Eridani b. It is located in a unique place in color-color and color-magnitude diagrams, which may be related to low-surface gravity and/or young age effects, but it also shares common properties with other late-T dwarfs. This analysis showed, however, the limit of empirical comparisons in the absence of a robust number of similar objects from clusters and young moving groups with spectral type later than T5. It also highlights the need to develop novel ways to compare outlying objects to a library of known spectra, as the concept of "similarity" has to be more carefully defined for any classification scheme to be meaningful.

Another part of my analysis was the direct comparison between mass estimates derived from evolutionary tracks at the system's age and the measured luminosity, with the mass determined directly from radius and surface gravity of the planet. Both methods are model dependent, e.g. the uncertainty in cloud physics may impact the surface gravity, but comparing the two approaches allows us to have an additional consistency check. I used the luminosity–mass relationship for populations derived from a core accretion population synthesis model under different accretion physics assumption (Mordasini et al., 2012a,b) resulting in a cold, warm, and hot population. Cold start models are clearly ruled out, and pure hot start models are also disfavored. Confirming expectations from theoretical models of accretion shocks, warm start models seem to be the most likely scenario (Marleau et al., 2017).

Chapter 3: Atmospheric characterization of other objects

PDS 70 b is the first planet discovered inside the gap of a transition disk (Keppler et al., 2018). I led the spectral characterization of the planet in Müller et al. (2018). The spectrum, with the exception of a weak water feature at 1.4 micron, is featureless and almost represents that of a black body. The spectrum is consistent with various cloudy atmospheric models

from the petitCODE (Mollière et al., 2015b, 2017) and Exo-REM (Charnay et al., 2018) using different physical assumptions for clouds, demonstrating the degeneracy of the problem. They also result in unrealistically large values for the radius (i.e. $> 2R_J$) indicating the need for more sophisticated models for very young planets. We suspect that there may be a substantial contribution from a circumplanetary disk (CPD, e.g. Zhu, 2015). Consequently, Wagner et al. (2018) has shown a significant H-alpha accretion signature at the position of PDS 70 b. Another potential explanation is the presence of dust extinction around the planet. The study shows the need for more detailed physical models of young, accreting exoplanets, as well as better data to solve the degeneracies from cloud/dust envelops and potential CPD contributions, e.g. through better wavelength coverage with JWST.

HIP 65426 b is the first planet discovered with the SPHERE instrument (Chauvin et al., 2017). I led the spectral characterization of the planet in Cheetham et al. (2019). While the planet has a similar spectral type, it differs significantly from PDS 70 b in actual properties. Its effective temperature is higher and there are less cloud opacities in the atmosphere, making the overall colors look bluer, while having approximately the same brightness in the J- and H-bands. The difference in opacities is most likely due to the age difference, with HIP 65426 b being about 10 Myr older than PDS 70 b, which still shows signs of accretion. Contrary to the T-type objects 51 Eridani b and GJ 504 b, for which the Ackerman & Marley (2001) cloud model worked well, the much hotter (∼1600 K) HIP 65426 b seems to be better described by the Rossow cloud model which is used in the BT-Settl models, the principle of which is based on iteratively "settling out" (removing) species following time-scale arguments until the atmosphere converges to a stable solution.

GJ 504 b is a companion that can either be a planet or a brown dwarf based on its true age, which is still under debate. With a temperature of 600 K, it is one of the coolest directly imaged companions. It was possible to explain the spectrum using the same set of cloudy atmospheric models as used for 51 Eridani b. However, the cloud parameters did not quite converge to their optimal solution, suggesting that a more extended grid or free-retrieval should be tested in the future. The derived surface gravity and radius would suggest a planetary mass rather than a brown dwarf mass. However, because GJ 504 b is outside the field-of-view of most IFUs, only photometry and no spectroscopy was used in the study. Therefore, further confirmation using more and better data and models is needed in the future.

6.1.2 Chapter 4 and 5: Development of novel high-contrast imaging algorithms

In addition to having extensively used established algorithms for speckle suppression (e.g., ADI/SDI, PCA, LOCI, ANDROMEDA), a major focus of my thesis work has been to develop new algorithms and explore ways to improve the fidelity of spectra, thereby increasing our discovery potential. There are two main projects that I have worked on with this goal in mind, which were described in detail in Chapters 4 and 5. The two corresponding papers are currently in preparation and to be submitted soon.

Chapter 4: A temporal, non-local systematics model for direct detection of exoplanets at very small angular separations

As the direct imaging field advances, we will technologically push towards detection of companions at smaller and smaller angular separations. In order to facilitate discoveries at very small angular separations, I have developed a new method of extracting signals of planets from high-contrast imaging data taken in pupil-stabilized mode, which is the standard observation mode for high-contrast imaging. The implementation of this algorithm is called TRAP (Temporal Reference Analysis for Exoplanets).

This new method constructs a temporal, non-local model of the systematic noise (speckles) in the data rather than a spatial, local model, and therefore differs from traditional approaches. Due to the construction of a systematics model based on non-local reference pixels, we can completely circumvent the need for a temporal exclusion criterion ("protection angle") that would usually be required to prevent contamination of the systematics model by the signal of interest. This would lead to over-fitting and self-subtraction and thus a reduction in signal-to-noise. In the traditional approach the temporal exclusion criterion is linked to the physical displacement of the signal on the detector. Given a certain rotation of the field-of-view, the time needed to significantly displace an object at smaller separations is therefore longer. At very small separations ($< 3\lambda/D$) the exclusion time can become on the order of 30 minutes, which is a significant fraction of usual observation sequences ($\sim$60 – 90 minutes for SPHERE), degrading the systematic model's accuracy for shorter time scale speckle variations significantly.

Using TRAP, I demonstrated that up to a factor of six can be gained in contrast at small separations. Furthermore, I have shown that the temporal sampling of the data has a significant impact. With higher sampling frequency a significant gain in contrast and signal-to-noise can be achieved. The impact of exposure time on the quality of the systematics model is underexplored in the literature, mainly because the performance of the algorithms seems to be limited by the exclusion time used in spatial systematics models, rather than by the exposure time (temporal sampling). With TRAP, however, we are sensitive to all time scales on which the systematics change, as long as the exposure time is short enough. This includes systematic noise that decorrelate on short time scales, such as turbulence-induced instrumental speckles. The next step will be to include spectral information in the systematics models, in order to significantly improve results for IFS instruments.

Adaptation of the CHARIS pipeline for SPHERE-IFS: Revisiting the spectrum of 51 Eridani b

The analysis of 51 Eridani b presented in Chapter 2 showed a discrepancy in the J-band between GPI and SPHERE data of about 40%. This motivated me to exploit the similarity between the new CHARIS instrument and SPHERE to adopt their advanced pipeline for extracting microspectra from raw data to use for SPHERE-IFS. This allows a completely independent evaluation of the data.

The CHARIS pipeline (Brandt et al., 2017) is completely open-source[16], written in Python and Cython, and has the potential to be adopted by all (lenslet-based) IFS instruments. For the purpose of adapting the pipeline to SPHERE-IFS, new SPHERE calibration data was

[16] `https://github.com/PrincetonUniversity/charis-dep`

needed to create high-quality PSF models for all lenslets at various wavelengths. These models can then be used to fit a forward model of the microspectra on the detector instead of using aperture photometry to extract the flux at each wavelength.

The lenslet array used by SPHERE consists of hexagons, therefore the native image geometry is based on a hexagon grid, which we preserve here. The image can also be resampled to a more familiar pixel-based geometry (introducing spatial correlations in the process) as is done in the official pipeline.

Using my adapted pipeline on 51 Eridani SPHERE-IFS data to extract images from the raw microspectra, we note that the speckles are less blurred. Furthermore, I extracted an image cube of 51 Eridani b in YH-band described in Chapter 2 using the "optimal extraction" method. Using the same ANDROMEDA post-processing, I confirmed the lower J-band flux.

6.2 Future Perspective

Observational capabilities as well as calibration techniques for direct observations and characterization of exoplanets have seen tremendous progress over the past decade. Yet the interpretation of such data by using 1D atmospheric models most often remains difficult. Exoplanets are highly complex entities: to constrain their properties, we have to make assumptions about their bulk composition, cloud properties and physics, chemistry, and atmospheric mixing (e.g., Ackerman & Marley, 2001; Morley et al., 2012). Additionally, some of these properties vary with time. Given these complexities, even 1D models tend to be highly degenerate. In order to draw conclusions about the formation and composition of exoplanets, we need to find ways to accurately disentangle parameters using a variety of techniques that are sensitive to different properties.

With this goal in mind, my future research will continue to have one main goal: to increase the reliability of the data in order to answer fundamental questions about exoplanet formation and composition by using statistical inference methods that rely on robust theoretical and data-driven models.

In addition to the various ways to further improve the TRAP algorithm as outlined in the conclusions in Chapter 4.9, and exploiting the synergies between the TRAP algorithm and the CHARIS pipeline outlined in Chapter 5.7, progress can be made in the following areas:

1. medium (> 1000) to high ($> 50\,000$) resolution studies, e.g. to directly constrain elemental abundances, and measure the planet rotation rate

2. time-domain variability observations, e.g. to constrain cloud coverage and properties

3. extended wavelength coverage, e.g. to constrain opacities, Mie-scattering and cloud properties

Some of these points will be revolutionized by instruments on the new generation of extremely large ($> 30\,\mathrm{m}$ aperture) telescopes currently under construction, which will extend our observations to longer wavelengths, fainter objects, and higher spatial and spectral resolutions. New powerful ground-based facilities such as the high-dispersion spectrograph VLT/CRIRES+, and the coupling of such instruments to high-contrast imagers equipped with extreme adaptive optics and coronagraphy (Snellen et al., 2015), as well as the development of entirely new facilities like VLT/METIS will achieve higher signal-to-noise and probe fainter objects. At the same time, new powerful space-based observatories like JWST will provide unique insights into planetary atmospheres at wavelengths inaccessible from the ground: in the absorption bands of the Earth's atmosphere, as well as longer wavelengths at which ground-based instruments lack the required sensitivity. In addition to the upcoming advancements in observations, an equal evolution in how we handle these data will have to take place. I aim is to improve the first two points noted above: to better use and interpret high resolution data, and the observations of time-domain variability of directly imaged exoplanets. Below is a short description of these topics.

6.2.1 Improving atmospheric model inference

Low spectral resolution or broad-band photometric observations often do not provide enough information to differentiate between competing models aimed at constraining the properties of exoplanet atmospheres. Resolutions of several thousands to a hundred thousand help resolve degeneracies in the models by picking up signatures of individual molecules like CO, H_2O, and CH_4, e.g. by matching line list templates to the spectra. Such high resolutions can also provide physical information on the molecules' environment. With the ELT/METIS instrument we could even distinguish isotopologue ratios, such as D/H, which could provide valuable insights into the history of icy-body enrichment and atmospheric evaporation processes (Mollière & Snellen, 2019). High resolution spectroscopy, however, is generally not used to constrain the overall pressure-temperature structure of the atmosphere, something low resolution is very useful for. It is therefore imperative to combine the best of both approaches to solve outstanding questions such as the extent of disequilibrium chemistry in exoplanet atmospheres. A first step in this direction has been taken by Brogi et al. (2017) demonstrating such an approach for day-side spectra of a hot Jupiter.

I will adopt this method for emission spectra of self-luminous planets and brown dwarfs. Taking a step further, I will work on not only fitting templates of molecular lines, but actual realizations of self-consistent atmospheres, such that we can link a non-detection of methane (Wang et al., 2018) quantitatively to specific non-equilibrium and vertical mixing models. I have already developed the BACON code for atmospheric parameter inference using MCMC and atmospheric model grids like petitCODE (Mollière et al., 2015b, 2017) to characterize 51 Eridani b and tested it on benchmark brown-dwarfs of known metallicity (see Chapter 2, and appendix therein). I have also included a free-retrieval approach in BACON (Mollière et al., 2019), which will help in identifying the correct physical parameter space. Adding a cross-correlation framework for the inclusion of high-dispersion data is the next logical step. In addition to using high-resolution data, interfacing the atmospheric model with the data reduction pipelines will further improve the achievable signal-to-noise and push the limits of atmospheric characterization.

It is common practice to extract a spectrum from the data, and in a subsequent step fit an atmospheric model to this extracted spectrum. This is the case for both low-resolution and high-resolution studies. However, fitting the planetary signal at the level of the data and optimizing the planetary model simultaneously with the (data-driven) instrumental systematics model should be preferred. This opens up a wide range of options, such as selectively including other wavelengths in the training set for the determination of the systematic noise model, which would lead to strongly biased results without an approach that simultaneous optimizes both models. Especially with the prospect of coupling high-contrast imagers and high-dispersion spectrographs on the horizon, this approach promises to be the way towards improving the detection limit.

6.2.2 Time-domain variability: constraining cloud properties of directly imaged planets

Recent space-based studies of the temporal variability of brown dwarfs have shown the time domain to be a powerful tool for constraining cloud models (e.g. Biller et al., 2015; Apai et al., 2017). In order to have a comprehensive understanding of directly imaged exoplanets, we will have to achieve similar results. As most facilities that are capable of directly imaging exoplanets are ground-based and influenced by the Earth's atmospheric conditions, achieving the needed precision will be difficult, but promising attempts have been made (Apai et al., 2016).

In the SPHERE consortium, I have been testing techniques to improve the quality and stability of the satellite spots for photometric calibration, e.g. by modulating the phase of the signal using the deformable mirror, as has been demonstrated for SCExAO (Jovanovic et al., 2015a), and alternating the orientation of the waffle pattern to subtract the underlying speckle field in each orientation, and thereby decreasing the calibration bias. In addition to the brightest planets (e.g., planets around HR 8799 and β Pic), 51 Eridani b may be a prime target for variability studies, as it has been shown to be affected strongly by clouds and is in a unique location in parameter space (Samland et al., 2017). By combining these ideas, as well as optimized post-processing algorithms for signal extraction, we may soon be able to achieve this feat and learn directly about the cloud physics and the atmospheric variability of directly imaged planets.

List of publications

These scientific publications have been submitted and published during my studies, or are in preparation for publication within this year. The following articles are partially or fully presented within this thesis:

1st AUTHOR PUBLICATION

1. **Samland, M.**; Mollière, P., Bonnefoy, M., Maire, A. -L., Cantalloube, F., Cheetham, A. C., Mesa, D., Gratton, R., Biller, B. A., Wahhaj, Z., Bouwman, J., Brandner, W., Melnick, D., Carson, J., Janson, M., Henning, T., Homeier, D., Mordasini, C., Langlois, M., Quanz, S. P., van Boekel, R., Zurlo, A., Schlieder, J. E., Avenhaus, H., Beuzit, J. -L., Boccaletti, A., Bonavita, M., Chauvin, G., Claudi, R., Cudel, M., Desidera, S., Feldt, M., Fusco, T., Galicher, R., Kopytova, T. G., Lagrange, A. -M., Le Coroller, H., Martinez, P., Moeller-Nilsson, O., Mouillet, D., Mugnier, L. M., Perrot, C., Sevin, A., Sissa, E., Vigan, A., Weber, L. 2017, A&A, 603, A57;
   "Spectral and atmospheric characterization of 51 Eridani b using VLT/SPHERE"

   [Part of Chapter 2 in this PhD Thesis]

2. **Samland, M.** et al., to be submitted
   "A temporal, non-local systematics model for direct detection of exoplanets at very small angular separations"

   [Chapter 4, will be the basis for a publication of the same or similar title]

3. **Samland, M.** et al., to be submitted
   "Adaptation of the CHARIS Pipeline for SPHERE-IFS"

   [Chapter 5, will be the basis for a publication of the same or similar title]

CONTRIBUTIONS TO PUBLICATIONS INCLUDED IN THIS THESIS

1. Cheetham, A. C., **Samland, M.** et al. 2018, A&A, 622, A80;
   "Spectral and orbital characterisation of the directly imaged giant planet HIP 65426 b"

   [Contribution summarized in Chapter 3]

2. Müller, A., Keppler, M., Henning, Th., **Samland, M.** et al. 2018, A&A, 617, L2;
   "Orbital and atmospheric characterization of the planet within the gap of the PDS 70 transition disk"

   [Contribution included in Chapter 3]

3. Bonnefoy, M., Perraut, K., Lagrange, A. -M., Delorme, P., Vigan, A., Line, M., Rodet, L., Ginski, C., Mourard, D., Marleau, G. -D., **Samland, M.** et al. 2018, A&A, 618, 32;

"The GJ 504 system revisited. Combining interferometric, radial velocity, and high contrast imaging data"

[Contribution summarized in Chapter 3]

PUBLICATIONS NOT INCLUDED IN THIS THESIS

1. Maire, A. -L., Rodet, L., Cantalloube, F., ... incl. **Samland, M.** et al. 2019, A&A, 624, 118;
"Hint of curvature in the orbital motion of the exoplanet 51 Eridani b using 3 yr of VLT/SPHERE monitoring"

2. Mesa, D., Bonnefoy, M., Gratton, R., ... incl. **Samland, M.** et al. 2019, A&A, 624, 4;
"Exploring the R CrA environment with SPHERE. Discovery of a new stellar companion"

3. Keppler, M., Teague, R., Bae, J., ... incl. **Samland, M.** et al. 2019, A&A, accepted;
"Highly structured disk around the planet host PDS 70 revealed by high-angular resolution observations with ALMA"

4. Lagrange, A. -M., Boccaletti, A., Langlois, M., ... incl. **Samland, M.** et al. 2019, A&A, 621, 8;
"Post-conjunction detection of β Pictoris b with VLT/SPHERE"

5. Sissa, E., Gratton, R., Garufi, A., ... incl. **Samland, M.** et al. 2018, A&A, 619, 160;
"High-contrast study of the candidate planets and protoplanetary disk around HD 100546"

6. Rodet, L., Bonnefoy, M., Durkan, S., ... incl. **Samland, M.** et al. 2018, A&A, 618, 23;
"Dynamical masses of M-dwarf binaries in young moving groups. I. The case of TWA 22 and GJ 2060"

7. Chauvin, G., Gratton, R., Bonnefoy, M., ... incl. **Samland, M.** et al. 2018, A&A, 617, 76;
"Investigating the young solar system analog HD 95086. A combined HARPS and SPHERE exploration"

8. Keppler, M., Benisty, M., MÃ¼ller, A., ... incl. **Samland, M.** et al. 2018, A&A, 617, 44;
"Discovery of a planetary-mass companion within the gap of the transition disk around PDS 70"

9. Maire, A. -L., Rodet, L., Lazzoni, C., ... incl. **Samland, M.** et al. 2018, A&A, 615, 177;
"VLT/SPHERE astrometric confirmation and orbital analysis of the brown dwarf companion HR 2562 B"

10. Boccaletti, A., Sezestre, E., Lagrange, A. -M., ... incl. **Samland, M.** et al. 2018, A&A, 614, 52;
"Observations of fast-moving features in the debris disk of AU Mic on a three-year timescale: Confirmation and new discoveries"

11. Peretti, S., Ségransan, D., Lavie, B., ... incl. **Samland, M.** et al. 2018, arxiv;
"Orbital and spectral analysis of the benchmark brown dwarf HD 4747B"

12. Lazzoni, C., Desidera, S., Marzari, F., ... incl. **Samland, M.** et al. 2018, A&A, 611, 43;
"Dynamical models to explain observations with SPHERE in planetary systems with double debris belts"

13. Ligi, R., Vigan, A., Gratton, R., ... incl. **Samland, M.** et al. 2018, MNRAS, 473;
"Investigation of the inner structures around HD 169142 with VLT/SPHERE"

14. Bonavita, M., D'Orazi, V., Mesa, D., ... incl. **Samland, M.** et al. 2017, A&A, 608, 106;
"Orbiting a binary. SPHERE characterisation of the HD 284149 system"

15. Chauvin, G., Desidera, S., Lagrange, A. -M., ... incl. **Samland, M.** et al. 2017, A&A, 605, 9;
"Discovery of a warm, dusty giant planet around HIP 65426"

16. Maire, A. -L., Stolker, T., Messina, S., ... incl. **Samland, M.** et al. 2017, A&A, 601, 134;
"Testing giant planet formation in the transitional disk of SAO 206462 using deep VLT/SPHERE imaging"

17. Konishi, M., Matsuo, T., Yamamoto, K., ... incl. **Samland, M.** et al. 2016, PASJ, 68, 92;
"A substellar companion to Pleiades HII 3441"

18. Olofsson, J., **Samland, M.** et al. 2016, A&A, 591, 108;
"Azimuthal asymmetries in the debris disk around HD 61005. A massive collision of planetesimals?"

19. Lagrange, A. -M., Langlois, M., Gratton, R. ... incl. **Samland, M.** et al. 2016, A&A, 591, 108; "A narrow, edge-on disk resolved around HD 106906 with SPHERE"

Bibliography

Ackerman, A. S., & Marley, M. S. 2001, ApJ, 556, 872 69, 79, 115, 118, 181, 184

Aime, C., & Soummer, R. 2004, ApJ, 612, L85 35

Allard, F. 2014, in IAU Symposium, Vol. 299, Exploring the Formation and Evolution of Planetary Systems, ed. M. Booth, B. C. Matthews, & J. R. Graham, 271–272 114, 118

Allard, F., Hauschildt, P. H., Alexander, D. R., Tamanai, A., & Schweitzer, A. 2001, ApJ, 556, 357 65, 78, 180

Allard, F., Homeier, D., & Freytag, B. 2012, Philosophical Transactions of the Royal Society of London Series A, 370, 2765 52, 58, 84

Allers, K. N., & Liu, M. C. 2013, ApJ, 772, 79 63

Amara, A., & Quanz, S. P. 2012a, MNRAS, 427, 948 38, 41, 132

—. 2012b, MNRAS, 427, 948 56

Andrews, S. M., Terrell, M., Tripathi, A., et al. 2018a, ApJ, 865, 157 118

Andrews, S. M., Huang, J., Pérez, L. M., et al. 2018b, ApJ, 869, L41 17, 18

Antichi, J., Dohlen, K., Gratton, R. G., et al. 2009, ApJ, 695, 1042 162

Apai, D., Kasper, M., Skemer, A., et al. 2016, ApJ, 820, 40 52, 186

Apai, D., Karalidi, T., Marley, M. S., et al. 2017, Science, 357, 683 186

Artigau, E., Nadeau, D., & Doyon, R. 2003, in IAU Symposium, Vol. 211, Brown Dwarfs, ed. E. Martín, 451 63

Bailey, V., Meshkat, T., Reiter, M., et al. 2014, ApJ, 780, L4 52

Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 2002, A&A, 382, 563 42

Baraffe, I., Chabrier, G., Barman, T. S., Allard, F., & Hauschildt, P. H. 2003, A&A, 402, 701 84

Baraffe, I., Homeier, D., Allard, F., & Chabrier, G. 2015a, A&A, 577, A42 84

—. 2015b, A&A, 577, A42 114, 118

Baudino, J.-L., Bézard, B., Boccaletti, A., et al. 2015, A&A, 582, A83 52

Bell, C. P. M., Mamajek, E. E., & Naylor, T. 2015, MNRAS, 454, 593 52

Benneke, B., & Seager, S. 2012, ApJ, 753, 100 45

Best, W. M. J., Liu, M. C., Magnier, E. A., et al. 2015, ApJ, 814, 118 62, 64

Beuzit, J.-L., Feldt, M., Dohlen, K., et al. 2008, in Proceedings of SPIE, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, 701418 52, 55, 161

Beuzit, J.-L., Vigan, A., Mouillet, D., et al. 2019, arXiv e-prints, arXiv:1902.04080 25, 128, 139, 160

Biller, B. A., Vos, J., Bonavita, M., et al. 2015, ApJ, 813, L23 61, 186

Binks, A. S., & Jeffries, R. D. 2014, MNRAS, 438, L11 52

Bloemhof, E. E., Dekany, R. G., Troy, M., & Oppenheimer, B. R. 2001, ApJ, 558, L71 36

Boccaletti, A., Abe, L., Baudrand, J., et al. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7015, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 1 55, 139

Bohlin, R. C. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 364, The Future of Photometric, Spectrophotometric and Polarimetric Standardization, ed. C. Sterken, 315 63

Bonnefoy, M., Zurlo, A., Baudino, J. L., et al. 2016, A&A, 587, A58 63, 65

Bonnefoy, M., Perraut, K., Lagrange, A. M., et al. 2018, A&A, 618, A63 113, 122, 179

Bonse, M. J., Quanz, S. P., & Amara, A. 2018, ArXiv e-prints, arXiv:1804.05063 132

Borysow, A. 1991, Icarus, 92, 273 65, 180

Bowler, B. P. 2016, PASP, 128, 102001 44, 47

Brandt, T. D., Kuzuhara, M., McElwain, M. W., et al. 2014, ApJ, 786, 1 128

Brandt, T. D., Rizzo, M., Groff, T., et al. 2017, Journal of Astronomical Telescopes, Instruments, and Systems, 3, 048002 160, 162, 163, 164, 169, 182

Brogi, M., Line, M., Bean, J., Désert, J. M., & Schwarz, H. 2017, ApJ, 839, L2 185

Brogi, M., Snellen, I. A. G., de Kok, R. J., et al. 2012, Nature, 486, 502 24

Burgasser, A. J. 2007, AJ, 134, 1330 82

Burgasser, A. J. 2014, in Astronomical Society of India Conference Series, Vol. 11, Astronomical Society of India Conference Series 62, 63

Burgasser, A. J., Burrows, A., & Kirkpatrick, J. D. 2006a, ApJ, 639, 1095 62

Burgasser, A. J., Cruz, K. L., Cushing, M., et al. 2010, ApJ, 710, 1142 63

Burgasser, A. J., Geballe, T. R., Leggett, S. K., Kirkpatrick, J. D., & Golimowski, D. A. 2006b, ApJ, 637, 1067 82

Burgasser, A. J., Kirkpatrick, J. D., Cruz, K. L., et al. 2006c, ApJS, 166, 585 63

Burgasser, A. J., Kirkpatrick, J. D., Reid, I. N., et al. 2000, AJ, 120, 473 63

Burgasser, A. J., Cushing, M. C., Kirkpatrick, J. D., et al. 2011, ApJ, 735, 116 65

Burningham, B., Pinfield, D. J., Leggett, S. K., et al. 2008, MNRAS, 391, 320 82

Burningham, B., Leggett, S. K., Homeier, D., et al. 2011, MNRAS, 414, 3590 65

Burrows, A., Sudarsky, D., & Hubeny, I. 2004, ApJ, 609, 407 22

Burrows, A., Marley, M., Hubbard, W. B., et al. 1997, ApJ, 491, 856 44

Caballero, J. A. 2018, Geosciences, 8, 362 42

Cagigal, M. P., & Canales, V. E. 2000, Journal of the Optical Society of America A, 17, 903 35

Cantalloube, F. 2016, Theses, Université Grenoble Alpes 29

Cantalloube, F., Mouillet, D., Mugnier, L. M., et al. 2015, ArXiv e-prints, arXiv:1508.06406 38, 41, 53, 56, 132, 135, 138, 156

Casagrande, L., Schönrich, R., Asplund, M., et al. 2011, A&A, 530, A138 100

Chabrier, G., Baraffe, I., Allard, F., & Hauschildt, P. 2000, ApJ, 542, 464 42

Chabrier, G., Baraffe, I., Leconte, J., Gallardo, J., & Barman, T. 2009, in American Institute of Physics Conference Series, Vol. 1094, 15th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, ed. E. Stempels, 102–111 77

Charnay, B., Bézard, B., Baudino, J.-L., et al. 2018, ApJ, 854, 172 115, 181

Chauvin, G., Desidera, S., Lagrange, A.-M., et al. 2017, A&A, 605, L9 47, 118, 181

Cheetham, A. C., Samland, M., Brems, S. S., et al. 2019, A&A, 622, A80 47, 49, 113, 118, 179, 181

Chilcote, J., Barman, T., Fitzgerald, M. P., et al. 2015, ApJ, 798, L3 52

Clarke, C. J., Gendrin, A., & Sotomayor, M. 2001, MNRAS, 328, 485 17

Claudi, R. U., Turatto, M., Gratton, R. G., et al. 2008, in Proceedings of SPIE, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, 70143E 55, 161

Conan, J. M., Madec, P. Y., & Rousset, G. 1994, in European Southern Observatory Conference and Workshop Proceedings, Vol. 48, 181 32

Currie, T., Fukagawa, M., Thalmann, C., Matsumura, S., & Plavchan, P. 2012, ApJ, 755, L34 40

Cushing, M. C., Marley, M. S., Saumon, D., et al. 2008, ApJ, 678, 1372 62, 79

Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003, 2MASS All Sky Catalog of point sources. 58

Cutri, R. M., Wright, E. L., Conrow, T., et al. 2013, Explanatory Supplement to the AllWISE Data Release Products, Tech. rep. 58

De Rosa, R. J., Nielsen, E. L., Blunt, S. C., et al. 2015, ApJ, 814, L3 52, 88, 96

de Wit, J., Gillon, M., Demory, B. O., & Seager, S. 2012, A&A, 548, A128 23

Delorme, P., Gagné, J., Malo, L., et al. 2012, A&A, 548, A26 65

Delorme, P., Dupuy, T., Gagné, J., et al. 2017a, ArXiv e-prints, arXiv:1703.00843 65

Delorme, P., Meunier, N., Albert, D., et al. 2017b, in SF2A-2017: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, ed. C. Reylé, P. Di Matteo, F. Herpin, E. Lagadec, A. Lançon, Z. Meliani, & F. Royer, 347–361 139, 175

Dohlen, K., Langlois, M., Saisse, M., et al. 2008, in Proceedings of SPIE, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, 70143L 55, 139, 161

Dou, J., Ren, D., Zhao, G., et al. 2015, ApJ, 802, 12 40

Einstein, A. 1905, Annalen der Physik, 322, 891 14

Faherty, J. K., Burgasser, A. J., West, A. A., et al. 2010, AJ, 139, 176 65

Faherty, J. K., Burgasser, A. J., Walter, F. M., et al. 2012, ApJ, 752, 56 63, 65

Feldt, M., Hippler, S., Obereder, A., Stuik, R., & Bertram, T. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9909, Adaptive Optics Systems V, 990961 31

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306 70, 114, 180

Foreman-Mackey, D., Montet, B. T., Hogg, D. W., et al. 2015, ApJ, 806, 215 135

Fried, D. L. 1966, Journal of the Optical Society of America (1917-1983), 56, 1372 29

Fusco, T., Sauvage, J.-F., Petit, C., et al. 2014, in Proceedings of SPIE, Vol. 9148, Adaptive Optics Systems IV, 91481U 32, 55, 139

Gagné, J., Burgasser, A. J., Faherty, J. K., et al. 2015, ApJ, 808, L20 63

Gagné, J., Lafrenière, D., Doyon, R., Malo, L., & Artigau, É. 2014, ApJ, 783, 121 63

Gaudi, B. S. 2010, Microlensing by Exoplanets, ed. S. Seager, 79–110 16

Gendron, E., & Lena, P. 1995, A&AS, 111, 153 32

Gibson, N. P. 2014, MNRAS, 445, 3401 158

Goebel, S. B., Guyon, O., Hall, D. N. B., et al. 2018, PASP, 130, 104502 36, 133

Gomez Gonzalez, C. A., Absil, O., Absil, P. A., et al. 2016, A&A, 589, A54 41

Gomez Gonzalez, C. A., Wertz, O., Absil, O., et al. 2017, AJ, 154, 7 132

Goodman, J., & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65 70

Goodman, J. W. 1975, Statistical properties of laser speckle patterns, ed. J. C. Dainty, Vol. 9, 9 35

Greco, J. P., & Brandt, T. D. 2016, ArXiv e-prints, arXiv:1602.00691 70, 115

Groff, T. D., Kasdin, N. J., Limbach, M. A., et al. 2015, in Proceedings of SPIE, Vol. 9605, Techniques and Instrumentation for Detection of Exoplanets VII, 96051C 25, 128, 160, 162

Guyon, O. 2004, ApJ, 615, 562 26

—. 2018, ARA&A, 56, 315 32

Guyon, O., Pluzhnik, E. A., Kuchner, M. J., Collins, B., & Ridgway, S. T. 2006, ApJS, 167, 81 34

Helling, C., Ackerman, A., Allard, F., et al. 2008, MNRAS, 391, 1854 118

Helou, G., & Walker, D. W., eds. 1988, Infrared astronomical satellite (IRAS) catalogs and atlases. Volume 7: The small scale structure catalog, Vol. 7, 1–265 58

Henning, T., & Stognienko, R. 1996, A&A, 311, 291 69

Hinkley, S., Oppenheimer, B. R., Soummer, R., et al. 2007, ApJ, 654, 633 36, 37

Hiranaka, K., Cruz, K. L., Douglas, S. T., Marley, M. S., & Baldassare, V. F. 2016, ArXiv e-prints, arXiv:1606.09485 65

Hoeg, E., Bässgen, G., Bastian, U., et al. 1997, A&A, 323 58

Hoeijmakers, H. J., Schwarz, H., Snellen, I. A. G., et al. 2018, ArXiv e-prints, arXiv:1802.09721 25, 179

Horne, K. 1986, PASP, 98, 609 164

Hugot, E., Ferrari, M., El Hadi, K., et al. 2012, A&A, 538, A139 161

Ingraham, P., Marley, M. S., Saumon, D., et al. 2014, ApJ, 794, L15 52

Jaeger, C., Molster, F. J., Dorschner, J., et al. 1998, A&A, 339, 904 69

Jensen-Clem, R., Millar-Blanchaer, M., Mawet, D., et al. 2016, ApJ, 820, 111 26

Jones, A., Noll, S., Kausch, W., Szyszka, C., & Kimeswenger, S. 2013, A&A, 560, A91 63

Jovanovic, N., Guyon, O., Martinache, F., et al. 2015a, ApJ, 813, L24 186

Jovanovic, N., Martinache, F., Guyon, O., et al. 2015b, PASP, 127, 890 162

Kao, M. M., Hallinan, G., Pineda, J. S., et al. 2016, ApJ, 818, 24 63

Keppler, M., Benisty, M., Müller, A., et al. 2018, A&A, 617, A44 47, 114, 180

King, R. R., McCaughrean, M. J., Homeier, D., et al. 2010, A&A, 510, A99 82

Knutson, H. A., Charbonneau, D., Cowan, N. B., et al. 2009, ApJ, 690, 822 23

Knutson, H. A., Lewis, N., Fortney, J. J., et al. 2012, ApJ, 754, 22 23

Kolmogorov, A. 1941, Akademiia Nauk SSSR Doklady, 30, 301 29

Konopacky, Q. M., Barman, T. S., Macintosh, B. A., & Marois, C. 2013, Science, 339, 1398 179

Kraus, A. L., Ireland, M. J., Martinache, F., & Lloyd, J. P. 2008, ApJ, 679, 762 83

Kuhn, J. R., Potter, D., & Parise, B. 2001, ApJ, 553, L189 26

Kuzuhara, M., Tamura, M., Kudo, T., et al. 2013, ApJ, 774, 11 122

Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., & Artigau, É. 2007, ApJ, 660, 770 38, 40, 132

Lagrange, A.-M., Gratadour, D., Chauvin, G., et al. 2009, A&A, 493, L21 139

Lagrange, A.-M., Bonnefoy, M., Chauvin, G., et al. 2010, Science, 329, 57 52, 56

Lagrange, A.-M., Boccaletti, A., Langlois, M., et al. 2019, A&A, 621, L8 139

Lang, D., Hogg, D. W., & Schlegel, D. J. 2016, AJ, 151, 36 73

Langlois, M., Vigan, A., Moutou, C., et al. 2013, in Proceedings of the Third AO4ELT Conference, ed. S. Esposito & L. Fini, 63 55

Lee, J.-M., Heng, K., & Irwin, P. G. J. 2013, ApJ, 778, 97 78

Leggett, S. K., Allard, F., Dahn, C., et al. 2000, ApJ, 535, 965 63

Line, M. R., Marley, M. S., Liu, M. C., et al. 2016, ArXiv e-prints, arXiv:1612.02809 45, 98, 100

Line, M. R., Teske, J., Burningham, B., Fortney, J. J., & Marley, M. S. 2015, ApJ, 807, 183 45, 79, 98, 100, 101

Line, M. R., Zhang, X., Vasisht, G., et al. 2012, ApJ, 749, 93 45

Liu, M. C., Dupuy, T. J., & Allers, K. N. 2016, ApJ, 833, 96 44, 63

Liu, M. C., Leggett, S. K., & Chiu, K. 2007, ApJ, 660, 1507 62

Lloyd, J. P., Graham, J. R., Kalas, P., et al. 2001, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4490, Multifrequency Electronic/Photonic Devices and Systems for Dual-Use Applications, ed. A. R. Pirich, P. L. Repak, P. S. Idell, & S. R. Czyzak, 290–297 33

MacDonald, R. J., & Madhusudhan, N. 2017, MNRAS, 469, 1979 45

Mace, G. N., Kirkpatrick, J. D., Cushing, M. C., et al. 2013, ApJS, 205, 6 62, 64

Macintosh, B., Poyneer, L., Sivaramakrishnan, A., & Marois, C. 2005, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 5903, Astronomical Adaptive Optics Systems and Applications II, ed. R. K. Tyson & M. Lloyd-Hart, 170–177 36

Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, Proceedings of the National Academy of Science, 111, 12661 52, 128

Macintosh, B., Graham, J. R., Barman, T., et al. 2015, Science, 350, 64 52, 58, 61, 62, 69, 74, 77, 79, 80, 81, 84, 86, 88, 105, 112, 139, 175, 176, 180

Macintosh, B. A., Graham, J. R., Palmer, D. W., et al. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7015, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 18 25, 88, 160

Maire, A.-L., Boccaletti, A., Rameau, J., et al. 2014, A&A, 566, A126 56

Maire, A.-L., Bonnefoy, M., Ginski, C., et al. 2016a, A&A, 587, A56 52, 56

Maire, A.-L., Langlois, M., Dohlen, K., et al. 2016b, in SPIE Conf. Ser., Vol. 9908, 990834 55, 57, 149

Maire, A. L., Rodet, L., Cantalloube, F., et al. 2019, A&A, 624, A118 88, 96, 97

Mamajek, E. E., & Bell, C. P. M. 2014, MNRAS, 445, 2169 52

Mancini, L., Giordano, M., Mollière, P., et al. 2016a, MNRAS, 461, 1053 69

Mancini, L., Kemmer, J., Southworth, J., et al. 2016b, MNRAS, arXiv:1603.08031 69

Marleau, G.-D., & Cumming, A. 2014, MNRAS, 437, 1378 77

Marleau, G.-D., Klahr, H., Kuiper, R., & Mordasini, C. 2017, ApJ, 836, 221 77, 180

Marley, M. S., Fortney, J. J., Hubickyj, O., Bodenheimer, P., & Lissauer, J. J. 2007, ApJ, 655, 541 77

Marley, M. S., Saumon, D., Cushing, M., et al. 2012, ApJ, 754, 135 44

Marley, M. S., Saumon, D., & Goldblatt, C. 2010, ApJ, 723, L117 81

Marocco, F., Day-Jones, A. C., Lucas, P. W., et al. 2014, MNRAS, 439, 372 65

Marois, C., Correia, C., Galicher, R., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9148, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 0 40, 56

Marois, C., Doyon, R., Nadeau, D., et al. 2005, PASP, 117, 745 25

Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, ApJ, 641, 556 38, 39, 40, 55, 132

Marois, C., Lafrenière, D., Macintosh, B., & Doyon, R. 2008a, ApJ, 673, 647 35, 138

Marois, C., Macintosh, B., Barman, T., et al. 2008b, Science, 322, 1348 52

Marois, C., Macintosh, B., & Véran, J.-P. 2010a, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7736, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 1 56

Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., & Barman, T. 2010b, Nature, 468, 1080 52

Martinache, F. 2019, arXiv e-prints, arXiv:1904.02087 34

Mawet, D., Milli, J., Wahhaj, Z., et al. 2014, ApJ, 792, 97 83, 138

Meeker, S. R., Mazin, B. A., Walter, A. B., et al. 2018, PASP, 130, 065001 26

Mermilliod, J. C. 2006, VizieR Online Data Catalog, 2168 58

Mesa, D., Gratton, R., Zurlo, A., et al. 2015, A&A, 576, A121 56, 98

Milli, J., Mouillet, D., Mawet, D., et al. 2013, A&A, 556, A64 21, 26

Milli, J., Banas, T., Mouillet, D., et al. 2016, in Proceedings of SPIE, Vol. 9909, Adaptive Optics Systems V, 99094Z 36, 37, 133, 134, 144

Mollière, P., & Snellen, I. A. G. 2019, A&A, 622, A139 185

Mollière, P., van Boekel, R., Bouwman, J., et al. 2017, A&A, 600, A10 69, 114, 181, 185

Mollière, P., van Boekel, R., Dullemond, C., Henning, T., & Mordasini, C. 2015a, ApJ, 813, 47 69, 114

—. 2015b, ApJ, 813, 47 181, 185

Mollière, P., Wardenier, J. P., van Boekel, R., et al. 2019, arXiv e-prints, arXiv:1904.11504 185

Monet, D. G., Dahn, C. C., Vrba, F. J., et al. 1992, AJ, 103, 638 63

Montet, B. T., Bowler, B. P., Shkolnik, E. L., et al. 2015, ApJ, 813, L11 52

Mordasini, C. 2013, A&A, 558, A113 42, 77

Mordasini, C., Alibert, Y., Georgy, C., et al. 2012a, A&A, 547, A112 77, 180

Mordasini, C., Alibert, Y., Klahr, H., & Henning, T. 2012b, A&A, 547, A111 77, 180

Mordasini, C., Klahr, H., Alibert, Y., Miller, N., & Henning, T. 2014, A&A, 566, A141 78

Mordasini, C., Marleau, G.-D., & Mollière, P. 2017, A&A, 608, A72 42, 115

Morley, C. V., Fortney, J. J., Marley, M. S., et al. 2012, ApJ, 756, 172 52, 69, 70, 82, 90, 93, 100, 105, 107, 110, 184

Müller, A., Keppler, M., Henning, T., et al. 2018, A&A, 617, L2 47, 49, 113, 179, 180

Müller, M., & Weigelt, G. 1985, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 556, International Conference on Speckle, ed. H. H. Arsenault, 270–273 38, 132

Naud, M.-E., Artigau, É., Malo, L., et al. 2014, ApJ, 787, 5 65

N'Diaye, M., Dohlen, K., Fusco, T., & Paul, B. 2013, A&A, 555, A94 32

N'Diaye, M., Vigan, A., Dohlen, K., et al. 2016, A&A, 592, A79 32

Noll, S., Kausch, W., Barden, M., et al. 2012, A&A, 543, A92 63

Olofsson, J., Samland, M., Avenhaus, H., et al. 2016, A&A, 591, A108 136

Palik, E. 2012, Handbook of Optical Constants of Solids No. Bd. 1 (Elsevier Science) 69

Pascucci, I., Testi, L., Herczeg, G. J., et al. 2016, ApJ, 831, 125 118

Patel, R. I., Metchev, S. A., & Heinze, A. 2014, ApJS, 212, 10 52

Paul, B., Sauvage, J. F., Mugnier, L. M., et al. 2014, A&A, 572, A32 32

Pavlov, A., Möller-Nilsson, O., Feldt, M., et al. 2008, in Proceedings of SPIE, Vol. 7019, Advanced Software and Control for Astronomy II, 701939 55, 139, 166

Perryman, M. 2011, The Exoplanet Handbook (Cambridge: Cambridge University Press), X, 410 S., includes bibliographical references and index 15, 19, 22

Peters, M. A., Groff, T., Kasdin, N. J., et al. 2012, in Proceedings of SPIE, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, 84467U 160, 162

Petit, C., Sauvage, J. F., Sevin, A., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8447, Adaptive Optics Systems III, 84471Z 161

Pinte, C., Price, D. J., Ménard, F., et al. 2018, ApJ, 860, L13 17

Prugniel, P., Soubiran, C., Koleva, M., & Le Borgne, D. 2007, ArXiv Astrophysics e-prints, astro-ph/0703658 58

Pueyo, L., & Kasdin, N. J. 2007, ApJ, 666, 609 25

Pueyo, L., Crepp, J. R., Vasisht, G., et al. 2012, ApJS, 199, 6 38, 40

Racine, R., Walker, G. A. H., Nadeau, D., Doyon, R., & Marois, C. 1999a, PASP, 111, 587 25

—. 1999b, PASP, 111, 587 56

Radigan, J., Jayawardhana, R., Lafrenière, D., et al. 2012, ApJ, 750, 105 61

Ragazzoni, R., & Farinato, J. 1999, A&A, 350, L23 31

Rajan, A., Rameau, J., De Rosa, R. J., et al. 2017, AJ, 154, 10 48, 88, 89, 90, 91, 92, 93, 94, 95, 175, 176

Rameau, J., Chauvin, G., Lagrange, A.-M., et al. 2013, ApJ, 772, L15 52

Ramírez, I., Allende Prieto, C., & Lambert, D. L. 2013, ApJ, 764, 78 100, 101

Ren, D., Dou, J., Zhang, X., & Zhu, Y. 2012, ApJ, 753, 99 40

Riviere-Marichalar, P., Barrado, D., Montesinos, B., et al. 2014, A&A, 565, A68 52

Rosenthal, E. D., Gurwell, M. A., & Ho, P. T. P. 1996, Nature, 384, 243 25

Rousset, G., Lacombe, F., Puget, P., et al. 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4839, Adaptive Optical System Technologies II, ed. P. L. Wizinowich & D. Bonaccini, 140–149 31

Ruffio, J.-B., Macintosh, B., Wang, J. J., et al. 2017, ApJ, 842, 14 132, 138

Samland, M., Mollière, P., Bonnefoy, M., et al. 2017, A&A, 603, A57 47, 48, 51, 88, 94, 113, 114, 122, 139, 156, 159, 166, 174, 175, 176, 186

Saumon, D., Marley, M. S., Cushing, M. C., et al. 2006, ApJ, 647, 552 100

Sauvage, J. F., Mugnier, L., Paul, B., & Villecroze, R. 2012, Optics Letters, 37, 4808 32

Savransky, D. 2015, ApJ, 800, 100 41

Schölkopf, B., Hogg, D. W., Wang, D., et al. 2016, Proceedings of the National Academy of Sciences, 113, 7391 133

Scott, A., & Duley, W. W. 1996, ApJS, 105, 401 69

Servoin, J. L., & Piriou, B. 1973, Physica Status Solidi B Basic Research, 55, 677 69

Shack, R. V. 1971, J. Opt. Soc. Am., 61, 656 30

Simon, M., & Schaefer, G. H. 2011, ApJ, 743, 158 52

Skemer, A. J., Hinz, P., Montoya, M., et al. 2015, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9605, Techniques and Instrumentation for Detection of Exoplanets VII, 96051D 160

Skemer, A. J., Morley, C. V., Zimmerman, N. T., et al. 2016, ApJ, 817, 166 78, 79

Snellen, I., de Kok, R., Birkby, J. L., et al. 2015, A&A, 576, A59 25, 184

Snellen, I. A. G., Brandl, B. R., de Kok, R. J., et al. 2014, Nature, 509, 63 24, 179

Snellen, I. A. G., de Kok, R. J., de Mooij, E. J. W., & Albrecht, S. 2010, Nature, 465, 1049 24

Snellen, I. A. G., de Mooij, E. J. W., & Albrecht, S. 2009, Nature, 459, 543 23

Soummer, R. 2005, ApJ, 618, L161 55, 139

Soummer, R., Pueyo, L., & Larkin, J. 2012, ApJ, 755, L28 38, 41, 56, 132

Sozzetti, A., Giacobbe, P., Lattanzi, M. G., et al. 2014, MNRAS, 437, 497 16

Sparks, W. B., & Ford, H. C. 2002, ApJ, 578, 543 25

Spiegel, D. S., & Burrows, A. 2012, ApJ, 745, 174 42

Spiegel, D. S., Burrows, A., & Milsom, J. A. 2011, ApJ, 727, 57 42

Stephens, D. C., Leggett, S. K., Cushing, M. C., et al. 2009, ApJ, 702, 154 63

Stolker, T., Bonse, M. J., Quanz, S. P., et al. 2019, A&A, 621, A59 132

Teague, R., Bae, J., Bergin, E. A., Birnstiel, T., & Foreman-Mackey, D. 2018, ApJ, 860, L12 17

Thalmann, C., Schmid, H. M., Boccaletti, A., et al. 2008, in Proceedings of SPIE, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, 70143F 55

Thiébaut, E., Denis, L., Mugnier, L. M., et al. 2016, in Adaptive Optics Systems V, ed. E. Marchetti, L. M. Close, & J.-P. Véran, Vol. 9909, Proc. Soc. Photo-Opt. Instrum. Eng., 990957–990957–10, conference date Jun. 2016, Edinburgh, UK 57

Tuthill, P. G., Monnier, J. D., Danchi, W. C., Wishnow, E. H., & Haniff, C. A. 2000, PASP, 112, 555 83

Venemans, B. P., Bañados, E., Decarli, R., et al. 2015, ApJ, 801, L11 73

Vigan, A., Gry, C., Salter, G., et al. 2015, MNRAS, 454, 129 171

Vigan, A., Moutou, C., Langlois, M., et al. 2010, MNRAS, 407, 71 55

Vigan, A., Bonnefoy, M., Ginski, C., et al. 2016a, A&A, 587, A55 52, 65, 79

—. 2016b, A&A, 587, A55 62

Vigan, A., Bonavita, M., Biller, B., et al. 2017, A&A, 603, A3 128, 166

Vigan, A., N'Diaye, M., Dohlen, K., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10703, Adaptive Optics Systems VI, 107035O 32

Wagner, K., Apai, D., Kasper, M., et al. 2016, Science, 353, 673 52

Wagner, K., Follete, K. B., Close, L. M., et al. 2018, ApJ, 863, L8 114, 115, 181

Wahhaj, Z., Cieza, L. A., Mawet, D., et al. 2015, A&A, 581, A24 40, 58

Waldmann, I. P., Rocchetto, M., Tinetti, G., et al. 2015, ApJ, 813, 13 45

Wang, D., Hogg, D. W., Foreman-Mackey, D., & Schölkopf, B. 2016, PASP, 128, 094503 134

Wang, J., Mawet, D., Fortney, J. J., et al. 2018, AJ, 156, 272 185

Wilby, M. J., Keller, C. U., Snik, F., Korkiakoski, V., & Pietrow, A. G. M. 2017, A&A, 597, A112 137

Yurchenko, S. N., & Tennyson, J. 2014, MNRAS, 440, 1649 118

Zahnle, K., Marley, M. S., Morley, C. V., & Moses, J. I. 2016, ApJ, 824, 137 75, 86

Zahnle, K. J., & Marley, M. S. 2014, ApJ, 797, 41 70

Zapatero Osorio, M. R., Béjar, V. J. S., Miles-Páez, P. A., et al. 2014, A&A, 568, A6 65

Zhang, K., Blake, G. A., & Bergin, E. A. 2015, ApJ, 806, L7 17

Zhu, Z. 2015, ApJ, 799, 16 181

Zuckerman, B., Song, I., Bessell, M. S., & Webb, R. A. 2001, ApJ, 562, L87 52

Zurlo, A., Vigan, A., Galicher, R., et al. 2016, A&A, 587, A57 52

Danksagung

Ich danke meinen Eltern, die mich immer unterstützt haben und dies immer noch tun. Ihr habt mich immer meinen Weg selbst entscheiden lassen und mir geholfen, diesen auch erfolgreich zu beschreiten. Ihr seid die besten.

Over the years, I have met too many people to thank you all individually. First of all I thank my office mates, ex-office mates, and extended adopted office mates. You're the best people that I could have spend the majority of my waking life with for the last almost four years. There are too many wonderful memories and experiences to count, and I hope they will continue to accumulate for the rest of our lives, regardless of which corner of the scientific or non-scientific world we end up in.

I thank all the people that frequent student coffee for the hours of great discussions after lunch, as well as my IMPRS dinner buddies for even more fun over dinner each Thursday. I'll miss the awkward silence that inevitably follows the question "where do we want to eat dinner tonight?", when no one has a clear opinion.

My friends outside of the institute deserve a special thanks for keeping in touch with me. My friends from Japan, my table tennis club, university, and wherever else we have met... It's not always easy to keep in touch with everyone, so for managing this: Yay, us! Banzai!

My dear sister and her husband also deserves a special thanks as well. You're awesome in general. Take care of your Scottish home!

I thank all the pianists on YouTube for providing me with the inspiration to try harder and practice, and for showing me what can be accomplished with effort and dedication.

I want to thank all my scientific collaborators, let's keep doing good science together! Science is a group endeavour, and you're making it worthwhile.

I especially want to thank my advisor, Wolfgang Brandner, for providing me with guidance for all these years, and even before then, during my master studies. Your office was always open to me, and I learned an incredible amount from you.

Lastly, I want to thank Thomas Henning and Sabine Reffert for agreeing to be my referees, as well as having been on my IMPRS Thesis Committee. You have been of great help to me, especially whenever I was in danger of getting too caught up in detail to remember the bigger picture.

Declaration of Originality

Declaration of Originality

I hereby declare that this thesis is my own work and that I have used no other than the stated sources and aids.

Declaration

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, der 24. Mai 2019

(Matthias Samland)