

Direct Imaging of Exoplanets Closer to Stars

Thesis by
Shubh Agrawal

In Partial Fulfillment of the Requirements for the
Degree of
Bachelor of Science in Physics

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2022
Defended June 1, 2022

© 2022

Shubh Agrawal

ORCID: 0000-0003-2429-5811

All rights reserved

ACKNOWLEDGEMENTS

This thesis was made possible with the utmost support and guidance of my advisors, Dimitri Mawet and Jean-Baptiste Ruffio. Throughout the last year, they have encouraged me to strive forward and answered every question I have had. They have made me excited about the work I do and helped me persevere through the problems that have come along the way. I believe that I will remain grateful to them throughout my academic career.

I thank our collaborators, Bruce Macintosh, Quinn Konopacky, Max Millar-Blanchaer, Kielan Hoch, and Eric Nielsen, among others, whose insight and skill have been invaluable to this project. I extend my gratitude to past advisors, including Jamie Bock, Bryan Steinbach, and Alessandro Schillaci, as well as, Alan Weinstein and Sunil Golwala. While enriching my Caltech experience, they also helped me develop the skills and temperament I needed to pursue the accomplishments of this thesis. I thank Alexandra Elbakyan for her efforts in promoting accessibility in science.

Finally, I want to thank my partner, Shreya Singh, for her unwavering encouragement, belief, and patience. I also thank my friends and family for their unending support.

ABSTRACT

Detecting exoplanets through direct imaging at lower angular separations, where more planets are expected to be, is limited by the variability of the stellar point spread function. Integral field spectrographs like OSIRIS at the Keck Observatory can leverage high spectral resolution to search for new planets at smaller separations (< 0.3 arcseconds) by detecting their distinct spectral signature compared to the diffracted starlight. In this thesis, we present the mid-survey results of a search for planets around 23 targets in the Ophiuchus and Taurus star-forming regions.

We use this pathfinder survey with Keck/OSIRIS to demonstrate our technique and compare the final sensitivities to other classical imaging techniques, particularly at separations of $0.05 - 0.3$ arcseconds. We detect a M dwarf companion around HD 148352 at a $\approx 34\sigma$ significance level. We measure this binary star companion to be at an angular separation of roughly 0.11 milliarcseconds, with a contrast of 0.38% , effective temperature $T_{\text{eff}} \approx 3200$ K, and radial velocity $RV \approx 12$ km/s. We also present other low-significance objects, along with detection maps and sensitivity limits around these 23 targets.

We use our open-source data analysis pipeline, called the Broad Repository for Exoplanet Analysis, Detection, and Spectroscopy (bread), as the framework for this planet search. bread operates on high spectral resolution data from existing and in-development instruments. Our code is based on a forward-modeling framework, which is statistically more accurate than classical cross-correlation techniques. It includes a built-in optimization and analytical marginalization of linear parameters in the forward model, therefore limiting the number of parameters to be explored by the posterior sampling method. We allow users to select forward models, parameters to detect and analyze, and fitting methods like Markov Chain Monte Carlo sampling, grid optimization, and gradient descent. bread provides a flexible framework to retrieve radial velocity, spin, and atmospheric parameters of high-contrast companions. We also describe wavelength and resolution calibration, transmission and spectra calculation, and bad pixel identification techniques.

Our work will be applicable to future integral field spectrographs like NIRSpec on the James Webb Space Telescope and other first light instruments on the future Extremely Large Telescopes, which are poised to become the next generation of exoplanet detection facilities.

TABLE OF CONTENTS

Acknowledgements	iii
Abstract	iv
Table of Contents	v
List of Illustrations	vii
List of Tables	xiii
Chapter I: Introduction	1
1.1 Indirect detection methods	1
1.2 Direct imaging	3
1.3 Theories of Formation and Migration of Gas-Giant Planets	4
1.4 Integral Field Spectroscopy and the OSIRIS instrument	5
1.5 Sensitivity of Moderate-resolution integral field spectroscopy	6
Chapter II: Approach and Methods	8
2.1 Observations and Data Collection	8
2.2 Constructing Datacubes through OSIRIS Data Reduction Pipeline	12
2.3 Developing Open Source Framework breads	13
2.4 Wavelength and Resolution Calibration	14
2.5 Extraction of Stellar Spectrum and Telluric Calibration	17
2.6 Forward Model, Noise Scaling, and Likelihood	18
2.7 Bad Pixel Identification, Pixel Bleeding, and Correction Strategies	23
2.8 Planet Search, Frame Combination, and PCA of Residuals	27
2.9 Noise Normalization for combined map	29
2.10 Throughput Correction	30
2.11 Contrast Curves and Sensitivity	34
2.12 Validation and Analysis of a Detection	35
Chapter III: Results and Discussion	37
3.1 Companion detection with combined signal-to-noise ratio maps	37
3.2 Validation of threshold SNR for detection	39
3.3 Current and Potential Sensitivity measured in Contrast Curves	41
3.4 Potential improvement with RV CCF Calibration	45
3.5 Factors limiting Sensitivity and Comparison with other methods	48
3.6 Detection of a binary companion around HD 148352	49
3.7 Possible $< 5\sigma$ detection in ROXs 35A data set	52
Chapter IV: Further Work	57
4.1 Further Optimization of Model	57
4.2 Future Observations, and Investigation of $< 5\sigma$ detections	58
4.3 Probing Characterization Capabilities	59
4.4 Investigating Capabilities with the James Webb Space Telescope and beyond	61
Bibliography	62

Appendix A: Appendix A: Combined Signal-to-Noise Ratio Maps before RV CCF Normalization	67
Appendix B: Appendix B: Combined Signal-to-Noise Ratio Maps after RV CCF calibration	75

LIST OF ILLUSTRATIONS

<i>Number</i>	<i>Page</i>
1.1 Detected exoplanets with measured or estimated mass and orbital period. (NASA Astrophysics Division, 2022) Color denotes the technique that detected the exoplanet for the first time.	2
1.2 Schematic of data reduction pipeline, drawn for target HD 148352. We discussed a detected binary companion in Section 3.6. From left to right: (1) The system with the OSIRIS field of view; the y axis points North; companion signal is hidden in the glare of the target star. (2) Raw two-dimensional data frame obtained from the OSIRIS detectors after observation; spectra are arranged as rows. (3) Data transformed by OSIRIS DRP into data cubes. The bright speckle to the left of the stellar PSF is not the companion signal, but rather a residue from pixel bleeding issues, described in Section 2.7 (4) Signal-to-noise ratio map, with detection of a companion, generated as described in Sections 2.8, 3.1, and 3.6.	6
2.1 Nominal examples of wavelength calibration performed at two different spatial locations. We take spectral data at each location (blue) for the long sky exposures, and fit a model for emission lines of the OH ⁻ radical (orange). Parameters fitted quantifies the broadening and wavelength-shift of the spectral lines.	16
2.2 Nominal two-dimensional heat map encoding values for the constant offset a_0 , computed as part of the wavelength calibration for 2022/06/26. As expected, the offset is of the order of 0.01 nm or 1/10 of a pixel.	16
2.3 (a) Nominal examples of extracted spectra of HIP 73049 and SR 3, both of which are A0 standard stars. As these are of the same spectral type, we expect the spectra to look similar. (b) Nominal examples of computed sky transmission, obtained by dividing extracted spectrum by a theoretical model of A0 star spectra. These data sets are from the same observation night, so we expect the transmission to be similar.	18

- 2.4 Schematic of the generation of the forward model for a spatial location. (Top Left) A model for the observed star spectrum (including a transmission component), computed from the data. (Bottom Left) A planet spectrum is computed from a theoretical model. (Top Right) We multiply the star spectrum with a continuum model (for our case, computed using a *spline interpolation*) to get diffracted starlight model. (Bottom Right) A linear combination ($M_\psi\phi$) of the diffracted starlight and planet models is used to fit the spectral data at the location. 19
- 2.5 Nominal example of the pixel bleeding issue. Left panel shows the collapsed two-dimensional spatial field of view of Keck/OSIRIS, with the center of the target PSF marked by a light blue cross. Right shows spectra at several locations (labelled by spatial indices), as detector counts for a given wavelength in microns. Pixel bleeding is exhibited in the atypical continuum behavior, at locations roughly 10 pixels to the left/right of the target, and within a few pixels in the vertical direction. The continuum at these locations is significantly higher at one edge of the wavelength range. 26
- 2.6 Algorithmic throughput t_a (ratio of flux recovered by our approach, to the flux that was injected as a simulated fake planet signal) as a function of separation from the star (in units of pixels, each spatial pixel corresponding to 20 milliarcseconds). As described in Section 2.10, we empirically note that throughput dominantly depends on separation from the star. We can interpolate between the t_a values computed at every pixel (as presented in this figure) to get t_a everywhere in the field of view. Notably, t_a tends to 80 – 90% (expected 100%) for most targets far away from the star, and to 10 – 20% (expected 0%) at the location of the star. Our model is able to distinguish most of the planetary signal from the diffracted starlight. 33
- 3.1 Probability distribution function of residual signal-to-noise ratio values in detection maps for all 23 targets, (a) before any noise normalization steps, and (b) after applying noise normalization, excluding RV CCF calibration. Black dotted line represents the ideal Gaussian PDF. We note that normalization reduces overestimation of the magnitude of SNR. 40

3.2	Probability distribution function of residual signal-to-noise ratio values in detection maps for two targets, before and after noise normalization steps. Black dotted line represents the ideal Gaussian PDF. We note that normalization reduces overestimation of the magnitude of SNR. The distribution for HD 148352 validates our detection threshold used in the discussion for Section 3.6.	41
3.3	Contrast curves, as a function of separation from the star, at 5σ significance for two targets. We include potential sensitivity (contrast before noise normalization) in orange, sensitivity after noise normalization (excluding RV CCF calibration) in green, and the stellar point spread function profile (quantifying speckle noise) in blue. Our sensitivity is well-below the speckle noise, with improvement possible to bring current sensitivity closer to potential values.	42
3.4	Contrast curves before noise normalization, mapping potential sensitivity at a 5σ significance level as a function of separation from star, for all 23 targets. Variability in sensitivity across targets is expected due to variations in amount of data, observing conditions, and spectral types or brightness. Scales of x and y axes are same as Figure 3.5.	43
3.5	Contrast curves after noise normalization (excluding RV CCF calibration), mapping current sensitivity at a 5σ significance level as a function of separation from star, for all 23 targets. Variability in sensitivity across targets is expected due to variations in amount of data, observing conditions, and spectral types or brightness. Scales of x and y axes are same as Figure 3.4.	44
3.6	Probability distribution function of residual signal-to-noise ratio values in detection maps for all 23 targets, after all steps of noise normalization, including RV CCF calibration. Black dotted line represents the ideal Gaussian PDF. We note that this additional step of normalization results in a distribution that is more consistent with the ideal Gaussian.	46

3.7	Contrast curves, as a function of separation from the star, at 5σ significance for two targets. We include potential sensitivity (contrast before noise normalization) in orange, final sensitivity after noise normalization (including RV CCF calibration) in green, and the stellar point spread function profile (quantifying speckle noise) in blue. Our sensitivity is well-below the speckle noise, with current sensitivity brought closer to potential values due to RV CCF calibration.	47
3.8	Contrast curves after all steps of noise normalization including RV CCF calibration, mapping final sensitivity at a 5σ significance level as a function of separation from star, for all 23 targets. Variability in sensitivity across targets is expected due to variations in amount of data, observing conditions, and spectral types or brightness. Note that this latter plot is graphed on a different scale on x and y axes than the earlier similar figures.	48
3.9	Sensitivity of our pathfinder survey with Keck/OSIRIS, compared to other direct imaging surveys. We estimate better sensitivity at separations of 50 to 300 milliarcseconds, with a potential to improve to even better contrasts.	50
3.10	Detection of a binary companion around HD 148352, with $\text{SNR} \approx 34$ (blue and purple dots) at $\Delta y = \Delta x = -4$ spaxels from the host star (red cross). This constrains the astrometry of the companion.	51
3.11	Our detection (red) overlaid on contrast curves for HD148352, similar to Figure 3.7b. Detection at 3.8% is above the sensitivity limits, while being significantly lower (roughly 1.5 orders of magnitude) than the stellar PSF profile. Our methods are able to model for diffracted starlight at separations as close as 5-6 spaxels (spatial locations).	52
3.12	Radial velocity cross correlation function for detected binary, at $\Delta y = \Delta x = -4$ spaxels from the star. We see a peak around $RV = 0$, which is expected for a real astrophysical signal. SNR is not normalized, up to a constant scaling factor.	53
3.13	Corner plot encoding outputs from an MCMC sampler used to characterize the T_{eff} , RV , $\log g$, and spin of the detected HD 148352 binary companion. breads can be similarly used to easily characterize companions in data sets.	54
3.14	Low-significance ($\approx 4\sigma$) blob (blue and purple dots) around ROXS 35A, at $\Delta y = 0$ or 1 and $\Delta x = -4$ spaxels from the host star (red cross).	55

3.15	Low significance blob (red) overlaid on contrast curves for ROXS 35A, similar to Figure 3.7b. If confirmed, this detection at just ≈ 4 spaxels from the star would demonstrate sensitivity capabilities of our technique.	56
3.16	Radial velocity cross correlation function for detected binary, at $\Delta y = \Delta x = -4$ spaxels from the star. We see a peak around $RV = 0$, which is expected for a real astrophysical signal. However, other peaks are appreciably high relative to the highest peak at $RV = 0$ SNR is not normalized, up to a constant scaling factor.	56
4.1	An initial example of setting the number of spline nodes to optimize our forward model. for four targets, we plot (a) throughputs and (b) uncertainty levels, which is directly proportional to the sensitivity, by a factor of 5 (threshold SNR level), as a function of splines nodes in our model. These values are computed at 16 different location at a separation of 60 milliarcseconds. We optimize the model by choosing setting this hard-coded parameter of our model to be one that increases throughput, while lowering noise levels.	58
4.2	An example of how to quantify the characterization capabilities of our model. Each color represents the T_{eff} of the simulated planet signal. If an SNR curve peaks around the same T_{eff} , we can claim that our method can accurately measure the effective temperature of a detected companion. In this example, we note that planets with $T_{\text{eff}} = 1100$ or 1300 K can be well-characterized (as the SNR curves peak around the same values), while there seems to be some degeneracy between planet signals with $T_{\text{eff}} = 1500, 1700,$ and 1900 K, which results in an SNR curve with two significant peaks.	60
A.1	Signal-to-noise ratio maps for AB_Aur	67
A.2	Signal-to-noise ratio maps for CW_Tau	67
A.3	Signal-to-noise ratio maps for DS_Tau	68
A.4	Signal-to-noise ratio maps for GM_Aur	68
A.5	Signal-to-noise ratio maps for HBC353	68
A.6	Signal-to-noise ratio maps for HBC354	69
A.7	Signal-to-noise ratio maps for HBC372	69
A.8	Signal-to-noise ratio maps for HBC388	69
A.9	Signal-to-noise ratio maps for HBC392	70
A.10	Signal-to-noise ratio maps for HD148352	70

A.11	Signal-to-noise ratio maps for HN_Tau	70
A.12	Signal-to-noise ratio maps for LkCa15	71
A.13	Signal-to-noise ratio maps for LkCa19	71
A.14	Signal-to-noise ratio maps for ROXs4	71
A.15	Signal-to-noise ratio maps for ROXs8	72
A.16	Signal-to-noise ratio maps for ROXs35A	72
A.17	Signal-to-noise ratio maps for ROXs43B	72
A.18	Signal-to-noise ratio maps for ROXs44	73
A.19	Signal-to-noise ratio maps for SR3	73
A.20	Signal-to-noise ratio maps for SR4	73
A.21	Signal-to-noise ratio maps for SR9	74
A.22	Signal-to-noise ratio maps for SR14	74
A.23	Signal-to-noise ratio maps for SR21A	74
B.1	Signal-to-noise ratio map, after RV CCF calibration, for CW_Tau . .	75
B.2	Signal-to-noise ratio map, after RV CCF calibration, for DS_Tau . .	75
B.3	Signal-to-noise ratio map, after RV CCF calibration, for GM_Aur . .	76
B.4	Signal-to-noise ratio map, after RV CCF calibration, for HBC353 . .	76
B.5	Signal-to-noise ratio map, after RV CCF calibration, for HBC372 . .	76
B.6	Signal-to-noise ratio map, after RV CCF calibration, for HBC388 . .	77
B.7	Signal-to-noise ratio map, after RV CCF calibration, for HBC392 . .	77
B.8	Signal-to-noise ratio map, after RV CCF calibration, for HD148352 .	77
B.9	Signal-to-noise ratio map, after RV CCF calibration, for HN_Tau . .	78
B.10	Signal-to-noise ratio map, after RV CCF calibration, for LkCa19 . . .	78
B.11	Signal-to-noise ratio map, after RV CCF calibration, for ROXs4 . . .	78
B.12	Signal-to-noise ratio map, after RV CCF calibration, for ROXs8 . . .	79
B.13	Signal-to-noise ratio map, after RV CCF calibration, for ROXs35A .	79
B.14	Signal-to-noise ratio map, after RV CCF calibration, for ROXs44 . .	79

LIST OF TABLES

<i>Number</i>	<i>Page</i>
2.1 Our 23 targets in Ophiuchus and Taurus star-forming regions. <i>RA</i> and <i>Dec</i> list right ascension (in hrs:mins:secs) and declination (in degrees, arcmins, arcsecs) respectively. <i>Type</i> , M/M_{\odot} , R_{mag} , K_{mag} list spectral type, stellar mass in units of solar mass, R-band magnitude (relevant for NGS), and K-band magnitude (relevant for observations). <i>Shorthand</i> lists the abbreviation used in this thesis to denote this specific target. Values are from Cheetham et al. (2015) for Ophiuchus and from Kenyon and Hartmann (1995) and ESA (1997) for Taurus targets. Note that HD 148352 is not in the Ophiuchus cluster, but rather is an interloper, as discussed in Section 3.6.	9
2.2 Science observations of the 23 targets. We include the date, number of sequences per target, frames per sequence, and exposure time (in seconds) per frame. Field of view is dithered between each sequence. For Taurus targets, the sequences that are in bold and <i>italicized</i> were taken after the field of view was rotated by ninety degrees (to counter pixel bleeding as described in Section 2.7).	10
2.3 Long exposure sky images taken to perform wavelength and resolution calibration using OH emission lines, as described in Section 2.4.	11
2.4 Observations of A0 standard stars, taken to calculate sky transmission and perform telluric calibration, as described in Section 2.5. We include date of observations, number of sequences, frames for each sequence, and exposure time per frame in seconds. Numbers in bold and <i>italics</i> represent that the sequence was taken after rotating the field of view by ninety degrees. Several of these observations also serve as science data for the corresponding target. We also list the targets calibrated using each set of observations.	11

Chapter 1

INTRODUCTION

Exoplanets are any planets outside our solar system, most of them found orbiting another star. The first exoplanets were discovered in the 1990s (Wolszczan and Frail, 1992; Mayor and Queloz, 1995). Over 5000 confirmed detections of exoplanets have now been made (NASA Astrophysics Division, 2022) using a variety of detection methods. About 95% of these have been discovered by one of two classical indirect methods: the radial velocity method and the transit method.

In Section 1.1, we briefly describe these two indirect methods: radial velocity and transit. While they are extremely powerful detection techniques, we detail their underlying mechanism to motivate some limitations in the detection and characterization of certain exoplanetary systems. Section 1.2 describes direct imaging and how it can address the limitations of indirect methods. This thesis focuses on direct imaging, specifically on how using this technique combined with high-resolution spectroscopy can improve our detection sensitivity at smaller separations from a star. Section 1.3 details how improved sensitivity gives us the information needed to explain the formation and migration mechanisms of directly imaged exoplanets. This remains a major scientific goal of our ongoing pathfinder survey of young stars, which is described in this thesis. Section 1.4 introduces the observation technique and instrument used for this survey, and Section 1.5 details the second main scientific goal of our survey - quantifying the improvement in sensitivity of direct imaging, when we complement classical techniques with moderate to high-resolution spectroscopy. The section introduces classical techniques and motivates our reasoning for expecting this better sensitivity closer to a host star.

1.1 Indirect detection methods

In this section, we describe the radial velocity method and the transit method. These are termed *indirect*, as they rely on observing light from the host star, instead of photons emitted by the exoplanet itself. Doppler's theory posits that stars that have motion along the line of sight of an observer, or velocities in the radial direction, would exhibit a change in color. By measuring shifts in the position of a specific absorption line in the spectrum of the host star, an observer can measure its radial velocity. Using Keplerian dynamics, Lovis and Fischer (2010) derive the radial

velocity semi-amplitude

$$K_1 \propto m_p \times \sin i \times (m_s + m_p)^{-2/3} \times P^{-1/3}, \quad (1.1)$$

where m_p and m_s are the masses of the planet and the star, P is the orbital period, and i is the inclination of the orbit. The method works best for detecting planets with shorter orbital periods, whose orbits make a small angle with the line of sight (a “side-view” or *edge-on* orbit).

The transit method detects exoplanets by measuring dips in the luminosity of the host star, as the planet moves in front or behind the star’s observed disk, one body shadowing the other. The probability of observing these transits or occultations is highest when the orbit is closely aligned with the line of sight (Winn, 2010).

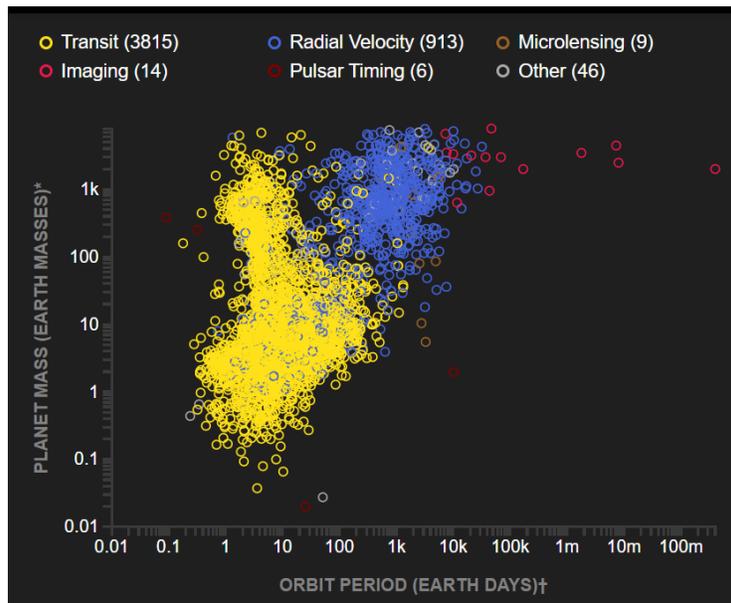


Figure 1.1: Detected exoplanets with measured or estimated mass and orbital period. (NASA Astrophysics Division, 2022) Color denotes the technique that detected the exoplanet for the first time.

To make a *credible* detection, an additional requirement, of measuring several radial velocity cycles or luminosity dips associated with transits, is expected. This criterion prefers exoplanets with shorter orbital periods.

To study theories of planet formation, and evolution, it is critical to know how the exoplanet mass, orbital parameters, or stellar characteristics affect the occurrence rates of planets. This can be described as attempting to populate Fig. 1.1 accurately.

As there are planetary systems that are not easily accessible to these indirect detection methods, techniques involving direct imaging come into play. An instance of this would be the argument of Boley (2009) that the dominant mechanisms for the formation of gas giant planets at radii $r < 100$ AU and $r > 100$ AU are core accretion and disk instability respectively. By Kepler's laws, planets with $r > 100$ AU have orbital periods of over 1000 years, which makes them more accessible to methods like direct imaging.

1.2 Direct imaging

Direct imaging methods attempt to spatially resolve the light of an exoplanet from that of its host star. Young planets, that recently stopped or are still accreting protoplanetary material, are hot enough to be significantly self-luminous. These offer the best contrast in brightness from their young host stars and are excellent candidates for direct imaging (Bowler, 2016). As they involve directly capturing photons that the planet emits, these techniques allow for spectroscopic analysis of the planet's atmosphere. This can provide information about bulk parameters of exoplanet atmospheres, including effective temperature, surface gravity, cloud coverage, and composition (Ruffio, Macintosh, et al., 2019), and also initial conditions, internal structure, or other physical properties (Bowler, 2016).

The main challenge to direct imaging is that planets are much fainter than their host stars, such that the planet image is "lost" in the blur of the much brighter stellar image. The ratio of planetary to stellar brightness, or *contrast*, is typically of the order of millionths to billionths ($\approx 10^{-6} - 10^{-9}$). Classical direct imaging techniques employ large telescopes with adaptive optics (AO) systems, which deform mirrors in real-time to correct for atmospheric turbulence, and coronagraphy, where stellar light is blocked to reduce required contrast from stray light. Extremely sophisticated high-contrast AO imaging has enabled the detection and study of planets with orbital separations $> \approx 10$ AU and masses $> \approx 1M_{\text{Jupiter}}$ (Bowler, 2016).

Most direct imaging surveys searching for new companions use photometric or low spectral resolution data. Higher resolution spectroscopy is a developing field, where instruments with high spectral resolutions R are generally used to characterize known exoplanets. High values of R allow for the resolution of spectral lines, which enables the study of atmospheres and the measurement of planetary radial velocity, spins, and surface cloud coverage (Q. M. Konopacky et al., 2013) (Wang et al., 2018).

Notably, high-resolution spectrographs are typically used for the characterization of *known* companions, but not for the detection of new ones. In this work, a blind search using an integral field high-resolution spectrograph is described. This is motivated by the suggestion that higher resolutions of $R \approx 4000$ might be able to detect planets at smaller separations from the host star, in regions where they are expected to be more common. These considerations relate to the survey’s scientific goals described next.

1.3 Theories of Formation and Migration of Gas-Giant Planets

Core accretion with gas capture and gravitational instability are the two main theories that each describe a mechanism for exoplanet formation from proto-planetary disks (Boley, 2009). Gas giants (like those in the outer solar system) are understood to be formed by a core slowly accreting solid material from the disk until it reaches a critical mass that triggers a runaway accretion of gas (Lissauer, 1987; Pollack et al., 1996). This remains the dominant mechanism for separation r under 100 AU (Boley, 2009). A likely alternative that can form planets faster at larger separations is gravitational instability in the outer disk or solar nebula, which can also form higher mass brown dwarfs (Boss, 1997). The process involves the rapid cooling of a massive protoplanetary disk, such that several planet-sized bodies emerge that can sustain themselves under their own gravity.

Directly imaged exoplanets (generally $\approx 10M_{\text{jupiter}}$, semi-major axis $\approx 10\text{--}1000$ AU) lie at the transition between these two methods (Bowler, 2016), and offer a unique problem of explaining how large planets form at wide orbits. For comparison, the vast majority of exoplanets detected by the radial velocity method are within 3 AU. Planets that have been directly imaged till now have been more massive than expected due to core accretion at their high separations, given that the low density of particles in a protoplanetary disk increases the timescales for accretion (Dodson-Robinson et al., 2009). On the other side, the occurrence rate of brown dwarfs in direct-imaging surveys is lower than what could indicate formation through gravitational instability (Nielsen et al., 2019). Increasing the sample size of directly imaged exoplanets through further surveys of younger stars can allow us to constrain and know more about the formation processes of young planets.

Another question that can be studied from more detailed surveys of young stars is how orbital migration affects the distribution of giant planets. Fewer planets than what is predicted by both theories of exoplanet formation have been discovered

around older stars (ages over about 20 Myr) in the dominant regions of formation, in orbits of about $\approx 10 - 100 au$). A likely explanation is that planets migrate into smaller orbits after formation over timescales of mega-years. Surveys of younger stars can test theories of orbital migration by mapping the exoplanet population over different ages of stars. It is also likely that more planets could be detected around younger stars, as planets around such a star could be still at larger separations (regions of formation) before they have migrated inwards. Additionally, direct imaging with high-resolution spectroscopy can probe abundance ratios such as the carbon-to-oxygen ratio, which are important tracers of planet formation because they can be related to the location of formation and the fraction of gas and solids it accreted from the protoplanetary disk (Ruffio, Quinn M. Konopacky, et al., 2021).

1.4 Integral Field Spectroscopy and the OSIRIS instrument

Integral field spectroscopy (IFS) is like 3D spectroscopy where we obtain spatially resolved spectra from a field of view on the sky. Unlike time-domain 3D spectroscopic techniques that scan through wavelengths over time, IFS takes in all spatial and spectroscopic information in one exposure simultaneously. This makes IFS less susceptible to time-variable systematics like sky or instrument conditions (Allington-Smith, 2006). 3D techniques like IFS produce a datacube with two spatial directions and one spectral direction, containing a scalar quantity related to the flux for each spatial location and each wavelength (Allington-Smith, 2006). 1D high-resolution spectrographs are typically used to characterize directly imaged exoplanets, by setting the location on the sky using astrometric predictions. For a survey searching for new planets around host stars, 3D techniques like IFS are preferable as accurate target acquisition is not necessary.

OSIRIS (OH-Suppressing InfraRed Imaging Spectrograph) is a near-infrared integral field spectrograph compatible with the Keck Adaptive Optics System, operating from 1 to 2.4 microns in wavelength at close to the Keck diffraction limit (Larkin et al., 2006). It can sample spectra from up to ≈ 3000 spatial locations from a region of the sky, with spatial resolution ranging discretely from 20 to 100 milliarcseconds. Each of these one-dimensional spectra at a given spatial location is termed a "spaxel", that is, each spaxel corresponds to a spatial location on the sky. The spectral resolution is around $\approx 3400 - 3800$, with the spectral range covering the z, J, H, or K bands. On its detector, OSIRIS measures a two-dimensional image, with a sequence of pixels in a row on the detector corresponding to a spectrum at a particular spatial location on the sky. Barring the edges of each spectrum, the

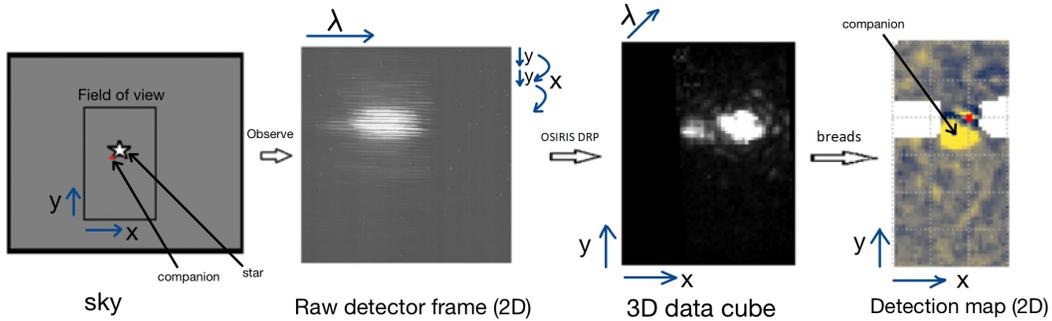


Figure 1.2: Schematic of data reduction pipeline, drawn for target HD 148352. We discussed a detected binary companion in Section 3.6. From left to right: (1) The system with the OSIRIS field of view; the y axis points North; companion signal is hidden in the glare of the target star. (2) Raw two-dimensional data frame obtained from the OSIRIS detectors after observation; spectra are arranged as rows. (3) Data transformed by OSIRIS DRP into data cubes. The bright speckle to the left of the stellar PSF is not the companion signal, but rather a residue from pixel bleeding issues, described in Section 2.7 (4) Signal-to-noise ratio map, with detection of a companion, generated as described in Sections 2.8, 3.1, and 3.6.

pixels horizontally adjacent to a specific pixel in the 2D map correspond to different wavelengths in the same spectrum (i.e., for the same spatial location), while the vertically adjacent pixels correspond to fluxes at different locations on the sky. The OSIRIS Data Reduction Pipeline (Keck Data Reduction Pipelines, 2022) then converts this two-dimensional detector map to a 3D data cube (with two spatial and one spectral dimension), which serves as the input to our data analysis. Figure 1.2 details a schematic of this data collection, reduction, and analysis pipeline.

1.5 Sensitivity of Moderate-resolution integral field spectroscopy

Direct imaging surveys have detected several exoplanets, generally at separations over 10 AU (Bowler, 2016), using classical techniques such as adaptive optics, coronagraphy, image processing, or a combination of these. These surveys typically use imagers (two-dimensional wavelength-collapsed pictures of the sky) or low-resolution spectrographs (such as the Gemini Planet Imager, Macintosh et al., 2014). These are combined with speckle (granular interference, diffracted starlight) subtraction algorithms such as Spectral (Marois, Doyon, et al., 2000, SDI) or Angular (Marois, Lafreniere, et al., 2006, ADI) or Referential (Lafrenière et al., 2009, RDI) Differential Imaging. These algorithms use the difference in characteristics of companion signal and speckle noise (such as wavelength-dependent magnification of the point spread function (PSF), time-domain rotation between signal and PSF, or

using other stars' PSF). These observing strategies become less effective at smaller separations, where self-subtraction of the companion signal becomes stronger. So, they become to become very less sensitive, starting at around 0.2 arcseconds (Hoeijmakers et al., 2018). Alternatives like non-redundant aperture masking are likely to be able to detect at smaller separations but are limited to brighter companions (Sallum and Skemer, 2019).

As noted before, radial velocity surveys have measured that the population of gas giants is maximized at separations close to 1-10 AU (Fulton et al., 2021), which implies the need to improve sensitivity at smaller separations. Having this capability would allow follow-up detections and confirmations from large-scale surveys using indirect methods, such as Gaia monitoring.

The development of infrared high-resolution spectroscopic facilities allows for the techniques we are proposing in this article, which involve moderate resolution integral field spectroscopy. With $R > 3000$, we can attempt to distinguish between the light from a companion planet or brown dwarf and the diffracted light from the star at a spatial location on the sky, by using our knowledge of the expected spectral features of the planetary atmosphere. Using cross-correlation techniques makes these high-resolution spectroscopy methods close to independent of the speckle noise, which was noted earlier to be the limiting factor for classical direct imaging at smaller separations. We expect our method's sensitivity to be solely dependent on the amount of starlight at the location of detection, which makes us closer to photon-noise limited.

Similar techniques that utilize narrow spectral features have been used to characterize previously discovered planets, in observations using the Keck/OSIRIS (Q. M. Konopacky et al., 2013) and VLT/SINFONI (Hoeijmakers et al., 2018) instruments. OSIRIS is extremely well suited to this approach due to its combination of moderate spectral resolution, a moderate field of view, and the Keck adaptive optics system. It can be used to characterize atmospheric abundances, carbon-to-oxygen ratios, and radial velocity of planets. The techniques we are developing will also apply to new integral field spectrographs on JWST and the extremely large telescopes, which are expected to significantly improve capabilities for direct imaging.

Chapter 2

APPROACH AND METHODS

In this chapter, we will emphasize the two main components of this thesis - an ongoing pathfinder survey of young stars in star-forming regions, and the development of data analysis modules that uses a forward modeling approach to reduce moderate to high-resolution integral field spectroscopy images.

Section 2.1 and 2.2 detail our observations of young stars in Taurus and Ophiuchus using the OSIRIS instrument at the Keck Observatory, and initial reduction using a provided data reduction pipeline, respectively. Section 2.3 presents an open-source framework *bread*. We develop our techniques described in the thesis within this framework, with the subsidiary goal of allowing other professionals to use them with their own high-resolution spectrographs and corresponding data sets. Within *bread*, additional calibration is performed on outputs from the OSIRIS DRP, as detailed in Section 2.4. Section 2.5 describes our process of using aperture photometry to compute sky transmission and stellar spectra. These are used to construct our model, as enumerated in Section 2.6, which also explains our forward modeling approach. Section 2.7 describes how we identify and correct bad pixels.

Generation of planet search maps, with noise calibration, is detailed in Sections 2.8 and 2.9. Quantification of sensitivity, after correcting for algorithmic systematics, is detailed in Sections 2.10 and 2.11. These final sections directly relate to the results discussed in Chapter 3.

2.1 Observations and Data Collection

For our survey, young stars without significant disks are attractive targets. Pre-main sequence stars in molecular cloud regions such as Taurus and Ophiuchus have been a standard for studying star formation (Kenyon and Hartmann, 1995; Cheetham et al., 2015). While substellar companions around stars in this region have been identified, these have been found at large separations or embedded in disks. Very young planets at smaller separations of $\approx 0.1''$ (corresponding to 10-20 AU) would allow for characterizing the physical and orbital parameters of planets at young ages.

Table 2.1 lists the 11 Ophiuchus targets and 12 Taurus targets observed using Keck/OSIRIS in the first and second Keck observing semesters of 2021 respectively;

Ophiuchus Target List							
Star	RA	Dec	Type	M/M_{\odot}	R_{mag}	K_{mag}	Shorthand
HD 148352	16:28:25.16	-24°45'00.9"	F2	1.52	7.3	6.511	HD148352
Em* SR 3	16:26:09.31	-24°34'12.1"	B6	5.20	10.6	6.504	SR3
Em* SR 21A	16:27:10.27	-24°19'12.7"	G1	3.69	13.25	6.719	SR21A
Em* SR 4	16:25:56.15	-24°20'48.1"	K4.5	1.35	12.11	7.518	SR4
ROXs 44	16:31:33.46	-24°27'37.3"	K3	2.07	12.35	7.61	ROXs44
ROXs 8	16:26:03.02	-24°23'36.0"	K1	2.95	9.34	6.227	ROXs8
ROXs 4	16:25:50.52	-24°39'14.5"	K5.5	1.15	13.88	8.33	ROXs4
ROXs 35A	16:29:33.97	-24°55'30.3"	K3	2.07	12.41	8.531	ROXs35A
ROXs 43B	16:31:20.19	-24°30'00.9"	K5	1.20	12*	7.089	ROXs43B
Em* SR 14	16:29:34.41	-24°52'29.2"	G4	3.64	10.04	8.878	SR14
Em* SR 9	16:27:40.28	-24°22'04.0"	K5	1.20	11.87	7.207	SR9
Taurus Target List							
AB Aurigae	04:55:45.85	+30°33'04.3"	A0Ve	4.00	7.05	4.23	AB_Aur
CW Tauri	04:14:17.00	+28°10'57.8"	K0Ve	1.40	12.36	7.127	CW_Tau
DS Tauri	04:47:48.60	+29°25'11.2"	K4Ve	0.97	12.3	8.036	DS_Tau
LkCa 15	04:39:17.79	+22°21'03.4"	K5Ve	0.97	12.03	8.163	LkCa15
LkCa 19	04:55:36.97	+30°17'55.1"	K0Ve	2.42	11.12	8.148	LkCa19
HBC 388	04:27:10.57	+17°50'42.7"	K1e	2.10	10.22	8.296	HBC388
GM Aurigae	04:55:10.98	+30°21'59.4"	K3Ve	1.40	13.1	8.283	GM_Aur
HN Tauri	04:33:39.36	+17°51'52.3"	K5e	0.97	13.4	8.384	HN_Tau
HBC 354	03:54:35.56	+25°37'11.2"	K3	1.40	13.79	11.095	HBC354
HBC 392	04:31:27.18	+17°06'24.8"	K5e	0.97	12.1	9.497	HBC392
HBC 372	04:18:21.48	+16°58'47.0"	K5	0.97	13.26	10.464	HBC372
HBC 353	03:54:30.18	+32°03'04.4"	G5	3.42	9.862	3.4188	HBC353

Table 2.1: Our 23 targets in Ophiuchus and Taurus star-forming regions. *RA* and *Dec* list right ascension (in hrs:mins:secs) and declination in (degrees, arcmins, arcsecs) respectively. *Type*, M/M_{\odot} , R_{mag} , K_{mag} list spectral type, stellar mass in units of solar mass, R-band magnitude (relevant for NGS), and K-band magnitude (relevant for observations). *Shorthand* lists the abbreviation used in this thesis to denote this specific target. Values are from Cheetham et al. (2015) for Ophiuchus and from Kenyon and Hartmann (1995) and ESA (1997) for Taurus targets. Note that HD 148352 is not in the Ophiuchus cluster, but rather is an interloper, as discussed in Section 3.6.

certain targets are listed in our reference material under other common identifying systems, which we compare using *Simbad*¹. Table 2.2 lists information about science observations of these targets. We aim to observe roughly 20 more stars in these two regions in 2022.

Our observing strategy is to take 3 to 5 sequences of images per target, with the Keck adaptive optics system operating and tracking using a natural guiding star (NGS) system (we use the bright target with high R_{mag} itself as the natural guiding star). Our exposure time for each frame is dependent on the brightness of the targets in K-band, denoted by K_{mag} . We observed targets in OSIRIS's KN5 filter, choosing this narrow band as it is expected to contain a higher density of planetary spectral lines, and OSIRIS covers a larger field of view with narrowband filters compared to

¹<http://simbad.u-strasbg.fr/>

Ophiuchus Science Observations				
Star	Date	Sequences	Frames	Exposure Time (s)
HD 148352	2021/26/06	3	21	30
Em* SR 3	2021/26/06	1, 3, 1	11, 21, 6	30
	2021/26/07	4, 2	6, 4	20
	2021/26/08	3, 4	3, 3	20
	2021/26/07	2	11	30
Em* SR 21A	2021/26/07	2	11	30
Em* SR 4	2021/26/07	4	5	90
ROXs 44	2021/26/07	4	5	90
ROXs 8	2021/26/07	4	11	20
ROXs 4	2021/26/07	5	4	120
ROXs 35A	2021/26/08	5	4	120
Em* SR 14	2021/26/08	4	4	120
ROXs 43B	2021/26/08	4	5	90
Em* SR 9	2021/26/08	4	4	90
Taurus Science Observations				
AB Aurigae	2021/10/18	2, 2, 3, 4	10	10, 10, 4, 4
	2021/10/19	3, 3, 3	6	4
	2021/10/20	3, 3	6	4
CW Tauri	2021/10/18	3, 3	10	30
DS Tauri	2021/10/18	3, 2	7	90
LkCa 15	2021/10/18	3, 3	5	90
LkCa 19	2021/10/19	3, 3	4	90
HBC 388	2021/10/19	3, 2	4	90
GM Aurigae	2021/10/19	3, 2	4	90
HN Tauri	2021/10/19	3, 1	4, 2	90
HBC 354	2021/10/20	3, 2	2	300
HBC 392	2021/10/20	3, 3	1	300
HBC 372	2021/10/20	3, 3	1	300
HBC 353	2021/10/20	3, 1	1	300

Table 2.2: Science observations of the 23 targets. We include the date, number of sequences per target, frames per sequence, and exposure time (in seconds) per frame. Field of view is dithered between each sequence. For Taurus targets, the sequences that are **in bold** and *italicized* were taken after the field of view was rotated by ninety degrees (to counter pixel bleeding as described in Section 2.7).

broad bands. We use a plate scale of 20 milliarcseconds (mas), which denotes our spatial resolution and is the finest resolution possible with OSIRIS. This provides us with a higher density of spatial locations close to the star ($< 0.3''$), which is the region of interest for our survey.

We offset our field of view by a few pixels in both axes, or *dither*, between each sequence, as a systematic check against instrumental artifacts like a bad spatial

Calibration Sky Observations		
Date	Frames	Exposure (s)
2021/26/06	2	600
2021/26/07	1	600
2021/26/08	2	600
2021/10/18	2	600
2021/10/19	2	600
2021/10/20	2	300

Table 2.3: Long exposure sky images taken to perform wavelength and resolution calibration using OH emission lines, as described in Section 2.4.

Ophiuchus Calibration Observations					
Star	Date	Sequences	Frames	Exposure (s)	Targets
HIP 73049	2021/26/06	6	1	2	SR3
	2021/26/07	7	1	1.5	SR4
	2021/26/08	7	1	1.5	ROXs35A
Em* SR 3	2021/26/06	1, 3	11, 21	30	HD148352, SR3, SR21A
	2021/26/06	1	6	30	-
	2021/26/07	4, 2	6, 4	20	ROXs44, ROXs8
	2021/26/07	2	4	20	ROXs4
	2021/26/08	3, 4	3, 3	20	SR14, ROXs43B
	2021/26/08	4	3	20	SR9
Taurus Calibration Observations					
AB Aurigae	2021/10/18	2, 2	10	10, 10	AB_Aur, CW_Tau
	2021/10/18	3	10	4	DS_Tau
	2021/10/18	4	10	4	LkCa15
	2021/10/19	3	6	4	LkCa19
	2021/10/19	3	6	4	HBC388, GM_Aur
	2021/10/19	3	6	4	HN_Tau
	2021/10/20	3	6	4	HBC354
	2021/10/20	3	6	4	HBC392, HBC372, HBC353

Table 2.4: Observations of A0 standard stars, taken to calculate sky transmission and perform telluric calibration, as described in Section 2.5. We include date of observations, number of sequences, frames for each sequence, and exposure time per frame in seconds. Numbers in **bold** and *italics* represent that the sequence was taken after rotating the field of view by ninety degrees. Several of these observations also serve as science data for the corresponding target. We also list the targets calibrated using each set of observations.

location. A real companion candidate will maintain the same position relative to the target star, while artifacts might not move consistently in our field of view. We take one sequence of skies per target with the same exposure time per image, after turning the adaptive optics off and putting the bright star out of our field of view. As our targets are bright stars, we attempt to acquire consecutive targets in alternating sections of our field of view (top and bottom halves) to avoid persistence artifacts between targets.

Analyzing the Ophiuchus data, we discovered a pixel bleeding issue as a consequence

of using bright targets, which is described in detail in Section 2.7. This problem results in specific sections of data on either spatially horizontal side of the star being unusable. To cover a larger region around the target, we amended our observing strategy for observations of Taurus, such that we rotate the field of view of OSIRIS by a right angle (90 degrees) for roughly half of our image sequences.

We additionally obtained two sets of calibration data. The details of how each calibration is performed are included in Sections 2.4 and 2.3. These data sets are listed in Tables 2.3 and 2.4 respectively. At the start of every observing night, we took 2-3 images of the sky with the adaptive optics system off, with long exposure times of 5 or 10 minutes. This data is used to perform wavelength and resolution calibration in the science data set using OH^- emission lines. These were taken in the OSIRIS's KN3 filter, as this narrow K-band contains more OH absorption lines. We additionally take data from standard stars of spectral type A0 to perform telluric calibration. A0 stars are expected to contain few spectral features, making them most suitable for telluric calibration. Table 2.4 also lists the targets that are calibrated using each calibration data set. We also obtained dark frames for the instrument for each observing run, for the same exposure duration as the wavelength and resolution calibration.

2.2 Constructing Datacubes through OSIRIS Data Reduction Pipeline

OSIRIS gives us two-dimensional raw data files, where the value at every detector location is a scalar related to the flux within a spectral bin from a location on the sky. We use the OSIRIS Data Reduction Pipeline (ODRP or OSIRIS DRP) to obtain three-dimensional data cubes.

The following settings were used for the wavelength and resolution calibration data:

- Reduction Type: `ARP_SPEC`
- Reduction Template: `basicARP_drftemplate.xml`
- Subtract Frame: Combined Dark Frame obtained as below, of the same exposure time as each calibration image.
- Extract Spectra: `s201217_c010___infl_Kn3_020.fits`
- all submodules turned on

Each respective combined dark was obtained by adding all frames of the same exposure time as the calibration sky image, from the set of darks taken for the particular observation run. These were combined using the OSIRIS DRP with Reduction Type and Reduction Template set to `CRP_SPEC` and `combine_skies_darks_drftemplate.xml` respectively.

The following settings were used for all the remaining data, taken with a target star in the field of view:

- Reduction Type: `ARP_SPEC`
- Reduction Template: `basicARP_drftemplate.xml`
- Subtract Frame: Sky Frame, taken specifically for each target
- Extract Spectra: `s201217_c010___infl_Kn5_020.fits`
- all submodules turned on

As mentioned before, we took one sequence of sky images for each target. For the n^{th} image in a sky sequence, we use the n^{th} sky image in the sky sequence. This is done to ensure that any pixel artifacts (that might be misidentified to be a planet) in a particular target or sky image are not propagated in a large portion of our data. The rectification matrices used in the Extract Spectra option for these settings are available on the Keck OSIRIS website ².

2.3 Developing Open Source Framework breads

`breads` stands for Broad Repository for Exoplanet Analysis, Detection, and Spectroscopy. It is designed to be an open-source flexible framework that can use our forward modeling techniques on data from high-resolution spectrographs, extending the accessibility of our methods in the scientific community. Hosted on GitHub ³ and Python Package Installer ⁴, `breads` employs object-oriented and inheritance principles. The philosophy is to let users make three main choices. First, users select a data class that encodes specific features of their instrument. We have defined classes for KPIC (Keck Planet Imager and Characterizer) and OSIRIS at Keck, as

²<http://tkserver.keck.hawaii.edu/osiris/>

³<https://github.com/jruffio/breads>

⁴<https://pypi.org/project/breads/>

well as JWST, with work on additional data classes planned to facilitate data formatting for common instruments. One can also define and use their own class for specific instrumental applications.

Users then select a forward model function with required linear and non-linear parameters. This distinction is important, as `bread`s performs analytical marginalization of linear parameters to reduce the total number of parameters and time taken in a fit, allowing for higher-dimensional models. Finally, users select the specific solving method to apply; typical examples and use-cases would include a simple optimizer, a Markov Chain Monte Carlo (MCMC) sampler like `emcee` (Foreman-Mackey et al., 2013) for planet characterization, a grid search over spatial directions for planet detection, a grid search in the radial velocity direction for obtaining a cross-correlation function. Certain forward models and solvers are predefined by us, publicly available on GitHub, and can be used reliably with the `bread`s framework. `bread`s also performs bad pixel removal at several steps of the analysis to enable robust results. Multiprocessing is also implemented through `multiprocessing`⁵ to enable faster run time.

We are developing documentation for `bread`s on ReadTheDocs⁶. Code that uses `bread`s to generate results and plots in this thesis are maintained publicly on GitHub⁷, with the `using-breads` repository meant to demonstrate how to use our code. These are projected to be formalized, as we work on developing a manuscript describing all the features on `bread`s .

2.4 Wavelength and Resolution Calibration

The three-dimensional data cubes outputted by the OSIRIS DRP (ODRP) have two spatial dimensions, representing the spatial positions of the sources, and one spectral direction, representing the wavelength space. For the narrow K-band filters we use, the instrument gives us one spectrum of length $N_\lambda = n_z = 465$ for each of the $n_y \times n_x = 66 \times 51$ spatial locations. The OSIRIS DRP calibrates and assigns each of the pixels on its detector field its spatial location and wavelength. This wavelength solution in the OSIRIS DRP is only derived roughly once every other year, when the instrument is serviced. Thus, the wavelength calibration is correct to $\approx 1/10$ of one pixel (that is, about 0.01 nm order of magnitude). For our science goals, we need to correct this arbitrary offset to be more precise than this error range. Notably, further

⁵<https://docs.python.org/3/library/multiprocessing.html>

⁶<https://bread.readthedocs.io/>

⁷<https://github.com/shubhagrawal30/using-breads>

calibration needs to be done independently for each spatial location, as there is no reason for the entire field of view of OSIRIS to have exactly the same offset error in wavelength calibration.

We also want to calibrate the resolution of the instrument at each spatial location in the field of view. Resolution is defined as

$$R = \frac{\lambda}{\Delta\lambda},$$

where λ is the central wavelength of a spectral bin, and $\Delta\lambda$ is the width of that spectral bin. Resolution is not necessarily constant over the wavelength range or at different spatial locations. The average resolution of OSIRIS is around 3800, with resolution at a spatial location lying somewhere within about 3400-4000 (Larkin et al., 2006).

Using long exposure sky images as described in previous sections, we attempt to establish a relation between the wavelengths and resolution recorded by the instrument and further corrected values. As we take long exposure skies at the start of every observing night, we can use this to calibrate science data taken around target stars each specific night.

We use emission lines of the OH^- radical present in the Earth's upper atmosphere, as a calibration reference. The Kn3 narrow band is dense in these lines; thus, we can locate these at a spatial location in our field of view and compare them to the known accurate theoretical wavelengths. The broadening of these lines from peaked delta functions also gives us a measure of the resolution of the instrument. Thus, these sky images can allow us to do field-of-view-dependent wavelength and resolution calibration.

We use code present in `breadsc.calibration`, specifically the `SkyCalibration` object, and `using-breads.get_sky_calibration`. Emission lines of OH^- are taken from Rousselot et al. (2000)⁸. We generate our model by selecting lines in the wavelength range of the filter, adding these lines with their respective relative intensities, and broadening them to a Gaussian function with FWHM $\Delta\lambda = \lambda/R$ (dependent on the resolution). The location of these lines in the model can be offsetted as $\lambda_r = a_0 + \lambda_0(1 + a_1)$, allowing for constant and linear offset terms a_0 and a_1 . We also add a polynomial term to the model to account for the background continuum. For each spatial location in the data cube, we fit this model to sky

⁸https://www.eso.org/sci/facilities/paranal/decommissioned/isaac/tools/oh/list_v2.0.dat

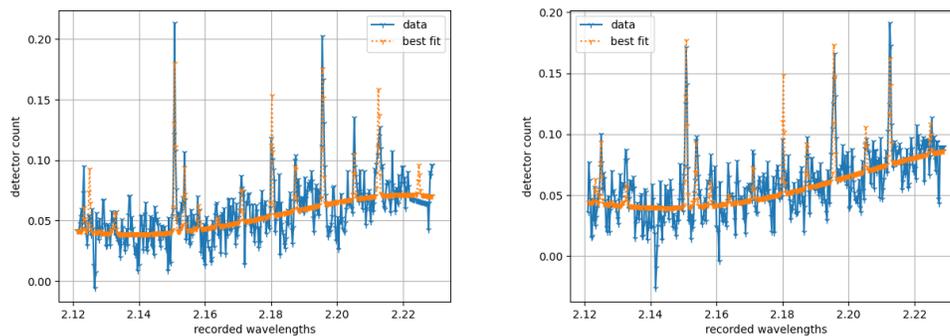


Figure 2.1: Nominal examples of wavelength calibration performed at two different spatial locations. We take spectral data at each location (blue) for the long sky exposures, and fit a model for emission lines of the OH^- radical (orange). Parameters fitted quantifies the broadening and wavelength-shift of the spectral lines.

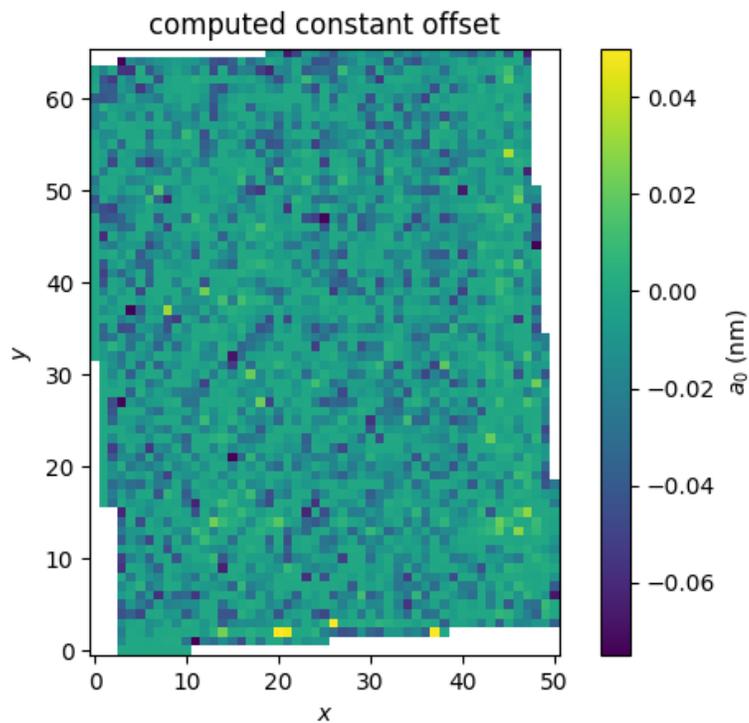


Figure 2.2: Nominal two-dimensional heat map encoding values for the constant offset a_0 , computed as part of the wavelength calibration for 2022/06/26. As expected, the offset is of the order of 0.01 nm or 1/10 of a pixel.

data to obtain values for a_0 , a_1 , and R . The fit performs analytical marginalization over linear parameters in the model (such as the overall scaling or the coefficient in background polynomial) to reduce the number of parameters in the curve-fitting and increase speed. Figure 2.1 shows nominal examples of this wavelength calibration fit for two spaxels, while Figure 2.2 plots a nominal two-dimensional heat map of constant offset values a_0 computed from a fit for a particular observation night.

The analysis described in the following sections uses this wavelength offset calibrations to improve the accuracy of our images before performing other relevant operations. We do not perform the resolution calibration for the results presented in this thesis, as we empirically noted that the calibration yields values consistent with the nominal $R = 4000$.

2.5 Extraction of Stellar Spectrum and Telluric Calibration

To detect an exoplanetary signal, we attempt to distinguish its expected spectral signature from the known spectral features of the target star. We can accurately ascertain these stellar spectral features from our OSIRIS data itself, by performing aperture photometry on the star. Relevant code can be found in `breadscalibration`, specifically `TelluricCalibration`, and using `breadscalibration.get_telluric_calibration`. After removing bad pixels from each science image, `breadscalibration` determines a two-dimensional Gaussian approximation to the wavelength-dependent point spread function, by curve-fitting each wavelength slice of the cube. The PSF model can be some other user-defined function. The center and width of this Gaussian fit vary with wavelength.

We then place an elliptical aperture on each wavelength slice concentric with the 2D Gaussian fit, of size dependent on the fitted width. `breadscalibration` allows users to set these apertures and their parameters. Using `photutils.aperture`⁹ (Bradley et al., 2021), we sum the flux inside this aperture for each wavelength slice to get a spectrum for the star, in units of detector count (electrons/second).

Additionally, several times during each observation night, we take data from standard A0 stars to perform telluric calibration, that is, to measure sky transmission during the time of our data collection. Particularly, we want the transmission as a function of wavelength, A0 stars have well-defined spectra (with very few spectral features due to their high temperatures) that facilitate accurately quantifying the amount of light absorbed when light (of a specific wavelength) passed through Earth’s atmosphere

⁹<https://photutils.readthedocs.io/en/stable/>

at observation time.

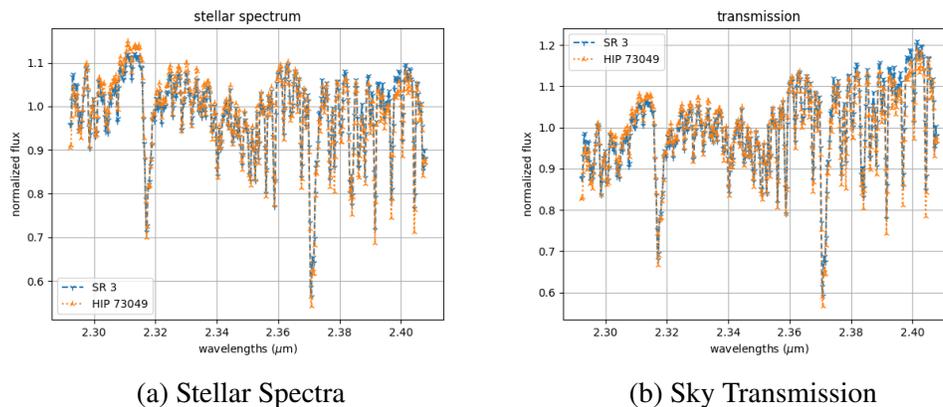


Figure 2.3: (a) Nominal examples of extracted spectra of HIP 73049 and SR 3, both of which are A0 standard stars. As these are of the same spectral type, we expect the spectra to look similar. (b) Nominal examples of computed sky transmission, obtained by dividing extracted spectrum by a theoretical model of A0 star spectra. These data sets are from the same observation night, so we expect the transmission to be similar.

This can be done by extending the spectral extraction process described above for science stars. We start with a theoretical description of the A0 star spectrum, broaden the absorption according to instrument resolution and offset wavelengths as per calibration results from the previous section. After obtaining a spectrum for an A0 star through aperture photometry, we divide it by this a theoretical model of the spectrum to get a model for sky transmission. Figure 2.3 shows nominal examples of extracted stellar spectra for two targets and computer transmission for 2022/06/26.

2.6 Forward Model, Noise Scaling, and Likelihood

Ruffio, Macintosh, et al. (2019) and Ruffio, Quinn M. Konopacky, et al. (2021) introduce and develop a forward modeling approach to analyzing high-contrast companions with high spectral resolution data, using it to explore the HR 8799 planetary system. Forward modeling relies on maximum likelihood estimation of the planet signal and does not need extensive pre-processing as opposed to cross-correlation methods. Jointly estimating the starlight and the planet signal reduces the possibility of over-subtraction of the planet signal while removing the starlight. Values and uncertainties for atmospheric parameters can also be directly estimated from the joint posterior probability distribution function.

We describe how our forward model functions for a particular location in the field of view, with the aid of schematics drawn in Figure 2.4. We repeat the modeling process for all locations in the field of view. Relevant code for the forward model is present in `breads.grid_search`, `breads.fm.hc_mask_splinefm`, and `using-breads.get_planet`. We first define `box_w` as 3, which quantifies the width (in pixels) of the box, centered at the given location, within which data is considered to get a model fit. Take the data in this $\text{box}_w \times \text{box}_w$ box to be d . Shape of d is thus $(\text{box}_w, \text{box}_w, N_\lambda)$, where N_λ is the number of bins in spectral dimension. We scale this data by noise vector s of same dimensions, as $d \rightarrow d/s$ element-wise. s is set to the higher of a base noise floor (detector noise limited) or the square root of the continuum (photon noise limited, as per Poisson statistics). The continuum is obtained by taking a moving average in the spatial direction for every spatial direction. The detector noise floor is obtained independently for each wavelength slice of each image’s data cube. We mask all locations where the continuum value is higher than the median of the continuum values for the spectral slice, and we set the floor to the standard deviation of the deviations of data from the continuum at the remaining locations in the spectral slice.

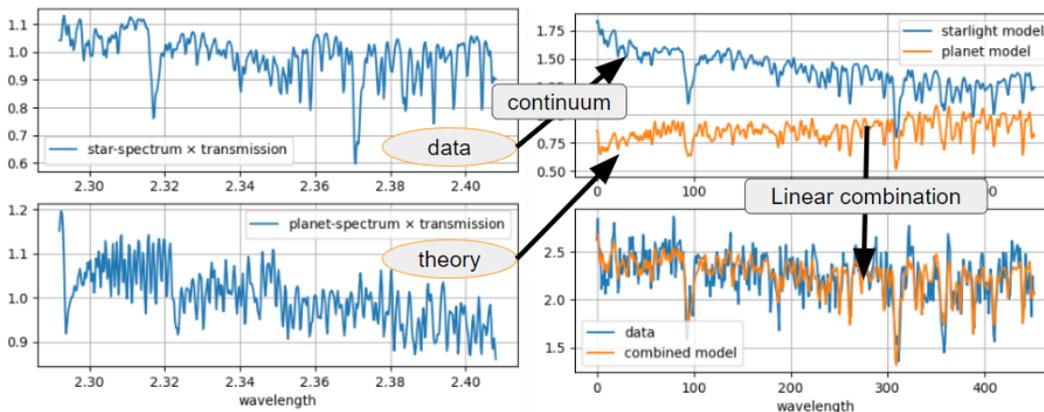


Figure 2.4: Schematic of the generation of the forward model for a spatial location. (Top Left) A model for the observed star spectrum (including a transmission component), computed from the data. (Bottom Left) A planet spectrum is computed from a theoretical model. (Top Right) We multiply the star spectrum with a continuum model (for our case, computed using a *spline interpolation*) to get diffracted starlight model. (Bottom Right) A linear combination ($M_\psi \phi$) of the diffracted starlight and planet models is used to fit the spectral data at the location.

Our forward modeling approach involves modeling each such d for every spatial location as $d = M_\psi \phi + n$, with n being a Gaussian noise vector, thereby yielding a maximum likelihood fit (Ruffio, Macintosh, et al., 2019). ψ and ϕ represent the

non-linear and linear parameters respectively. Making a distinction between the two allows us to analytically marginalize over linear parameters in the model (for this case, examples would include planet or starlight contribution to data), after having obtained M_ψ for a specific set of non-linear parameters ψ (for our case, these include radial velocity and position on the sky). Marginalizing over linear parameters allows a faster estimation of best-fit values ϕ for a particular M_ϕ , and we then optimize over non-linear parameters ϕ . Similar to how the data is scaled by the noise as described above, the model $m = M_\psi \phi$ described as follows is also scaled by the noise s as $m \rightarrow m/s$, after all the components of m have been computed to construct $M_\psi \phi$. Noise scaling is necessary for a statistically accurate Gaussian likelihood problem with uncorrelated noise, with the division by s equalizing the information contained in each data point. Data with a higher uncertainty are reduced more, so as to result in lesser weight in computing the χ^2 likelihood fit, than data with lesser uncertainty.

We qualitatively describe the model $M_\phi \psi$ of shape $(\text{box_w}, \text{box_w}, N_\lambda)$ as follows. The contribution comes from a planet signal component, a starlight component, and a transmission component. The planet signal submodel is constructed for a particular exposure as follows. In the data cube, we take a $\text{box_w} \times \text{box_w}$ stamp around the star center. For each slice in the spectral direction of this stellar PSF stamp, we normalize the flux such that the total count for each spectral slice in the resultant stamp cube is unity. We construct a theoretical spectrum of the planet for the same wavelength values as in the stamp cube. Then, we element-wise multiply this spectrum model, stamp cube, and the sky transmission for each wavelength slice (obtained from telluric calibration described in previous sections) to get a planet model. The total flux inside this planet model stamp is multiplied by a factor that makes it equal to the total flux inside the initial stellar PSF stamp; this property helps later in planet search and sensitivity.

To get a theoretical planetary spectrum, we start with a `BT-Sett1-CIFIST2011c`¹⁰ atmospheric model (Allard, Homeier, and Freytag, 2010; Allard, 2014) for a nominal directly imaged exoplanet (typically taken to be temperature $T = 1800$ K and specific gravity $\log g = 5$). `BT-Sett1` is a set of one-dimensional cloud models for exoplanets (as well as stars and brown dwarfs), where timescales of condensation, coalescence, gravitational settling, and mixing are used to determine the abundance and size of ≈ 55 types of atmospheric grains (Lafrenière et al., 2009), with the final spectrum determined by the spectral and radiative transfer modeling done by the

¹⁰<https://phoenix.ens-lyon.fr/Grids/BT-Sett1/README>

¹¹<https://phoenix.ens-lyon.fr/Grids/BT-Sett1/CIFIST2011c/SPECTRA/>

PHOENIX code (Husser et al., 2013). CIFIST2011c accounts for calibrations due to mixing length theory (Freytag et al., 2010) as well as non-equilibrium chemistry for certain atmospheric components. As it offers a temperature and specific gravity range of $T_{\text{eff}} = 300 - 7000$ K and $\log g = 2.5 - 5.5$, this set is sufficient to model the planetary signal we expect from directly imaged companions in young clusters like Ophiuchus and Taurus. We select the model with required planet characteristics, extract data from within the wavelength range of the Kn5 filter data, and broaden absorption lines consistent with the spectral resolution of OSIRIS. After construction of the needed planetary spectrum, the wavelengths of this spectrum are adjusted to account for redshift due to the radial velocity of the planet (which is provided as a non-linear parameter):

$$\frac{\lambda_f}{\lambda_0} = 1 - \frac{RV - RV_0}{c}, \quad (2.1)$$

where λ_0 and λ_f are initial and final wavelengths, RV is the provided radial velocity of the planet, c is the speed of light, and RV_0 is the barycentric radial velocity (conventionally used as a reference, to account for Earth's variable motion due to rotation or revolution).

The diffracted starlight model is constructed independently for each of the $\text{box}_w \times \text{box}_w$ spatial locations. For each location, we set 10 nodes at roughly equispaced locations along the spectral direction. The density of nodes is higher by a set factor in the region where the slope of the continuum is steeper, as this allows for a better overall fit to data. A *spline interpolation model* is constructed for the spectrum at this location. The nodes constrain this model to output a value close to the counts in the data cube, at the wavelength for each of the 10 nodes; in other words, if we plot the spectrum data and the spline model, they must approximately intersect at these 10 nodal locations. Between each pair of nodes, the model interpolates using a low degree polynomial. We multiply these by factors of the stellar spectrum (obtained from the same exposure) and sky transmission, as derived from procedures discussed before. Combining these spline models at each of the locations in the $\text{box}_w \times \text{box}_w$ region gives us a diffracted starlight model.

The reason a spline interpolation model is an appropriate choice for modeling diffracted starlight is the wavelength dependence of the point spread function associated with several instruments, including Keck/OSIRIS. Ignoring factors of sky transmission and stellar spectrum, the counts at a particular location can still vary because the point source function at that location for that wavelength has a brighter region (termed a *speckle*) than for another wavelength. Thus, due to the wavelength

dependence of the PSF, if we consider a specific location and vary the wavelength, the continuum value (ignoring spectral features) varies, that is, the continuum is not uniform and flat. A low degree polynomial spline interpolation can be used to model this behavior, with the spline nodes providing an adequate constrain to prevent over-fitting.

An empirical stellar spectrum is used in generating the diffracted starlight model. The process for obtaining this stellar spectrum is different from aperture photometry described in previous sections, and the output depends on the spatial location we are generating a model for. In order to prevent self-subtraction of planet signal, we mask a $\text{box_w} \times \text{box_w}$ region centered at the location at which we are modeling the data and then obtain a spectrum by summing flux inside a rectangular aperture. This improves our sensitivity, because (assuming that in case there is a planetary signal at the location under consideration) the planetary signal is not taken as a significant part of the stellar spectrum and thus almost absent in the starlight model.

The general schematic to construct both the planet and diffracted starlight models is to start with the product of a PSF model and sky transmission close to time to observation, and then add in a model for either planet spectrum or stellar spectrum and continuum respectively. Notably, both models rely on sky transmission extracted from the telluric calibration process, resulting in some common spectral features in either component of the model. These common features are mostly not spectral lines from the stellar or planetary atmospheres, but rather they are absorption lines due to the upper layers of the Earth's atmosphere. Empirically, we noted that this can result in a forward model predicting higher than reasonable contributions from planetary models, especially given that an extracted stellar spectrum is more susceptible to noise than a theoretically generated planetary model. We add in a final third component to the forward model, in the form of a product of the PSF stamp model and sky transmission. We start with a $\text{box_w} \times \text{box_w}$ stamp, taken from the center of the stellar PSF, normalize this stamp cube such that every wavelength slice has unity flux, and then multiply the wavelength slices with the sky transmission.

These components (a planet model, several diffracted starlight splines (for several locations), and a transmission stamp cube) compose different columns in M_ψ , with the rows denoting wavelengths. ϕ denotes contributions from these different components, and thus the matrix product $M_\psi \phi$ gives the needed forward model. Code in `bread.f.m.hc_mask_spline.f` optimizes ϕ for every choice of ψ through analytical marginalization of linear parameters, with the non-linear parameter space

ψ (generally radial velocity or spatial location) left to be separately defined and optimized. As an example, if these non-linear parameters ψ are a particular spatial location and radial velocity expected for a planet, different elements of ϕ give us the contribution from each of the components to the best fit. This is useful, as the contribution from the planet model ϕ_p as well as the uncertainty on that value σ_{ϕ_p} give us a measure of how much planetary signal is needed to explain the spectrum observed at the location. A high signal ϕ_p relative to the uncertainty σ_{ϕ_p} indicates that our forward modeling approach is predicting that the data cannot be explained in the best manner possible without a planetary component, thereby pointing to a detection.

For completeness, we note that other forward models were also developed as part of this thesis. The initial version of the forward model is present in `bread.s.fm.hc_splinefm`, which was structured similarly to the model used for developing the results. This older version does not implement pixel bleeding identification (described below), does not optimize the placing of spline nodes based on continuum slope, and constructs the planet component using a simple two-dimensional Gaussian model (instead of a more accurate stellar PSF stamp). We used best-fit values from 2D Gaussian fits for the stellar PSF (computed as part of the `TelluricCalibration` and spectra extraction) to model the planetary signal. We expect the newer forward model to be a more accurate representation of the planetary PSF, as it uses a more accurate model for the stellar PSF (by taking an exact stamp, instead of a rough 2D Gaussian approximation). We discuss in a later section that *algorithmic throughput* is a quantifier of the performance of the forward model that maps the sensitivity and how accurately a model can recover the signal from a planet. The forward model we use, `hc_mask_splinefm`, outperforms the older version based on this quantifier.

2.7 Bad Pixel Identification, Pixel Bleeding, and Correction Strategies

We apply bad pixel identification methods at various steps in our data reduction. Data cubes reduced by the OSIRIS DRP are imported as `bread.instruments.OSIRIS` objects. At this stage, a rough fit is performed for spectra at every spatial location, and outliers, deviated by over 3σ are identified as bad pixels. The rough fit consists of only the diffracted starlight component from the model m described above, and this component is also simplified to get a faster runtime. We use 20 nodes in the low degree spline interpolation and, instead of placing more nodes in regions where the spectrum continuum has a larger slope, we place nodes at equal separations. For this identification stage, a stellar spectrum is not necessary, `bread` can construct

a rough spectrum by taking a median of data around the star (used for science or transmission data sets) or over the entire field of view (used for sky images). This allows for bad pixel removal for non-target images, including for the OH line and telluric calibration data or data which was reduced using pair subtraction. We also identify the edge of the field of view in the Kn5 filter by determining the pixels where the count value is NaN (not a number, as set by the OSIRIS DRP beyond OSIRIS's field of view), and identify all these spatial locations, as well as immediately adjacent pixels in all three directions, as bad pixels, thereby taking care of instrumental edge issues.

Analyzing residuals from fitting spectra at a single spatial location, we observed that fit deviates from data at certain deeper telluric features. While the fit can model the location of these absorption lines due to Earth's upper atmosphere, it is unable to predict well the depth of these features. It is possible that this is a result of a detector undersampling and interpolation or uncertainty in wavelength calibration, and further work can identify the validity of these claims. In the absence of a component accounting for this behavior, our model would attempt to explain this deviation by unnecessarily increasing the absolute magnitude of contribution from the planetary component, significantly skewing our results. We hard-coded a mask to remove wavelength ranges that contain these telluric features. We similarly remove edges in the spectral direction, to nullify instrumental edge effects. The regions removed were between the wavelength indices: $0 \rightarrow 5$ (edge), $312 \rightarrow 318$, $343 \rightarrow 349$, $366 \rightarrow 370$, $373 \rightarrow 378$, $384 \rightarrow 388$, $396 \rightarrow 402$, $418 \rightarrow 422$, $446 \rightarrow 465$ (edge).

As noted in the sections above where observations with Keck/OSIRIS are described, Ophiuchus data during the first observation run was taken without rotating the field of view. we introduced rotating by 90 degrees between different sequences as an observation strategy after identifying a pixel bleeding issue. As described above, the physical detectors in the instrument measure a two-dimensional image, with each row capturing a sequence of spectra (for individual spatial locations). So, while typically the horizontally adjacent pixels on the 2D image are adjacent in the spectral direction to the pixel we are considering (flux at a different wavelength but the same spatial location), the vertically adjacent pixels are unrelated in wavelength or spatial location. Note that the distinction or boundary between different pixels is not strictly perfectly defined, and, due to instrumental diffraction, light can *leak* from or contaminate one pixel to the other (with the amount leaked decreasing with

distance from the center pixel). This effect is more pronounced for bright spatial locations, and the targets we observe are significantly bright; our exposure times push us close to saturation limits. While the OSIRIS DRP considers and corrects this pixel bleeding phenomenon, we still observe resulting effects in our data that we account for as part of bad pixel correction.

Consider the two-dimensional spatial field of view of OSIRIS on the sky, obtained when we collapse the data cube along the wavelength axis. Say the location of the star is (y_s, x_s) , which is generally (after removal of bad pixels) the brightest spatial location. We observed pixel bleeding significantly affecting our results in two extended spatial regions, that are around the same vertical location as the star. Along the horizontal direction, the centers of these regions were separated from the star by roughly 10 pixels, and they are present on either side of the star. Pixel bleeding at these spatial locations caused our forward model to predict a high *negative* planet model contribution or *planet flux* (relative to the uncertainty in this contribution or flux), which is expected to be a non-astrophysical systematic we attempt to correct for. Investigating these regions showed that the continuum of the spectra at these spatial locations is neither flat nor modulated by passing speckles as typical. In fact, the continuum is much higher at one edge of the wavelength axis and falls rapidly to a flatter uniform closer to the center of our wavelength range. This is demonstrated for one target in Figure ??, where we plot spectra at several spaxels.

We observe that atypical continuum behavior in several images for several targets. This indicates that the issue does not originate from a specific observational systematic. Given that the regions are dependent on the location of the star in the field of view, we claim that the issue is not limited to a specific bad region of the OSIRIS instrument. We deduce that this behavior is a result of contamination from adjacent pixels due to bright starlight. While the spatial location of the star and these affected regions are separated in the data cube, the locations can be close to each other in the two-dimension raw frame that OSIRIS records from its detectors. Because of the high brightness of our targets, the OSIRIS DRP is unable to determine between real data and leaked light.

This bleeding problem cannot be accurately modeled, as the pixels (the light is bleeding from) correspond to wavelength values that are unrelated to the wavelength values of the pixels their light bleeds into. In other words, vertically adjacent pixels on the two-dimensional OSIRIS raw frame are not related in wavelength. Thus, bleeding light introduces spectral features at inconsistent wavelength locations. We

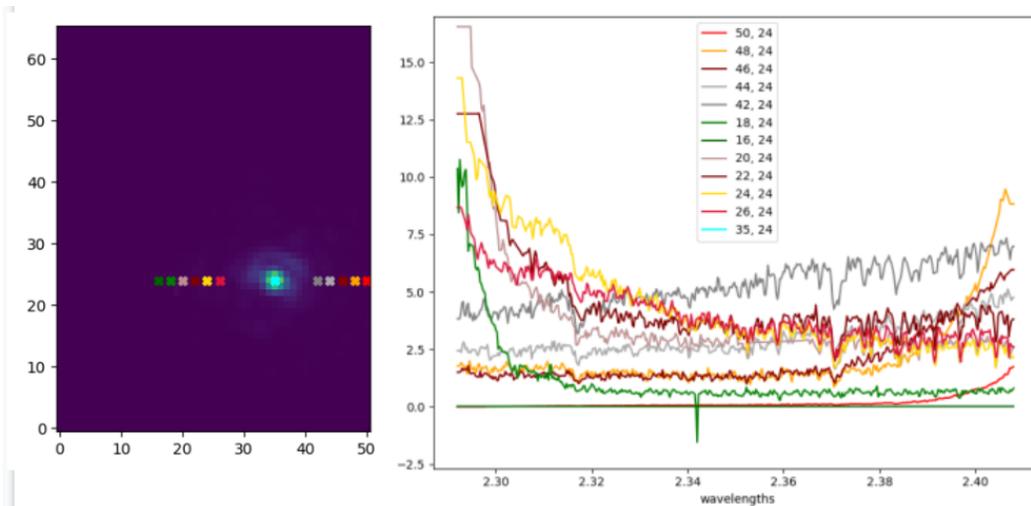


Figure 2.5: Nominal example of the pixel bleeding issue. Left panel shows the collapsed two-dimensional spatial field of view of Keck/OSIRIS, with the center of the target PSF marked by a light blue cross. Right shows spectra at several locations (labelled by spatial indices), as detector counts for a given wavelength in microns. Pixel bleeding is exhibited in the atypical continuum behavior, at locations roughly 10 pixels to the left/right of the target, and within a few pixels in the vertical direction. The continuum at these locations is significantly higher at one edge of the wavelength range.

attempted varying the settings in the OSIRIS DRP (as described above) to get a more accurate data cube itself, but the result remained sufficiently inaccurate.

It is unfeasible to model this behavior accurately enough for our science purposes, so we treat it as a bad pixel identification problem. As described before while discussing the diffracted starlight model, we optimize the placement of spline nodes to get a higher density of nodes in regions (in wavelength direction) where the continuum has a high slope, or, is less flat. This was initially added to attempt to model the behavior but proved to be insufficient due to the spectral line inconsistency discussed above. Instead, the data object code in `breads.instruments.OSIRIS` was modified to consider the slope of the continuum of the spectra at every spatial location in the field of view. The code detects if the ratio of the counts at either edge of the spectrum to the median of all the counts at the location is higher than a certain threshold; if yes, the code masks away the region of the spectrum a factor above the median to the high edge. This removes data that is artificially higher than astrophysical reality due to pixel bleeding while attempting to retain as much data as possible at every location in the field of view.

With the Ophiuchus data set, we noted that this code identified a large spatial

region as unusable (which is a significant disadvantage to the probability of new detections), which resulted in a different observation strategy for the Taurus and remaining Ophiuchus data set. We rotate the field of view by 90 degrees for roughly half of the data, thereby having different parts of the sky be covered by the unusable regions in either half. Combining these data frames gives us a much larger field of view on the sky, which improves our chances of detecting a companion in the survey.

The forward model code additionally identifies any wavelength indices where the planet, stellar, or transmission spectra are NaN as bad pixels. The results from these bad pixel identification methods is stored internally in the `breadsticks.OSIRIS` data object as the `bad_pixels` cube, and the forward model code ignores any pixels recognized to be unreliable.

2.8 Planet Search, Frame Combination, and PCA of Residuals

Take linear parameter ψ_p to be the element of the ψ which is multiplied by the planetary component, that is, ψ_p denotes the estimated contribution of planetary signal to the best fit. While constructing the planetary component of the forward model $m = M_\phi\psi$, we normalized it to get the flux in the planetary stamp cube to be the same as the flux in the initial stellar PSF stamp cube. Thus, ϕ_p is in units of total stellar flux, or, ϕ_p gives the ratio of estimated planet brightness to stellar brightness, which is termed *contrast*. The ratio between ϕ_p and its uncertainty σ_{ϕ_p} gives us a signal-to-noise ratio (SNR), which is an extremely useful value to estimate. For a set of non-linear parameters, if this ratio $S/N = \phi_p/\sigma_{\phi_p}$ is higher than a threshold (typically 5), it indicates that the signal is inconsistent with zero with a related high probability and that the forward model predicts a detection with a related certainty.

Code in `using-breads.get_planet` and `breadsticks.grid_search` runs the described forward modeling approach for a given single image for all possible spatial locations. This is done using a *grid search*, that is, by selecting the set of non-linear parameters such that every combination of (RV, y, x) has the same $RV = 0$, and y, x covering a large enough range to analyze every spatial location in the field of view. These give us single-frame SNR maps, containing the signal-to-noise ratio at every location that we can estimate based on just one single image. Code in `using-breads.SNRmaps_contrast` or `using-breads.combine_frames` can use several single frame SNR maps from all sequences taken for a particular target, to yield a combined SNR map, which is a final signal-to-noise map for a target from

our survey.

In the process of generating these single frame SNR maps, we perform principal component analysis (PCA) of the residuals to better our forward model of diffracted starlight. An initial faster model is computed for a small region of the field of view spatially above the star, taking $y = 1 \rightarrow 11, x = -5 \rightarrow 6$. These best-fit predictions are subtracted from the data to obtain residuals over this smaller data cube. Performing PCA, by computing the eigenvectors of the residuals' covariance matrix, gives us the orthonormal basis composing the principal components, which are then used in the forward model for the lower half of the field of view, with $y = -40 \rightarrow 0, x = -20 \rightarrow 20$. Similarly, the principal components obtained from PCA on a region from below the star, $y = -10 \rightarrow 0, x = -5 \rightarrow 6$, are used to get a better model for the top half of the field of view $y = 0 \rightarrow 40, x = -20 \rightarrow 20$. The reason for using a disjoint region of the sky, to get principal components for analyzing each half, lies in noting that a planetary signal could also appear as a residual, which would contribute to its own principal component. If we were to use an overlapping region of the sky to generate principal components, we could model for the planetary signal (as a principal component of the residuals) and subtract it away assuming it is residual from diffracted starlight modeling. Having separate regions for PCA prevents self-subtraction of the planetary signal.

Given that we dither the detector's field of view by a few pixels between sequences and, for some targets, rotate it by 90 degrees, the combination algorithm works as follows. We use the location of the star as the physically constant reference position. In the single frame reduction code, we set the spatial location with the highest flux as the center of the stellar PSF, and the forward model interprets the non-linear parameters for spatial location (y, x) as deviations (in pixels) from this reference. So, if the forward model is provided $y = x = 0$, it interprets it as the location of the star and attempts to estimate linear parameters, including ϕ_p (and thus SNR) at that location. Similarly, the forward model interprets $y = 2, x = 5$ as 2 pixels vertically up and 5 pixels horizontally right from the center of the stellar PSF. The range of non-linear parameters provided to the grid search is $y = -40 \rightarrow 40, x = -20 \rightarrow 20$, which is sufficient to cover the field of view covered in the full survey, accounting for dithering between sequences.

Thus, the single-frame SNR maps we get have consistent indices, up to the 90-degree rotation of the field of view, which we can also align by centering on the star. So, our code rotates the needed sequences around the star, and then adds the planet flux

ϕ_p and error σ_{ϕ_p} at every spatial location for all single frame maps, taking care of possible NaN values at some locations in some maps. If $\phi_{p,j}$ and $\sigma_{\phi_{p,j}}$ are the (non-NaN) flux and error at a certain location in the j^{th} single frame, then we can get the total ϕ_p and σ_{ϕ_p} as:

$$\phi_p = \frac{\sum_j \frac{\phi_{p,j}}{\sigma_{\phi_{p,j}}^2}}{\sum_j \frac{1}{\sigma_{\phi_{p,j}}^2}} \quad (2.2)$$

$$\sigma_{\phi_p} = \left(\sum_j \frac{1}{\sigma_{\phi_{p,j}}^2} \right)^{-1/2} \quad (2.3)$$

2.9 Noise Normalization for combined map

We use some theoretical assumptions about the noise in planetary signal σ_{ϕ_p} to perform several stages of normalization on the values of σ_{ϕ_p} obtained above in the combined SNR map. Code used is `using-breads.SNRmaps_contrast`. we first compute the standard deviation of the single-frame SNR values at every location (taking care of possible NaN values in some maps), and multiply the final noise σ_{ϕ_p} by this standard deviation at every spatial location:

$$\sigma_{\phi_p}(y, x) \rightarrow \sigma_{\phi_p}(y, x) \times \text{std}_j \left(\frac{\phi_{p,j}(y, x)}{\sigma_{\phi_{p,j}}(y, x)} \right), \quad (2.4)$$

where std_j is standard deviation taken while varying j , with a set (y, x) . The final noise in the planetary flux is underestimated and must be scaled by the noise in the SNR values.

We additionally normalize the noise in the radial direction. We make a theoretical assumption that the standard deviation of SNR values within a thin circular shell (ring or *annuli*) should be 1, based on the fact that there is no physical reason (in the absence of a planet) for the noise and flux to behave differently due to angular orientation. We consider annuli of radius r and thickness dr (of the order of a pixel), take the standard deviation of SNR values within this ring, and then multiply all noise values in the ring by the standard deviation:

$$\sigma_{\phi_p}(r, \theta) \rightarrow \sigma_{\phi_p}(r, \theta) \times \text{std}_\theta \left(\frac{\phi_p(r, \theta)}{\sigma_{\phi_p}(r, \theta)} \right), \quad (2.5)$$

where $m(r, \theta)$ is value of m at radius r (between r and $r + dr$) and angular orientation θ and standard deviation is taken while varying over θ .

As part of further work, we are working on including a normalization in the radial velocity direction, termed a *RV CCF calibration*. This normalization has not been applied to and optimized for the entire data set, so we cannot claim its efficiency for all targets. As it was developed as an option (to be tested and included in future reductions) as part of this thesis work, we discuss the targets it is applied to in the Results Chapter 3. We also include an entire reduction where this normalization is not performed.

If we vary the third non-linear parameter, radial velocity, in addition to the spatial non-linear parameters for the forward model, we obtain a map of how the SNR depends on the radial velocity, which we term a *radial velocity cross-correlation function (RV CCF)*. Data from a real astrophysical object would correspond to a radial velocity of a few km/s instead of several hundreds or thousands of km/s. So, the RV CCF should peak strongly around zero RV, and be of low absolute magnitude (between -1 and 1, for the signal to remain within 1σ) far away from 0 km/s. We use a grid search at every spatial location for 41 steps of RV from -4000 to 4000 km/s. Removing the central region around 0, we compute the median and the standard deviation (while varying RV) of the SNR values at a spatial location. We normalize all SNR values at that location by subtracting the median and dividing by the standard deviation, as a way to have the SNR be between -1 and 1 in the RV CCF far away from 0 km/s:

$$\sigma_{\phi_p}(y, x) \rightarrow \sigma_{\phi_p}(y, x) \times \text{std}_{RV} \left(\frac{\phi_p(RV, y, x)}{\sigma_{\phi_p}(RV, y, x)} \right), \quad (2.6)$$

with a set spatial location (y, x) .

2.10 Throughput Correction

We quantify how effective our algorithm is at recovering the contrast (or brightness, in units of stellar brightness) of a planet. We term this measure, *throughput*. It is important to note that we are not referring to instrumental throughput (which is the fraction of light an instrument can capture from a source on the sky). Rather, we consider *algorithmic* throughput, which measures the ratio of the flux our algorithm estimates for a companion (*recovered flux*) to the real astrophysical brightness of the companion. We expect throughput to mostly depend on the distance from the target star, with the theoretical value being 1 far away from the star (we are fully sensitive to planet signal where diffracted starlight is negligible), and 0 immediately adjacent and on top of the star (in which case, any model will assume that the planet signal and its spectral features are part of the star signal and spectrum).

In order to estimate this algorithmic throughput, we need to perform our forward modeling approach on data where the presence and brightness of a companion are already known, which is not possible for a blind search data set (where the goal is to detect and characterize new companions). We use the code in `breadc.injection` and `using-breadc.throughput_maps` to perform fake injection and recovery. We inject, with a set combination of non-linear parameters (specific spatial location and radial velocity), a simulated planet into each data cube for a specific target. We can then test our forward model by attempting to recover the simulated companion with the same non-linear parameter setup.

We need the simulated planets to mimic astrophysical sources as accurately as possible. For the injection process, we start with a data cube outputted by the OSIRIS DRP, instead of the raw two-dimensional frames collected from the detectors, because it is easier to model a companion with the cube, instead of mapping the 3D cube into the 2D frame. We are provided a set of non-linear parameters, such as radial velocity and spatial location, describing the planet signal, as well as the brightness of the fake planet relative to the star, or `contrast`.

We first extract a stamp cube centered on the center of stellar PSF from the data cube. This is similar to the construction of the planetary component of the forward model, except we take a much larger stamp size of `stamp_w = 51` spatial locations (instead of `box_w = 3`). In the forward model, the only relevant portion of the planetary signal was within the `box_w`, as that is also the size of the data stamp d we fit to. However, for the injection of a simulated planet, we need a larger stamp, because the aperture used for computing the stellar spectrum is large. A real planet signal would extend to every pixel in the field of view, thus the computed starlight spectrum would be biased by planet contributions at every pixel. Thus, to best simulate a planet signal, the injection should also affect every pixel within the aperture used to estimate the stellar spectrum. Our injection code considers edge cases where a portion of this (much larger) `stamp_w × stamp_w` stamp would lie outside the field of view, by managing padding pixels that are discarded after injection.

Similar to the forward model construction, for every wavelength slice of this larger stellar PSF stamp, we normalize the flux such that the total count for each wavelength is unity. Having constructed a theoretical spectrum from a `BT-Settl-CIFIST2011c` atmospheric model for a nominal directly imaged planet with $T_{\text{eff}} = 1800$ K and $\log g = 5$, we modify the wavelengths of this spectrum to account for redshift due to radial velocity. Then, we perform an element-wise

multiplication over this spectrum, sky transmission (from telluric calibration), and the PSF model. The total flux over this planetary stamp is multiplied by a factor to make it `contrast` times the total flux with the initial stellar PSF stamp. we add this to our overall data cube at the specified spatial location, to get a simulated exoplanet with a given `contrast`.

For a complete description of the thesis work, an initial version of the injection code was also developed that used a 2D Gaussian model for the PSF, with wavelength-dependent position and widths, as obtained from the process of telluric calibration. This was rejected because a stellar PSF stamp is a more accurate representation of a real astrophysical source than a functional approximation.

We obtain a throughput map for every image of every target as follows. We select a range of spatial locations that covers the field of view, $y = -40 \rightarrow 40, x = -20 \rightarrow 20$. For each spatial location, we inject a planet of `contrast` = 0.01 with a fixed spectrum (1800 K, $\log g = 5$) and $RV = 0$. The contrast used to be neither too high (such that it would interfere with the algorithm by being comparable in brightness to the star) nor too low (such that we can get a good estimate of throughput and are not highly limited by the noise floor). We run our forward model at the same location to estimate the planetary flux ϕ_p . Note that this *recovered flux* is also in units of stellar brightness, so it in fact, gives us a value for the estimated contrast. Thus, directly taking a ratio of the recovered to the injected flux, as $t_a = \phi_p / \text{contrast}$, gives us the algorithmic throughput at that spatial location. Repeating this over the entire field of view gives us a throughput map.

Throughput maps are essential to understanding systematic errors that our algorithm introduces in the reduction. Notably, it can be used to quantify the true brightness of a detected companion based on the estimated brightness from the forward model. Hence, if our approach predicts a planet signal of contrast 1/10 near the star, it might be underestimating the signal more significantly than if it predicted the contrast 1/10 at a location spatially far from the star. Throughput correction does not however affect signal-to-noise (and, hence, are talked about after the section on combined SNR map generation), as both ϕ_p and σ_{ϕ_p} are scaled by the same factor t_a in the correction, resulting in the same ratio. Contrast curves, that we discuss in the next section, do need a throughput correction because they proportionally depend on the estimated uncertainty in planet flux σ_{ϕ_p} . Because of how we construct these contrast curves, we do not need to merge single frame throughput maps into a combined map for each target, because we can apply the correction to each image using the same

frame’s throughput map.

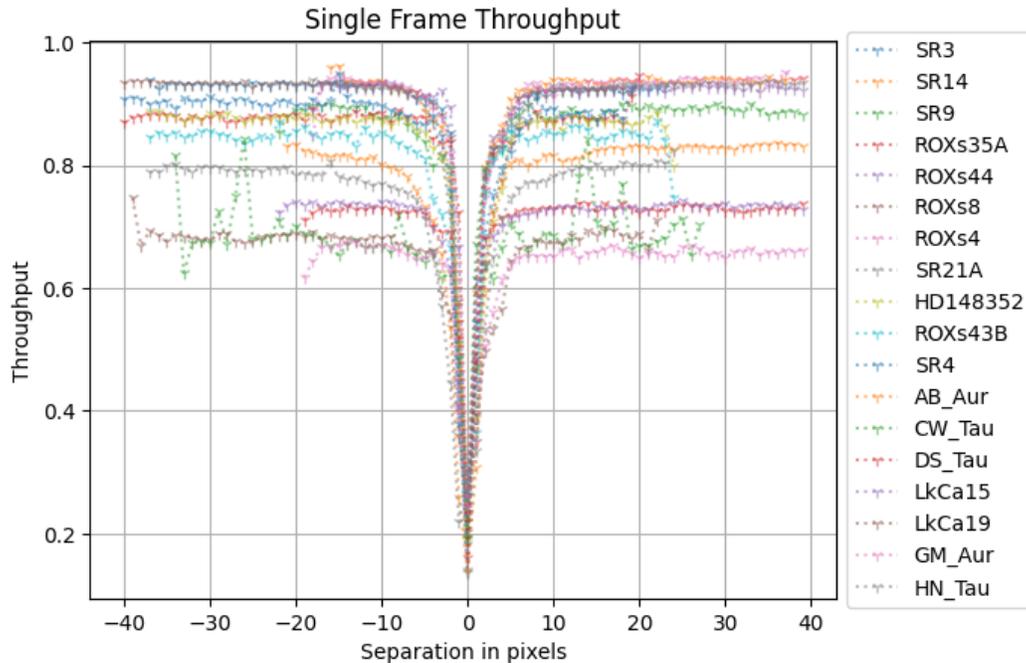


Figure 2.6: Algorithmic throughput t_a (ratio of flux recovered by our approach, to the flux that was injected as a simulated fake planet signal) as a function of separation from the star (in units of pixels, each spatial pixel corresponding to 20 milliarcseconds). As described in Section 2.10, we empirically note that throughput dominantly depends on separation from the star. We can interpolate between the t_a values computed at every pixel (as presented in this figure) to get t_a everywhere in the field of view. Notably, t_a tends to 80 – 90% (expected 100%) for most targets far away from the star, and to 10 – 20% (expected 0%) at the location of the star. Our model is able to distinguish most of the planetary signal from the diffracted starlight.

Empirically, we noted that our two-dimensional throughput maps (representing t_a at all spatial locations in the field of view) indicate that throughput depends on separation from the star (with throughput decreasingly rapidly closer to the star) and the amount of useful data at a location (edges or pixel bleeding regions have a lower throughput as some portion of the data was removed in bad pixel identification). t_a does not significantly depend on the angular orientation of a location from the star. This provides us with a significant opportunity to cut down on the time it takes to compute the throughput correction for contrast curves. Instead of using a two-dimensional map that gives us t_a values for the entire field of view, we can instead compute t_a along a one-dimensional strip on the map. Using a range of $x = 0, y = -40 \rightarrow 40$, we get t_a at integral separations from the star (in units of

pixels) along a vertical strip. These are shown for a subset of our targets in Figure 2.6. We use spline interpolation to get t_a at separations between these discrete integral values. We perform throughput correction using this faster spline interpolation for t_a to obtain our contrast curves.

2.11 Contrast Curves and Sensitivity

Contrast curves encode the sensitivity of our survey for each target. We use a 5σ threshold for a credible detection. To be counted as a candidate, a real astrophysical companion should be at least five times brighter than the estimated uncertainty at that location. Contrast curves plot this required planet contrast threshold (brightness, in units of stellar brightness) as a function of separation from the star. Our survey will detect, at $> 5\sigma$ significance, a companion with separation and contrast that lies above this curve. These also allow comparison to other surveys in the exoplanetary field, which is extremely useful in discussing one of our main science goals of testing the sensitivity of our high-resolution spectroscopy methods.

We computed σ_{ϕ_p} at all spatial locations using `using-breads.get_planet` while generating single frame SNR maps. We scale this by the throughput correction to get a estimation of the real uncertainty in planet flux $\sigma_{\phi_p}(r) \rightarrow \sigma_{\phi_p}(r)/t_a(r)$, where t_a is computed using the spline interpolation described above. Combining these two pre-computed quantities gives us a corrected uncertainty for all spatial locations. Using the known location of the star in the field of view, we compute the separations of each spatial location from the star (each spatial location corresponds to 20 milliarcseconds on the sky). This gives us the planet uncertainty as a function of radial separation on the sky. Multiplying this by the detection threshold of 5 gives us the needed contrast curve, denoting the planet flux required for detection.

The code that generates contrast curves is also in `using-breads.SNRmaps_contrast`, which also generates combined SNR maps and performs noise normalization. We can easily compute two contrast curves, one each for before and one for after noise normalization that scales the values of σ_{ϕ_p} . Comparing these two curves is interesting because it encodes information about systematics introduced by our forward model and the under-utilized potential of the technique. All the noise normalization steps decrease the signal-to-noise ratio by increasing the noise or uncertainty σ_{ϕ_p} . We need to artificially inflate the noise levels to match the theoretical expectations we have about the behavior of σ_{ϕ_p} . If our forward model was more capable of exactly modeling spectral data (for each exposure for each combination of non-linear

parameters), an artificial noise scaling would not be necessary, and we would be only photon and detector noise limited. Thus, the contrast curve computed without any noise normalization gives us an estimate of the method’s potential sensitivity with a better forward model.

2.12 Validation and Analysis of a Detection

In this section, we detail the steps we plan to perform when we observe a possible detection in any of the combined SNR maps generated as in Section 2.8. Our forward model, as previously described, would yield an estimate for the planet flux and its uncertainty, at a specific location. We directly compute the brightness of the companion relative to the host star, as well as astrometric data, including the separation from the host star.

The SNR maps encode signal-to-noise ratios at different spatial locations, varying two non-linear parameters y, x while keeping RV at a fixed value. Given the location of a possible detection, we can instead vary the RV while keeping the spatial location constant. This computes SNR as a function of radial velocity, which is a radial velocity cross-correlation function. We already compute these values to perform the RV CCF calibration, as part of the noise normalization described in Section 2.9. For real astrophysical objects, the RV CCF should peak around $RV = 0$, as radial velocities of bound companions would not be of the order of hundreds of kilometers per second.

We have already computed SNR values at different RV settings, as part of the RV CCF calibration. The plots presented in this thesis were computed using these values, plotting SNR at 41 equispaced values between -4000 to 4000 km/s. These values of ϕ_p and σ_{ϕ_p} are computed independently using the forward model for each exposure of the target, and then combined using Equations 2.2 and 2.3, as detailed in Section 2.8. As only the overall structure of the CCF and peak around $RV = 0$ are used in the validation of detection, we perform only two steps of noise normalization described in Section 2.9. We normalize while varying over different frames and varying radial velocity, but do not perform normalization over radial annuli.

Code was also developed, in the form of a new fitter `breadsc.f.m.hc_atmgrid_hpffm` and `using-breadsc.analyze_planet`, to ascertain other characteristics of a possible candidate. These still construct different components of the forward model M_ψ similar to `hc_mask_splinefm`, and optimize over linear parameters ϕ using analytical marginalization. However, instead of optimizing spatial location (y, x)

with a fixed planet model and RV , it assumes a fixed location and optimizes over parameters that describe the spectrum of a planet. `BT-Sett1-CIFIST2011` offers spectra models for planets of varying effective temperature T_{eff} and specific gravity $\log g$, while we can model for planetary spin and RV in terms of broadening and Doppler shifts of absorption lines, respectively.

Our code uses a grid of theoretical atmospheric spectral models for planets of varying T_{eff} , $\log g$, RV , and spin. Due to the high complexity of the optimization problem, we replace the grid-search with a Markov Chain Monte Carlo (MCMC) solver, which utilizes data from a single image of a target. We use `emcee`¹² with 512 walkers, 1000 steps for burn-in, and 1000 real samples. We plot the results as a `corner`¹³ plot, encoding the probability distribution for each variable, as well as the correlation between different parameters in the form of two dimensional PDFs.

¹²<https://emcee.readthedocs.io/en/stable/>

¹³<https://corner.readthedocs.io/en/latest/>

RESULTS AND DISCUSSION

3.1 Companion detection with combined signal-to-noise ratio maps

We present two-dimensional heat maps encoding signal-to-noise ratios at different spatial locations in our combined field of view, with the forward model using $RV = 0$ and a `BT-Sett1-CIFIST2011c` theoretical spectrum for an 1800 K and $\log g = 5$ companion. These are developed as described in Section 2.8, which uses the forward model described in Sections 2.6 and 2.7. We include SNR maps both before and after noise normalization as described in Section 2.9 (excluding the RV CCF calibration). Real astrophysical signal should be present in either map. Due to a large number of maps (2 for each of our 23 targets), these plots are included in Appendix A. In the SNR direction, as mapped by the color of different locations in the heat map, the maps computed before noise normalization range from -25 to 25, while the maps after range from -5 to 5 (which is our threshold limit). The vertical y and horizontal x axes of the heat map represent the separations of the spatial location (Δx and Δy in the x and y direction) from the center of the stellar PSF, in units of spaxels. Each spaxel is separated by 20 milliarcseconds on the sky from the immediately adjacent spaxel. $+x$ direction represents negative change in Right Ascension (towards West), while $+y$ direction represents positive change in Declination (towards North).

On all of these maps, the red cross marks the location of the star, while the blue dot marks the location of the highest SNR we compute. Not only does a planetary signal at a spatial location implies a high SNR value at that location, but we would also expect an extended region that is roughly circular, or *blob*, of high SNR. This is because the PSF of a planetary signal is extended, and we use a non-unity stamp cube (of width `bow_w`) in our forward model. We note the following features in the SNR maps:

1. Other than the two targets `Em*` SR 9 and HD 148352, we do not see any extended regions or blobs of signal-to-noise ratio of over 5, which is our set threshold for detection.
2. Generally, we do not see features of high SNR in these maps. This is reassuring because a correctly computed SNR map does not have several false detections

or extended features of very high positive or negative SNR. We observed features like those in earlier versions of the forward model (for example, with `breadths_fm.hc_spline_fm`), especially in the regions with significant pixel bleeding (as discussed in Section 2.7, pixel bleeding resulted in our forward model predicting a highly negative planet flux in these regions, which would result in extended feature with highly negative SNR). We do not see features with the current version of the forward model, which increases our confidence in the correctness of this forward model.

3. We extensively discuss the detection in the HD 148352 data in a Section 3.6, which we believe is a binary star companion. We use this detection as a validation of our sensitivity, as well as a test case to perform follow-up analysis on. This allows for an opportunity for us to demonstrate the methods we would use in case of a detection.
4. For target Em* SR 9, we note an extended feature, centered at roughly 30 pixels to the north of the star. Em* SR 9 is a well-studied wide binary star system, and Ghez, Neugebauer, and Matthews (1993) lists the binary separation and position angle as 0.59 ± 0.01 arcseconds and $350^\circ \pm 1^\circ$. This is also what we roughly measure, with ≈ 30 pixels corresponding to $\approx 30 \times 0.02 = 0.6$ arcseconds and position angle roughly towards the north. This is not a new detection, but still a validation of our forward modeling technique.
5. We lose a significant portion of our field of view, especially close to the star which is the scientifically interesting region to analyze, for the Ophiuchus data set without the 90-degree rotation. Our newer observation strategy fixes this issue.
6. We observe interesting blob-like features in a few SNR maps, such as those for ROXs 35A or SR 21A. However, none of these cross over our threshold detection SNR of 5. As discussed in Sections 2.9, 3.4, and 4.1, our sensitivity can further be improved, which might yield a $> 5\sigma$ detection from these currently low-significance blobs.
7. Similarly, we observed some extended features in some SNR maps after noise calibration is performed, that are not planetary signals but possibly residual errors from the forward modeling and noise normalization process. The Further Work Chapter 4 describes additions to the noise normalization

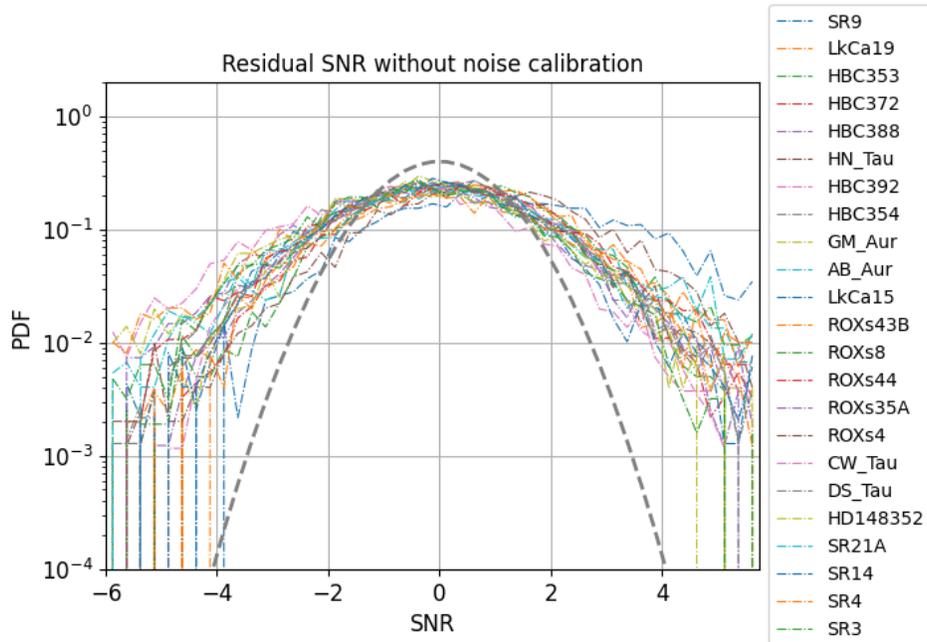
process, which are also discussed in Section 2.9 and are possible improvements to our model that we will continue to develop.

3.2 Validation of threshold SNR for detection

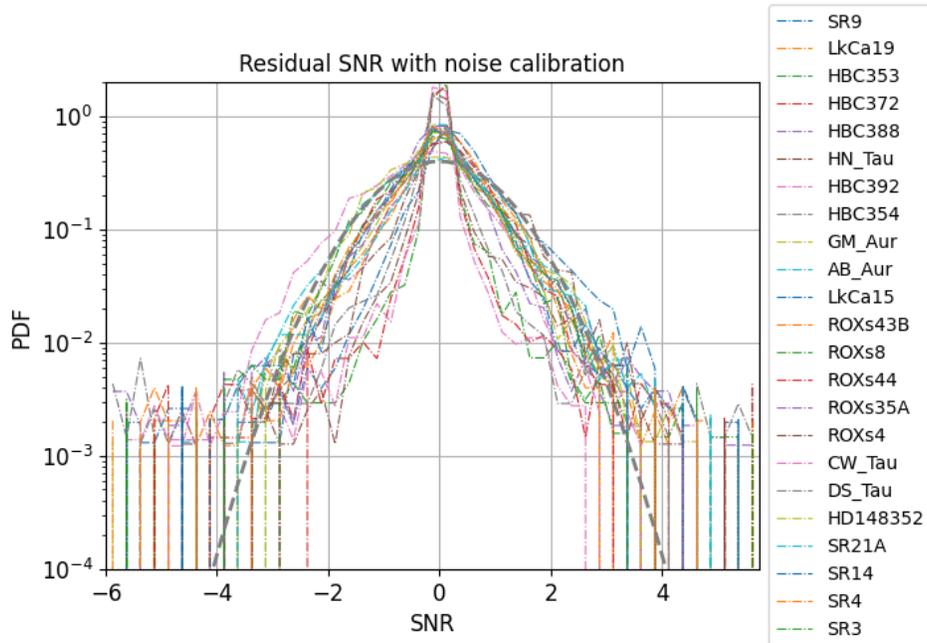
We present histograms of the combined signal-to-noise ratio spatial maps, that encode the probability distribution function (PDF) of the SNR values presented in Section 3.1. The top panels of Figure 3.1 show these histograms for all data sets before and after noise normalization. These residual SNR values should be theoretically distributed according to a Gaussian PDF, which is represented by the black dotted line. For an efficient model, the curves for each target will be approximately consistent with this Gaussian ideal. If the data is a wider bell curve than the theoretical, the probability of obtaining high SNR regions is higher than it should be ideally, which increases the probability of false detections. A narrower-than-theoretical curve for a target implies that there is a higher probability of missed detections (false negatives), and that true candidates (>5 SNR) are being underestimated as lower SNR (3-4) regions. While we strongly aim for attaining consistency between the empirical and theoretical bell curves for the PDF, if that is not possible, we prefer a model that has yields a lower probability of false positives than of false negatives.

The top panel of Figure 3.1 shows that the PDF for signal-to-noise ratios before noise normalization is significantly wider than the ideal, with the empirical probability of an SNR lying between -4 and 4 (very roughly) equal to the ideal probability of the range from -2 to 2. Without noise normalization, we overestimate our SNR, which is expected because the normalization generally decreases SNR values by inflating the uncertainty for a better agreement with its theoretically expected behavior. Appendix A plots heat maps for combined SNR frames without normalization over a larger range ($-25 \rightarrow 25$) because of this overestimation of SNR. The PDF improves after we perform noise normalization, as in the bottom panel of Figure 3.1, with the empirical curves contained within the ideal, reducing the probability of false detections. For a subset of targets, the noise normalization is too severe such that empirical bell curves are too narrow, which increases the probability of missing candidates. Possible improvements to the forward model and noise normalization are discussed in the Further Work Chapter 4 and Sections 2.9 and 3.4.

Figures 3.2a and 3.2b demonstrate the improvement in residual SNR distribution for two of our targets, HD 148352 and AB Aurigae. We select these two, because the



(a) without noise calibration



(b) with noise calibration, excluding RV CCF calibration

Figure 3.1: Probability distribution function of residual signal-to-noise ratio values in detection maps for all 23 targets, (a) before any noise normalization steps, and (b) after applying noise normalization, excluding RV CCF calibration. Black dotted line represents the ideal Gaussian PDF. We note that normalization reduces overestimation of the magnitude of SNR.

former is the target with a binary star detection, while the latter is a nominal example of a data set with some sequences rotated by ninety degrees. In both these cases, we note a significant improvement in the consistency between the empirical and theoretical probability distribution of the residual SNR. The curve for HD 148352 with noise normalization is additionally interesting, as it validates the detection threshold used for this target.

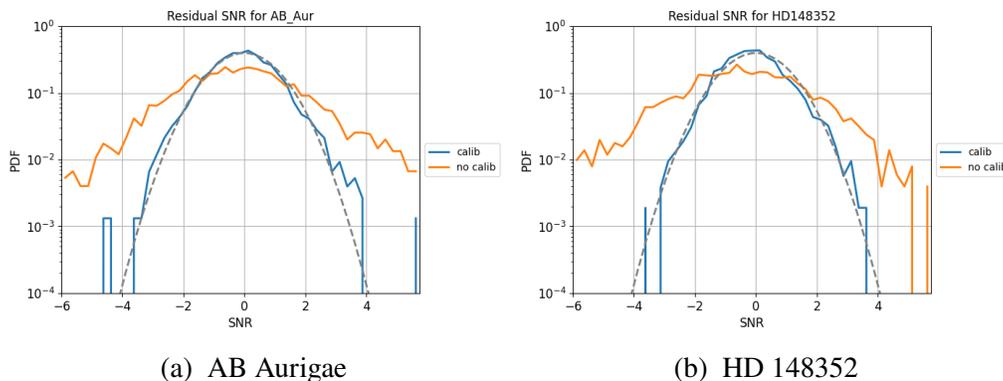


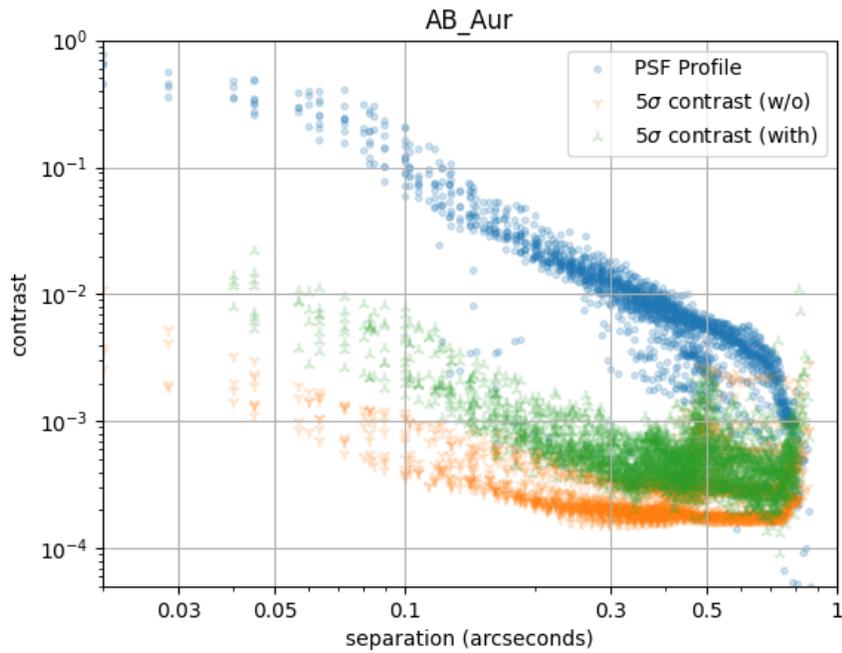
Figure 3.2: Probability distribution function of residual signal-to-noise ratio values in detection maps for two targets, before and after noise normalization steps. Black dotted line represents the ideal Gaussian PDF. We note that normalization reduces overestimation of the magnitude of SNR. The distribution for HD 148352 validates our detection threshold used in the discussion for Section 3.6.

3.3 Current and Potential Sensitivity measured in Contrast Curves

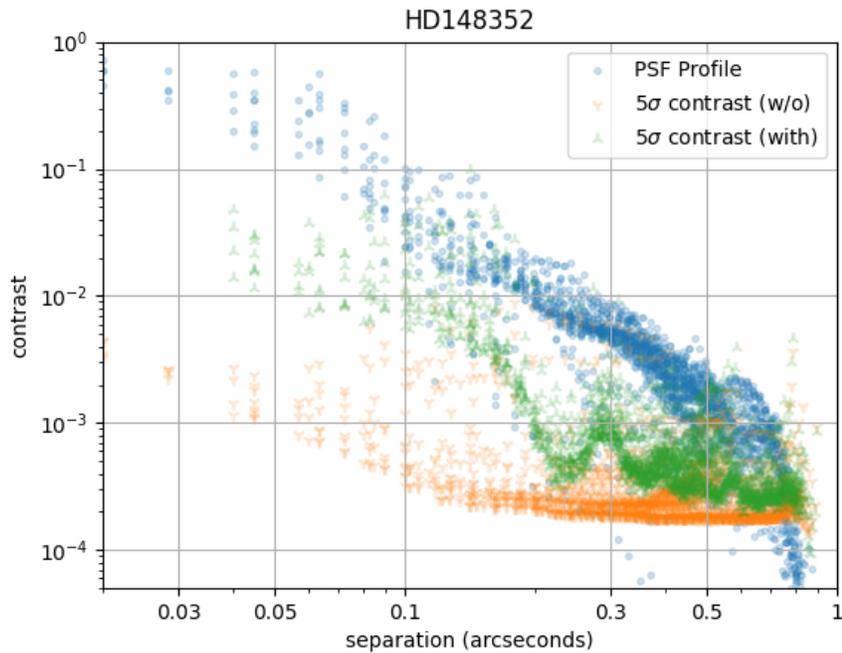
We present 5σ sensitivity curves for each target in the form of two kinds of contrast curves as developed in Sections 2.10 and 2.11, one before and one after we apply noise normalization (excluding RV CCF calibration) as described in Section 2.9. We use the forward model, described in Section 2.6, using $RV = 0$ and a BT-Settl-CIFIST2011c theoretical spectrum for a 1800 K and $\log g = 5$ companion.

Figures 3.3a and 3.3b overlay these two kinds of contrast curves, as well as a profile of the stellar point spread function, for two of our targets. The stellar PSF profile denotes, as a function of separation from the star, the stellar component's contribution at different spatial locations, as estimated by the forward model. It is graphed in the units of peak stellar flux, and demarcates an estimate of the speckle noise, due to diffracted starlight, at different separations from the star.

As described in Section 2.11, the potential sensitivity of our methods is estimated by the 5σ contrast curve before normalization. The curves after normalization estimate



(a) AB Aurigae



(b) HD 148352

Figure 3.3: Contrast curves, as a function of separation from the star, at 5σ significance for two targets. We include potential sensitivity (contrast before noise normalization) in orange, sensitivity after noise normalization (excluding RV CCF calibration) in green, and the stellar point spread function profile (quantifying speckle noise) in blue. Our sensitivity is well-below the speckle noise, with improvement possible to bring current sensitivity closer to potential values.

the current sensitivity, which is weaker than the former because noise normalization inflates our uncertainty levels after our forward model has been independently applied. We can improve our current sensitivity (to the potential contrast) by improving the capability of our model to remove diffracted starlight or by performing more robust noise calibration, like RV CCF calibration that is discussed in Sections 2.9 and 3.4.

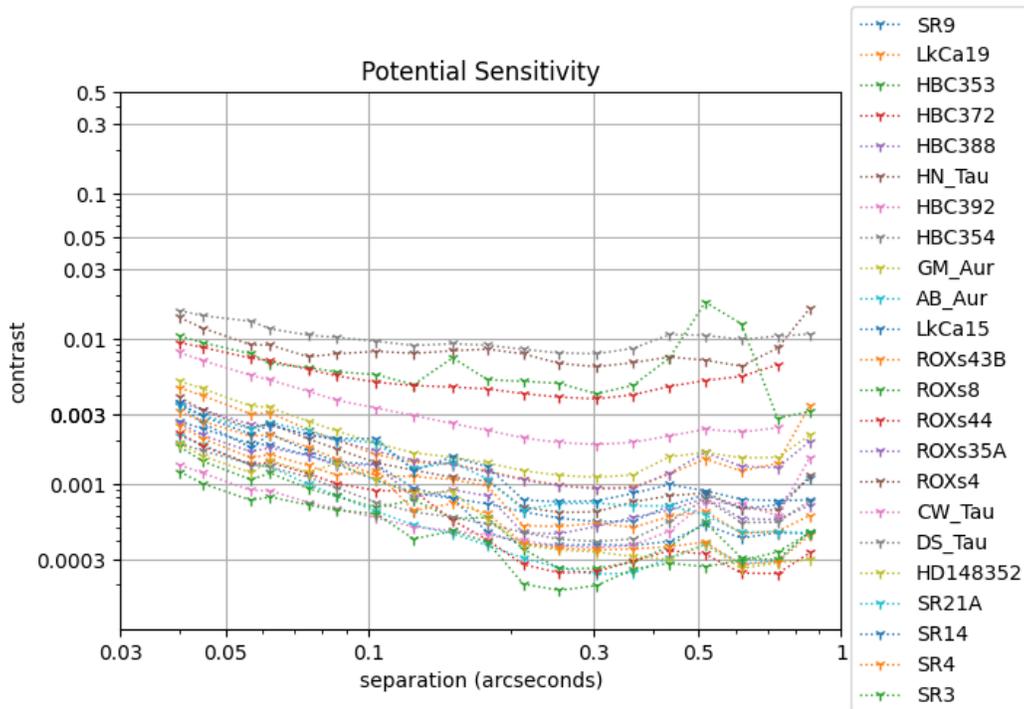


Figure 3.4: Contrast curves before noise normalization, mapping potential sensitivity at a 5σ significance level as a function of separation from star, for all 23 targets. Variability in sensitivity across targets is expected due to variations in amount of data, observing conditions, and spectral types or brightness. Scales of x and y axes are same as Figure 3.5.

Figures 3.3a and 3.3b present results for AB Aurigae and HD 148352, nominal examples for data sets with and without the ninety-degree rotation, respectively. Figures 3.4 and 3.5 plot contrast curves for all of our 23 targets before and other noise normalization. We note the following features in these contrast curves:

1. For the single target plots, both curves are well below the PSF profile, which demonstrates that we are able to model for diffracted starlight to get more sensitive than speckle noise.

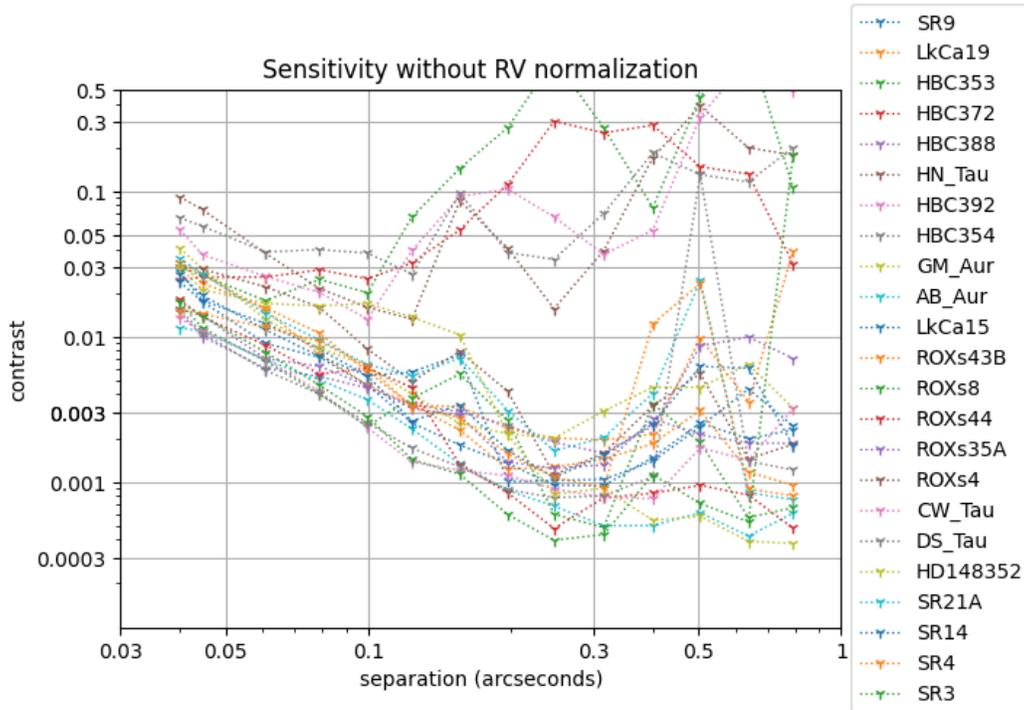


Figure 3.5: Contrast curves after noise normalization (excluding RV CCF calibration), mapping current sensitivity at a 5σ significance level as a function of separation from star, for all 23 targets. Variability in sensitivity across targets is expected due to variations in amount of data, observing conditions, and spectral types or brightness. Scales of x and y axes are same as Figure 3.4.

2. As discussed in Section 1.5, the separations of $\lesssim 0.3$ arcseconds are important for our science goal of demonstrating improved sensitivity closer to the star. We are well below speckle-noise at those separations.
3. We see features in the noise-calibrated contrast curves over ≈ 0.5 arcseconds, which represents the edge of the field of view, where less usable data is present.
4. Scatter for a given separation is low, which implies that: as expected, sensitivity is predominantly a function of separation from star and amount of usable data and is only weakly dependent on angular orientation.
5. Based on the potential sensitivity estimates, optimizing the model and calibration methods can improve our current sensitivity by roughly $1/2 - 1$ order of magnitude. Possible improvements with RV CCF calibration are discussed in Section 3.4, with further ideas developed in Chapter 4.

6. Difference in sensitivities for different targets is expected due to variations in spectral type, observing condition, and amount of exposure time.

3.4 Potential improvement with RV CCF Calibration

We present detection maps, PDF of residual SNRs, and sensitivity curves, similar to Sections 3.1, 3.2, and 3.3, for 14 out of our 23 targets, after we add in the RV CCF noise calibration described in Section 2.9. These results are generated with the forward model, using a BT-Settl-CIFIST2011c theoretical spectrum for a 1700 K and $\log g = 5$ companion. We vary RV to 41 points between -4000 km/s and 4000 km/s and normalize the signal-to-noise ratio values to have them be more consistent with theoretical expectations. Logistically, this newer method takes significant time to reduce each target. We present these results separately from the earlier sections because we have not reduced the entire data set using RV CCF calibration, and thus, we have not optimized the technique for all targets. Sections 3.1, 3.2, and 3.3 detail a full reduction of the entire data set.

As we have already presented SNR maps before noise normalization in Appendix A, we only include SNR maps after RV CCF calibration in Appendix B. We also generate Figure 3.6 as a histogram of these combined signal-to-noise ratio spatial maps, presenting the probability distribution function of these SNR values computed using the RV CCF calibration. As before, these empirical curves should be ideally consistent with a Gaussian PDF, denoted by a black dotted line.

We compare these figures with previous results when we did not apply this radial velocity calibration:

1. By including the RV CCF calibration, we reduce the frequency and intensity of extended features in the SNR maps.
2. We still observe a detection in the HD 146352 data set; in fact, we compute a higher SNR with the calibration than without. We still observe interesting $< 5\sigma$ blob-like features around ROXs 35A, which are discussed in Sections 3.7 and 4.2. We detected the well-known binary star of the Em* SR 9 system earlier, but this target has not been reduced using the RV CCF calibration yet.
3. The histogram encoding the PDF of the SNR values is more consistent with the Gaussian ideal for the reduction that uses the RV CCF calibration. This represents an improvement in the reduction, particularly enhancing the validation of the threshold signal-to-noise. This implies that the probability of both

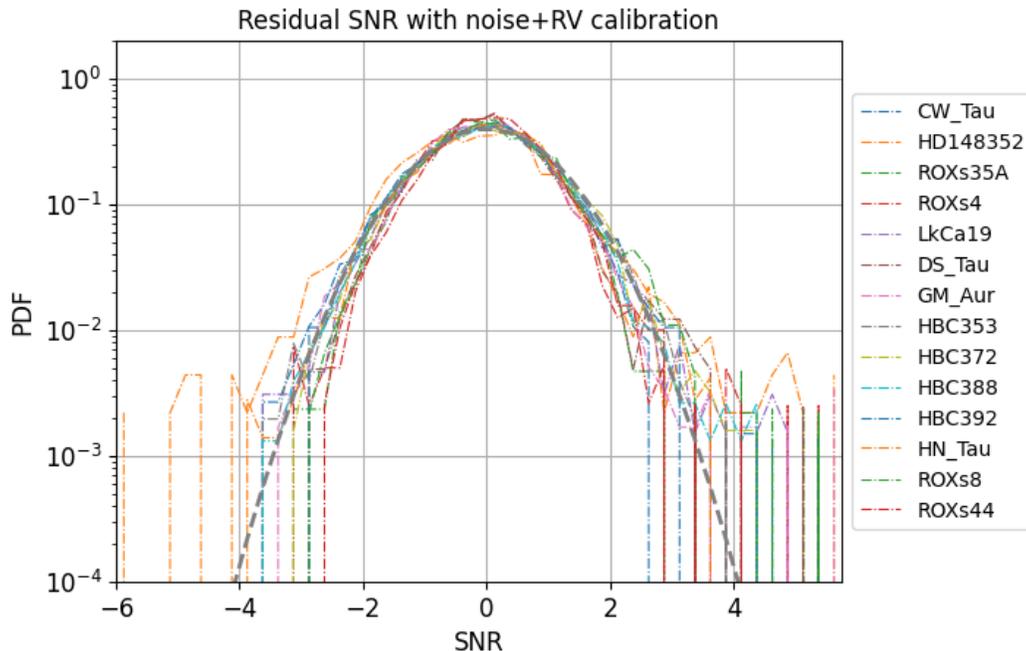
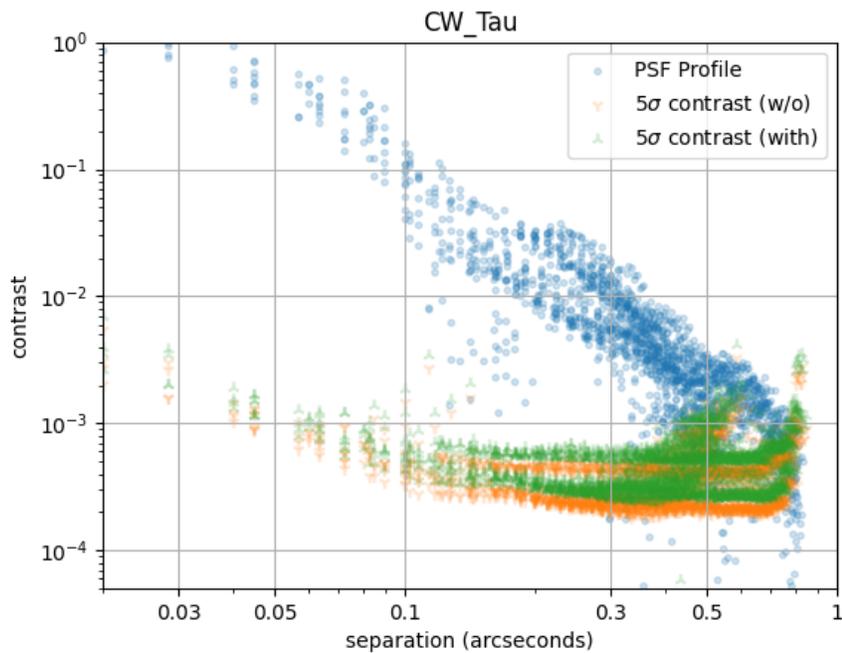


Figure 3.6: Probability distribution function of residual signal-to-noise ratio values in detection maps for all 23 targets, after all steps of noise normalization, including RV CCF calibration. Black dotted line represents the ideal Gaussian PDF. We note that this additional step of normalization results in a distribution that is more consistent with the ideal Gaussian.

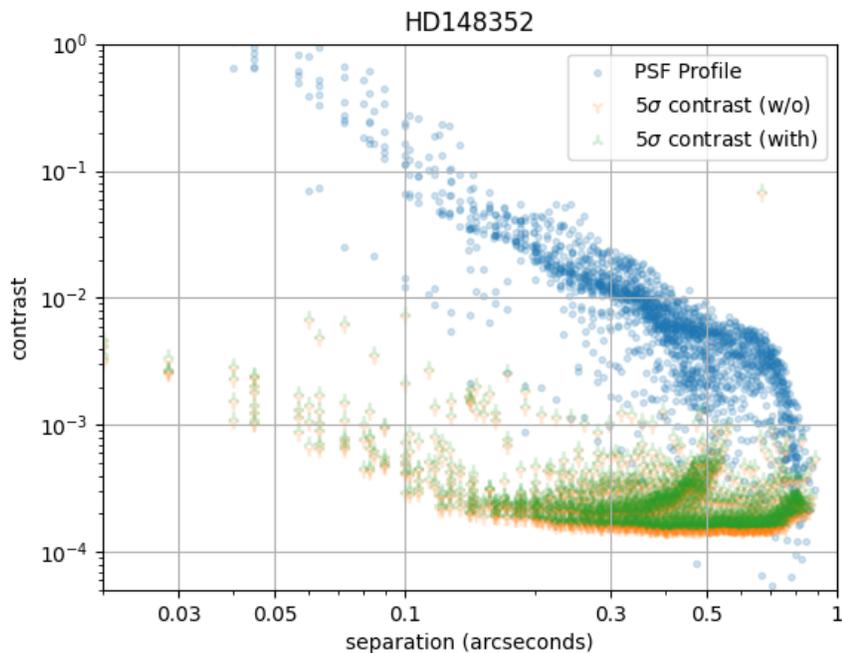
false and missed detections is low. Particularly, the probability of detection at a $> 5\sigma$ significance is closer to true value, after RV CCF calibration.

With Figures 3.7a and 3.7b, we similarly present 5σ sensitivity curves, one before and one after noise normalization (including RV CCF calibration), for CW Tauri and HD 148352 (nominal examples of data sets with and without the 90-degree rotation). We also include the stellar PSF profile in units of peak stellar flux, denoting an estimate of speckle noise. Figure 3.8 plots contrast curves for all targets reduced using the RV CCF calibration. We note the following in these figures:

1. We are still well below the PSF profile, so we are able to model for diffracted starlight. We see features due to the edge of the field of view, but they are less prominent.
2. Sensitivity after and before noise normalization is closer when we include the the RV CCF calibration. This implies that the calibration improves our method's sensitivity, and brings it closer to the best possible value.



(a) CW Tauri



(b) HD 148352

Figure 3.7: Contrast curves, as a function of separation from the star, at 5σ significance for two targets. We include potential sensitivity (contrast before noise normalization) in orange, final sensitivity after noise normalization (including RV CCF calibration) in green, and the stellar point spread function profile (quantifying speckle noise) in blue. Our sensitivity is well-below the speckle noise, with current sensitivity brought closer to potential values due to RV CCF calibration.

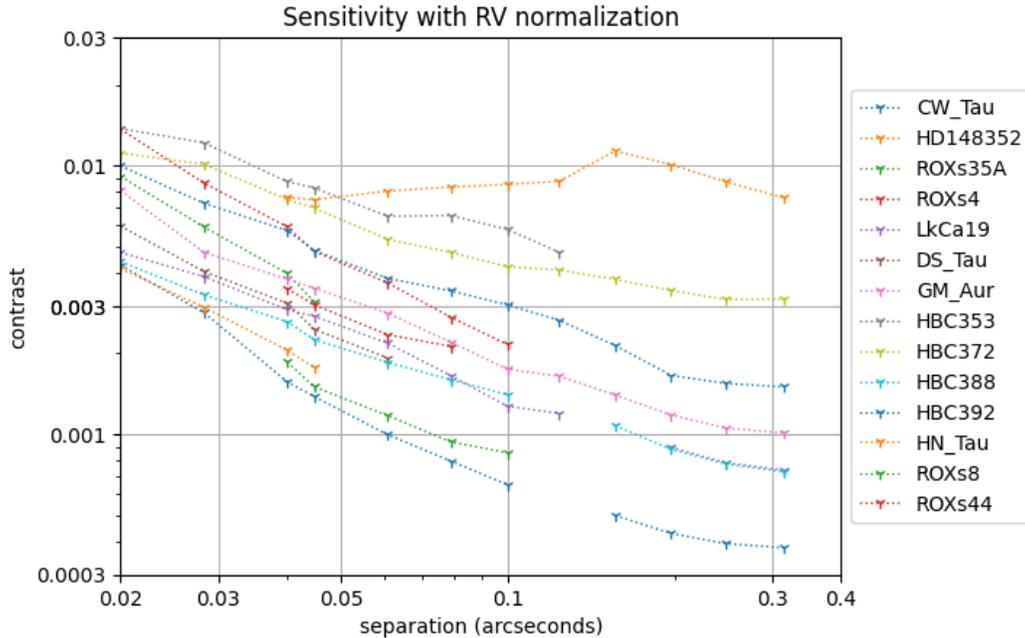


Figure 3.8: Contrast curves after all steps of noise normalization including RV CCF calibration, mapping final sensitivity at a 5σ significance level as a function of separation from star, for all 23 targets. Variability in sensitivity across targets is expected due to variations in amount of data, observing conditions, and spectral types or brightness. Note that this latter plot is graphed on a different scale on x and y axes than the earlier similar figures.

- Figure 3.8 plots sensitivity in the region of interest, at separations lower than roughly 300 milli-arcseconds. *We will use these values in the following sections as our final contrast curves, mapping the sensitivity of the survey.*

3.5 Factors limiting Sensitivity and Comparison with other methods

Besides being limited by systematics in our forward model for the diffracted starlight (even after these optimization steps), we are also limited by the instrumental structure of Keck/OSIRIS. In Sections 1.4 and 2.7, we describe that the detectors of OSIRIS lie on a two-dimensional optical plane, and the raw two-dimensional image taken by OSIRIS contains several rows of spectra corresponding to different spatial locations on the sky. The OSIRIS DRP converts this two-dimensional raw frame into the data cube that is used in our forward modeling approach. Our sensitivity can also be limited by the contamination of these microspectra into another, resulting in systematic correlated noise at different spatial locations. This can be addressed by more spacing between pixels on the two-dimensional detector field or by more calibration in the reduction pipeline. Spatially non-uniform properties of the instrument

that the DRP does not account for can also limit our sensitivity.

Figures 3.9a and 3.9b present our current sensitivity (that is, after all the noise normalization steps, including RV calibration) and potential sensitivity (that is, contrast curves before noise normalization) as compared to other direct imaging surveys. We select three representative instruments/surveys to contextualize our K-band sensitivity.

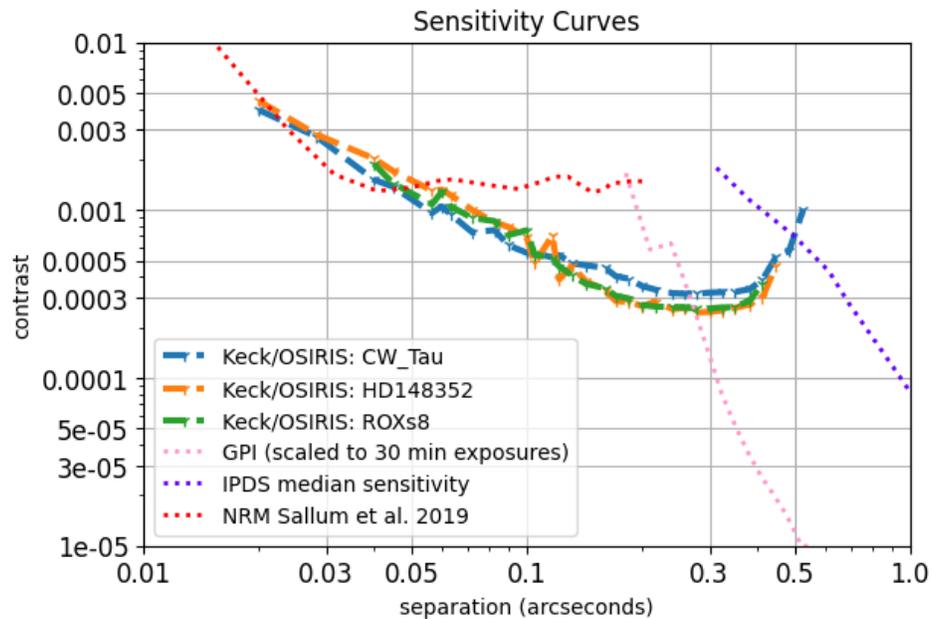
The Gemini Planet Imager (GPI) (Macintosh et al., 2014) is a high-contrast imager at the Gemini South telescope, which operates adaptive optics, diffraction control, and a low resolution integral field spectrograph with $R \approx 30 - 70$. We scale sensitivities of GPI to exposures of thirty minutes, which is roughly the exposure time (per target) for our survey. Sallum and Skemer (2019) lists sensitivity for non-redundant masking, which converts a typical telescope into an interferometric array, using a pupil-plane mask. The International Deep Planet Survey (Galicher et al., 2016) used the NIRC2 Imager at the Keck Observatory, and is another example of a classical direct imaging survey.

Within the separation range of 0.05 to 0.3 arcseconds, we are estimating that our survey is more sensitive than its counterparts, by about half an order of magnitude. As discussed in Section 1.5, our scientific goal focuses on demonstrating improved sensitivity at these separations, of under 300 milliarcseconds. The potential sensitivity contrast curves encode possible improvements to our sensitivity with improvements to the forward model. Moderate resolution integral field spectroscopy is able to detect planets closer to the star than classical high-contrast imaging instruments and deeper than non-redundant masking.

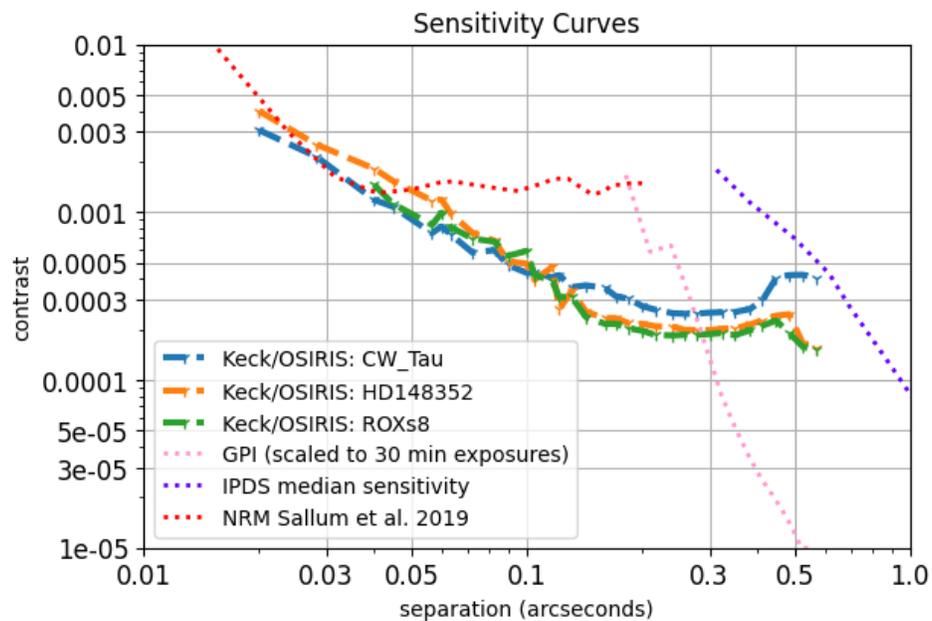
3.6 Detection of a binary companion around HD 148352

In Appendices A and B, we presented signal-to-noise maps for the target HD 148352. We observe a roughly elliptical region of high SNR, which we believe to be a binary star companion to HD 148352. This region is highlighted in Figure 3.10. The detection SNR is about 34, after all the steps of noise calibration. Figure 3.11 plots this detection with the sensitivity curves, before and after these noise normalization (including the RV CCF calibration), as well as the PSF profile. We note that the contrast of the binary is $\approx 0.0038 = 0.38\%$, while the separation from the host star is ≈ 113 milliarcseconds. We are able to detect a $< 1\%$ companion at close separation to the star (under 0.3 arcseconds), demonstrating the sensitivity of our technique.

We perform some procedures to validate this detection, as described in Section 2.12.



(a) Current Sensitivity, with all steps of noise normalization



(b) Potential Sensitivity (computed before noise calibration)

Figure 3.9: Sensitivity of our pathfinder survey with Keck/OSIRIS, compared to other direct imaging surveys. We estimate better sensitivity at separations of 50 to 300 milliarcseconds, with a potential to improve to even better contrasts.

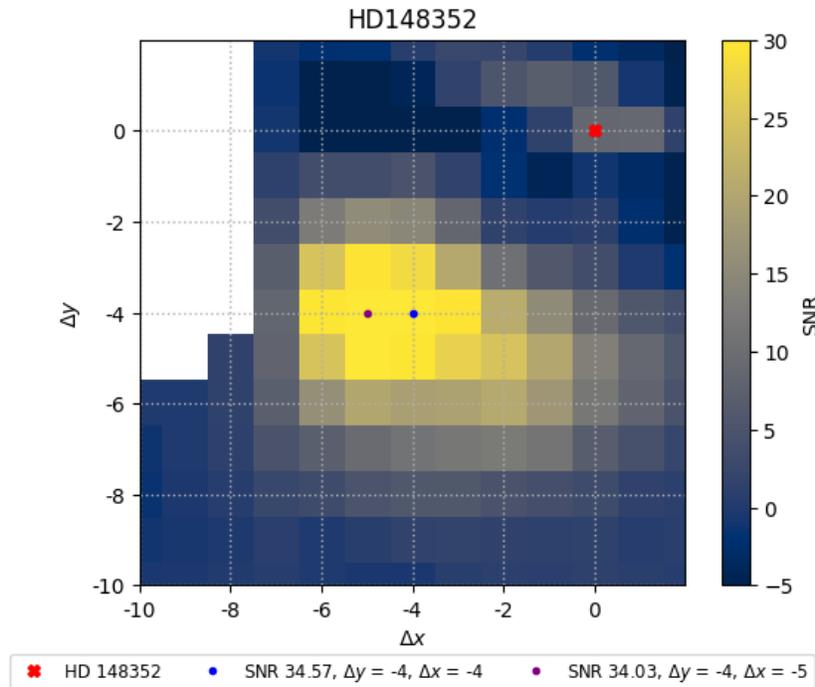


Figure 3.10: Detection of a binary companion around HD 148352, with $\text{SNR} \approx 34$ (blue and purple dots) at $\Delta y = \Delta x = -4$ spaxels from the host star (red cross). This constrains the astrometry of the companion.

Figure 3.12 plots the signal-to-noise ratios at the location of the detection blob, as a function of the RV, which is the radial velocity cross-correlation function. These SNR values do not include the noise normalization over the radial annuli, as an overall scaling factor is not relevant to the validation. For real astrophysical objects, the RV CCF should peak around $RV = 0$, which it does, because radial velocities of bound companions would not be deviated by several kilometers per second.

We incorrectly included HD 148352 as an Ophiuchus star in our survey. While it lies in the same region of the sky as the Ophiuchus star-forming cluster, it is much closer to us (< 100 pc) than the cluster (≈ 140 pc). Mamajek (2008) note that the proper motion of HD 148352 is much larger than expected for members of the Ophiuchus region, with the values being instead consistent with it being a foreground F dwarf with high proper motion.

As it is not a member of a star-forming region, but instead just an interloper that happens to be projected onto the same region of the sky, the probability of HD 148352 having a hot young companion (that can be directly imaged) is lower than we expected for a typical Ophiuchus target. Simultaneously, the probability of a

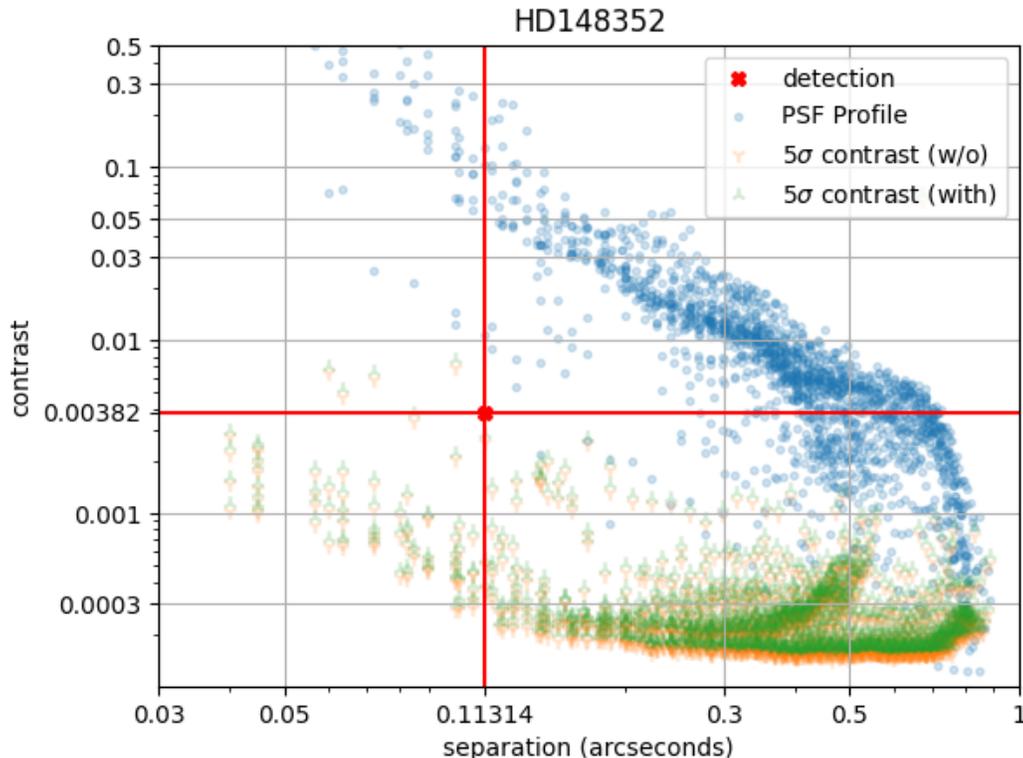


Figure 3.11: Our detection (red) overlaid on contrast curves for HD148352, similar to Figure 3.7b. Detection at 3.8% is above the sensitivity limits, while being significantly lower (roughly 1.5 orders of magnitude) than the stellar PSF profile. Our methods are able to model for diffracted starlight at separations as close as 5-6 spaxels (spatial locations).

binary companion, like an M dwarf, is increased.

In Figure 3.13, we present the results of an MCMC sampler as described in Section 2.12. we use a grid covering $T_{\text{eff}} = 2000 \rightarrow 4000$ K, $\log g = 3.5 \rightarrow 5.5$, and spin from $0 \rightarrow 50$ km/s. We estimate the temperature of the companion to be ≈ 3200 K with an RV of ≈ -12 km/s and spin of ≈ 8 km/s. As these values are based on data from one exposure, we do not report error bars on these crude estimates. Based on this T_{eff} (Morrell and Naylor, 2019; uni.edu, 2022), we believe that the binary star companion is indeed an M dwarf.

3.7 Possible $< 5\sigma$ detection in ROXs 35A data set

In SNR maps for ROXs 35A presented in Appendices A and B, we observe a blob-like positive SNR feature that is below the detection threshold of 5σ . This region is highlighted in Figure 3.14, and could be a potential detection. The detection SNR is about 4.2, after all steps of noise calibration, with a model using $T_{\text{eff}} = 1700$ K,

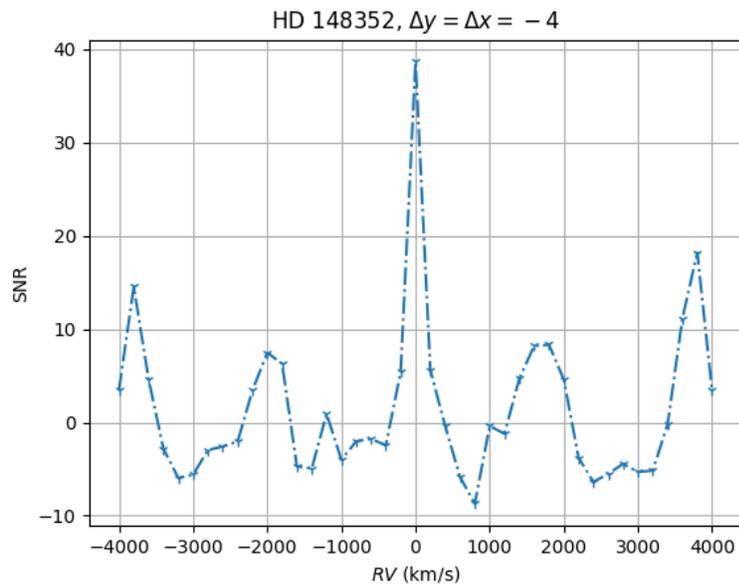


Figure 3.12: Radial velocity cross correlation function for detected binary, at $\Delta y = \Delta x = -4$ spaxels from the star. We see a peak around $RV = 0$, which is expected for a real astrophysical signal. SNR is not normalized, up to a constant scaling factor.

$\log g = 4$, and $RV = 0$. We note that these parameters are not optimized, and a spectrum model that is more consistent with the real astrophysical object (if it exists) could yield a higher SNR than the current estimate.

Figure 4.2 plots the location of this blob with the sensitivity curves and the PSF profile. Notably, if detected with a higher significance, the separation of a candidate at this location would be ≈ 82 milliarcseconds, or close to 4 pixels with the current plate scale of Keck/OSIRIS. This could be an excellent demonstration of the sensitivity of our methods, and, thus, further work is extensively discussed in Chapter 4. Currently, contrast is estimated to be $\approx 0.00194 = 1.94\%$, though changes to the model that yields a 5σ detection would vary this estimate. Figure 3.16 plots the radial velocity cross-correlation function, as described in Section 2.12. We observe a peak around $RV = 0$, which is consistent with expectations for a real astrophysical signal.

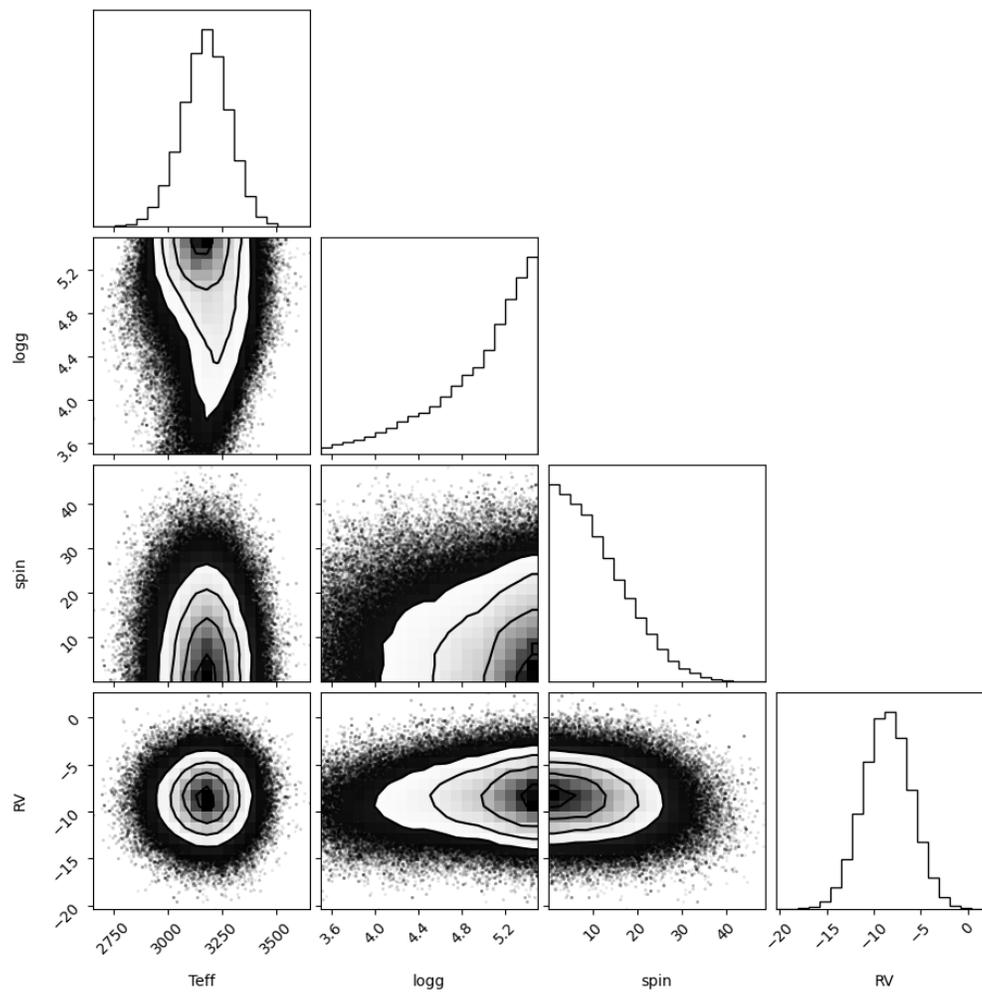


Figure 3.13: Corner plot encoding outputs from an MCMC sampler used to characterize the T_{eff} , RV , $\log g$, and spin of the detected HD 148352 binary companion. breads can be similarly used to easily characterize companions in data sets.

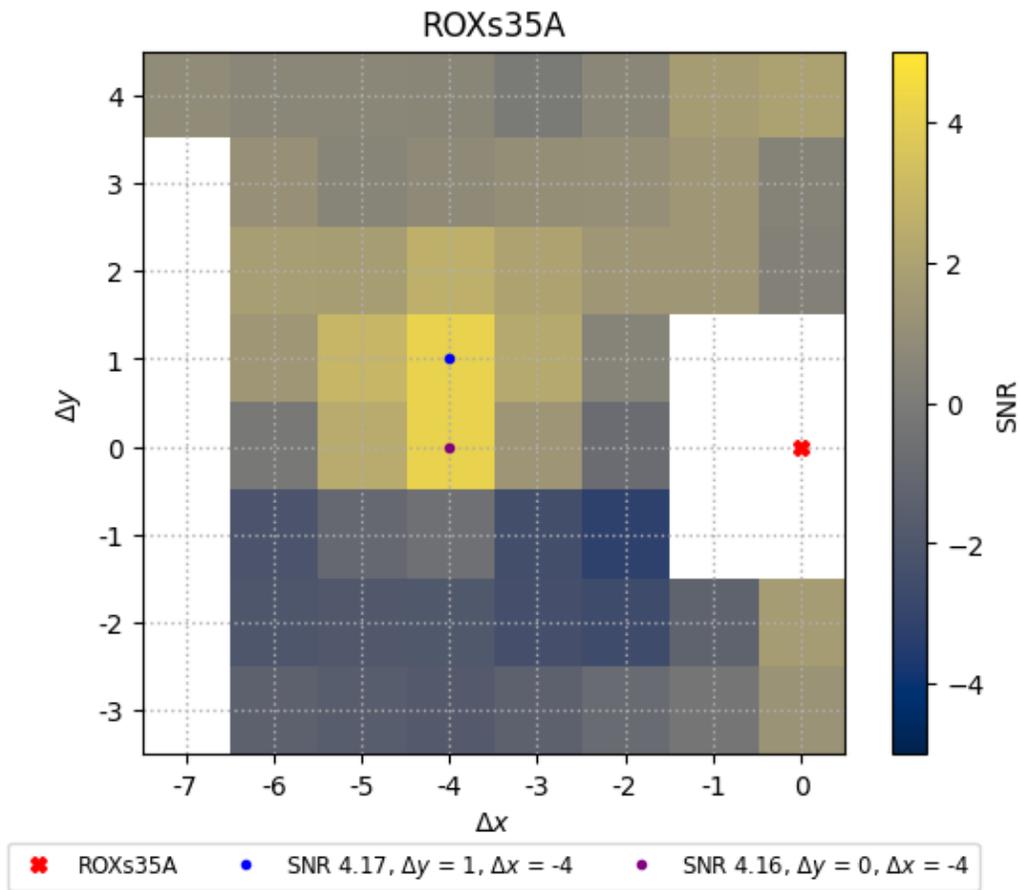


Figure 3.14: Low-significance ($\approx 4\sigma$) blob (blue and purple dots) around ROXS 35A, at $\Delta y = 0$ or 1 and $\Delta x = -4$ spaxels from the host star (red cross).

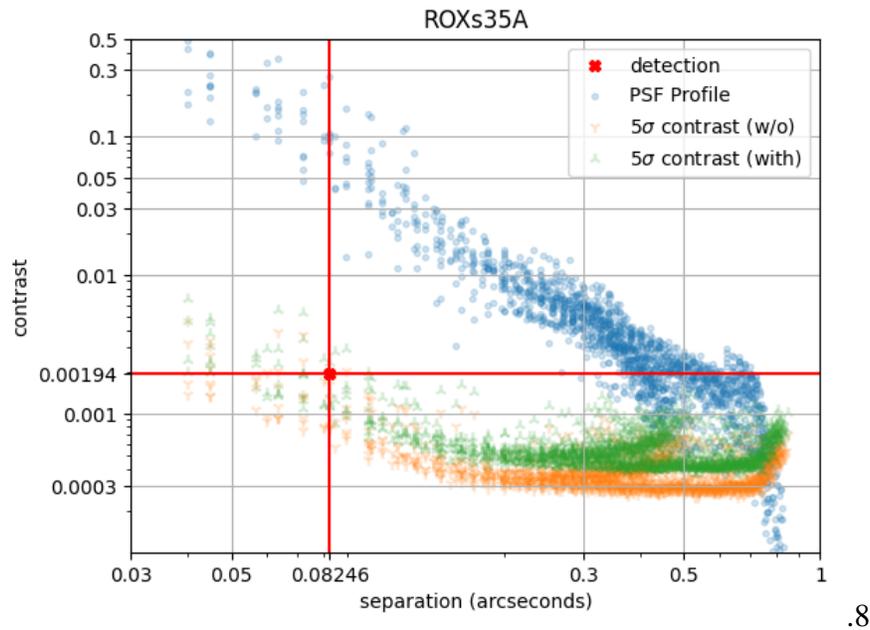


Figure 3.15: Low significance blob (red) overlaid on contrast curves for ROXS 35A, similar to Figure 3.7b. If confirmed, this detection at just ≈ 4 spaxels from the star would demonstrate sensitivity capabilities of our technique.

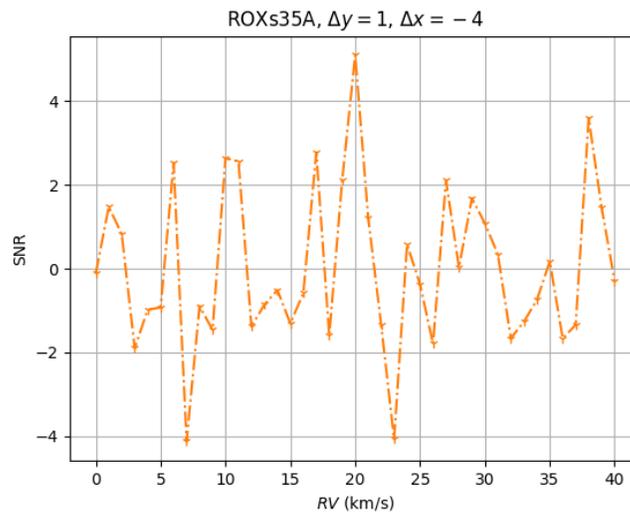


Figure 3.16: Radial velocity cross correlation function for detected binary, at $\Delta y = \Delta x = -4$ spaxels from the star. We see a peak around $RV = 0$, which is expected for a real astrophysical signal. However, other peaks are appreciably high relative to the highest peak at $RV = 0$. SNR is not normalized, up to a constant scaling factor.

*Chapter 4***FURTHER WORK****4.1 Further Optimization of Model**

While there could be several approaches to improving our model, we will discuss two potential steps, which are relevant to the discussion in previous Chapters. First, we will reduce data from all the targets using the RV CCF calibration. In Section 3.4, we present results for 14 targets, out of the total of 23 we observed. As visible from Figures 3.6 and 3.8, the normalization yields varied results for different targets. Reducing the remaining targets could reveal that the normalization is not optimized or does not work well for some particular targets. Optimizing the RV CCF calibration would include using a different range of radial velocity values or a finer grid. While our sensitivity is improved after the RV CCF calibration, we can note from Figures 3.7a and 3.7b that more improvements can be made to push us closer to the estimated potential sensitivity limits.

The second optimization we discuss in detail is related to the parameters that described our forward model. These are hard-coded values that define some form of threshold limit or model characteristic. Examples of these values include the threshold deviation for bad pixel identification, the specific `BT-Set t1` model to use in SNR map reductions, thresholds used to define the intensity of masking for pixel bleeding issues, sizes or shapes of apertures used to compute stellar spectra, or the factor used to optimize placement of spline nodes.

Using 4.1, we specifically discuss the effects of changing the number of spline nodes in the forward model, as a nominal example. As described in Section 2.6, while generating the diffracted starlight model, we use a low-degree polynomial spline interpolation to model for a non-uniform continuum. The nodes represent the constraints, as well as the number of different functional pieces, on this piece-wise fit. The number of nodes determines the complexity of the fit, which affects the capacity of the model to fit the star-light, while also affecting the probability of over-fitting the planetary signal.

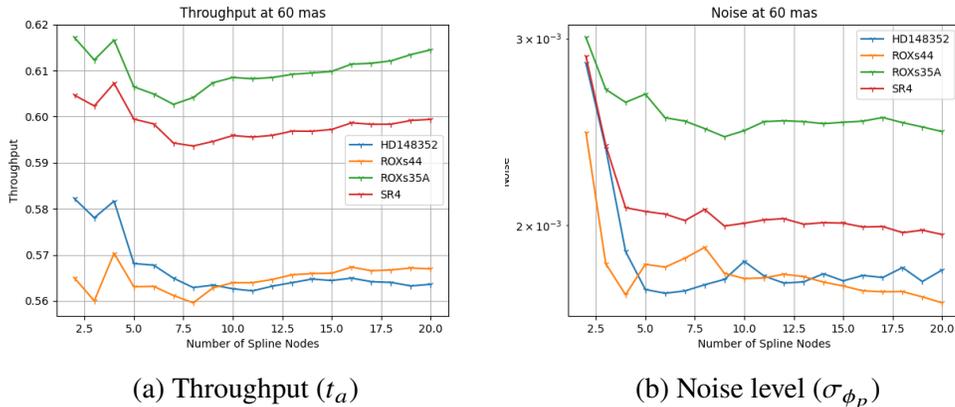


Figure 4.1: An initial example of setting the number of spline nodes to optimize our forward model. For four targets, we plot (a) throughputs and (b) uncertainty levels, which is directly proportional to the sensitivity, by a factor of 5 (threshold SNR level), as a function of splines nodes in our model. These values are computed at 16 different locations at a separation of 60 milliarcseconds. We optimize the model by choosing setting this hard-coded parameter of our model to be one that increases throughput, while lowering noise levels.

4.2 Future Observations, and Investigation of $< 5\sigma$ detections

As mentioned in Section 2.1, we have been awarded more time on the OSIRIS instrument at the Keck Observatory for the two upcoming semester cycles. We aim to observe roughly 10 more targets each in the Ophiuchus and Taurus regions. Observing more targets increases the probability of detecting an exoplanet or brown dwarf companion, as well as demonstrating the sensitivity of our methods over a wide array of stars of different spectral types and brightness.

In Section 3.7, we described a $\approx 4\sigma$ feature that we detect in our combined signal-to-noise ratio map, generated using data from observations around ROXs 35A. This is a nominal example of positive SNR blobs, which do not cross the threshold of 5σ , that we observe in maps presented in Appendices A and B. Investigating these $3-5\sigma$ features is important, not only because they could encode the discovery of a new companion, but also because it is directly relevant to our scientific goal of estimating and demonstrating the sensitivity of high-resolution integral field spectrography.

We compute signal-to-noise ratio with a set of given planet characteristics, using $T_{\text{eff}} = 1700 \text{ K}$ or 1800 K , $\log g = 4$, and $RV = 0$. Varying these values, as well as further optimizing the model as in Section 4.1, can yield a model that is more consistent with the astrophysical signal, resulting in a different SNR. This, combined with validation checks described in Section 2.12, enable us to select targets (where

we see these $< 5\sigma$ features) that we might want to observe more as part of the remaining time we have on Keck/OSIRIS. For example, we will collect more data around ROXs 35A to investigate the feature we see at $\Delta x = -4, \Delta y = 1$.

Another target worth revisiting would be AB Aurigae. Currie et al. (2022) proposes a Jovian protoplanet candidate around the star, embedded in a planet-forming disk, which can yield significant science relevant to planet formation theories. As this candidate is at a wide separation (about 30 pixels at our plate scale of Keck/OSIRIS) and this finding was published after our observation run, the proposed location of the candidate is not in our field of view for most of our exposures. Thus, we do not have enough combined sensitivity to perform a follow-up detection. Collecting more data (keeping the proposed location in the field of view) could allow us to test the finding of Currie et al. (2022), and, if it exists, confirm the detection of a Jovian protoplanet.

4.3 Probing Characterization Capabilities

There are several linear (ϕ) and non-linear (ψ) parameters associated with our forward model. Our approach should be able to accurately and precisely estimate the characteristics of a planet signal. These characteristics include the planet's astrometry (spatial location), radial velocity, contrast or brightness, T_{eff} , spin, and $\log g$. Testing how well our technique is able to analyze planets is important, especially in case we detect a new companion. Section 2.10 describes how a fake planet signal can be simulated in our data cubes, while Section 2.12 details techniques to characterize detections.

Our throughput calculations in Section 2.10 already identify the amount of flux our model is able to recover from an injected simulated planet signal. Thus, this determines how well our algorithm can estimate the brightness or contrast of the planet (which is a linear parameter in our model). We can similarly employ injection-and-recovery routines to probe other characterization capabilities of our approach, particularly the non-linear parameters. Injecting a planet at a given spatial location, and then estimating its astrometry from a combined SNR map (as in Section 2.8) can quantify how well we can estimate the location of a companion. Similarly, injecting a planet signal with a specific Doppler shift corresponding to a specific radial velocity, and recovering the signal with a grid of models with varying radial velocities (using a grid search or an MCMC solver such as the one described in Section 2.12) can quantify our ability to recover the radial velocity of companions.

Parameters including T_{eff} , spin, RV , and $\log g$ are used to construct a theoretical spectrum from a BT-Settl model. We can simulate a fake planet by injecting a planet with a set combination of these parameters, and then attempt to recover this signal by varying the theoretical spectra we use to generate the forward model. Figure ?? demonstrates this for a single parameter, effective temperature T_{eff} . We independently inject planets of 5 varying T_{eff} , represented by different colors, at 16 locations each with separation of 100 milliarcseconds. Then, we attempted to recover planet flux ϕ_p at these locations using several forward models ($T_{\text{eff}} = 1000 - 2000$ K, in steps of 100 K) constructed using BT-Settl of different T_{eff} . For a specific injected T_{eff} , we want the SNR obtained using a model with the same T_{eff} to be highest among all models.

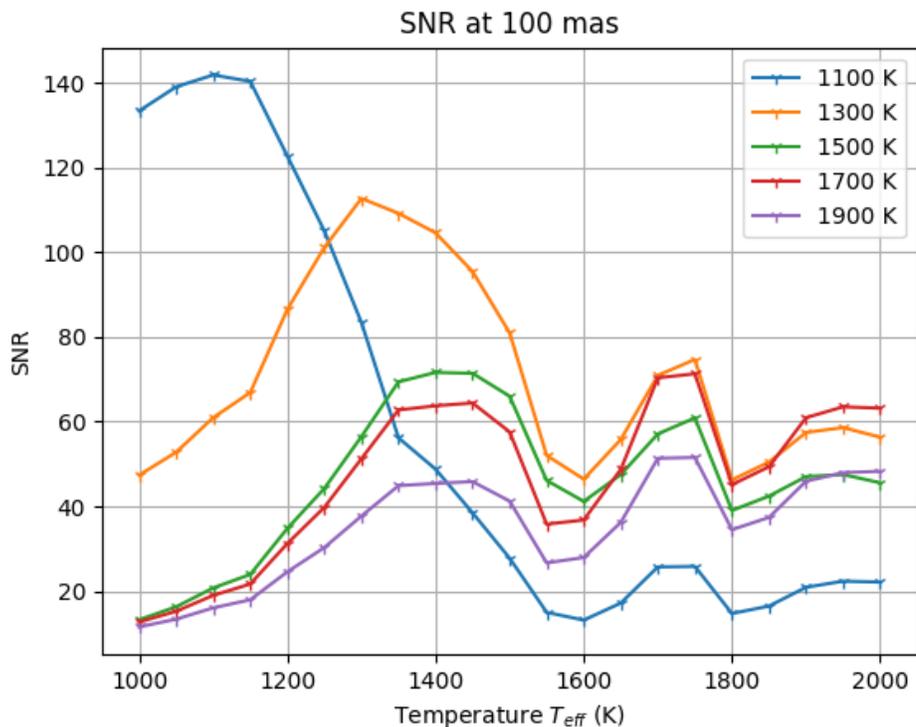


Figure 4.2: An example of how to quantify the characterization capabilities of our model. Each color represents the T_{eff} of the simulated planet signal. If an SNR curve peaks around the same T_{eff} , we can claim that our method can accurately measure the effective temperature of a detected companion. In this example, we note that planets with $T_{\text{eff}} = 1100$ or 1300 K can be well-characterized (as the SNR curves peak around the same values), while there seems to be some degeneracy between planet signals with $T_{\text{eff}} = 1500, 1700,$ and 1900 K, which results in an SNR curve with two significant peaks.

4.4 Investigating Capabilities with the James Webb Space Telescope and beyond

Near-Infrared Spectrograph, or NIRSpec, is an infrared integral field spectrograph on board the James Webb Space Telescope, operating over 0.6 to 5 microns, with science data to be released within the next year. An approved early release discretionary program involves high contrast imaging data taken using instruments including NIRSpec ¹. The breads has been tested to be compatible with available simulations of NIRSpec data, using the instrument class `breads.instruments.jwstnirpsec`.

Techniques developed in this thesis will be applicable to this new integral field spectrograph on the JWST, as well as in-development IFS instruments on the extremely large telescopes. These telescopes are expected to become the next generation of planet detection facilities, pushing down sensitivity limits for direct imaging.

¹<https://www.stsci.edu/jwst/science-execution/approved-ers-programs>

BIBLIOGRAPHY

- Allard, F. (Jan. 1, 2014). “The BT-Settl Model Atmospheres for Stars, Brown Dwarfs and Planets”. In: 299. Conference Name: Exploring the Formation and Evolution of Planetary Systems ADS Bibcode: 2014IAUS..299..271A, pp. 271–272. DOI: 10.1017/S1743921313008545. URL: <https://ui.adsabs.harvard.edu/abs/2014IAUS..299..271A> (visited on 05/16/2022).
- Allard, F., D. Homeier, and B. Freytag (Nov. 24, 2010). “Model Atmospheres From Very Low Mass Stars to Brown Dwarfs”. In: DOI: 10.48550/arXiv.1011.5405. URL: <https://arxiv.org/abs/1011.5405v1> (visited on 05/16/2022).
- Allington-Smith, Jeremy (June 1, 2006). “Basic principles of integral field spectroscopy”. In: *New Astronomy Reviews*. Integral Field Spectroscopy: Techniques and Data Production 50.4, pp. 244–251. ISSN: 1387-6473. DOI: 10.1016/j.newar.2006.02.024. URL: <https://www.sciencedirect.com/science/article/pii/S1387647306000157> (visited on 04/28/2022).
- Boley, Aaron C. (Mar. 2009). “THE TWO MODES OF GAS GIANT PLANET FORMATION”. In: *The Astrophysical Journal* 695.1. Publisher: American Astronomical Society, pp. L53–L57. ISSN: 0004-637X. DOI: 10.1088/0004-637X/695/1/L53. URL: <https://doi.org/10.1088/0004-637x/695/1/L53> (visited on 04/09/2022).
- Boss, Alan P. (June 20, 1997). “Giant Planet Formation by Gravitational Instability”. In: *Science* 276.5320. Publisher: American Association for the Advancement of Science, pp. 1836–1839. DOI: 10.1126/science.276.5320.1836. URL: <https://www.science.org/doi/10.1126/science.276.5320.1836> (visited on 04/09/2022).
- Bowler, Brendan P. (Oct. 1, 2016). “Imaging Extrasolar Giant Planets”. In: *Publications of the Astronomical Society of the Pacific* 128.968, p. 102001. ISSN: 0004-6280, 1538-3873. DOI: 10.1088/1538-3873/128/968/102001. arXiv: 1605.02731. URL: <http://arxiv.org/abs/1605.02731> (visited on 04/08/2022).
- Bradley, Larry et al. (Mar. 2021). *astropy/photutils: 1.1.0*. Version 1.1.0. DOI: 10.5281/zenodo.4624996. URL: <https://doi.org/10.5281/zenodo.4624996>.
- Cheetham, Anthony C. et al. (Oct. 2015). “MAPPING THE SHORES OF THE BROWN DWARF DESERT. IV. OPHIUCHUS”. In: *The Astrophysical Journal* 813.2. Publisher: American Astronomical Society, p. 83. ISSN: 0004-637X. DOI: 10.1088/0004-637X/813/2/83. URL: <https://doi.org/10.1088/0004-637x/813/2/83> (visited on 04/15/2022).

- Currie, Thayne et al. (Apr. 4, 2022). “Images of embedded Jovian planet formation at a wide separation around AB Aurigae”. In: *Nature Astronomy*. Publisher: Nature Publishing Group, pp. 1–9. ISSN: 2397-3366. DOI: 10.1038/s41550-022-01634-x. URL: <https://www.nature.com/articles/s41550-022-01634-x> (visited on 05/22/2022).
- Dodson-Robinson, Sarah E. et al. (Dec. 10, 2009). “THE FORMATION MECHANISM OF GAS GIANTS ON WIDE ORBITS”. In: *The Astrophysical Journal* 707.1, pp. 79–88. ISSN: 0004-637X, 1538-4357. DOI: 10.1088/0004-637X/707/1/79. URL: <https://iopscience.iop.org/article/10.1088/0004-637X/707/1/79> (visited on 04/28/2022).
- ESA (Jan. 1, 1997). “The HIPPARCOS and TYCHO catalogues. Astrometric and photometric star catalogues derived from the ESA HIPPARCOS Space Astrometry Mission”. In: *ESA Special Publication* 1200. ADS Bibcode: 1997ESASP1200....E. ISSN: 1609-042X. URL: <https://ui.adsabs.harvard.edu/abs/1997ESASP1200....E> (visited on 05/19/2022).
- Foreman-Mackey, Daniel et al. (Feb. 25, 2013). “emcee: The MCMC Hammer”. In: *Publications of the Astronomical Society of the Pacific* 125.925. Publisher: IOP Publishing, p. 306. ISSN: 1538-3873. DOI: 10.1086/670067. URL: <https://iopscience.iop.org/article/10.1086/670067/meta> (visited on 05/06/2022).
- Freytag, B. et al. (Apr. 2010). “The role of convection, overshoot, and gravity waves for the transport of dust in M dwarf and brown dwarf atmospheres”. In: *Astronomy and Astrophysics* 513, A19. ISSN: 0004-6361. DOI: 10.1051/0004-6361/200913354. URL: <https://ui.adsabs.harvard.edu/abs/2010A&A...513A..19F/abstract> (visited on 05/16/2022).
- Fulton, Benjamin J. et al. (July 2021). “California Legacy Survey. II. Occurrence of Giant Planets beyond the Ice Line”. In: *The Astrophysical Journal Supplement Series* 255.1. Publisher: American Astronomical Society, p. 14. ISSN: 0067-0049. DOI: 10.3847/1538-4365/abfcc1. URL: <https://doi.org/10.3847/1538-4365/abfcc1> (visited on 04/29/2022).
- Galicher, R. et al. (Oct. 1, 2016). “The International Deep Planet Survey - II. The frequency of directly imaged giant exoplanets with stellar mass”. In: *Astronomy & Astrophysics* 594. Publisher: EDP Sciences, A63. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201527828. URL: <https://www.aanda.org/articles/aa/abs/2016/10/aa27828-15/aa27828-15.html> (visited on 05/24/2022).
- Ghez, A. M., G. Neugebauer, and K. Matthews (Nov. 1, 1993). “The Multiplicity of T Tauri Stars in the Star Forming Regions Taurus-Auriga and Ophiuchus-Scorpius: A 2.2 Micron Speckle Imaging Survey”. In: *The Astronomical Journal* 106. ADS Bibcode: 1993AJ....106.2005G, p. 2005. ISSN: 0004-6256. DOI: 10.1086/116782. URL: <https://ui.adsabs.harvard.edu/abs/1993AJ...106.2005G> (visited on 05/17/2022).

- Hoeijmakers, H. Jens et al. (Aug. 2018). “Atomic iron and titanium in the atmosphere of the exoplanet KELT-9b”. In: *Nature* 560.7719, pp. 453–455. ISSN: 1476-4687. DOI: [10.1038/s41586-018-0401-y](https://doi.org/10.1038/s41586-018-0401-y). URL: <http://dx.doi.org/10.1038/s41586-018-0401-y>.
- Husser, T.-O. et al. (May 2013). “A new extensive library of PHOENIX stellar atmospheres and synthetic spectra”. In: *Astronomy and Astrophysics* 553, A6. ISSN: 0004-6361. DOI: [10.1051/0004-6361/201219058](https://doi.org/10.1051/0004-6361/201219058). URL: <https://ui.adsabs.harvard.edu/abs/2013A&A...553A...6H/abstract> (visited on 05/16/2022).
- Keck Data Reduction Pipelines (Apr. 17, 2022). *Keck OSIRIS Spectroscopic Data Reduction Pipeline*. original-date: 2015-07-08T21:06:20Z. URL: <https://github.com/Keck-DataReductionPipelines/OSIRISDRP> (visited on 04/29/2022).
- Kenyon, Scott J. and Lee Hartmann (Nov. 1995). “Pre-Main-Sequence Evolution in the Taurus-Auriga Molecular Cloud”. In: *The Astrophysical Journal Supplement Series* 101, p. 117. ISSN: 0067-0049, 1538-4365. DOI: [10.1086/192235](https://doi.org/10.1086/192235). URL: <http://adsabs.harvard.edu/doi/10.1086/192235> (visited on 04/14/2022).
- Konopacky, Q. M. et al. (Mar. 2013). “Detection of Carbon Monoxide and Water Absorption Lines in an Exoplanet Atmosphere”. In: *Science* 339.6126, pp. 1398–1401. ISSN: 1095-9203. DOI: [10.1126/science.1232003](https://doi.org/10.1126/science.1232003). URL: <http://dx.doi.org/10.1126/science.1232003>.
- Lafrenière, David et al. (Apr. 1, 2009). “HST/NICMOS Detection of HR 8799 b in 1998”. In: *The Astrophysical Journal* 694. ADS Bibcode: 2009ApJ...694L.148L, pp. L148–L152. ISSN: 0004-637X. DOI: [10.1088/0004-637X/694/2/L148](https://doi.org/10.1088/0004-637X/694/2/L148). URL: <https://ui.adsabs.harvard.edu/abs/2009ApJ...694L.148L> (visited on 04/29/2022).
- Larkin, James et al. (June 1, 2006). “OSIRIS: a diffraction limited integral field spectrograph for Keck”. In: 6269. Conference Name: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series ADS Bibcode: 2006SPIE.6269E..1AL, 62691A. DOI: [10.1117/12.672061](https://doi.org/10.1117/12.672061). URL: <https://ui.adsabs.harvard.edu/abs/2006SPIE.6269E..1AL> (visited on 04/28/2022).
- Lissauer, Jack J. (Feb. 1, 1987). “Timescales for planetary accretion and the structure of the protoplanetary disk”. In: *Icarus* 69.2, pp. 249–265. ISSN: 0019-1035. DOI: [10.1016/0019-1035\(87\)90104-7](https://doi.org/10.1016/0019-1035(87)90104-7). URL: <https://www.sciencedirect.com/science/article/pii/0019103587901047> (visited on 04/09/2022).
- Lovis, C. and D. Fischer (Dec. 1, 2010). *Radial Velocity Techniques for Exoplanets*. Pages: 27-53 Publication Title: Exoplanets ADS Bibcode: 2010exop.book...27L. URL: <https://ui.adsabs.harvard.edu/abs/2010exop.book...27L> (visited on 04/04/2022).

- Macintosh, Bruce et al. (Sept. 2, 2014). “First light of the Gemini Planet Imager”. In: *Proceedings of the National Academy of Sciences* 111.35. Publisher: Proceedings of the National Academy of Sciences, pp. 12661–12666. DOI: 10.1073/pnas.1304215111. URL: <https://www.pnas.org/doi/10.1073/pnas.1304215111> (visited on 04/29/2022).
- Mamajek, E.E. (2008). “On the distance to the Ophiuchus star-forming region”. In: *Astronomische Nachrichten* 329.1. _eprint: <https://onlinelibrary.wiley.com/doi/10.1002/asna.200710827>, pp. 10–14. DOI: <https://doi.org/10.1002/asna.200710827>.
- Marois, Christian, René Doyon, et al. (Jan. 2000). “Efficient Speckle Noise Attenuation in Faint Companion Imaging”. In: *Publications of the Astronomical Society of the Pacific* 112.767. Publisher: IOP Publishing, pp. 91–96. ISSN: 1538-3873. DOI: 10.1086/316492. URL: <https://doi.org/10.1086/316492> (visited on 04/29/2022).
- Marois, Christian, David Lafreniere, et al. (Apr. 2006). “Angular Differential Imaging: A Powerful High-Contrast Imaging Technique”. In: *The Astrophysical Journal* 641.1. Publisher: American Astronomical Society, pp. 556–564. ISSN: 0004-637X. DOI: 10.1086/500401. URL: <https://doi.org/10.1086/500401> (visited on 04/29/2022).
- Mayor, Michel and Didier Queloz (Nov. 1995). “A Jupiter-mass companion to a solar-type star”. In: *Nature* 378.6555, pp. 355–359. ISSN: 1476-4687. DOI: 10.1038/378355a0. URL: <https://www.nature.com/articles/378355a0> (visited on 04/03/2022).
- Morrell, Sam and Tim Naylor (Aug. 2019). “Exploring the M-dwarf Luminosity–Temperature–Radius relationships using Gaia DR2”. In: *Monthly Notices of the Royal Astronomical Society* 489.2. _eprint: <https://academic.oup.com/mnras/article-pdf/489/2/2615/29809298/stz2242.pdf>, pp. 2615–2633. ISSN: 0035-8711. DOI: 10.1093/mnras/stz2242. URL: <https://doi.org/10.1093/mnras/stz2242>.
- NASA Astrophysics Division (2022). *Exoplanet Exploration: Planets Beyond our Solar System*. URL: <https://exoplanets.nasa.gov/> (visited on 04/03/2022).
- Nielsen, Eric L. et al. (June 12, 2019). “The Gemini Planet Imager Exoplanet Survey: Giant Planet and Brown Dwarf Demographics from 10 to 100 au”. In: *The Astronomical Journal* 158.1. Publisher: IOP Publishing, p. 13. ISSN: 1538-3881. DOI: 10.3847/1538-3881/ab16e9. URL: <https://iopscience.iop.org/article/10.3847/1538-3881/ab16e9/meta> (visited on 04/28/2022).
- Pollack, James B. et al. (Nov. 1, 1996). “Formation of the Giant Planets by Concurrent Accretion of Solids and Gas”. In: *Icarus* 124. ADS Bibcode: 1996Icar..124..62P, pp. 62–85. ISSN: 0019-1035. DOI: 10.1006/icar.1996.0190. URL: <https://ui.adsabs.harvard.edu/abs/1996Icar..124..62P> (visited on 04/28/2022).

- Rousselot, P. et al. (Feb. 2000). “Night-sky spectral atlas of OH emission lines in the near-infrared”. In: *Astronomy and Astrophysics* 354, pp. 1134–1150. ISSN: 0004-6361. URL: <https://ui.adsabs.harvard.edu/abs/2000A&A...354.1134R/abstract> (visited on 05/07/2022).
- Ruffio, Jean-Baptiste, Quinn M. Konopacky, et al. (Oct. 2021). “Deep exploration of the planets HR 8799 b, c, and d with moderate resolution spectroscopy”. In: *Submitted to AJ, preprint*.
- Ruffio, Jean-Baptiste, Bruce Macintosh, et al. (Oct. 2019). “Radial Velocity Measurements of HR 8799 b and c with Medium Resolution Spectroscopy”. In: *The Astronomical Journal* 158.5, p. 200. ISSN: 1538-3881. DOI: 10.3847/1538-3881/ab4594. URL: <http://dx.doi.org/10.3847/1538-3881/ab4594>.
- Sallum, Steph and Andy Skemer (Jan. 2019). “Comparing nonredundant masking and filled-aperture kernel phase for exoplanet detection and characterization”. In: *Journal of Astronomical Telescopes, Instruments, and Systems* 5.1. Publisher: SPIE, p. 018001. ISSN: 2329-4124, 2329-4221. DOI: 10.1117/1.JATIS.5.1.018001. URL: <https://www.spiedigitallibrary.org/journals/Journal-of-Astronomical-Telescopes-Instruments-and-Systems/volume-5/issue-1/018001/Comparing-nonredundant-masking-and-filled-aperture-kernel-phase-for-exoplanet/10.1117/1.JATIS.5.1.018001.full> (visited on 04/29/2022).
- uni.edu (2022). *Spectral Types*. URL: <https://sites.uni.edu/morgans/astro/course/Notes/section2/spectraltemps.html> (visited on 05/20/2022).
- Wang, Jason J. et al. (Oct. 2018). “Dynamical Constraints on the HR 8799 Planets with GPI”. In: *The Astronomical Journal* 156.5, p. 192. ISSN: 1538-3881. DOI: 10.3847/1538-3881/aae150. URL: <http://dx.doi.org/10.3847/1538-3881/aae150>.
- Winn, Joshua N. (Jan. 1, 2010). *Transits and Occultations*. Publication Title: arXiv e-prints ADS Bibcode: 2010arXiv1001.2010W Type: article. URL: <https://ui.adsabs.harvard.edu/abs/2010arXiv1001.2010W> (visited on 04/04/2022).
- Wolszczan, A. and D. A. Frail (Jan. 1992). “A planetary system around the millisecond pulsar PSR1257 + 12”. In: *Nature* 355.6356, pp. 145–147. ISSN: 1476-4687. DOI: 10.1038/355145a0. URL: <https://www.nature.com/articles/355145a0> (visited on 04/03/2022).

*Appendix A***APPENDIX A: COMBINED SIGNAL-TO-NOISE RATIO MAPS
BEFORE RV CCF NORMALIZATION**

Included below are two combined SNR maps for each of the 23 targets, one before and one after noise normalization (excluding RV CCF normalization). These maps are discussed in detail in Section 3.1.

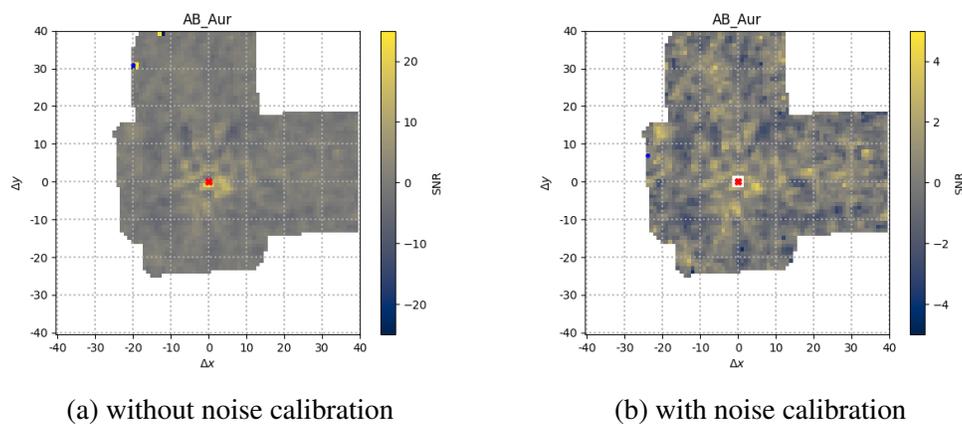


Figure A.1: Signal-to-noise ratio maps for AB_Aur

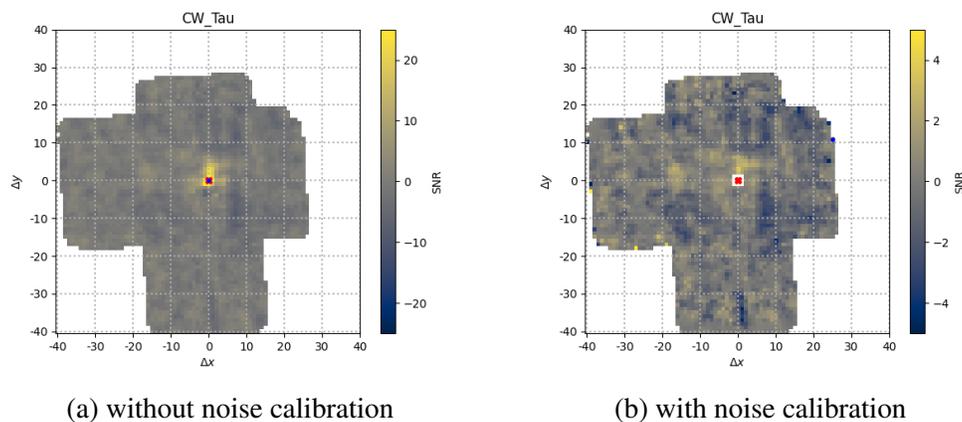


Figure A.2: Signal-to-noise ratio maps for CW_Tau

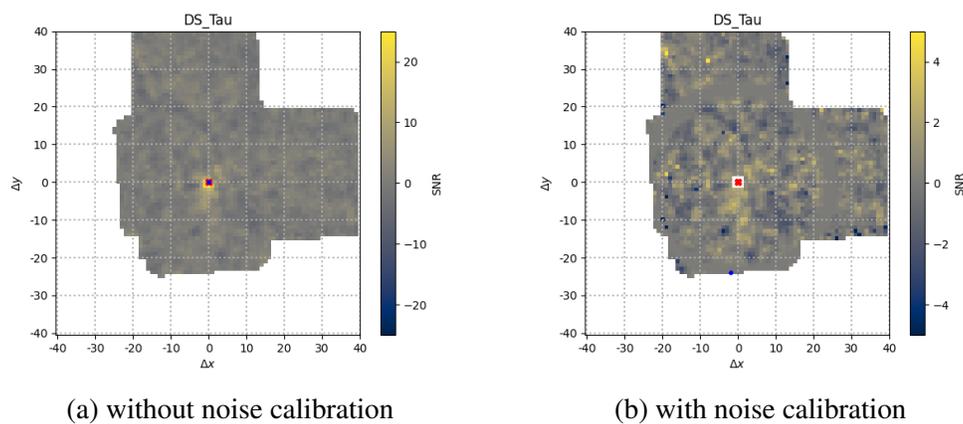


Figure A.3: Signal-to-noise ratio maps for DS_Tau

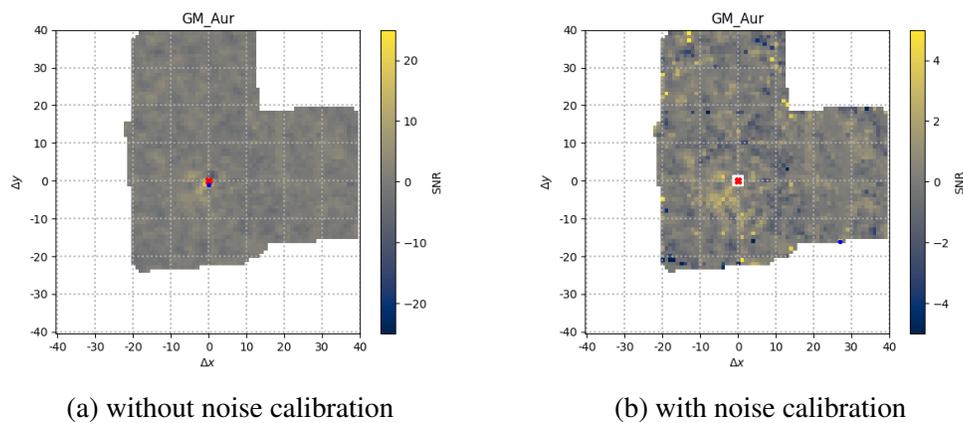


Figure A.4: Signal-to-noise ratio maps for GM_Aur

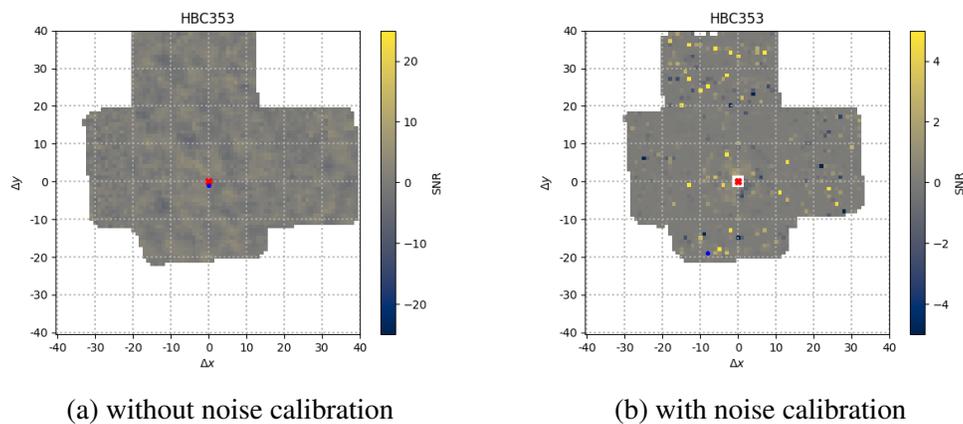


Figure A.5: Signal-to-noise ratio maps for HBC353

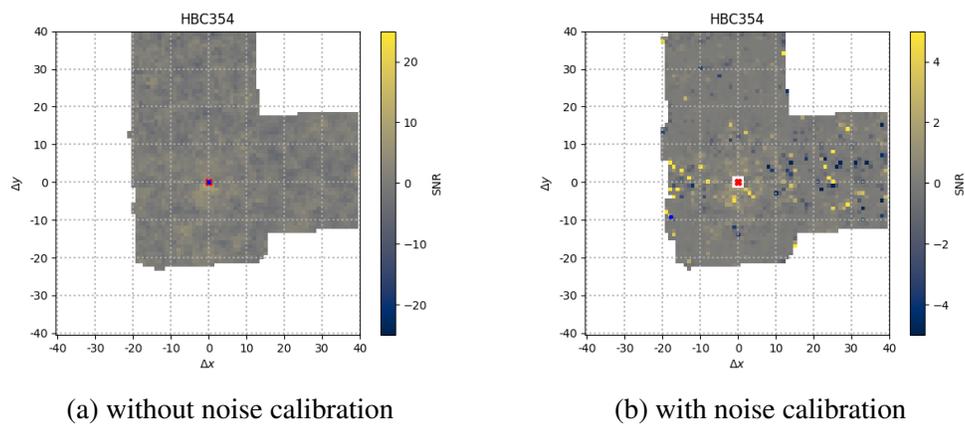


Figure A.6: Signal-to-noise ratio maps for HBC354

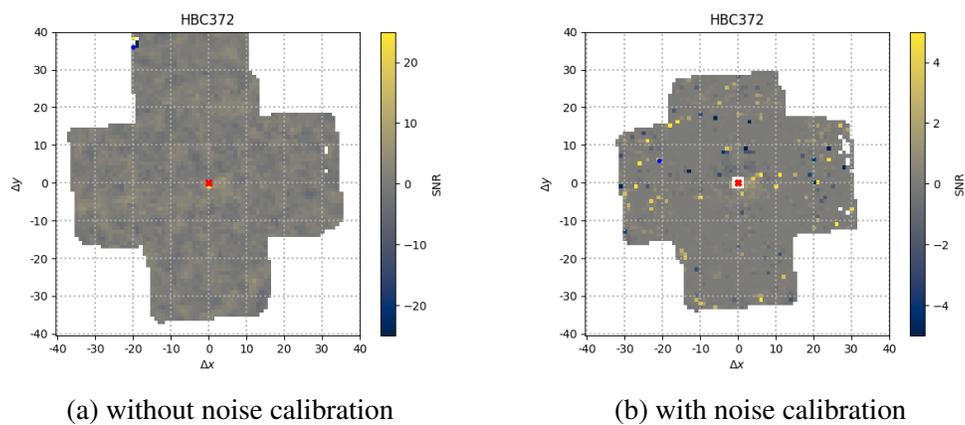


Figure A.7: Signal-to-noise ratio maps for HBC372

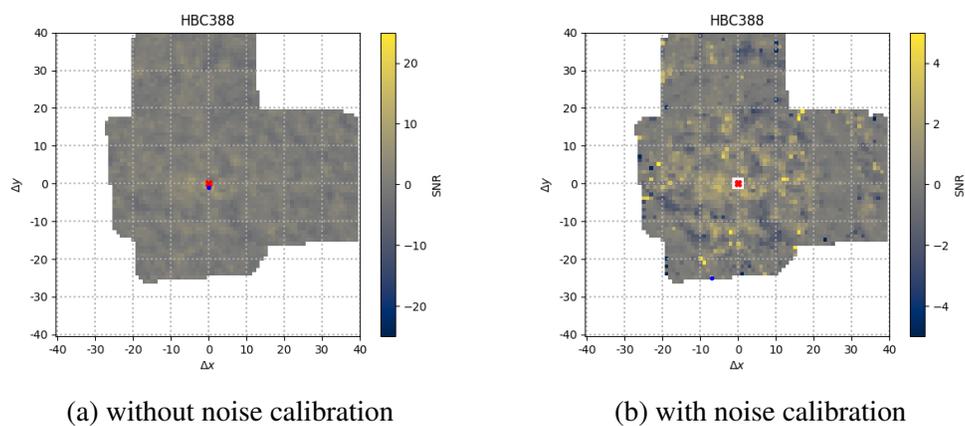


Figure A.8: Signal-to-noise ratio maps for HBC388

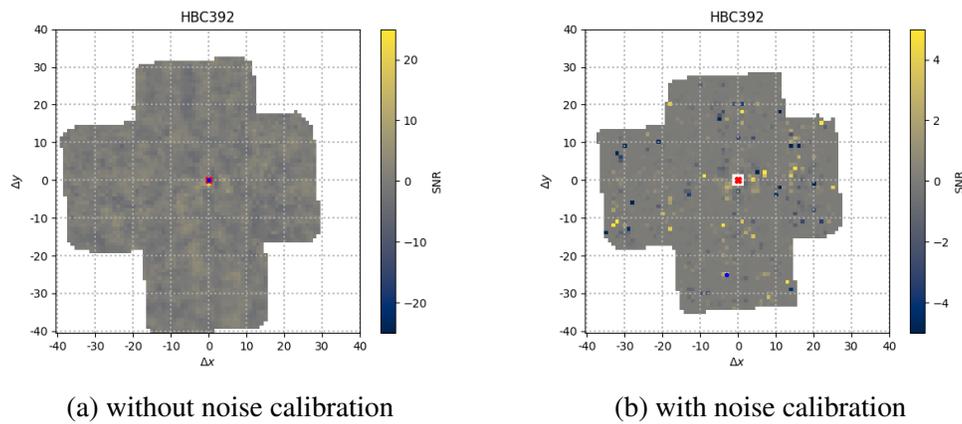


Figure A.9: Signal-to-noise ratio maps for HBC392

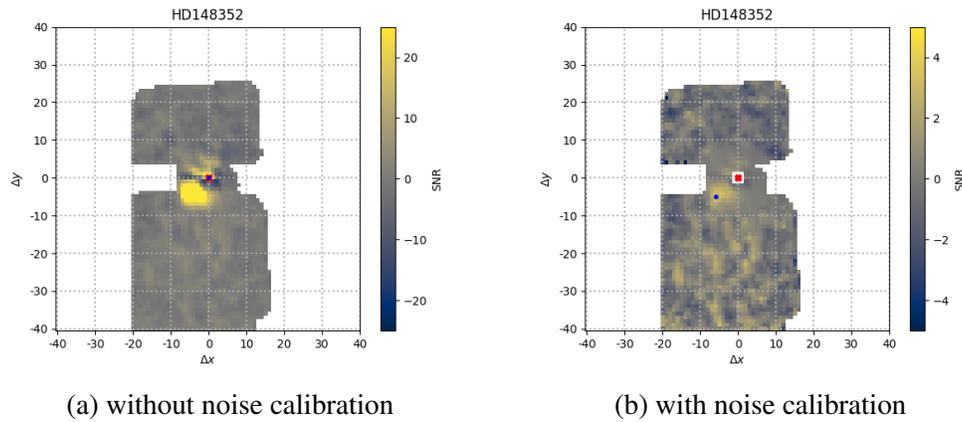


Figure A.10: Signal-to-noise ratio maps for HD148352

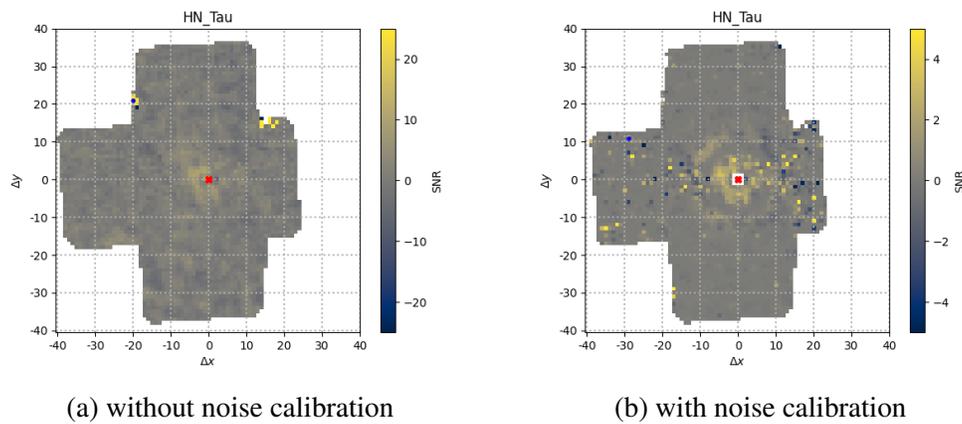


Figure A.11: Signal-to-noise ratio maps for HN_Tau

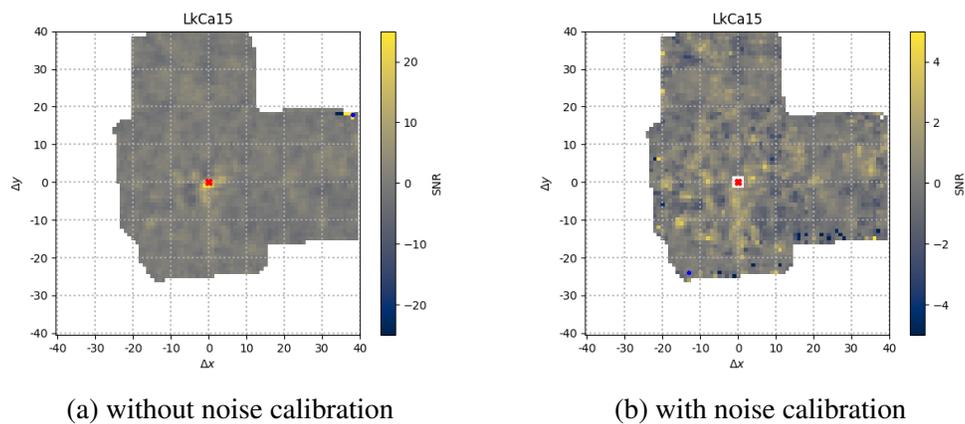


Figure A.12: Signal-to-noise ratio maps for LkCa15

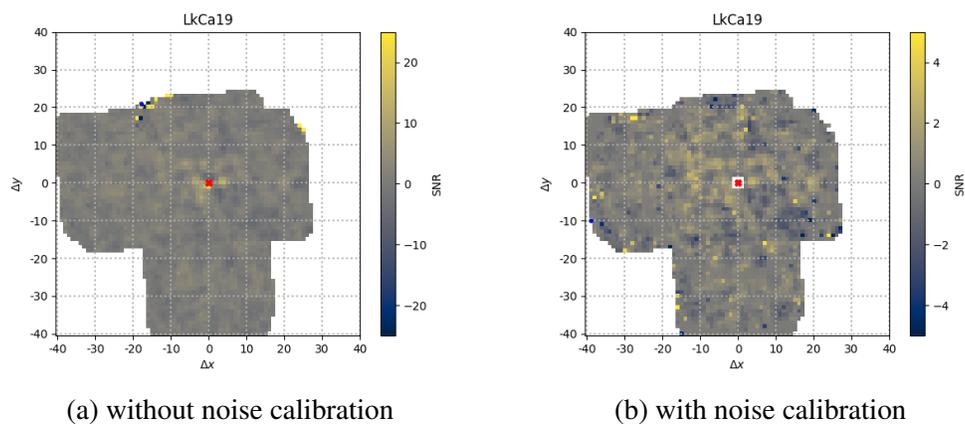


Figure A.13: Signal-to-noise ratio maps for LkCa19

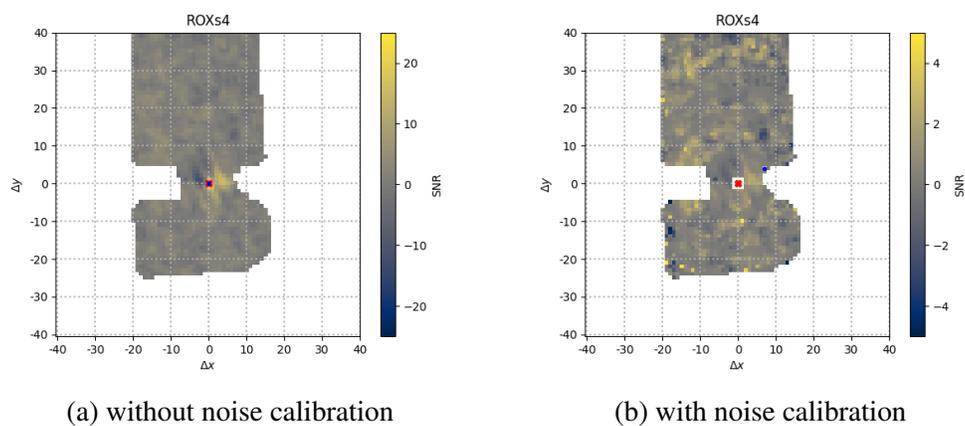


Figure A.14: Signal-to-noise ratio maps for ROXs4

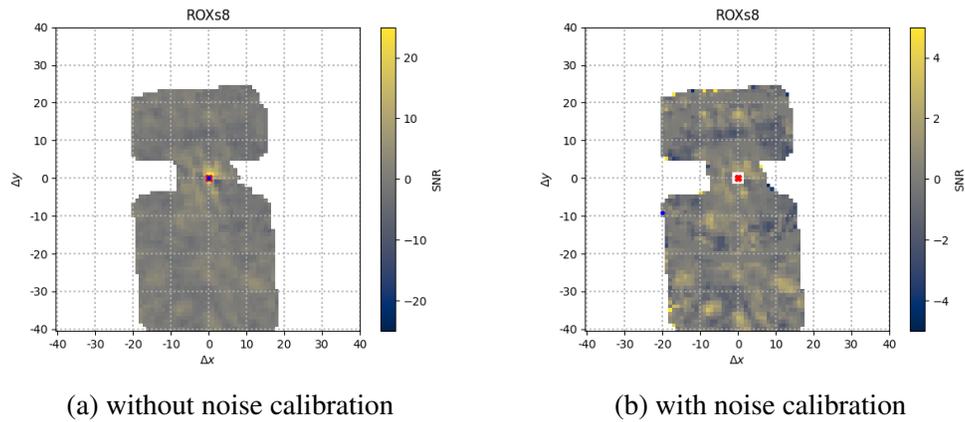


Figure A.15: Signal-to-noise ratio maps for ROXs8

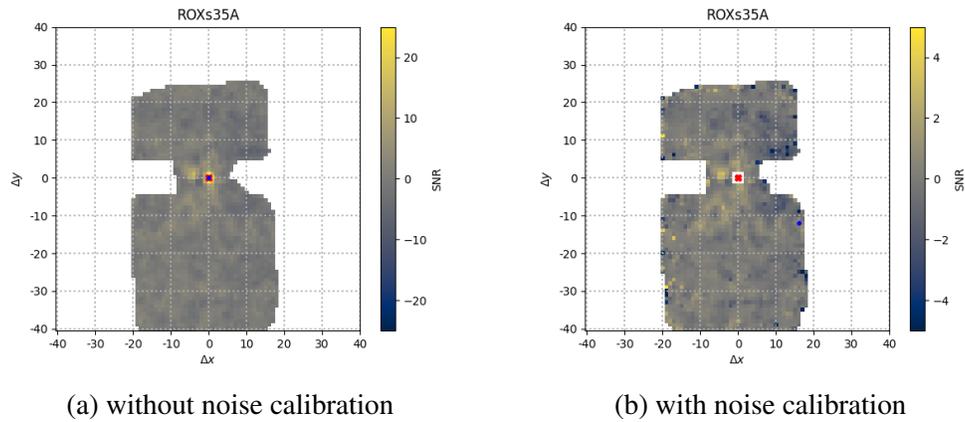


Figure A.16: Signal-to-noise ratio maps for ROXs35A

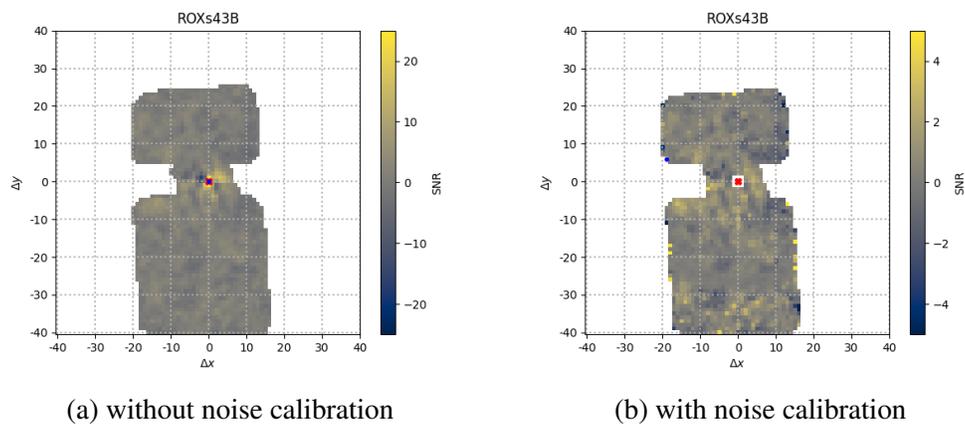


Figure A.17: Signal-to-noise ratio maps for ROXs43B

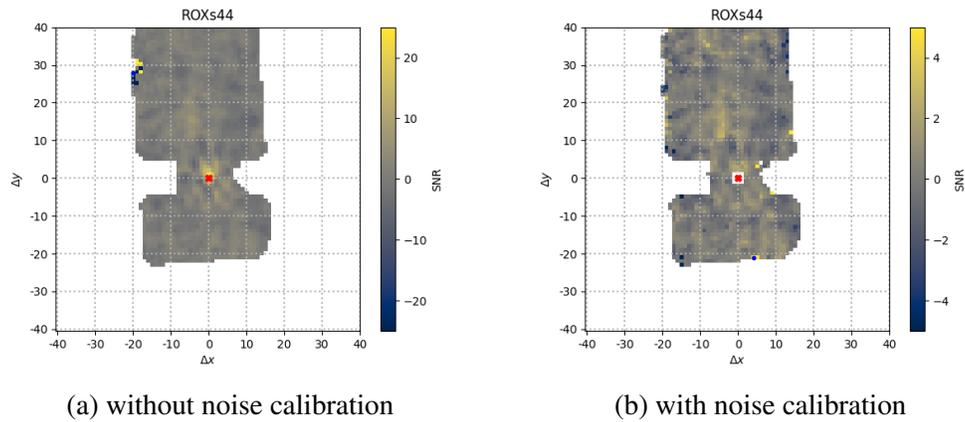


Figure A.18: Signal-to-noise ratio maps for ROXs44

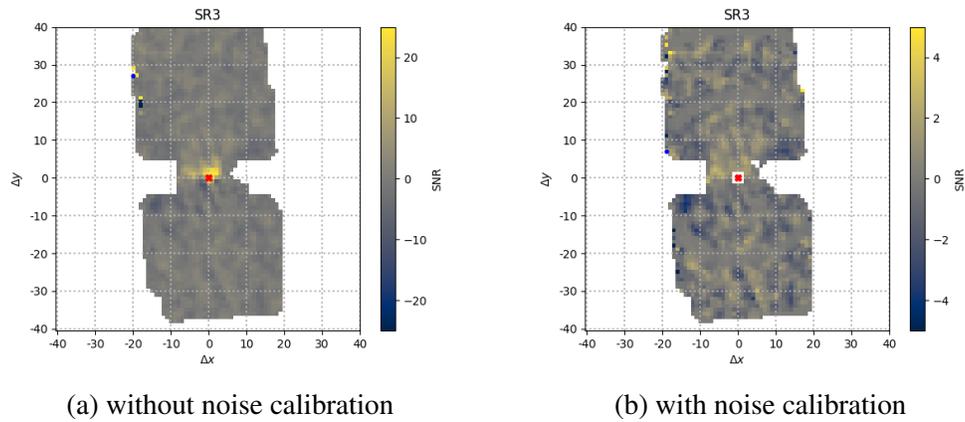


Figure A.19: Signal-to-noise ratio maps for SR3

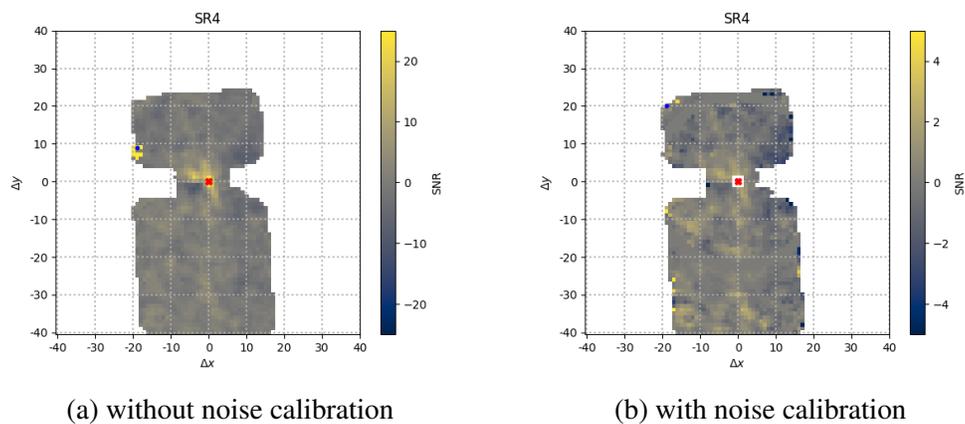


Figure A.20: Signal-to-noise ratio maps for SR4

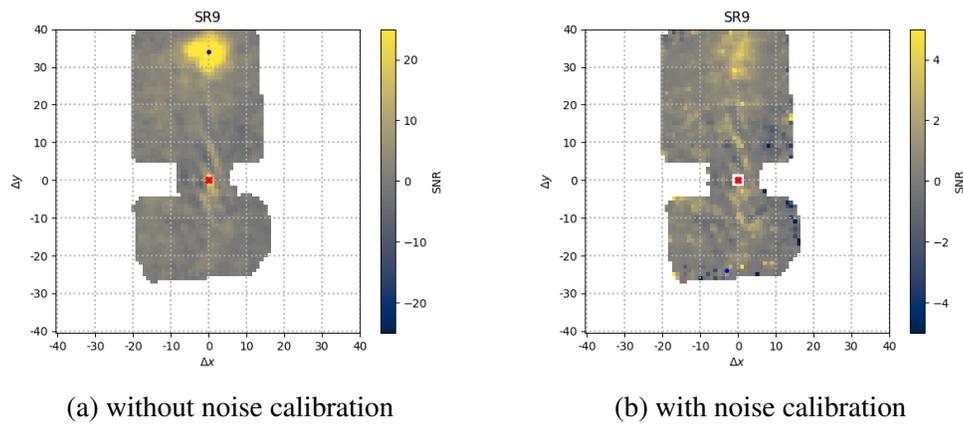


Figure A.21: Signal-to-noise ratio maps for SR9

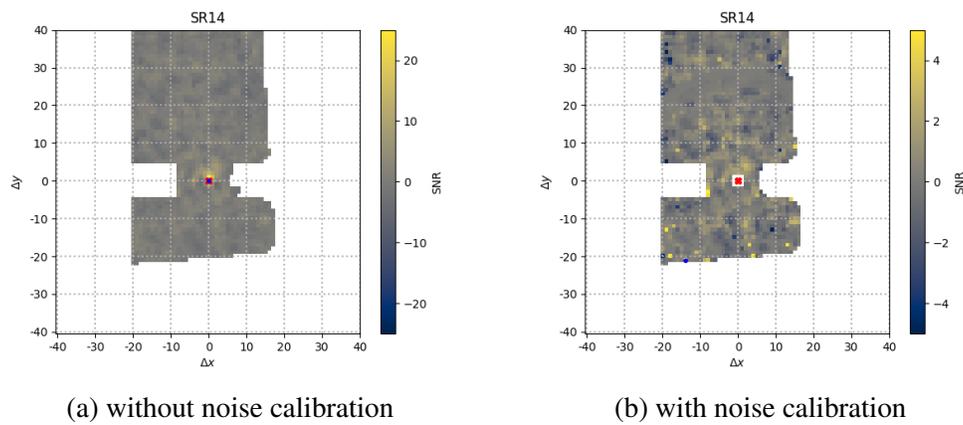


Figure A.22: Signal-to-noise ratio maps for SR14

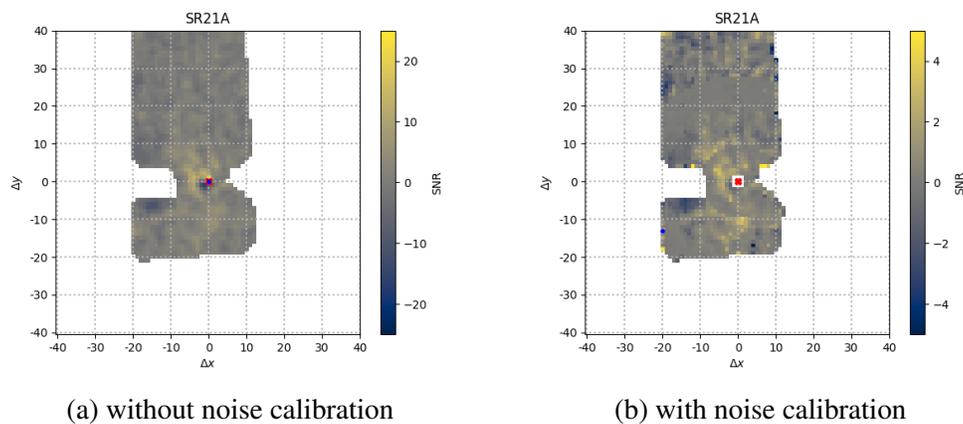


Figure A.23: Signal-to-noise ratio maps for SR21A

*Appendix B***APPENDIX B: COMBINED SIGNAL-TO-NOISE RATIO MAPS
AFTER RV CCF CALIBRATION**

Included below are combined SNR maps, one for each target, computed after all steps of noise normalization (including RV CCF calibration). These maps are discussed in detail in Section 3.4.

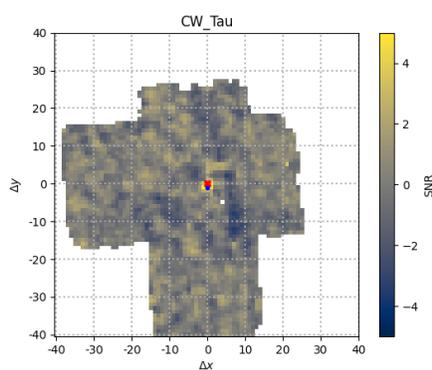


Figure B.1: Signal-to-noise ratio map, after RV CCF calibration, for CW_Tau

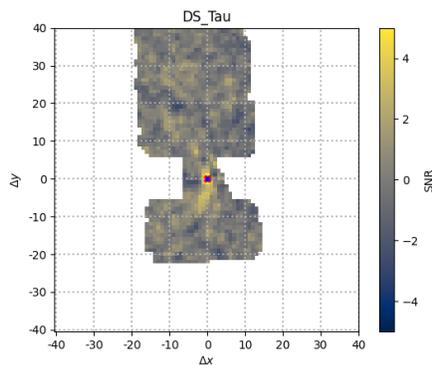


Figure B.2: Signal-to-noise ratio map, after RV CCF calibration, for DS_Tau

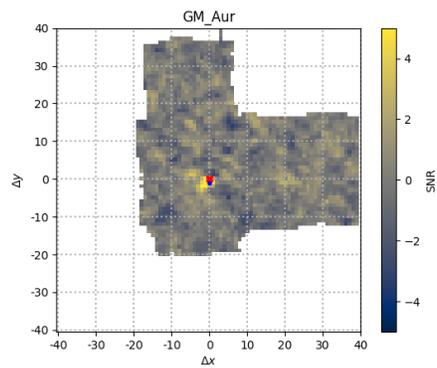


Figure B.3: Signal-to-noise ratio map, after RV CCF calibration, for GM_Aur

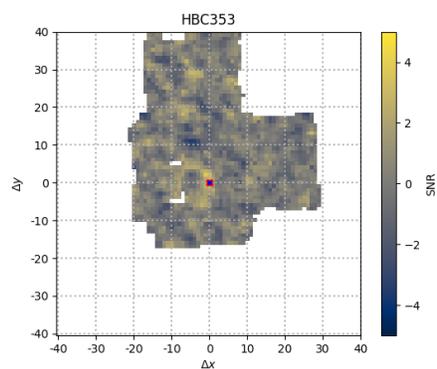


Figure B.4: Signal-to-noise ratio map, after RV CCF calibration, for HBC353

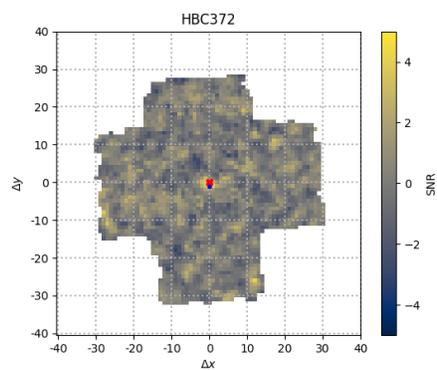


Figure B.5: Signal-to-noise ratio map, after RV CCF calibration, for HBC372

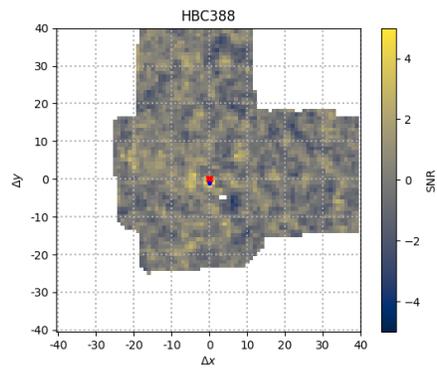


Figure B.6: Signal-to-noise ratio map, after RV CCF calibration, for HBC388

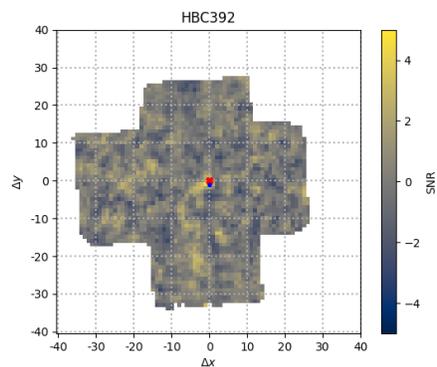


Figure B.7: Signal-to-noise ratio map, after RV CCF calibration, for HBC392

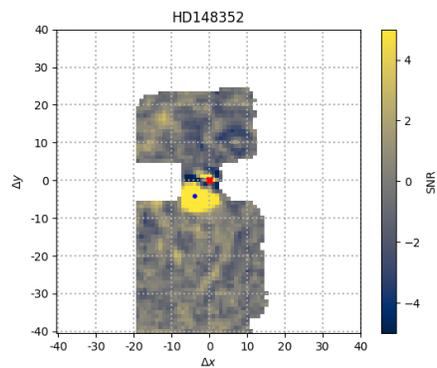


Figure B.8: Signal-to-noise ratio map, after RV CCF calibration, for HD148352

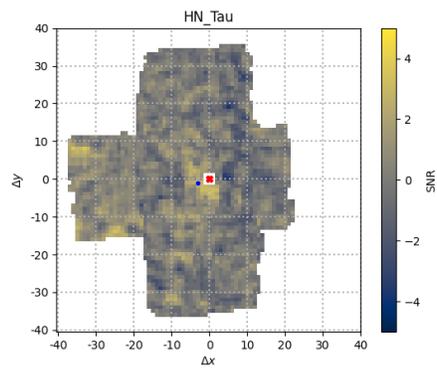


Figure B.9: Signal-to-noise ratio map, after RV CCF calibration, for HN_Tau

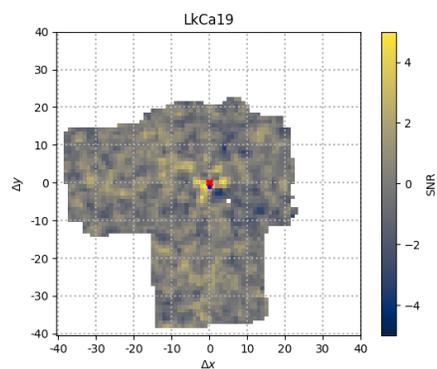


Figure B.10: Signal-to-noise ratio map, after RV CCF calibration, for LkCa19

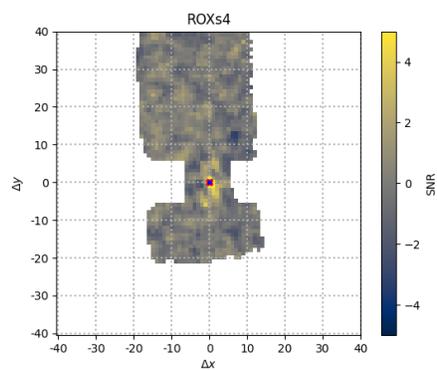


Figure B.11: Signal-to-noise ratio map, after RV CCF calibration, for ROXs4

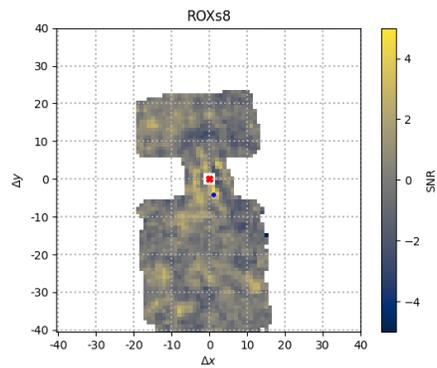


Figure B.12: Signal-to-noise ratio map, after RV CCF calibration, for ROXs8

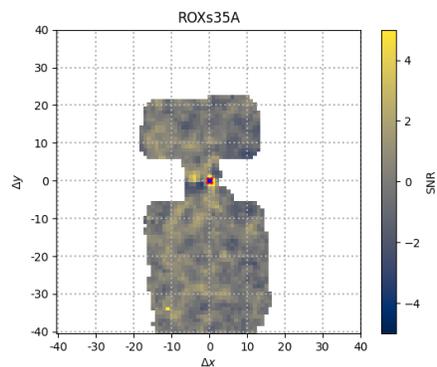


Figure B.13: Signal-to-noise ratio map, after RV CCF calibration, for ROXs35A

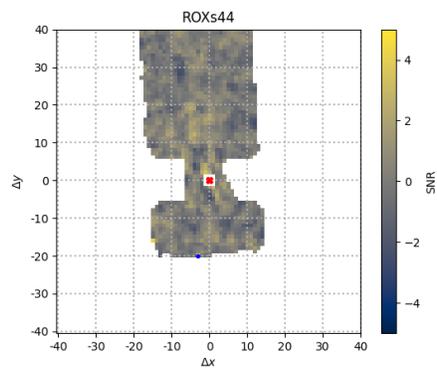


Figure B.14: Signal-to-noise ratio map, after RV CCF calibration, for ROXs44