CHARACTERIZING THE EXOPLANET POPULATION AROUND MID-TO-LATE M DWARFS

CHARACTERIZING THE EXOPLANET POPULATION AROUND MID-TO-LATE M DWARFS

By ERIK DIEGO GILLIS, BSc

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University

MASTER OF SCIENCE (2024)

Hamilton, Ontario, Canada (Physics and Astronomy)

TITLE:	Characterizing the Exoplanet Population around Mid-to-Late M dwarfs
AUTHOR:	Erik Diego Gillis BSc (Astronomy and Astrophysics), University of Toronto, Toronto, Canada
SUPERVISOR:	Ryan Cloutier
NUMBER OF PAGES:	xvi, 78

Lay Abstract

With thousands of exoplanet discoveries over the last three decades, exoplanet astronomy has revolutionized our understanding of the planet population. Successive surveys have shown that the two most common types of planets are rocky super-Earths with radii between 1 and 1.7 Earth radii and sub-Neptunes with radii spanning 2 to 5 Earth radii. Between these two populations is dearth of planets dubbed the Radius Valley. Super-Earths and sub-Neptunes are unrepresented in the solar system and models that can reproduce their population and the Radius Valley have sharpened our understanding of the planet formation process. We cannot reach a full understanding of the planet formation process without extending our understanding to the full population of planet-forming stars. The smallest of these stars are mid-to-late M dwarfs spanning from 8% to 40% the mass of our Sun and previous surveys have been largely insensitive to their planets at a population level. In this work I present my planet finding pipeline, and its deployment to survey 9,131 mid-to-late M dwarfs observed by the Transiting Exoplanet Survey Satellite (TESS) for transiting planet signals. By combining a population of 73 planets from this survey with a detailed, empirically established understanding of my pipeline's detection sensitivity to planets of various properties, this survey finds that each mid-to-late M dwarf hosts 1.3 planets within 30 days on average. This population is dominated by super-Earths, with relatively few sub-Neptunes. This result is in agreement with theoretical predictions of the planet population around the lowest mass stars and provides strong evidence that the radius valley disappears in this mass stellar regime.

Abstract

Around Sun-like stars and early M dwarfs alike, super-Earths and sub-Neptunes form a bimodal distribution separated by a dearth of planets between 1.6 and 1.9 Earth radii known as the radius valley. Modeling these planet populations and the radius valley have refined planet formation models but full understanding of the planet formation process requires a complete picture of the planet population extending to the lowest mass stars. As of yet, transiting exoplanet surveys have been largely insensitive to planets around mid-to-late M dwarfs. Fortunately, NASA's Transiting Exoplanet Survey Satellite (TESS) has opened a window into the exoplanet population around mid-to-late M dwarfs. I have led a systematic search for small transiting planets around 9,131 mid-to-late M dwarfs observed by TESS to characterize the planet population. I will present my pipeline to process TESS light curves and to detect and vet signals from transiting planets. Over the set of targets, this survey recovers a population of 73 manually vetted transiting planet signals. Using injection-recovery tests, I characterize the sensitivity of my pipeline to transiting planets around stars in the sample as a function period, instellation and radius. Using the recovered planet population combined with my survey completeness, I measure an occurrence rate of $1.326^{+0.210}_{-0.208}$ planets per star, with radii $< 6.5R_\oplus$ and orbital periods within 30 days, dominated by a population of sub- and super-Earths with very few sub-Neptunes

compared to more massive M dwarfs. This result is in agreement with previous work in this regime, while surveying a factor of 25 more stars, and aligns with theoretical predictions of the planet population. Along with our occurrence rate calculation, we provide strong evidence that the radius valley disappears in this stellar mass regime.

Acknowledgements

This thesis was completed at McMaster University I would like to recognize that McMaster is located on the traditional territories of the Mississauga and Haudenosaunee nations, and within the lands protected by the "Dish with One Spoon" wampum agreement.

I would like to acknowledge and thank my parents who instilled me with a love of science and astronomy, and continue to support me in my academic journey; My dear friends and family that surround me and give my life welcome and necessary depth; My patient supervisor Ryan Cloutier who has provided invaluable guidance, along with and the McMaster physics and astronomy student community, which has welcomed me these past two years and which I am proud to call myself a member of.

Acknowledgements to the institutions which made this work possible follow: This work was partially funded by an Ontario Graduate Fellowship awarded by the Government of Ontario. This paper made use of data collected by the TESS mission and are publicly available from the Mikulski Archive for Space Telescopes (MAST) operated by the Space Telescope Science Institute (STScI). Funding for the TESS mission is provided by NASA's Science Mission Directorate. We acknowledge the use of public TESS data from pipelines at the TESS Science Office and at the TESS Science Processing Operations Center. Resources supporting this work were provided by the

NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center for the production of the SPOC data products. This research has made use of the NASA Exoplanet Archive and the Exoplanet Follow-up Observation Program (ExoFOP; DOI: 10.26134/ExoFOP5) website, which are operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

Table of Contents

Lay Abstract **iii**

Abstract **v**

Acknowledgements **vii**

Notation, Definitions, and Abbreviations **xv**

1 Introduction **1**

1.1 The Transit Method 3

1.2 Occurrence Rates and The Radius Valley 6

1.3 Investigating the Radius Valley around the Lowest Mass Stars 13

2 Finding Transiting Exoplanets **17**

2.1 TESS Light Curve Data and Stellar Parameters 18

2.2 Detrending 20

2.3 Transit Least Squares (TLS) 29

2.4 Characterizing Planet Candidates 34

3 Pipeline Performance & Sensitivity Characterization **40**

3.1 Finding TOIs 40
3.2 Injection-Recovery 41

4 The Occurrence Rate of Planets around mid-to-late M Dwarfs 48

4.1 Transit Survey Results 48
4.2 Survey Completeness 51
4.3 Smooth Occurrence Rate Calculation 59

5 Discussion 62

5.1 Comparisons to Previous Calculations 62
5.2 Comparisons to Theoretical Predictions 67

6 Conclusions 69

A Visual Diagnostic Tests 71

List of Figures

1.1 Cumulative exoplanet discoveries to date 2

1.2 Sketch of the transit method 4

1.3 Occurrence rates of planets around Sun-like stars (Fulton et al., 2017) 8

1.4 Occurrence rates of planets around early M type stars (Cloutier & Menou, 2020) 9

1.5 Distribution of planets in mass/radius space simulated by Venturini et al. (2024) 13

1.6 Distributions of the survey's targets 14

1.7 Targets plotted as a function of TESS Magnitude and Radius with fiducial sensitivity curves. 16

2.1 Median model light curve detrending illustration. 20

2.2 Flares identified using consecutive 3σ outliers and excised using the change-point algorithm 22

2.3 Lomb-Scargle periodigram characterizing stellar rotation. 23

2.4 Rotation periods, and rotational model ΔBIC tests over my sample of stars. 25

2.5 Gaussian process rotation kernel light curve detrending illustration. . 28

2.6 Transit Least Squares spectrum and highlighted transit signal. 31

2.7 An illustration of the pipeline's iterative TLS search. 32
2.8 Illustration of sets of results used to characterize planet candidates. . 35
2.9 Fit models for the two planets around TIC 98796344 36
2.10 Evolution of the Monte Carlo Markov Chain walkers fitting the 5.36 day planet candidate around TIC 98796344. 38
2.11 Posterior distributions for the planet parameters of the 5.36 day planet candidate around TIC 98796344. 38
3.1 Found and not Found TESS TOIs. 42
3.2 Sensitivity of my pipeline to transiting planets around non-rotating stars 45
3.3 Sensitivity of my pipeline to transiting planets around non-rotating stars 46
4.1 Planets found by my survey. 51
4.2 Full sensitivity of my pipeline to transiting planets 52
4.3 Geometric transit probability of planets around stars in the survey's sample 53
4.4 Completeness of the survey 54
4.5 Occurrence rate of planets around mid-to-late M dwarfs as a function of period and radius 55
4.6 Occurrence rate of planets around mid-to-late M dwarfs as a function of instellation and radius 56
4.7 Occurrence rate of planets in radius space with uncertainties. 58
4.8 Smoothed sensitivity calculation 60
4.9 Smooth occurrence rate in period/radius space 61
A.1 Diagnostic figure showing the fit transit model used to inspect the goodness of fit. 72

A.2 Diagnostic figure showing the transits in the raw and detrended lightcurves. 72
A.3 Diagnostic plot illustrating the TLS results which contributed to the detection of the planet candidate. 73

List of Tables

2.1 Summary of the parameters used in the Gaussian Process used to remove the strong stellar rotation signals from light curves. 28

3.1 TESS targets of interest missed by my pipeline's planet survey. 47

4.1 Table of non-TOI credible signals 50

4.2 Number of planets per star versus planet radius and orbital period . . 58

4.3 Number of planets per star versus planet radius and instellation . . . 59

5.1 Comparison of my occurrence rate calculation in planet radius and instellation to Dressing & Charbonneau (2015) and Ment & Charbonneau (2023). 64

5.2 Comparison of my occurrence rate calculation in planet radius and period to Dressing & Charbonneau (2015) and Ment & Charbonneau (2023). 65

Notation, Definitions, and Abbreviations

Notation

T_0	The baryocentric Julian date of the first recorded transit
P	The period of the transiting planet
R_p/R_*	The ratio of the planet's radius to its host star's radius
b	The impact parameter of the transiting planet
BIC	Bayesian Information Criterion
$M_\odot$, $R_\odot$	Solar Mass and Radius
$M_\oplus$, $R_\oplus$	Earth Mass and Radius

Definitions

Super-Earth	A planet larger than the Earth with a radius between $1 - 1.7R_\oplus$.

Sub-Neptune A planet smaller than Neptune with a radius between $2 - 5R_{\oplus}$.

Radius Valley

The deficit in the occurrence rate of planets between $1.6 - 2R_{\oplus}$ Separating super-Earths and sub-Neptunes (Fulton et al. (2017))

Abbreviations

TESS Transiting Exoplanet Survey Sattellite

TLS Transit Least Squares

SDE Signal Detection Efficiency

SNR Signal to Noise Ratio

TOI Target of Intrest

ExoFOP Exoplanet Follow-up Observing Program

TFOPWG TESS follow-up observing program working group

FPP False-Positive Probability

Chapter 1

Introduction

Astronomy's modern history spans hundreds of years, proceeding millennia of night sky observations from all peoples through all time. The movement of the inner six planets of the solar system across the sky have been observable with the naked eye since prehistory. Explaining their motions across the background stars has motivated cosmological theories which eventually allowed astronomers to connect Kepler's laws to universal gravitation—a revolutionary connection that underpins modern physics. The first addition to the original six was William Hershel's discovery of Uranus with observations in 1781 (Herschel, 1783). Neptune was discovered some decades later when physically inconsistent measurements of Uranus' orbit led astronomers to an outer perturber (Smyth, 1846). These eight planets motivated our planet formation theories, but could never offer a window into the full population of all planets over all stars. Looking outside the solar system has only been made possible recently, and we have only been able to observe extra-solar planets in the three decades since Mayor & Queloz (1995) discovered a Jupiter-mass planet around 51 Pegasi. Since their discovery, the number of conclusively confirmed and published exoplanets has grown

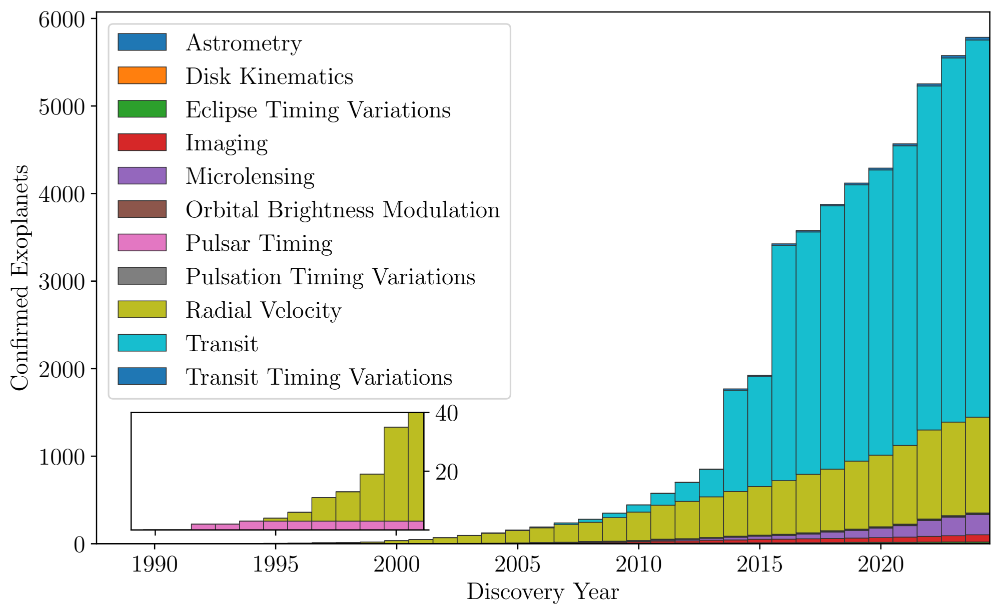

Figure 1.1: Cumulative exoplanet discoveries to date labeled by discovery method. Retrieved from the Exoplanet Archive on November 3rd 2024.

to 5,759 as of November 3rd 2024 (according to the NASA Exoplanet Archive at `https://exoplanetarchive.ipac.caltech.edu/`), with thousands of unpublished candidate planets and thousands more expected from coming surveys (see Perryman et al. (2014), Johnson et al. (2020), and Rauer et al. (2014)). This wealth of recent discovery has given us a robust understanding of the planet population, and in turn challenged and refined our understanding of the astrophysics of planet formation.

The worlds revealed by successive searches have shown us populations of planets completely unrepresented in the solar system. Of particular importance to this work are sub-Neptunes and super-Earths. Super-Earths are rocky planets with iron cores, larger than the Earth without an extended envelope like the gas giants. Their radii span from $1 - 1.7R_{\oplus}$. Sub-Neptunes are planets smaller than Neptune, likely with a rocky core and a gasseous envelope. Their bulk composition can also be very high in volatiles like water and their radii span $2 - 5R_\oplus$. Exoplanet surveys have shown super-Earths and sub-Neptunes represent the most common type of planet around Sun-like stars (Youdin, 2011).

Despite our extensive and growing catalog of exoplanets, there are still gaps and limitations to our understanding of the planet population. While we have a wide window into the planet population around Sun-like stars, our understanding of the planet population around smaller stars is more limited. There is relatively little understanding of the planet population around the smallest stars in the universe for reasons we will discuss. These stars—called mid-to-late M dwarfs—are anywhere from $0.1 - 0.5R_\oplus$, and are the most common type of star in the galaxy (Winters et al., 2015). This work aims to fill this gap in our understanding with the deepest comprehensive search for transiting planets around mid-to-late M dwarfs to date, and provide a robust understanding of the planet population around these stars.

1.1 The Transit Method

The exoplanet detection method key to this work is the transit method. If an exoplanetary system is aligned edge-on relative to the observer's line of sight, the planet will temporarily occult its host star as it passes in front of the star in a transit. This phenomenon is sketched in figure 1.2. By carefully recording the brightness of stars over time, we can detect dips which can be matched to physical transit-curves to provide strong evidence of the existence of an exoplanet. Since it was first employed to detect the transiting gas giant HD 209458 by Charbonneau et al. (2000), the transit has been responsible for the discovery of more than three-quarters of all known

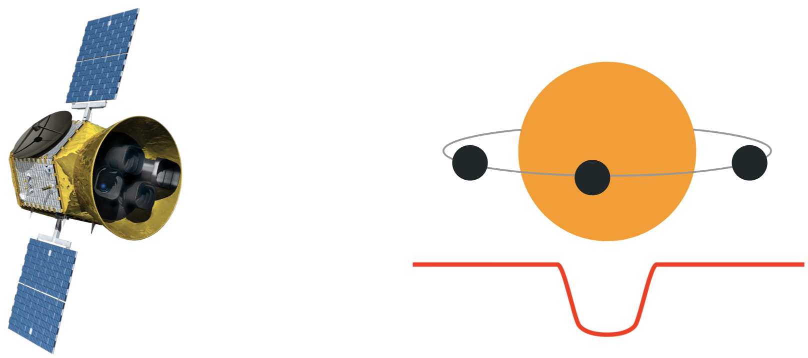

Figure 1.2: Sketch of the transit method: A telescope observing a star's brightness over time sees a decrease in brightness as an orbiting exoplanet transits the star.

exoplanets, illustrated in Figure 1.1.

Through the transit method, we can measure the planet's period and the ratio between its radius as the host star's radius expressed as R_p/R_*: The signal observed by the transit method, or more specifically the depth of the transit Δ, scales with the stellar and planet radii as $\Delta = (R_p/R_*)^2$. This observed ratio, combined with the stellar radius of the target, allows the planet's radius to be determined. Smaller stars produce deeper transit signals because a transiting planet of a given size blocks relatively more of their light. This means that the planets around mid-to-late M stars produce the deepest transit signals. However, their lower luminosity due to their cool temperatures and small radii mean that the photometric precision with which we can observe these stars is lower on average compared to earlier type stars. This means that only nearby M dwarfs make suitable targets for a transit search.

NASA's space-based transit survey satellites have contributed the bulk of discovered exoplanets using the transit method. The Kepler space telescope observed over 150,000 stars in a 16° diameter patch of sky from 2009 to 2013 (Koch et al., 2010). Following the failure of two of its reaction wheels, Kepler's primary mission ended but it continued to observe as K2 until 2018. Although it's pointing precision was greatly decreased because of the mechanical failure it was able to continue pointing along the ecliptic plane using radiation pressure from the Sun to slow its drift, along with periodic thruster firings to realign the observatory. (Howell et al., 2014). The telescope continued to discover exoplanets through the second phase of the mission. Both of these phases were very successful exoplanet discovery missions, with Kepler discovering 2,778 confirmed exoplanets with 1,982 yet to be confirmed. K2 carried on to discover 547 more confirmed exoplanets with 975 yet to be confirmed. The primary goal of the Kepler mission was to discover an Earth-like planet around a Sun-like star. This motivated its long deep stare to detect what would be yearly transits from an Earth double. This limited its access to the smallest M stars, because their low magnitude meant that relatively few were accessible in the small region of sky that Kepler surveyed (Cloutier & Menou, 2020).

In 2018, the Transiting Exoplanet Survey Satellite (TESS) launched. TESS uses four cameras to image the sky, each with a 24° by 24° field of view, stacked to form an observing area of 24° by 96° which spans from 6° above or below the ecliptic to the over the ecliptic pole (Ricker et al., 2015). TESS observes the sky in sectors which last 28 days where it observes the brightness of the stars in its field of view. After each sector, the telescope rotates 27° about the ecliptic pole shifting its field of view. With this strategy, TESS observed the southern hemisphere over its first 13 sectors, and the following 13 sectors observed the northern hemisphere. While each sector represents a different pointing of the telescope, the top camera is always aligned with one of the ecliptic poles meaning that TESS has two continuous viewing zones (CVZs). The CVSs are observed for 13 sectors consecutively and targets near either ecliptic pole can overlap with multiple sectors of viewing. At time of writing in November 2024, TESS continues to observe with its 86th sector of observation. It has also been an incredibly successful mission, with 589 confirmed exoplanets, and 4,709 candidates which have yet to be confirmed.

TESS enables the investigation of the planet population around mid-to-late M dwarfs in two key ways: Its large field of view and full sky access allows it to observe every nearby M dwarf where Kepler was limited to the few in its field of view. TESS's bandpass is also significantly redder than Kepler's, spanning 600-1000 nm (Ricker et al., 2015). Compared to Kepler's 420-900 nm bandpass (Koch et al., 2010), this range provides better photometric precision by more fully capturing the spectral profile of these cool, red stars.

1.2 Occurrence Rates and The Radius Valley

The wealth of planets discovered by the transit method have provided a strong sample to build demographic studies of exoplanets. By combining the population of discovered planets with the likelyhood of having detected each planet, exoplanet surveys with the transit method can build an understanding of the occurrence rates of planets. These surveys have shown that the occurrence rate of close-in, sub-Neptune-sized planets forms a bimodal size distribution around Sun-like stars known as the Radius Valley separating super-Earths and sub-Neptunes (e.g. Fulton et al. (2017), Van Eylen et al. (2018)) in the following subsection, I will more thoroughly outline these observations.

Many theories are able to replicate the Radius Valley: Photoevaporative theories proposed by Owen & Wu (2013) and elaborated on in Owen & Wu (2017) argue photoevaporation selectively strips enveloped cores, creating a population of H/He enveloped sub-Neptunes and rocky super-Earths. Alternatively Gupta & Schlichting (2019) introduce core-powered mass loss that rocky super-Earths and enveloped sub-Neptunes can carve out the Radius Valley, but argue that the cooling luminosity of the planet can drive away atmosphere to carve the valley. Planet synthesis models offer an alternate window into the Radius Valley, and in particular Burn et al. (2024) and Venturini et al. (2024) show that water-rich formation plays an important role in shaping the Radius Valley. I will discuss these theories in depth following the discussion of the observational Radius Valley.

1.2.1 The Observed Radius Valley

The first robust observation of the Radius Valley comes from Fulton et al. (2017) which updated planet radii and provided an occurrence rate calculation based on the sensitivity of the Kepler mission to each of the 2,025 exoplanets in its sample from Sun-like FGK-type stars. These methods allowed the true size distribution to be sharply defined and compared to the predicted distribution from models of the Radius Valley's emergence proposed in Owen & Wu (2013). Their results extended past a one-dimensional characterization of the Radius Valley, and showed the occurrence of planets in period/radius space, establishing that Sun-like stars generally host close-in super-Earths and further-out sub-Neptunes. Because of this dichotomy, these two populations can be separated by a line with positive slope in radius/instellation space. This is illustrated in Figure 1.3. This positive slope is consistent with predictions of thermally driven mass-loss models, with the difference in planetary radii owed to the presence or lack of a thick hydrogen/helium (H/He) envelope.

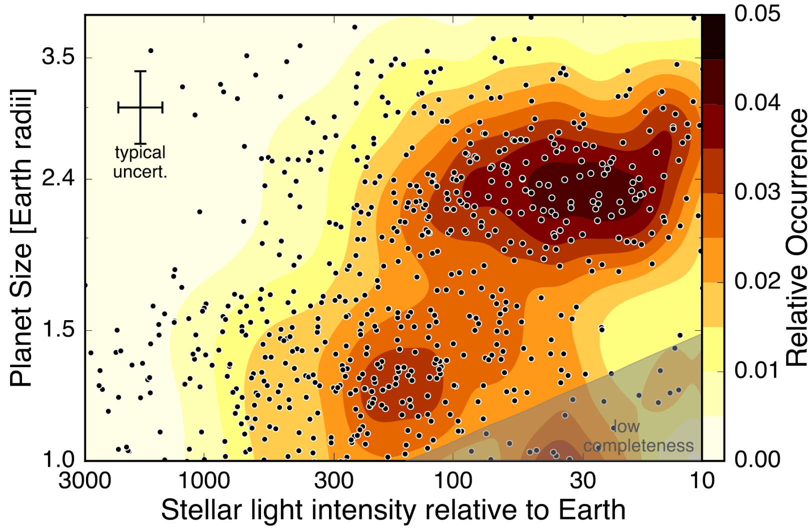

Figure 1.3: Occurrence rates of planets around Sun-like stars from Fulton et al. (2017) in Instellation (interchangeable with period)/radius space. These populations are separated by a negatively sloped curve in Period/radius space.

More careful data analysis by Van Eylen et al. (2018) restricted the sample of planets to those whose radius could be tightly constrained with detailed stellar radii derived with astroseismography and detailed transit modeling for each planet to best-fit their radii. Their analysis finds that the slope which separates super-Earths and sub-Neptunes in period/radius space is $m = -0.09^{+0.02}_{-0.04}$ with 68% confidence. Their sample is significantly smaller than the sample analyzed in Fulton et al. (2017), and the average star in their sample is larger and older, but these biases do not significantly effect the validity of their slope measurement because of the accuracy with which they were able to measure the radii of the planets included in their sample.

Recall that a detailed analysis of planets around late type stars was intractable with the Kepler mission. But despite these shortcomings, and with additional planets discovered by the follow-up mission K2, Cloutier & Menou (2020) was able to detect the Radius Valley in the population of planets around late K to early M type stars.

Using 350 planets, again with detailed considerations of their survey's completeness, they found that while the Radius Valley still exists around smaller stars, the slope in period-radius space is different. This is illustrated in Figure 1.4, and shows that the slope in period/radius space is positive. This inconsistency suggests that the dominant formation mechanism driving the Radius Valley around early M stars is different than dominant sculptor of the Radius Valley the around Sun-like stars.

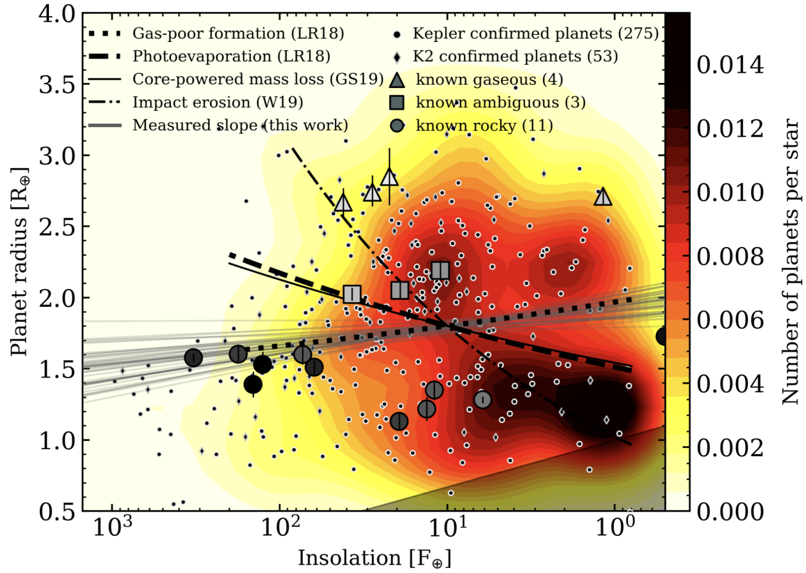

Figure 1.4: Occurrence rates of planets around early M type stars from Cloutier & Menou (2020). The best fit slope which separates these two populations is computed and various Radius Valley models' slopes are overplotted.

There is currently very little observational understanding of the Radius Valley in planets around the lowest mass M dwarf stars. Using data from TESS, Ment & Charbonneau (2023) present seven planets around mid-to-late M dwarfs within 15 parsecs of the Earth to derive an occurrence rate. Based on occurrence rate calculations, they find that sub-Neptunes are significantly less abundant in this stellar mass regime, being outnumbered at least 14-to-1 by super-Earths. They report an occurrence rate of $0.61^{+0.24}_{-0.19}$ terrestrial planets per star, but their sample is far too small to confirm the existence of a Radius Valley in this stellar mass regime.

1.2.2 Theoretical Radius Distributions

The first model which predicted the Radius Valley comes from Owen & Wu (2013). The then newly discovered populations of super-Earths and sub-Neptunes already showed hints of two planet populations years before Fulton et al. (2017) were able to compile their robust occurrence calculation. These two populations motivated a theory which could produce both. The photoevaporation theory proposed in Owen & Wu (2013) and refined in Owen & Wu (2017) begins by assuming that each planet emerges from the protoplanetary disk as a rocky core surrounded with a hydrogen/helium envelope. During the first 100 Myrs of their lives, stars are XUV active emitting strong radiation beyond their red blackbody spectrum, and the envelopes of these planets are subject to stripping by the strong radiation from their host star. After the XUV active phase ends, planets are either stripped rocky cores observable as super-Earths or have retained their envelopes and are observable as sub-Neptunes. In this way, a smooth initial distribution of planets is split into the two populations separated by the Radius Valley. While this photoevaporation theory is consistent with the Fulton et al. (2017) Radius Valley, it's predictions are not completely consistent with the refined Radius Valley outlined in Van Eylen et al. (2018) indicating that this picture may not be sufficient.

Gupta & Schlichting (2019) provide an alternate channel which can also explain the Radius Valley through core-powered mass loss. As planets form, their gravitational energy superheats their cores to between 10,000 K and 100,000 K and this energy is radiated away as the planets cool. Gupta & Schlichting (2019) present a comprehensive argument that this cooling luminosity is significant enough to carve out the Radius Valley without the effects of photoevaporation. Loosely, the ability of core-powered mass loss to strip planets of their envelopes depends on the comparison between the liberated thermal energy and the gravitational binding energy of the atmosphere. By dynamically evolving a population of enveloped Earth-like cores and without invoking photoevaporation, Gupta & Schlichting (2019) shows that core-powered mass loss is sufficient to replicate the slope and shape of the Radius Valley around Sun-like stars, matching Van Eylen et al. (2018) measurements very well.

As we've seen however, the slope found by Cloutier & Menou (2020) in Figure 1.4 is inconsistent with theories of thermally-driven atmospheric escape such as photoevaporation and core-powered mass loss suggesting that the planet population cannot be completely explained by appealing to evaporation. The theories discussed thus far have attempted to re-create the Radius Valley from a population of enveloped rocky cores, and their models do not investigate the role that water plays in forming planets and differentiating super-Earths from sub-Neptunes.

The Bern New Generation Planetary Population Synthesis (NGPPS) model (Emsenhuber et al. (2021a), Emsenhuber et al. (2021b), Schlecker et al. (2021), Burn et al. (2021)) aims to provide a complete understanding of the exoplanet population by synthesizing planets from planetessimal in a protoplanetary disk. Their suite of simulations completely evolves the protoplanetary disk and the planets formed from planetesimals therein, providing a full description of the exoplanetary systems it produces. These simulated planets' composition can include a large contribution from water accreted from ice-rich planetessimals. Using the Bern model Burn et al.

(2024) shows that including water-rich planets with condensed water cannot accurately reproduce the Radius Valley. However, by accurately treating water vapour mixed with H/He in a steam atmosphere, a population of water-rich sub-Neptunes matching the sub-Neptune population observed by Fulton et al. (2017) emerges. According to the Bern model, the dominant driver of the Radius Valley is composition, not post-formation photoevaporation.

Using their own pebble-based planet synthesis model Venturini et al. (2024) simulated the formation of planets around stars with 0.1, 0.4, 0.7, 1.0, 1.3, and 1.5 $M_{\odot}$. Their model considers the coagulation, fragmentation, drift, diffusion and ice sublimation of pebbles based on Birnstiel et al. (2011) and Drazkowska et al. (2016). In agreement with Burn et al. (2024), Venturini et al. (2024) find steamy water-rich sub-Neptunes are responsible for the second peak in the Radius Valley. Their model find a much more prominent bimodal distribution in density illustrated in Figure 1.5 with two distinct populations of rocky and water-rich planets, separated in density along the highlighted curves. Without the effects of atmospheric steam puffing up their atmospheres, these two populations of planets overlap in radius space, and the Radius Valley only manifests because of the inflated sub-Neptunes. This effect weakens in lower-mass stars, which causes a so-called fading Radius Valley towards the lowest mass stars and they predict that the Radius Valley is filled in around stars with 0.1-0.4$M_{\odot}$.

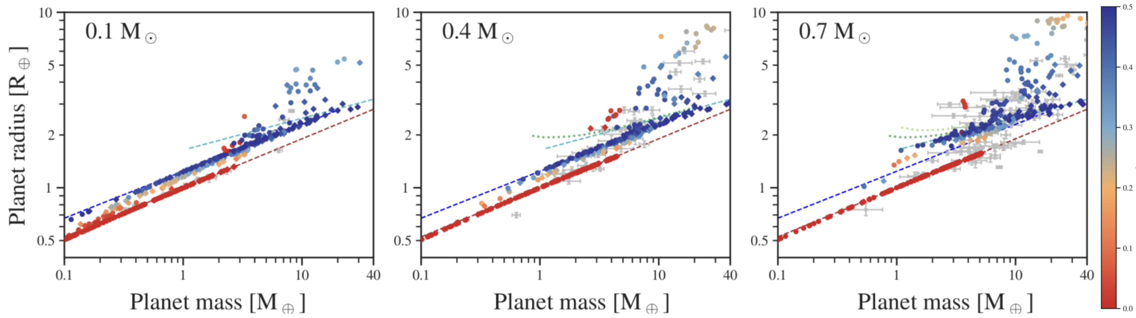

Figure 1.5: Figure 1 from Venturini et al. (2024). Distribution of planets in mass/radius space generated by their simulations, showing that a water-rich and water-poor population can reproduce a bimodal distribution in radius, which is prominent around larger stars, and fades in mid-to-late M dwarfs.

1.3 Investigating the Radius Valley around the Lowest Mass Stars

To characterize the Radius Valley around mid-to-late M dwarfs, we survey a sample of 9,131 nearby stars. Data for these targets was acquired through a TESS Guest Investigator-approved programs G03274, G04214, G05152, and G05152 authored by Ryan Cloutier to observe the survey's targets from sectors 27-69. The programs' goal was to observe a uniform sample of mid-to-late M dwarfs and determine the central radius of the valley around this population of stars. The experiment was designed to determine which theoretical model is the dominant driver of the Radius Valley by comparing their model predictions to the population of planets in nature. The distribution of the targets' distances and radii is plotted in Figure 1.6. Each target's brightness was recorded with a cadence of two minutes for a minimum of 28 days. This set of targets was selected based on the detectability of transiting planets whose radii span the Radius Valley. This detectability can be assessed based

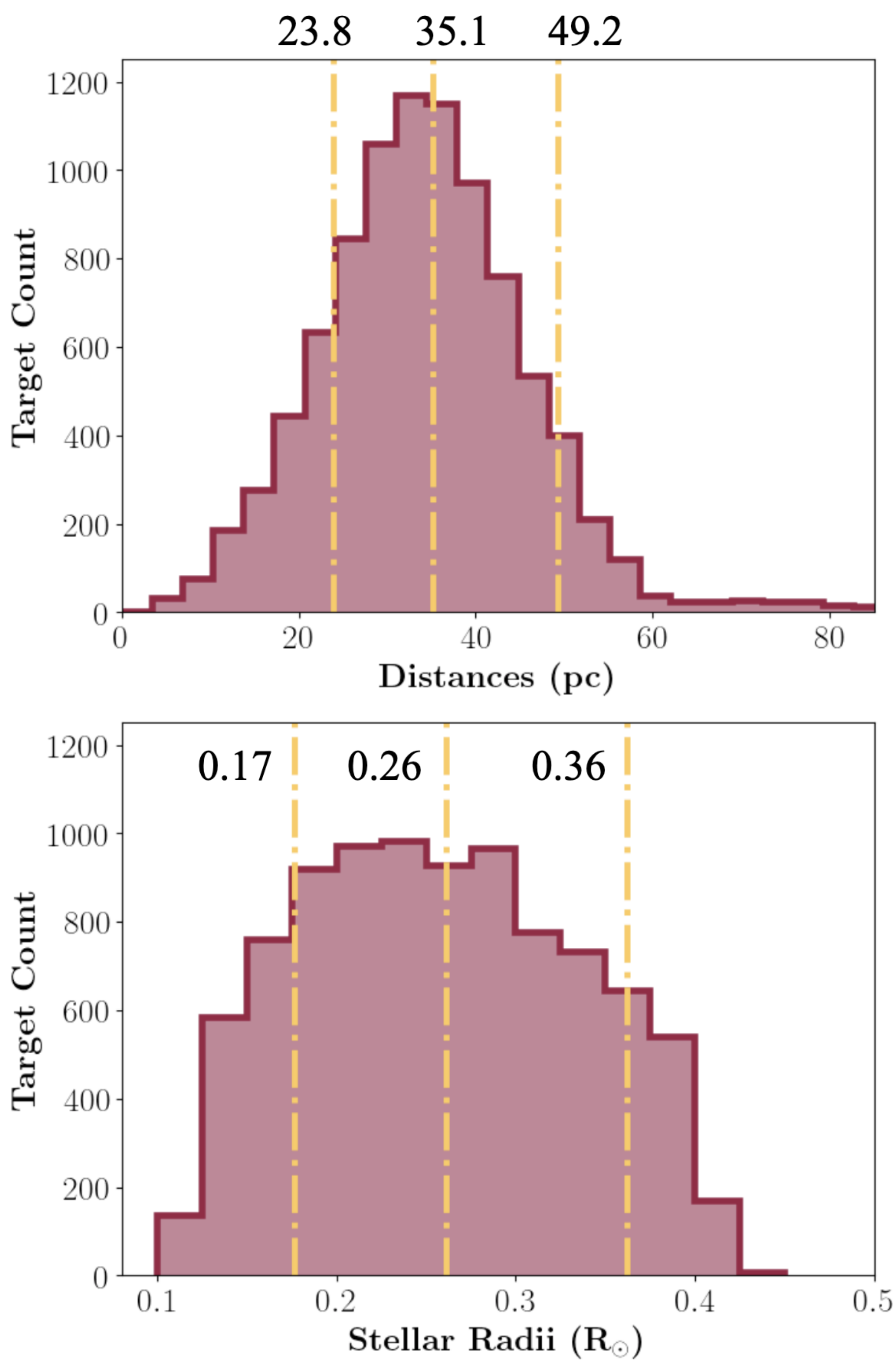

Figure 1.6: Distribution of the survey's targets in radius and distance with the distribution's mean and 1σ intervals marked.

on the expected signal-to-noise ratio (SNR). This is the ratio between the expected photometric noise, and the signal of transiting planets with varying radii and periods. In the ideal case, the photometric noise—expressed as $\sigma_{\rm lc}$, is a function of the target's TESS magnitude, was computed using `ticgen` (Barclay, 2017). The photometric noise is calculated assuming that the signal is integrated over the transit duration. The transit signal is the transit depth Δ, computed as:

$$\Delta = \left(\frac{R_p}{R_*}\right)^2, \tag{1.3.1}$$

where R_* and R_p are the radii of the host star and planet respectively. The ratio between Δ and $\sigma_{\rm lc}$ provides the SNR of a single transit. By combining a number of transits $N_{\rm transits}$, the SNR can be enhanced and for a given planet. With some $N_{\rm transits}$ in a light curve, the SNR can be expressed as:

$$\mathrm{SNR} = \frac{\Delta}{\sigma_{\rm lc}}\sqrt{N_{\rm transits}}. \tag{1.3.2}$$

Using the SNR, calculated from the host stars' properties and planet parameters we can determine which planets ought to be detectable around a given star. This is shown in Figure 1.7, which illustrates that this survey should have compete sensitivity to sub-Neptunes, and good sensitivity to super-Earths based on the target's properties and the planet's period. Each of the sensitivity curves in the figure represent the maximum magnitude a fiducial planet would be observable at. The true sensitivity of this survey can only be computed by empirically testing a particular pipeline which is outlined in Chapter 3.

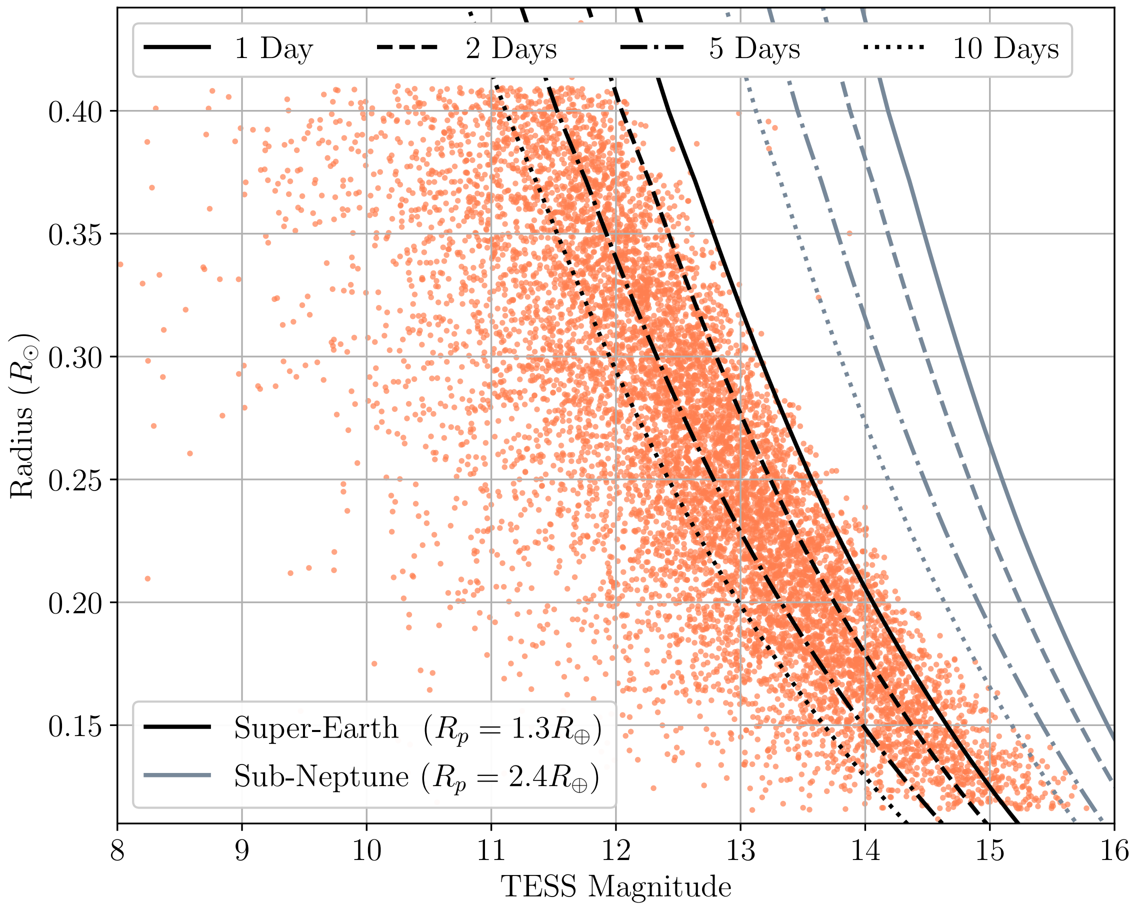

Figure 1.7: Targets plotted as a function of TESS magnitude and radius with fiducial sensitivity curves for a 6σ detection with a 28 day light curve. Each of the sensitivity curves in the figure represent the maximum magnitude a fiducial planet would be observable at. In ideal conditions, the set of targets with a detectable planet of the given characteristics are left of the respective curve.

Chapter 2

Finding Transiting Exoplanets

This chapter is a description of the pipeline I have used to search my set of targets for transiting exoplanets. Building a pipeline was necessary to characterize the exoplanet population, because a complete understanding can only be established when the found population of planets is combined with the completeness of the survey. The latter is only possible to deduce by testing the pipeline that was used to find the planets. So, while there are known planets around stars in my sample, the completeness of the searches that discovered them is unknown. My pipeline was custom built for this survey and its development, testing and deployment at scale represents a large portion of the work in my Master's thesis.

2.1 TESS Light Curve Data and Stellar Parameters

All of the light curve data used in this survey is hosted on the Mikulski Archive for Space Telescopes (MAST) (`https://dx.doi.org/10.17909/T9RP4V`) which comprises all of TESS's observations of our targets up to its 85th sector of observation. Each target was observed with a 2-minute cadence over the sectors where they were in-frame from sectors 27 to 69 because of the aforementioned TESS Guest Investigator proposals. Many targets also have additional 2-minute cadence light curves from sectors 1-27 and 70-85 because of TESS's primary mission and other GI programs, respectively.

TESS records the brightness of each target using Simple Aperture Photometry (SAP) using an aperture of pixels subtended by the star on TESS's detector. SAP derives the flux from a target by summing the dark-subtracted, debiased pixels within the TESS predetermined photometric aperture. This flux is subject to instrumental variations, but the Presearch Data Conditioning Corrected SAP (PDCSAP) flux works to correct these variations. The PDCSAP uses principal component analysis over all light curves from the detector to remove large-scale correlated variations exhibited in all the sources on the detector. The PDCSAP flux best preserves the astrophysical variability of each target while removing instrumental variability exhibited by every star on the detector.

For each sector in which a given target was observed, my pipeline retrieves the light curve and supplementary data products, which comprise the following series: The BJD timestamp of each observation, the PDCSAP flux, the 1σ error on the PDCSAP flux, the exposure time, the data quality flags, and the sector of the observation. Any data with a nonzero quality flag (indicating some issue) is excised from each time series.

My pipeline saves these data as a set of light curves by concatenating consecutive sectors of data up to a limit of four consecutive sectors. This light curve length limit reduces the computational cost of processing each target's data, because some steps of the pipeline scale super-linearly with the length of the time series.

For each target, my pipeline also retrieves the effective temperature, radius, radius uncertainty, mass, and mass uncertainty from the TESS Input Catalog (TIC, Stassun et al. (2019)). The effective temperature is calculated in the TIC with established colour relations, or values from literature where available. The masses of the targets are derived using parallax measurements characterizing the distance combined with the luminosity where available. Without parallax, mass is estimated using colours as well as effective temperature. For targets missing mass or radius measurements, my pipeline estimates the missing parameter using the other quantity based on the relationships from Pecaut & Mamajek (2013). Targets with both mass and radius missing were removed from the original sample requested by the aforementioned TESS GI proposal. My pipeline also saves the target's quadratic limb-darkening parameters for each target from Claret (2017), which characterizes the source intensity from the center to the limb, which is necessary to accurately model transit light curves.

2.2 Detrending

Before each light curve can be searched for transit signals, we must remove any systematic variability which was not removed by the PDCSAP correction. This variability can be due to the star's changing brightness residual instrumental systematics. In particular, photometric variability caused by stellar rotation and flares must be treated before each light curve can be searched for evidence of transiting planets. Trends are removed from each light curve by normalizing each data point by the sliding 12-hour median of the light curve. This detrending is illustrated in Figure 2.1. The median filter is insufficient however, in the case of the aforementioned flare and rotation signals. These both require treatment before the median filter is applied. These prescriptions are outlined in the following subsections, with plots illustrating situations where they are relevant.

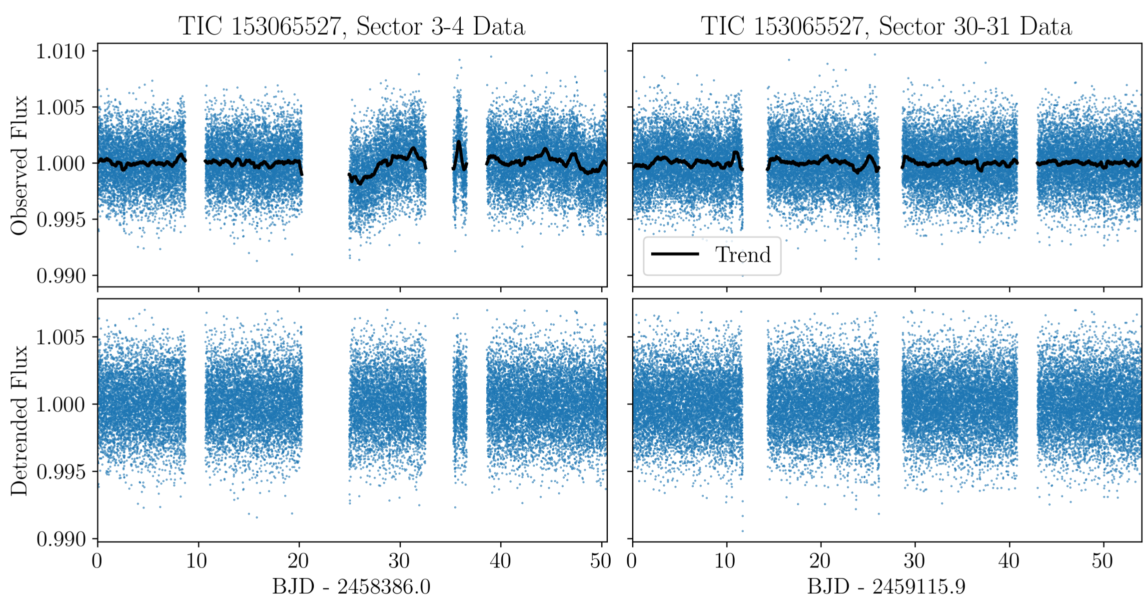

Figure 2.1: Top: TESS light curve from each observed sector of TIC 170636897 with consecutive sectors concatenated and trend overplotted. Bottom: Detrended light curve with rolling 12-hour median trend removed.

2.2.1 Flares

Stellar flares from our targets can be seen in many light curves as sharp increases in brightness with a decaying exponential tail (see Figure 2.2). For our purposes, flares are nuisance signals that we identify and mask from our light curves. Since they span very short sections of our light curves, we identify and remove them from the data. This step is done before any detrending methods are applied.

This process of flare removal follows Chang et al. (2015). We start by identifying data points which are ≥ 3 median absolute deviations from the median of the light curve. If there are three or more consecutive outlying points, the 30 data points on either side of the identified flare are taken as part of the series which may contain the flare. With this sequence of data we apply the change-point algorithm described in Chang et al. (2015). With a sequence of flux measurements $x_1, x_2, ..., x_n$ with some mean $\bar{x}$ we define the k^{th} cumulative sum as

$$S_k = \sum_{i=0}^{k} (x_i - \bar{x}) \, . \tag{2.2.1}$$

A change-point corresponds to the maximum value of $|S_k|$ and its significance is determined by boostrapping and comparing its value to the cumulative sum from shuffled x_i series. If its magnitude exceeds 90% of the boostrapped change-points, it is significant, and the series is divided into two sets on either side of the change-point. This process continues recursively on those subsets until no new change-points are identified. Identifying these change-points allows the pipeline to identify the extent of a flare, and the change-points, along with the data in between them are excised. Flares identified with this algorithms and the data points the pipeline would excise are illustrated in Figure 2.2.

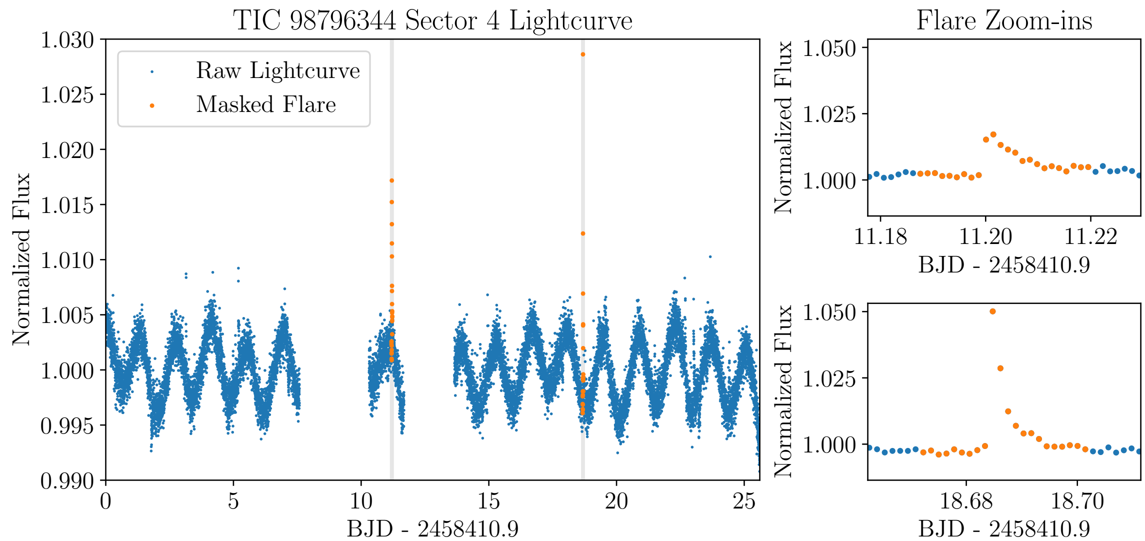

Figure 2.2: Left: Normalized TESS light curve with two flares highlighted and the excised data has been marked in orange, identified by looking for sets of 3 consecutive 3σ outliers, and masked by using the change-point algorithm described in Chang et al. (2015). Right: Zoom-in plots of the two flares showing their extent.

2.2.2 Stellar Rotation

It is particularly common for M dwarfs to show strong quasi-periodic photometric variability because of rotation. Large star spots can rotate in and out of view introducing a quasi-periodic signal into their light curves. This signal is illustrated by the light curve in Figure 2.3 from TIC 98796344. We identify this variability using a Lomb-Scargle periodigram implemented by Schröter et al. (2019). Before computing the periodigram, I bin-down the light curve's points to 0.05 days to reduce the expense of the computation. The Lomb-Scargle periodigram for TIC 98796344 is illustrated in Figure 2.3 showing evidence of strong 1.39 day periodicity.

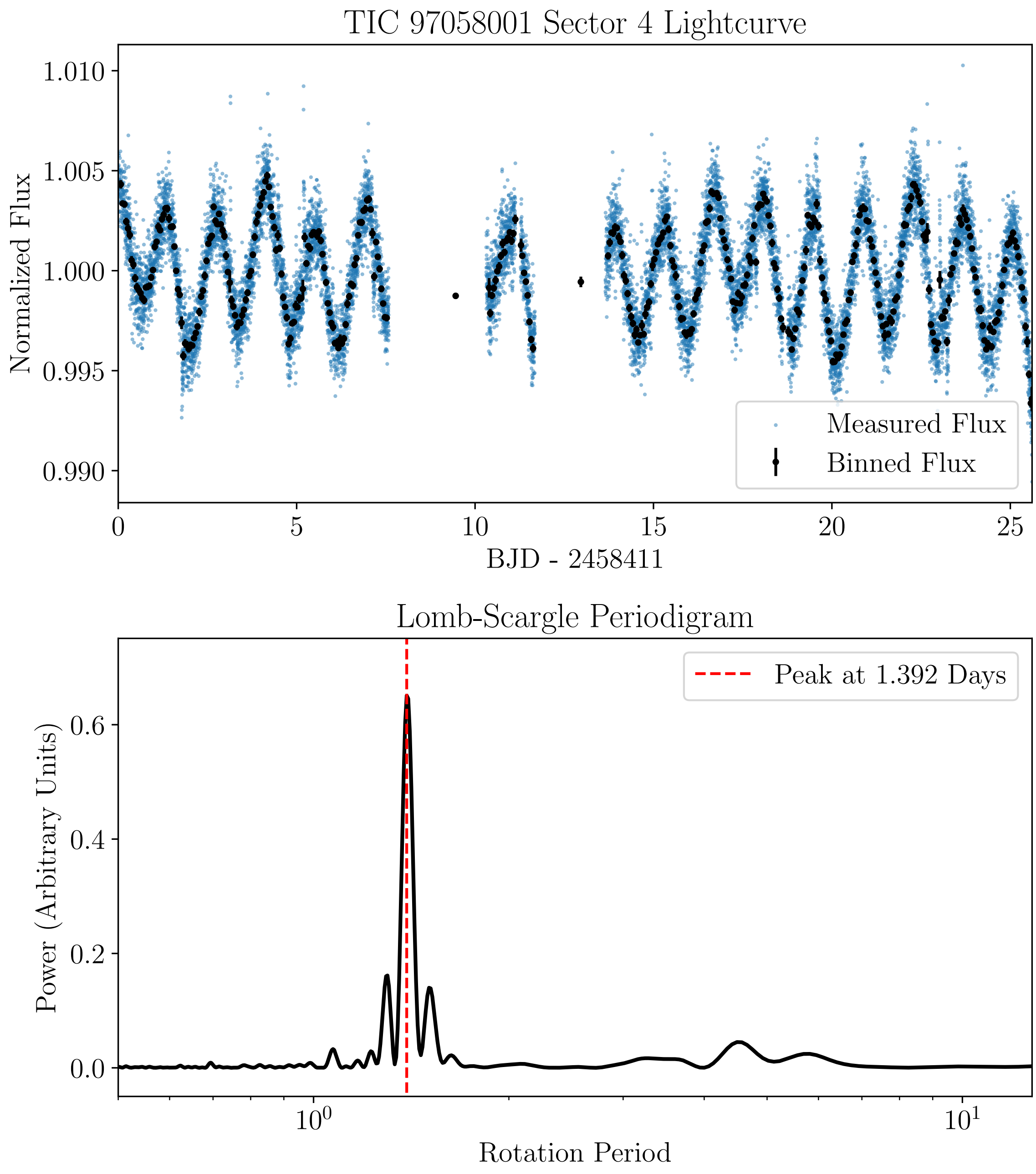

Figure 2.3: Top: Normalized light curve from sector 4 observations of TIC 98796344 with binned points overplotted. Bottom: Lomb-Scargle periodigram of the binned points with the peak period highlighted.

The primary peak of the periodigram prescribes the period, phase and amplitude of the dominant periodic signal. To test whether the rotation signal is significant we use a Bayesian Information Criterion (BIC) to compare a sinusoidal model with a median model. Where $\hat{L}$ is the likelihood of a model, n is the number of data points, and k is the number of free parameters in the model, we compute the criterion as:

$$\mathrm{BIC} = k \ln(n) - 2\ln(\hat{L}). \tag{2.2.2}$$

Taking the difference between the BICs for the sinusoidal and median models characterizes how well one fits over the other and is expressed as the ΔBIC. Should the BIC favour the rotation model with $\Delta\mathrm{BIC} \leq -50$, we consider the rotational signal significant. Additionally, we require rotation signals with a period of more than 1 day to have an amplitude greater than the median reported photometric error. This extra condition is imposed because of the 12 hour length of our median filter, which deals significantly worse with short period variability. The rotational characteristics of each target in our sample of stars are illustrated in Figure 2.4.

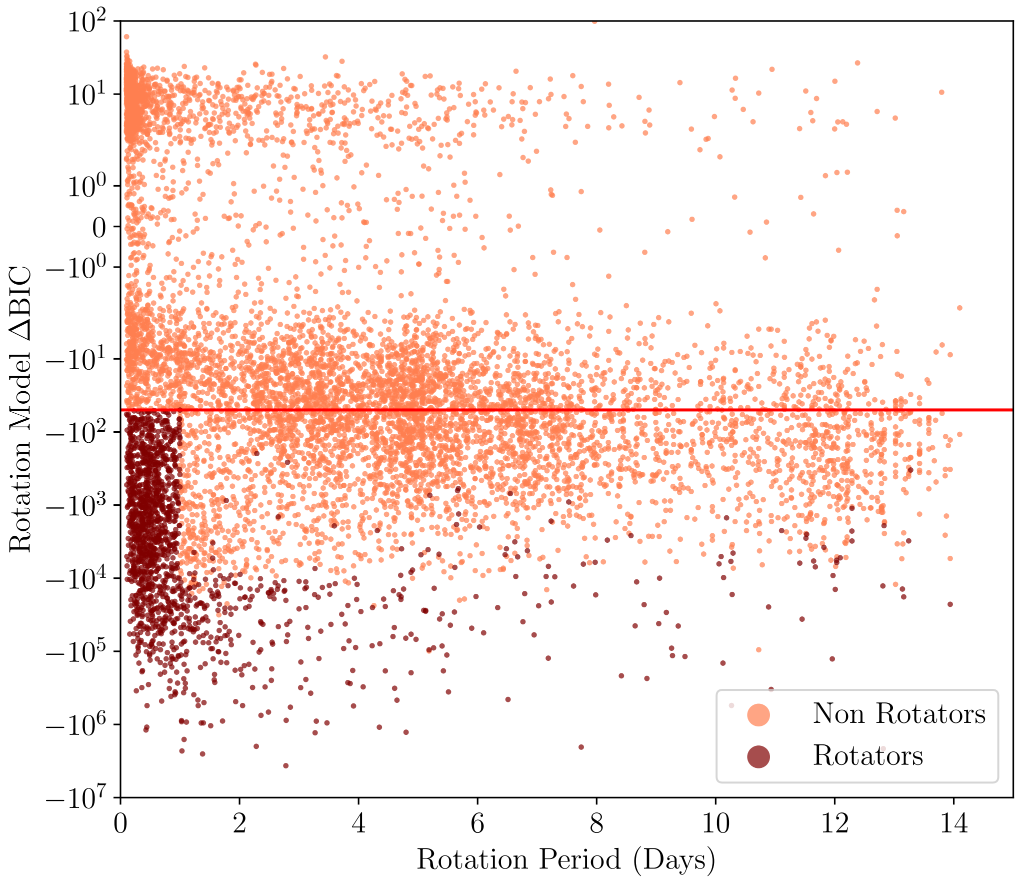

Figure 2.4: Detected rotation period of each of the stars in my sample, plotted alongside how well a ΔBIC test favours a rotational model over a flat median model. Rotators are favoured by the ΔBIC test by at least 50, and must have a rotational amplitude greater than the median photometric error.

Gaussian processes (Rasmussen & Williams, 2006) produce a nondeterministic model of a physical process that is difficult to model physically. It has been shown by Angus et al. (2018) that Gaussian processes can remove the quasi-periodic signals introduced by stellar rotations. Pure sinusoids are an inaccurate model for these rotation signals because of other stochastic and non-periodic contributing factors, and physical models require a large and degenerate parameter space to accurately model the signals we observe, and are unsuitable for deployment at the scale of this survey.

Each light curve with significant rotation is detrended using a Gaussian process with the rotation term kernel provided by `celerite2` (Foreman-Mackey, 2018). This kernel is a combination of two simple harmonic oscillator kernels, with two modes in Fourier space: one at the rotation period and one at one-half the rotation period. The parameters used in this model are summarized in Table 2.1. The power spectral density of the simple harmonic oscillator is given by:

$$S(\omega) = \sqrt{\frac{\pi}{2}} \frac{S_0 {\omega_0}^4}{\left(\omega^2 - {\omega_0}^2\right)^2 + \omega^2 {\omega_0}^2 / Q^2}, \tag{2.2.3}$$

where ω_0 is the undamped angular frequency, Q is the quality factor and S_0 characterizes the power at ${\omega_0}^2$. The rotation term kernel combines two harmonic oscillators with the parameters:

$$Q_1 = 1/2 + Q_0 + \Delta Q \tag{2.2.4}$$

$$\omega_{0,1} = \frac{4\pi Q_1}{P\sqrt{4{Q_1}^2 - 1}} \tag{2.2.5}$$

$$S_{0,1} = \frac{\sigma^2}{(1+f)\omega_{0,1} Q_1} \tag{2.2.6}$$

and

$$Q_2 = 1/2 + Q_0 \tag{2.2.7}$$

$$\omega_{0,2} = \frac{8\pi Q_2}{P\sqrt{4{Q_2}^2 - 1}} \tag{2.2.8}$$

$$S_{0,2} = \frac{f\sigma^2}{(1+f)\omega_{0,2}Q_2}. \tag{2.2.9}$$

The free parameters in equations 2.2.4 through 2.2.9 are outlined in Table 2.1. The effectiveness of this Gaussian process is illustrated by Figure 2.5, where it smoothly removes significant stellar variability and produces a flat light curve.

While Gaussian processes are effective at modeling stellar rotation signals, they can be expensive to run, and can degrade transit signals because the Gaussian process does not distinguish stellar variability from planetary transits (Garcia et al., 2024). For these reasons, my pipeline uses the Gaussian process to detrend where it is deemed necessary by the period, ΔBIC test, and amplitude of the rotational signal.

Parameter	Explanation	Prior
μ	Light curve mean	$\mathcal{N}(0, 10)$
$\log(\sigma_{\mathrm{jitter}})$	Excess white noise	$\mathcal{U}(-3, 2)$
$\log(\sigma_{\mathrm{gp}})$	Spread of initial data	$\mathcal{U}(-3, \log(\sigma_{\mathrm{rms}}))$
σ_{lc}	Measurement uncertainty	$\mathrm{Lognormal}(\log(\bar{\sigma}), 0.1)$
P_{rot}	Rotation period	$\mathcal{N}(P, 0.1)$
Q	Secondary oscillation quality factor	$\mathcal{N}(7.5, 2)$
ΔQ	Difference between modes' quality factors	$\mathrm{Lognormal}(0, 1)$
f	Secondary mode fractional amplitude	$\mathcal{U}(0.8, 1)$

Table 2.1: Summary of the parameters used in the Gaussian Process used to remove the strong stellar rotation signals from light curves. Here $\mathcal{N}$ denotes a normal distribution, $\mathcal{U}$ denotes a uniform distribution, Lognormal denotes a log-normal distribution, σ_{rms} is the standard-deviation of the light curve, P is the period found by the Lomb-Scargle periodigram and $\bar{\sigma}$ is the median flux error.

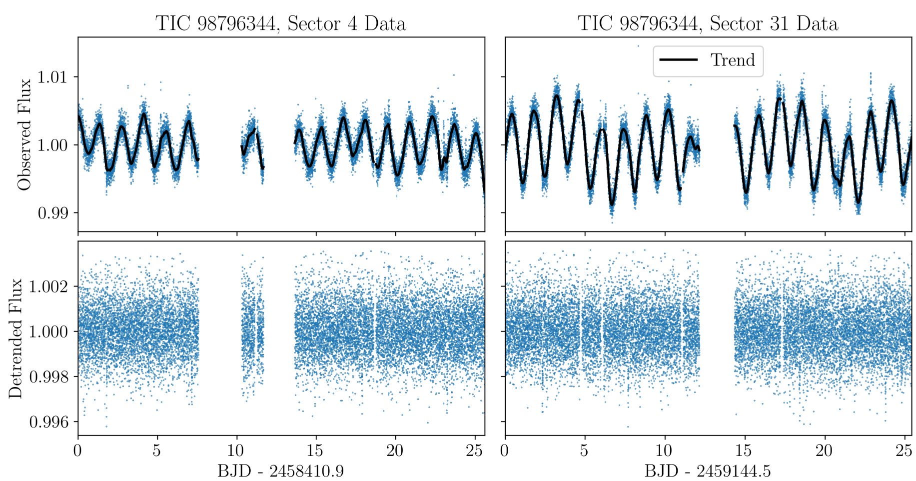

Figure 2.5: Top: TESS light curve from each observed sector of TIC 98796344 with Gaussian Process trend overplotted. Bottom: Detrended light curve with Gaussian Process trend removed.

2.3 Transit Least Squares (TLS)

After each light curve for a target has been cleaned and detrended, my pipeline searches the time series for signals from transiting planets. My pipeline employs the Transit Least Squares (TLS) algorithm developed by Hippke & Heller (2019) implemented in Python with the `transitleastsquares` package. This method improves upon the Box Least Squares (BLS) algorithm developed by Kovács et al. (2002) which has been employed in previous work such as Ment & Charbonneau (2023).

Both the BLS and the TLS employ a similar technique for detecting transit signals. I will sketch out the BLS algorithm and outline the key improvements that the TLS algorithm offers over the BLS. Let us denote a time series of fluxes given by $\{t_1, t_2, ..., t_n\}$ and $\{I_1, I_2, ..., I_n\}$ where (t_i, I_i) is a measurement at some time. In this case, we assume that the median of $\{I_i\}$ is 1, which is the case for our normalized time series. These methods fold the time series on a period P by transforming the time measurements by $t_i{}' = t_i \mod P$. If P is the period of the transiting planet with transit depth Δ, we expect that all of the in-transit points will fall between the values of some $t_j{}'$ and $t_k{}'$. Since all of the points between these values are in transit, we expect that the means of the flux in and out of transit should differ by Δ, or that over indices l where $l < j$ or $l > k$ and m where $j \leq m \leq k$, we expect $\overline{I_l}$ - $\overline{I_m} = \Delta$. The BLS finds j and k and Δ by the least-squares method, finding a two-valued function which takes values of 1 out of transit and $1 - \Delta$ in transit.

The Transit Least Squares algorithm improves over the BLS's approach by fitting a physically modeled transit model to the phase folded time series. The TLS model is also informed by the host star's radius, mass and limb-darkening parameters. This allows for a greater detection efficiency, because the best-fit model better physically resembles the transit signals we are searching for.

The approach outlined above based Kovács et al. (2002) is idealized, but without optimizations the multidimensional search space is very large. Transit Least Squares searches an array of transit periods, mid-transit epochs, and an array of transit durations between a maximum and a minimum transit time. The inner and outer ranges of the duration search are set assuming a circular orbit, but allowing a large range of planet and host star masses. The range of durations that TLS samples encompasses all known exoplanets at the time that Hippke & Heller (2019) was published. Accurately sampling the period is particularly important because a proper fit must line up each transit in the same folded time range. In a light curve spanning D days we expect to be able to find transit signals with periods anywhere from 0.1 to $D/2$ days. A planet with a period of 2 days around a 0.3 $M_{\odot}$ star has an expected transit duration of about 1 hour. In one sector of TESS data, spanning 28 days, we expect 14 transits from a 2 day planet. If we require our chosen period to fold all of the transits within the transit duration with a tolerance of 25%, we require the chosen period to be sampled within 8×10^{-4} days of the true period of 2 days. The consequences of misaligning the transits in the folding process compound with longer light curves. The accuracy required grows linearly with the length of the light curve, as does the upper limit of the period search. This scaling of computational cost motivates limiting the maximum number of concatenated consecutive sectors to four.

Once the TLS is run, it produces a signal detection efficiency (SDE) spectrum illustrated in Figure 2.6 along with a best-fit period, mid-transit time, transit depth, and transit duration. The best-fit values represent the peak in the signal detection efficiency of the TLS. My pipeline considers this result significant if it corresponds to

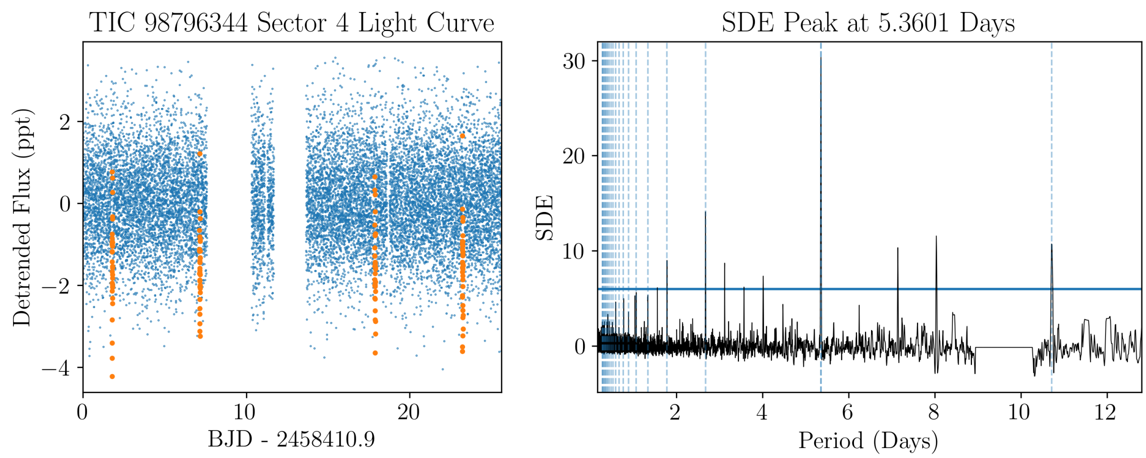

Figure 2.6: Left: Single sector detrended light curve with the in-transit points found by the Transit Least Squares (TLS) highlighted in orange. Right: TLS spectrum indicating a strong 5.3601-day periodic transit-like signal. Harmonics of the signal are marked with dashed lines.

a peak in the SDE with SDE>6. According to Hippke & Heller (2019), a peak in the SDE spectrum greater than 6 corresponds to a 0.05 false alarm probability, in tests with white noise. While a higher SDE threshold would result in fewer false positive TLS results, it would also reduce the sensitivity of my pipeline to transiting planets.

Many of the targets in my sample are known to host multiple planets, so my pipeline must be able to find multiple planet signals in a given light curve. Should the TLS return a result with SDE ≥ 6 for a light curve, the search continues iteratively, but if the search returns an insignificant result the search ends. When a significant signal be found, my pipeline uses the TLS's best-fit parameters to fit a physical transit model using scipy's optimization routine (Virtanen et al., 2020) and `batman` (Kreidberg, 2015) to generate transit curves. This physical transit model is subtracted from the light curve to remove the signal that the TLS has identified. Once the signal has been removed, the TLS is run again on the new light curve. This process continues iteratively until the TLS fails to find a significant result or a maximum number of four iterations is reached. I choose four as the maximum number of consecutive iterations because no targets within my sample have more than three confirmed transiting planets within the period range which we are sensitive to. This iterative TLS step is illustrated in Figure 2.7.

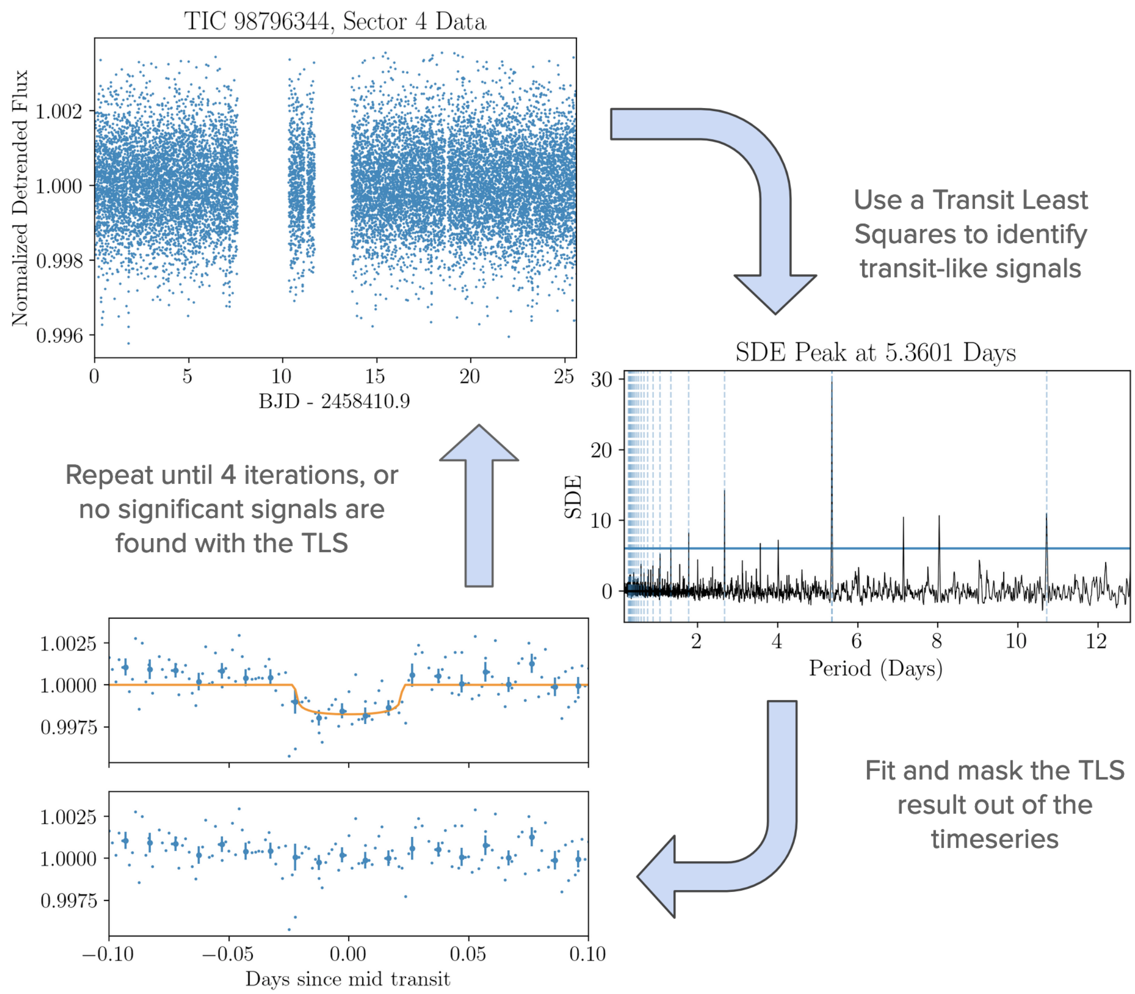

Figure 2.7: An illustration of the pipeline's iterative TLS search. Starting with a detrended light curve (Top), a TLS is run to generate an SDE spectrum (Left). If a significant signal is found with SDE≥ 6, the signal is modeled and masked out of the light curve (Bottom). This process continues iteratively until a maximum or four iterations are reached, or no significant signals are found.

Many targets in our sample have multiple light curves corresponding to different sectors of observations. For each of the light curves from a given target, this iterative TLS search yields a set of results. At this point, many of these results represent astrophysical and statistical false positives, and my pipeline employs a vetting procedure to flag and remove false positive results that would otherwise contaminate our occurrence rate inferences. Using the vetting procedures described in Ment & Charbonneau (2023) as a reference, each TLS result must pass the following four vetting conditions:

1. The best-fit physical `batman` transit model must have a signal-to-noise ratio (SNR) greater than three calculated using Equation 1.3.2.
2. The TLS result cannot have the same period, or be an integer multiple of the star's rotation period within a 0.3% tolerance. TLS results with SDE >15 are exempt from this condition.
3. The period of a TLS result cannot match the period of another TLS result from the same light curve within a tolerance of 0.1%.
4. In targets with four or more light curves, the TLS result's period must match the period of results from half the light curves or more, rounded down.

To compute the signal-to-noise ratio to fulfill vetting condition 1, with an empirical characterization of σ_{lc} using the detrended light curve. The number of transits N_{transits} is counted based on the number of in-transit points times the cadence, divided by the transit duration. σ_{lc} is calculated by taking the detrended light curve and binning it with bin lengths equal to the transit duration. σ_{lc} is then calculated by taking the standard deviation of this binned light curve. These vetting conditions do not rule out all false positive TLS results, but it reduces their number, meaning that the next steps of the pipeline are less polluted. These vetting conditions are necessary to reduce the final number of false positives from this survey to make manual vetting possible.

2.4 Characterizing Planet Candidates

The next step of the processing pipeline involves transforming TLS results from each light curve into a set of planet candidates characterized using all light curve data from the target at once. This was not possible in earlier steps due to the prohibitive cost of the TLS over very long light curves. From the sets of TLS results the iterative search generates, the pipeline collects sets of results with common depths within 20%, durations within 40% and periods whose ratio is within 0.01% of an integer. The period condition is motivated by the fact that transit signals can be picked up as period aliases by the TLS and long-period transiting planets in short light curves may still be detected at shorter period harmonics. The large ranges on durations and depth are necessary to account for these misfolded signals. This process of collecting sets of results is illustrated in Figure 2.8 using the results from the TIC 98796344 as an example.

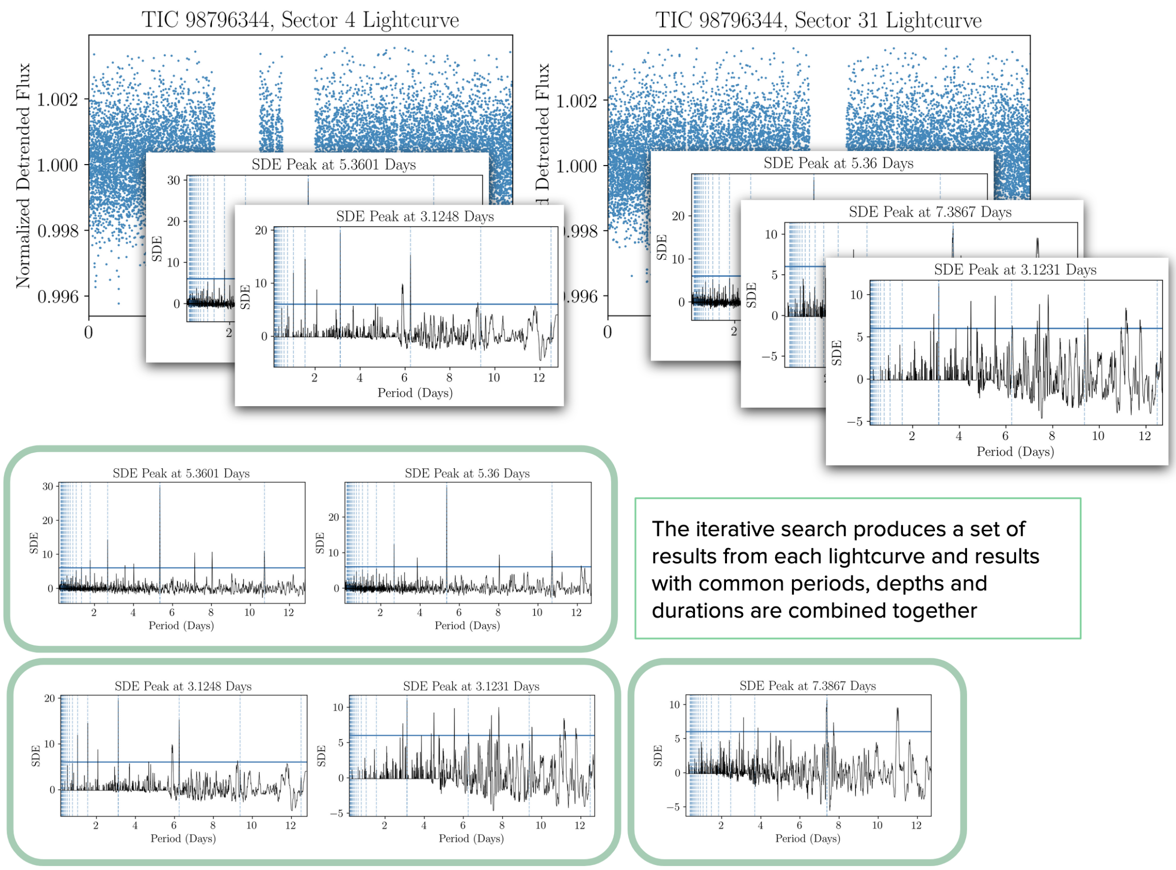

Figure 2.8: Illustration of the sets of results from each light curve being combined to characterize planet candidates.

Each set of results is used to characterize a planet candidate for which the pipeline finds a best-fit model using the concatenated light curve data from the target. First a very narrow TLS search is run within the period uncertainty of the result with the highest SDE. Generally, a search over a long concatenated light curve is prohibitively expensive, but in this case the very narrow range makes it tractable because only a few hundred trial periods must be searched. Using the period fit by this narrow TLS allows the pipeline to be confident that the transits will be folded correctly. With this period, the pipeline then uses the folded light curve to fit the remaining transit parameters T_0, R_p/R_*, b as well as an offset parameter to better characterize the transit baseline using `scipy`'s `curve_fit` (Virtanen et al., 2020) with `batman` transit models (Kreidberg, 2015). This fit and subsequent vetting and characterization is only run on the data within 2 transit durations of the folded transit's midpoint, because the transit only represents a small fraction of the folded data. The best-fit model for the two planets around TIC 98796344 is illustrated in Figure 2.9.

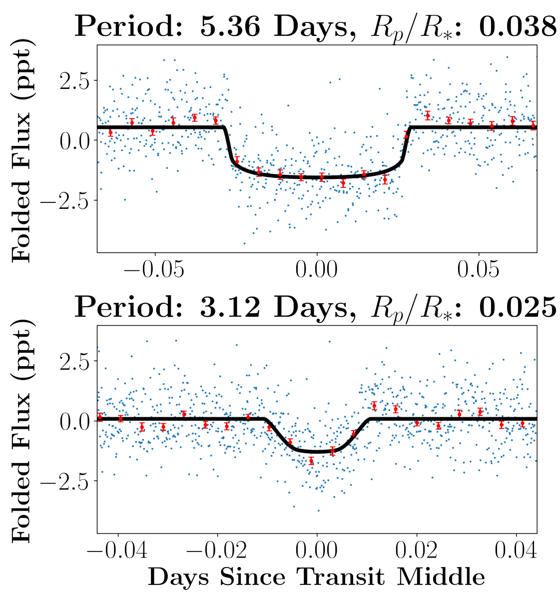

Figure 2.9: Fit models for the two planets around TIC 98796344, fit using the combined results laid out in Figure 2.8.

At this stage, the best-fit model must have an SNR > 6 calculated according to equation 1.3.2, derived using the same process described in the TLS vetting conditions. The pipeline further characterizes these planet candidates by running a Markov Chain Monte Carlo (MCMC) with the `emcee` package (Foreman-Mackey et al., 2013) to model the posterior distributions of the 5 fit quantities. These quantities and their priors are:

- T_0: Gaussian centered at the T_0 found by the narrow TLS, with width given by the TLS's T_0 uncertainty.
- P: Gaussian centered at the period found by the narrow TLS, with width given by the TLS's period uncertainty.
- R_p/R_*: Log-uniform from 0.005 to 0.5.
- b: Uniform between 0 and 0.95.
- Offset: Gaussian centered at 0 with width given by the standard deviation of the light curve within 2 durations of the transit midtime.

Before running the MCMC, the fit planet models of the other planet candidates are removed from the light curve by subtracting their modeled transit curves. The MCMC runs 48 walkers for 4,000 steps to explore the posterior of the fit parameters. The evolution of the walkers for a single planet candidate is illustrated in Figure 2.10. The burn-in and number of steps were calibrated based on tests with targets hosting known transiting planets. Fewer steps and a shorter burn-in are likely sufficient but the MCMC does not dominate the runtime, so a long run and burn-in are favoured. The walkers' initial positions are sampled from the priors, with the exception of R_p/R_* which is sampled from a Gaussian centered at $\sqrt{\Delta}$, with width $\sqrt{\Delta}/20$ where Δ is the narrow TLS's found depth. The full MCMC chains are saved, as well as a truncated chain with 1,000 steps of burn-in removed which characterizes the planet parameters' posterior distributions. One of these distributions is illustrated in Figure 2.11.

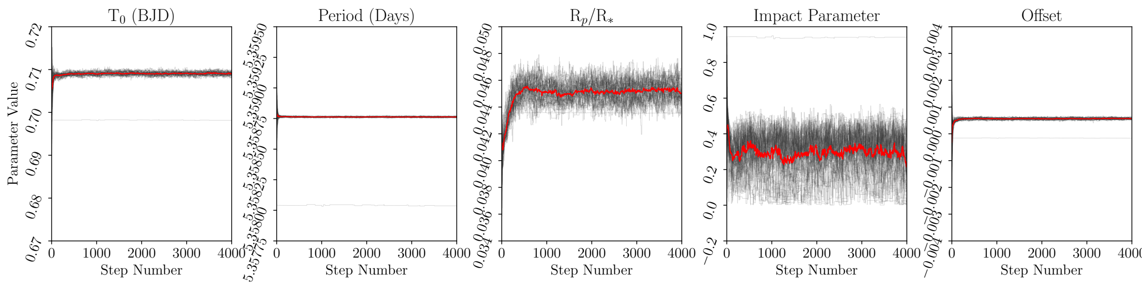

Figure 2.10: Evolution of the Monte Carlo Markov Chain walkers fitting the 5.36 day planet candidate around TIC 98796344. The black lines trace out each walker's path, with the red line tracing the median walker position at each step.

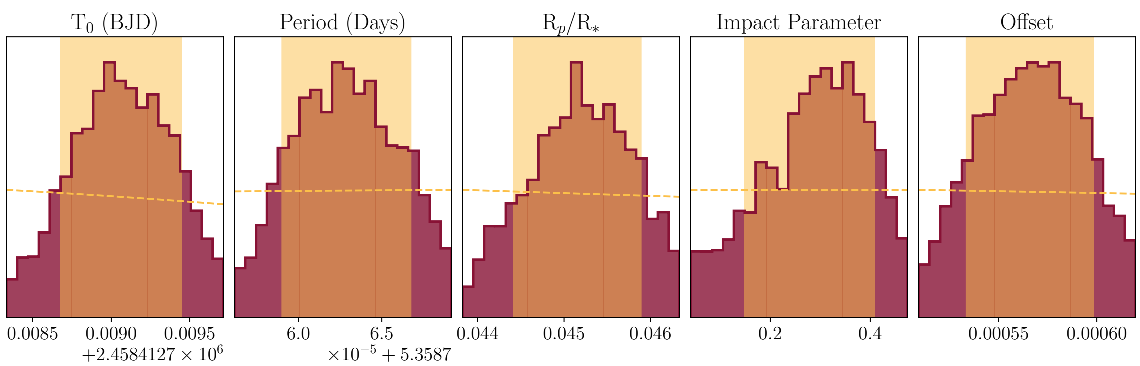

Figure 2.11: Posterior distributions for the planet parameters of the 5.36 day planet candidate around TIC 98796344. The prior distributions are overplotted as dashed lines, and the 1σ confidence interval is shaded.

After each planet candidate is characterized by the MCMC, it must pass a final set of statistical vetting conditions, calibrated to remove false positives in searches of lightcuves without planet signals:

1. The fit model must pass a χ^2 test with a value of 1.5 or less.
2. A ΔBIC test must favour the fitted transit model over a median model, which is constant and equal to the median of the folded light curve within 1 transit duration of the transit midpoint by -10 or more.

3. A ΔBIC test must favour the fitted transit model over a null model, which is constant and equal to the fit offset within 1 transit duration of the transit midpoint by -10 or more.

These statistical tests do not rule out all false positives, but they reduce their numbers enough that a visual verification of every planet candidate is feasible over our a large number of. If a planet candidate passes all of these conditions it must finally pass a visual check using the diagnostic plots outlined in Appendix A. The planets identified by my pipeline are outlined in Chapter 4.

Chapter 3

Pipeline Performance & Sensitivity Characterization

The set of detected planets is on its own insufficient to characterize the true population of planets. Accurately characterizing the planet population requires combining the found population with detailed understanding of my pipeline's sensitivity to different types of transiting planets. The latter of these necessary parts is outlined in this chapter, along with my pipeline's ability to find known planet signals in the targets of my survey.

3.1 Finding TOIs

All of the data in this survey have been previously processed by the proprietary TESS pipeline, which has identified a set of known and probable planet signals called TESS Objects of Interest (TOIs). Each TOI is recorded by the Exoplanet Follow-up Observing Program (ExoFOP) and has been assigned a disposition by the TESS follow-up program working group (TFOPWG). There are 86 credible TOIs hosted by the targets in this survey labeled as a *Confirmed Planet, Planet Candidate* or *Known Planet*, with TOIs labeled *False Alarm* or *False Positive* disregarded.

As part of my pipeline's testing, I use this set of TOIs to determine how well my pipeline is able to recover transiting planet signals, without an exhaustive, extensive set of injection-recovery tests such as those described in section 3.2. In its current state, my pipeline is able to recover 70 out of the 86 TOIs, and 34 out of the 36 confirmed planets. These found and missed TOIs are illustrated in Figure 3.1 with the confirmed planets highlighted. I have also determined why each missing TOI was not found, which is tabulated in Table 3.1. The vetting condition that a TLS result of a given period or period harmonic must result be found in half of the light curves or more was responsible for more than half of the missed TOIs. Relaxing this condition for targets with many light curves will result in fewer missed TOIs. This is not currently feasible because maintaining the consistency of this survey and my sensitivity calculations would also require rerunning the transit search and all injection-recovery tests on targets which the new condition may affect.

3.2 Injection-Recovery

I have empirically measured my pipeline's sensitivity using injection-recovery tests. Starting with light curves without planet signals, I inject a transit signal, and process the synthetic light curve using my end-to-end pipeline to determine whether the signal is recovered. The recovery fraction of these injected signals, over many targets and injected planets reveals my pipeline's sensitivity to transiting planets as a function of their injected periods and radii.

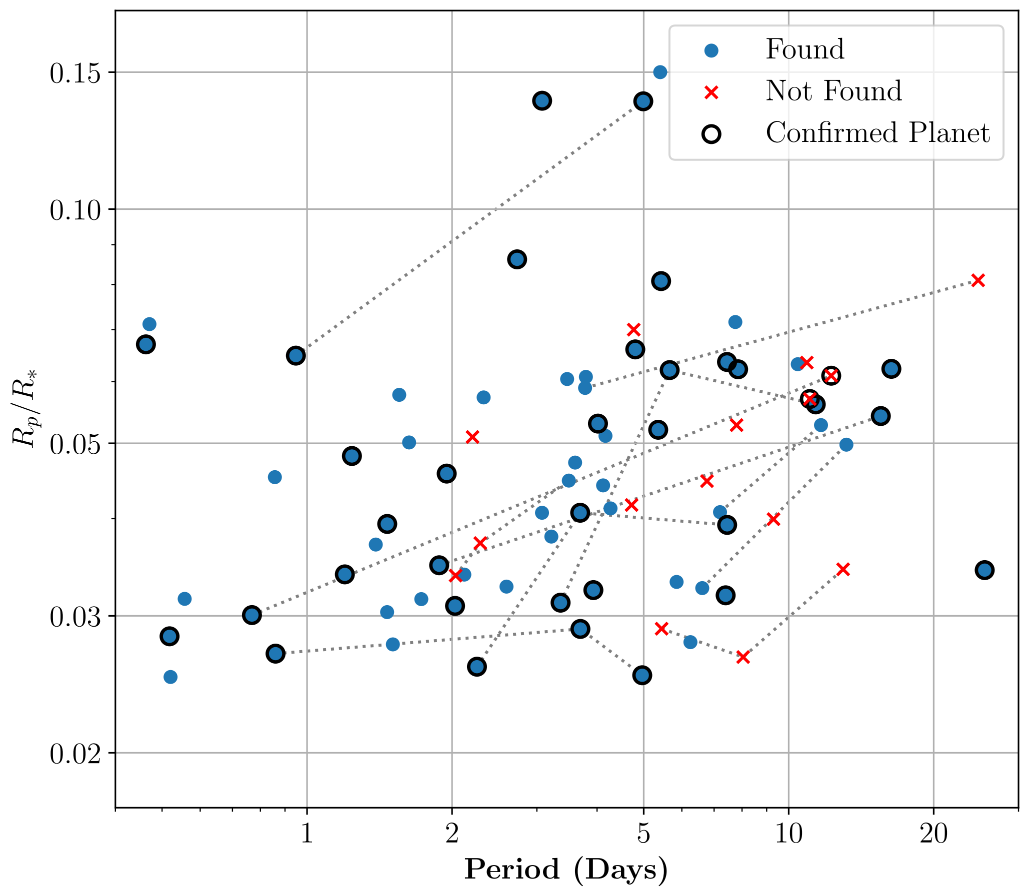

Figure 3.1: Found and not Found TESS TOIs from the stars in our sample, plotted as a function of period and radius. Multiplanet systems are connected with dashed lines, and confirmed planets are encircled.

Before running injection-recovery tests, I split up my sample into three sets based on their TESS magnitude. These sets have TESS magnitude <12, 12-13 and >13 respectively, and roughly represent one third each of my sample of stars. As TESS magnitude is one of the main factors contributing to photometric noise, splitting up the sample of stars will allow me to more sharply resolve the boundary between recoverable and non-recoverable planet populations. Because my pipeline treats stars with strong rotating signals differently, I further subdivide these sets into rotating and non-rotating stars using the criteria described in Chapter 2.

Before injecting a signal in a target's light curve, my injection-recovery pipeline masks any known planet signals in the light curve by cross-referencing the chosen light curve's TIC with ExoFOP's TOI table and removing the data associated with the TOIs in transit times. Each injected signal is generated by using `batman` (Kreidberg, 2015) with some prescribed T_0, b, period P, and R_p/R_*, and are added to the pre-detrended light curve. After the injection, the entire pipeline is run to recover the signal as a planet candidate. Following Ment & Charbonneau (2023), a planet signal is considered recovered if the pipeline finds and vets a planet candidate with recovered period P' which matches the injected period P, Pn or P/n where $n \in \{2, 3, 4, 5, 6\}$ within a tolerance of 0.2%. This validation strategy follows Ment & Charbonneau (2023). For each injection, the sampled TIC, T_0, b, P and R_p/R_* are saved alongside the recovered T_0', b', P' and R_p'/R_*.

Within each TESS magnitude bin, I am primarily interested in characterizing my pipeline's sensitivity as a function of period from 0.2 to 30 days and R_p/R_* from 0.01 to 0.2. Our injection strategy follows Ment & Charbonneau (2023) which explores the recovery space recursively using a quadtree. This process starts with 200 injections over four quadrants in period vs R_p/R_* space to create four regions of sensitivity. These injections are drawn from log-uniform distributions in R_p/R_* and period, with T_0 and b drawn uniformly from 0 to the sampled period and 0 to 1 respectively. This process is repeated recursively for any quadrant whose recovery rate differs from a neighboring area by more than 20%. This sampling scheme means that we sample regions with a high gradient in recovery fraction most finely, without dedicating many samples to regions where the recovery fraction is relatively uniform.

In total, I have injected 131,662 planet signals total over the six sets of targets described above using the resources provided by the Digital Research Alliance of Canada. While injection-recovery tests are still ongoing (subject to resource availability) the calculations completed so far provide a nearly complete picture of my pipeline sensitivity. I do not expect the forthcoming refinements to my sensitivity calculation to have a major effect on my planet occurrence rate inferences. The recovery fraction as a function of period and R_p/R_* in the recursively sampled regions, alongside a scatter-plot of injected and recovered planet signals for each set of targets is plotted in Figures 3.2 and 3.3. Two main features dominate the shape of the sensitivity in each plot: An upward slope from the bottom left to the top right comes from the SNR cutoff imposed by the pipeline, and the drop-off at large radii and periods comes from my pipeline's inability to detect long period planets in single sectors of TESS data.

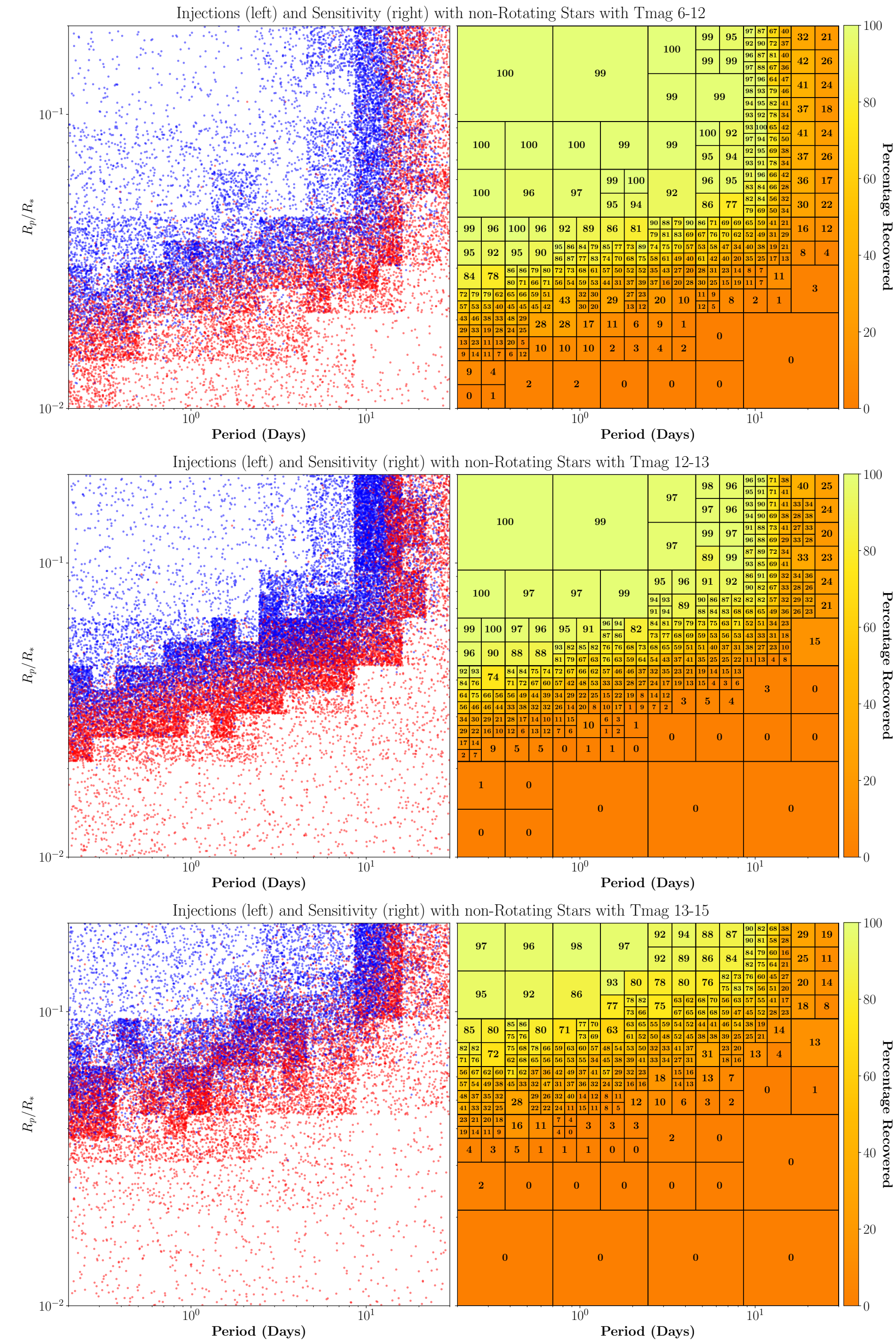

Figure 3.2: Left: Sensitivity of my pipeline to transiting planets around non-rotating stars as a function of period and R_p/R_*. Right: All injection-recovery tests to characterize the left plot, coloured blue if recovered, red if not recovered.

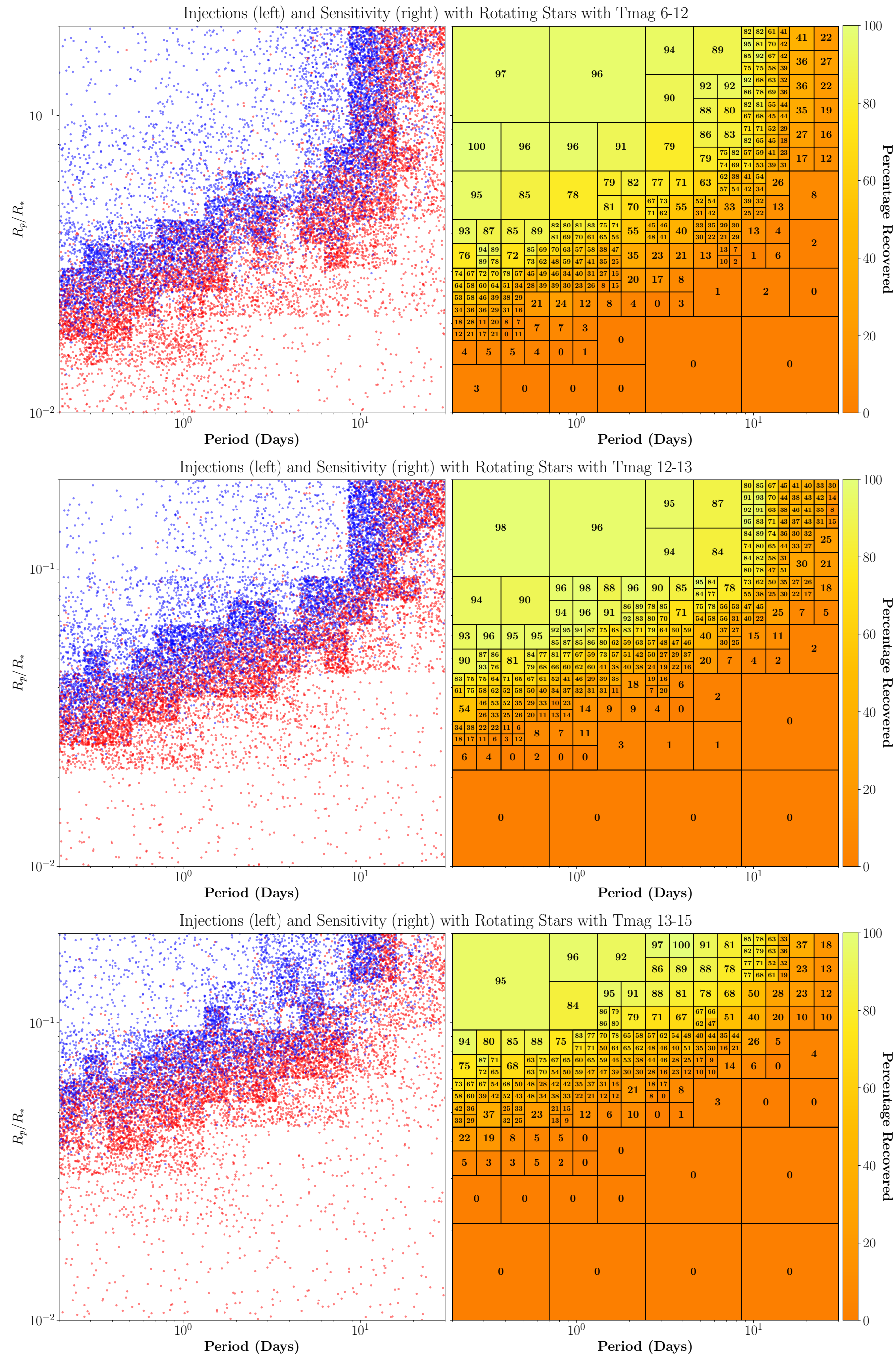

Figure 3.3: Same as Figure 3.3 but for non-rotating stars

TOI	TLS Result	TFOPWG Disposition	Reason for non-detection
256.01	×	KP	Known planet, possibly not recoverable from TESS light curve
732.02	✓	CP	Planet candidate fit failed the χ^2 test
789.01	✓	PC	Found TLS results fail the half light curves or more condition
789.02	✓	PC	Found TLS results fail the half light curves or more condition
789.03	✓	PC	Found TLS results fail the half light curves or more condition
1452.01	✓	CP	Found TLS results fail the half light curves or more condition
2142.01	✓	PC	Median model favored by ΔBIC test
2267.02	✓	PC	Found TLS results fail the half light curves or more condition
2267.03	✓	PC	Found TLS results fail the half light curves or more condition
2495.01	✓	PC	Planet Candidate SNR < 6
5716.01	✓	PC	Found TLS results fail the half light curves or more condition
6002.01	×	PC	10 day TOI not found by TLS
6255.01	✓	PC	Found TLS results fail the half light curves or more condition
6599.01	✓	PC	Planet Candidate SNR < 6
6714.01	×	PC	TOI signal not found by TLS
6716.01	×	PC	TOI signal not found by TLS

Table 3.1: TESS targets of interest missed by my pipeline's planet survey. Left-middle column indicates where the TOI signal was found by the TLS, right-middle column records the TFOPWG Disposition, and the rightmost column describes why the TOI was not found.

Chapter 4

The Occurrence Rate of Planets around mid-to-late M Dwarfs

In this chapter, I describe the results of my transit survey followed by a calculation of the occurrence rate of close-in planets around mid-to-late M dwarfs. These results represent the main scientific findings of my thesis and the occurrence rate around the smallest and most common stars. In the discrete calculations following, I use an 18x18 log-log grid with a period range of 0.2 to 30 days, a radius range of $0.5R_\oplus$ to $6.5R_\oplus$ and an instellation range of $0.1S_\oplus$ to $1000S_\oplus$.

4.1 Transit Survey Results

Using my pipeline described in Chapter 2, I have processed data from 9,131 mid-to-late M dwarfs observed in the first 85 sectors of TESS observation. This survey was run using the resources available from the Digital Research Alliance of Canada. After each target was processed, every planet candidate identified was manually vetted by visually inspecting the best-fit folded transit curve, the TLS spectra which led to the planet candidate and the individual transits in the light curves.

In total, I have found and characterized 80 signals which pass all of the aforementioned vetting conditions. Of these, 71 are consistent with TOIs or known planets. For the nine non-TOI signals, I have used `TRICERATOPS`, (Giacalone et al. (2021) and Giacalone & Dressing (2020)) which is a tool for statistically validating signals from TESS light curves, to compute the probability that the signal may be an astrophysical false positive. `TRICERATOPS` accomplishes this by retrieving the properties of all of the stars within the photometric aperture used to generate the target's light curve, and determines the proportional flux contributed from each star. `TRICERATOPS` uses this information to calculate the marginal likelihood of each transit-producing scenario and determines the probability of each scenario. These comprise eclipsing binary signals from background stars or unresolved bound companions, as well as transiting planets around background stars or unresolved bound companions. Applying `TRICERATOPS` to the set of found TOIs, I find that none have a false positive probability (FPP) greater than 8%. For this survey, I will include any planet candidate signal with a `TRICERATOPS` FPP less than 10%. The `TRICERATOPS` FPPs and other reasons for rejecting or accepting the 9 non-TOI signals are summarized in Table 4.1. The signal from TIC 420112587 is strong, but its period closely matches that of TOI 1452.01, a confirmed planet around another star in the aperture.

TIC	Measured R_p/R_*	Period (Days)	FPP	Status
420112587	0.0544 ± 0.0019	$11.0619748 \pm 7.9 \times 10^{-6}$	-	Background TOI
170636897	0.0213 ± 0.001	$0.6927215 \pm 6.5 \times 10^{-7}$	8.7%	Accepted
355815567	0.0487 ± 0.0028	$3.519462 \pm 1.9 \times 10^{-5}$	10.6%	Rejected
417732194	0.0594 ± 0.003	$5.3074071 \pm 7.4 \times 10^{-6}$	17.3%	Rejected
96202086	0.0237 ± 0.001	$0.63048452 \pm 3.1 \times 10^{-7}$	11%	Rejected
149927512	0.0311 ± 0.0018	$1.4646799 \pm 1.8 \times 10^{-6}$	8.4%	Accepted
61710094	0.0278 ± 0.0014	$1.8389543 \pm 1.4 \times 10^{-6}$	12%	Rejected
171796577	0.0715 ± 0.0046	$3.64946 \pm 7.2 \times 10^{-4}$	10.2%	Rejected

Table 4.1: Table of credible signals found through my survey which are not confirmed planets or TESS TOIs. Signals with a `TRICERATOPS` FPP$<$ 10% are included in my survey.

For each accepted signal, I compute the physical radius of the transiting planet using the host star's radius from the TIC. I also compute each planet's instellation as:

$$S_p = S_\oplus \left(\frac{T_*}{T_\odot}\right)^4 \left(\frac{P}{1\text{ year}}\right)^{-4/3} \left(\frac{R_*}{R_\odot}\right)^2 \left(\frac{M_*}{M_\odot}\right)^{-2/3} \tag{4.1.1}$$

using the star's effective temperature, mass and radius from the TIC. All of the planets found by my survey are plotted in Figure 4.1, based on their radius, period, and instellation. At this point, we can already see that my survey have found a large number of super-Earths (1-1.7 $R_\oplus$), some sub-Earths (0.5-1 $R_\oplus$) but very few sub-Neptunes (2-5 $R_\oplus$). This data is not sufficient to characterize the planet population, however, as my pipeline's likelihood of detecting these planets differs over the range of our survey which is computed in the following section.

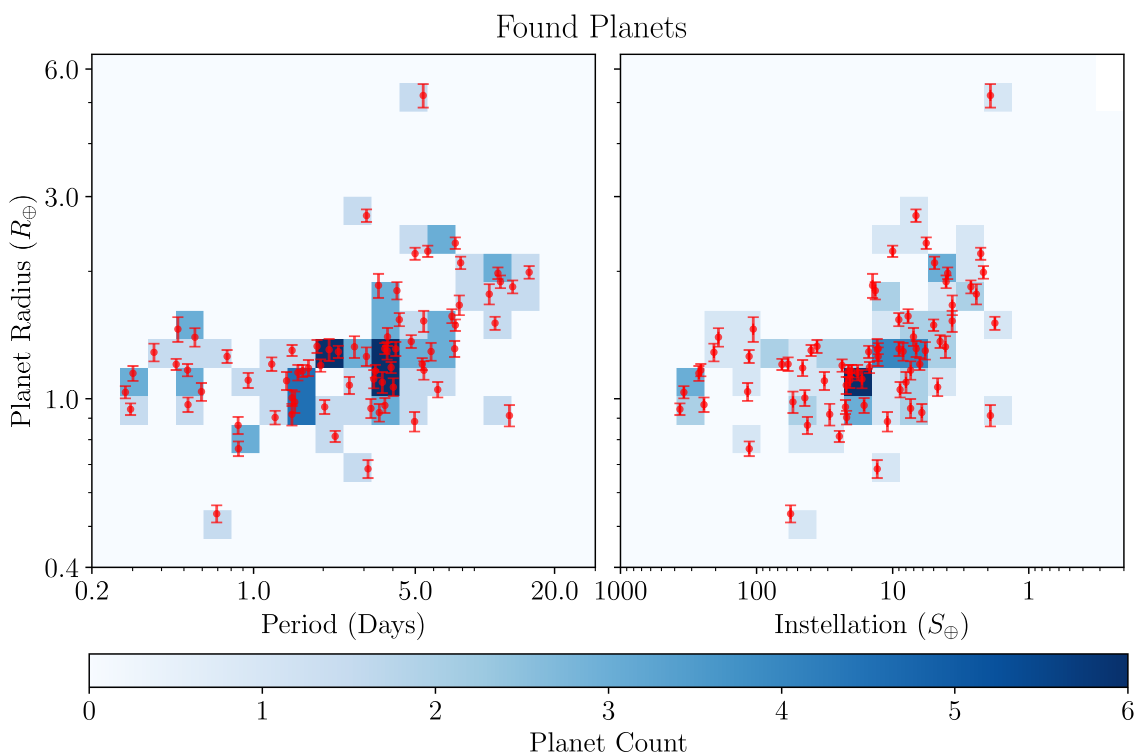

Figure 4.1: Planets found by my survey plotted on 18x18 log/log grids as a function of their periods, radii and instellation.

4.2 Survey Completeness

The completeness of my survey describes the average probability that my pipeline has of detecting any planet around any star in my sample. Computing the completeness requires the pipeline's sensitivity and the geometric transit probability, i.e. the probability that a planet of a given period and radius will transit its host star. I compute the completeness as a function of period and R_p (i.e. physical planet radius) and as a function of radius and instellation. The former allows my calculation to be most easily compared to other established occurrence rates. The latter allows for

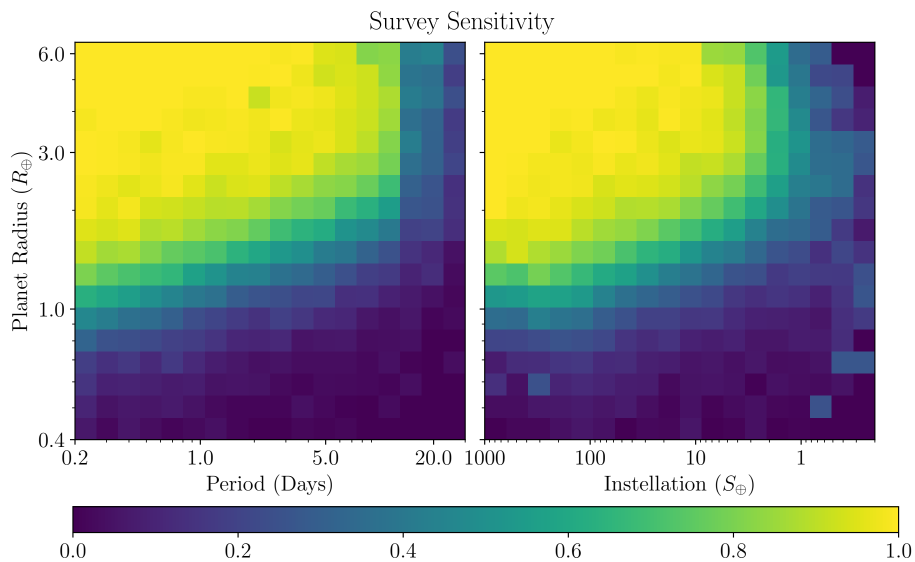

Figure 4.2: Sensitivity of my pipeline to transiting planets on 18x18 log/log grids as a function of period and radius (Left) and as a function of instellation and radius (Right). This map was generated by transforming the injections illustrated in Figures 3.2 and 3.3 and weighting the sensitivity map for each bin by the number of targets in each bin.

an occurrence rate calculation that can link the physical properties of the planets to radiation they receive from their host star, and roughly determine which planets may experience Earth-like levels of instellation.

The sensitivity of my pipeline is computed using the injection-recovery tests outlined in Chapter 3. Each set of injections illustrated in Figures 3.2 and 3.3 is transformed and binned in period/radius space and instellation/radius space. These sensitivity maps are then combined together, weighted by the proportion of stars they represent in the sample. The combined sensitivity maps are illustrated in Figure 4.2.

The transit probability is the probability that a planet at an orbital distance from its star with a random inclination will transit. Assuming our orbits circular, the distance between the planet and it's host is given by:

$$a = 1\text{AU}\left(\frac{P}{1\text{ year}}\right)^{2/3}\left(\frac{M_*}{M_\odot}\right)^{1/3}, \tag{4.2.1}$$

and the transit probability is computed as the ratio of the host star's radius with the orbital distance, following Dressing & Charbonneau (2015) as

$$P_{\text{transit}} = R_*/a \tag{4.2.2}$$

The transit probability is illustrated in Figure 4.3.

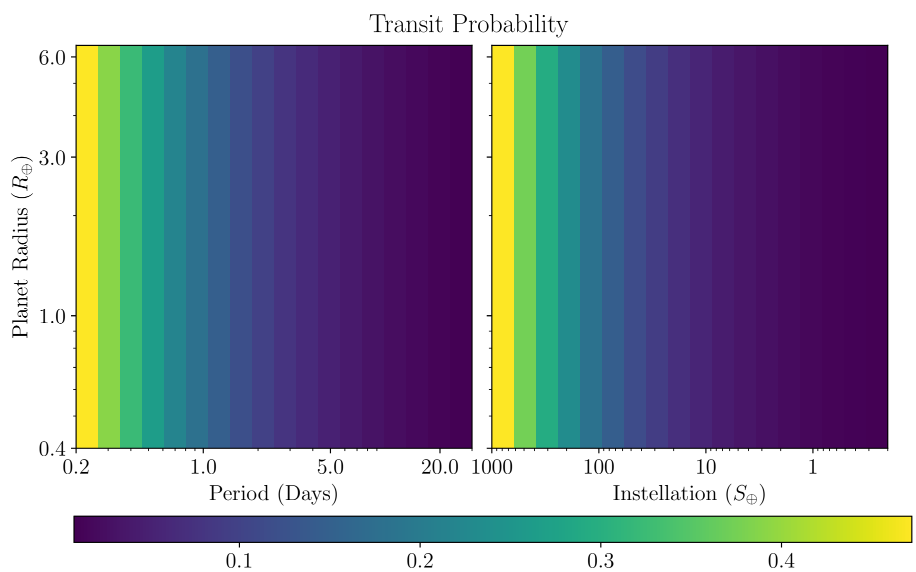

Figure 4.3: Average geometric transit probability of planets around stars in the survey's sample on 18x18 log/log grids as a function of period and planet radius (Left) and as a function of instellation and planet radius (Right).

The survey completeness is calculated as the product of the transit probability and detection sensitivity as a function of period and radius or instellation and radius, which is illustrated in Figure 4.4.

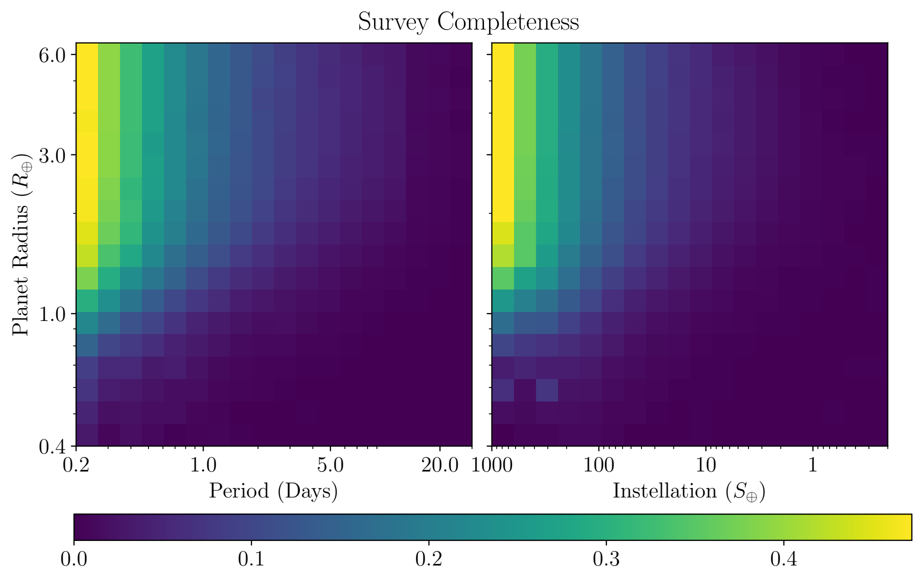

Figure 4.4: Completeness of the survey on 18x18 log/log grids, which is the product of my pipeline's sensitivity (Figure 4.2) and the transit probability of planets over the survey's sample (Figure 4.3).

The completeness c is combined with the planet count $N_{\text{planets}}(R_p, P)$ and the target count N_{stars} to compute the occurrence rate f as:

$$f(R_p, P) = \frac{N_{\text{planets}}(R_p, P)}{c(R_p, P) N_{\text{stars}}} \tag{4.2.3}$$

The binned occurrence rate calculated using the data in Figures 4.1 and 4.4 is plotted in Figure 4.5 as a function of period and radius, and in Figure 4.6 as a function of instellation and radius.

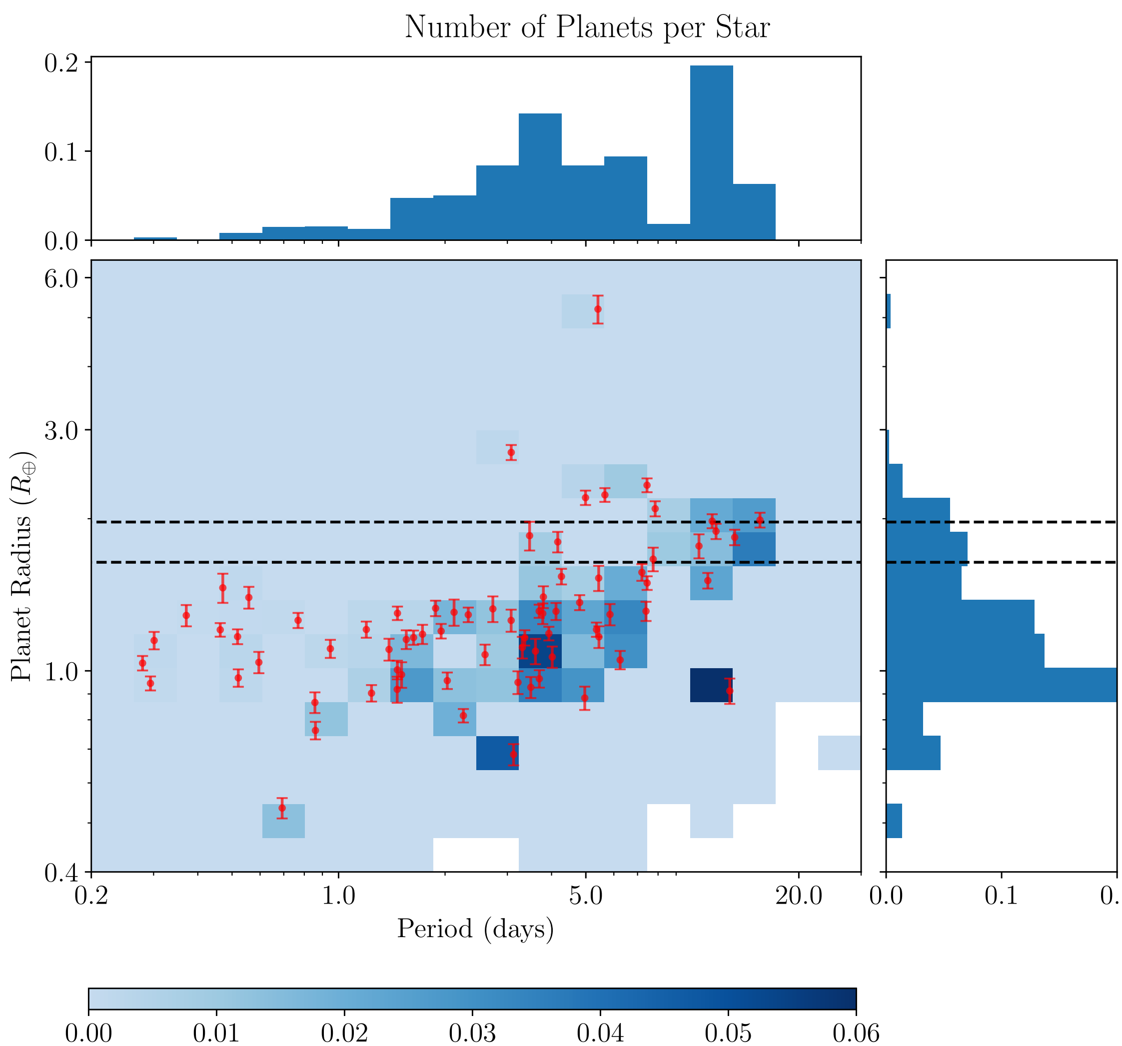

Figure 4.5: The occurrence rate of planets around mid-to-late M dwarfs as a function of period and radius computed by combining the planets found in Figure 4.1 and with the survey's completeness in Figure 4.4. Found planets are marked in red, and the histograms represent the cumulative distribution over their respective axes. The Fulton et al. (2017) Radius Valley highlighted between 1.64 and 1.97 $R_\oplus$.

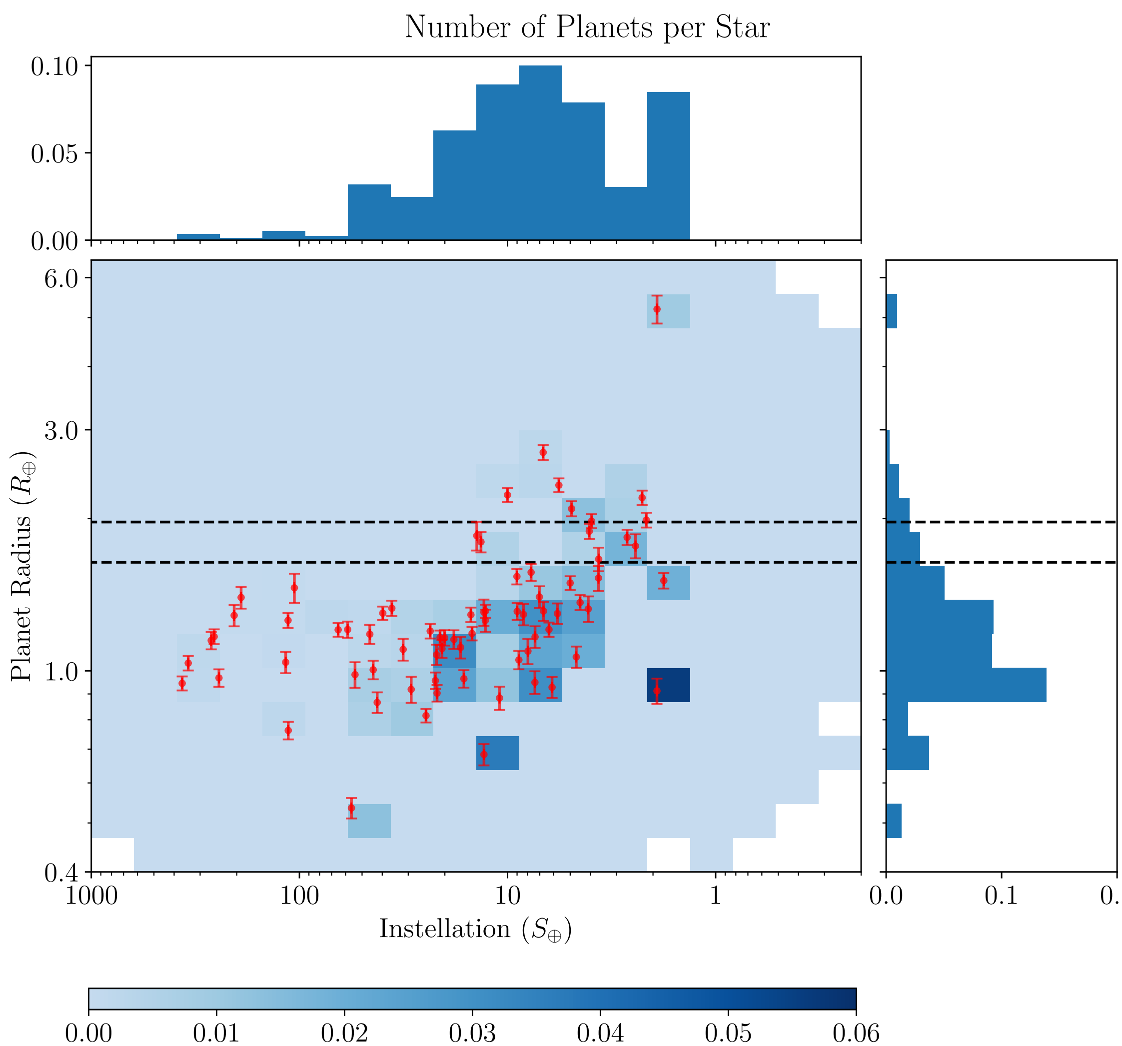

Figure 4.6: The occurrence rate of planets around mid-to-late M dwarfs as a function of instellation and radius computed by combining the planets found in Figure 4.1 and with the survey's completeness in Figure 4.4. Found planets are marked in red, and the histograms represent the cumulative distribution over their respective axes. The Fulton et al. (2017) Radius Valley highlighted between 1.64 and 1.97 $R_{\oplus}$.

It is clear from these occurrence rate calculations that the Radius Valley does not exist around mid-to-late M dwarf stars. The close-in planet population is made up almost entirely of super- and sub-Earths which were overall less likely to be detected than the few sub-Neptunes my survey finds.

To analyze the uncertainty in this distribution we adopt binomial statistics following Dressing & Charbonneau (2015). This is valid because my occurrence rate calculation can be viewed as a set of Bernoulli experiments in each bin whose probability of success depends on the occurrence rate. The posterior probability distribution for the occurrence in each cell is given by:

$$P(\mathrm{Occ.} = f) = \mathrm{Bin}(k, n, fc) \tag{4.2.4}$$

where n is the total number of targets searched, c is the completeness in the cell, and k is the number of planets found in that cell. By sampling this posterior for each cell, I generate many realizations of the binned occurrence rate in period/radius space and radius/instellation space. I do not sample from cells with $k = 0$, because many distributions with a non-zero tail could add an arbitrarily large amount of occurrence to the cumulative rate; however, the distributions in regions with zero occurrence can still be used to set an upper limit on the occurrence of planets therein. Using these occurrence rate realizations, I plot the occurrence rate of planets around mid-to-late M dwarfs in Figure 4.7.

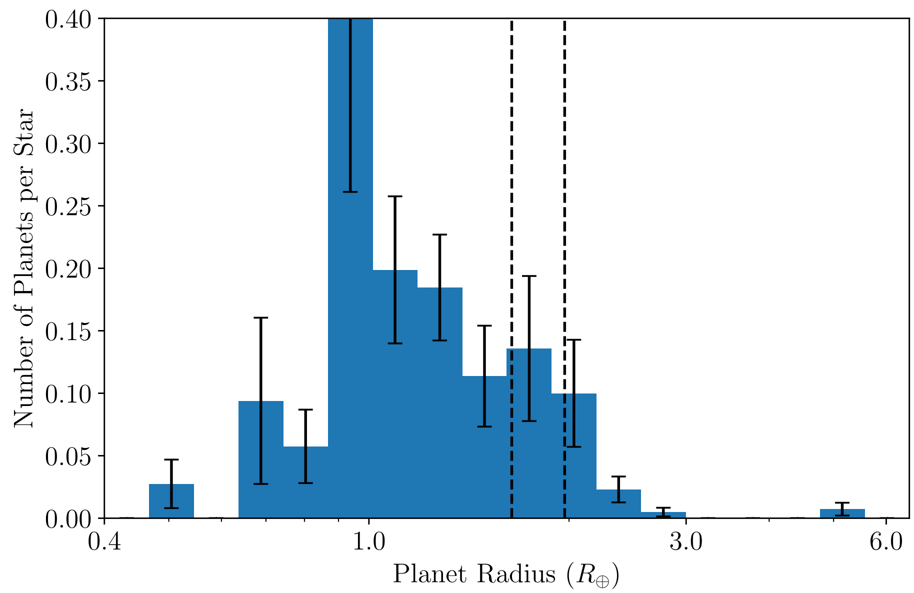

Figure 4.7: Occurrence rate of planets in radius space, realized by sampling the posterior distribution of each cell in Figure 4.5. Fulton et al. (2017) Radius Valley highlighted between 1.64 and 1.97 $R_\oplus$.

Using the 18x18 occurrence rate grids I can also present the cumulative occurrence rate in 6x6 sub-grids, dividing the space into 9 equally spaced log bins in the period/radius and instellation/radius map. In bins where non-detections make an occurrence rate calculation impossible, I report the 1σ upper limit for the occurrence rate. These calculations are summarized in Tables 4.2 and 4.3.

	0.4-1.01$R_\oplus$	1.01-2.57$R_\oplus$	2.57-6.5$R_\oplus$
0.2-1.06 Days	0.051 ± 0.021	0.021 ± 0.006	< 0.03
1.06-5.65 Days	0.318 ± 0.095	0.313 ± 0.055	0.012 ± 0.006
5.65-30.0 Days	0.226 ± 0.157	0.387 ± 0.093	< 0.72*

Table 4.2: Number of planets per star versus planet radius and orbital period calculated by sampling the posterior distributions of the occurrence rate illustrated in Figure 4.5. *Unreliable upper limit.

	0.4-1.01$R_\oplus$	1.01-2.57$R_\oplus$	2.57-6.5$R_\oplus$
0.2-3.42 $S_\oplus$	0.1 ± 0.071	0.051 ± 0.02	0.019 ± 0.014
3.42-58.48 $S_\oplus$	0.228 ± 0.067	0.3 ± 0.051	0.005 ± 0.004
58.48-1000.0 $S_\oplus$	0.007 ± 0.004	0.009 ± 0.003	< 0.026

Table 4.3: Number of planets per star versus planet radius and instellation calculated by sampling the posterior distributions of the occurrence rate illustrated in Figure 4.6.

4.3 Smooth Occurrence Rate Calculation

In this section, I present a smoothed occurrence rate calculation to parallel the binned occurrence rate calculation carried out above. To generate a smooth completeness map, each injection is added as a Gaussian with width equal to one third of the width of the quadtree cell in which the injection was sampled (see Figures 3.2 and 3.3). A distribution of injected planets and a distribution of recovered planets are created this way, and the former is normalized by the latter to produce the map on left-hand side of Figure 4.8. The transit probability is calculated exactly as before on a grid to match the resolution of the smooth occurrence map.

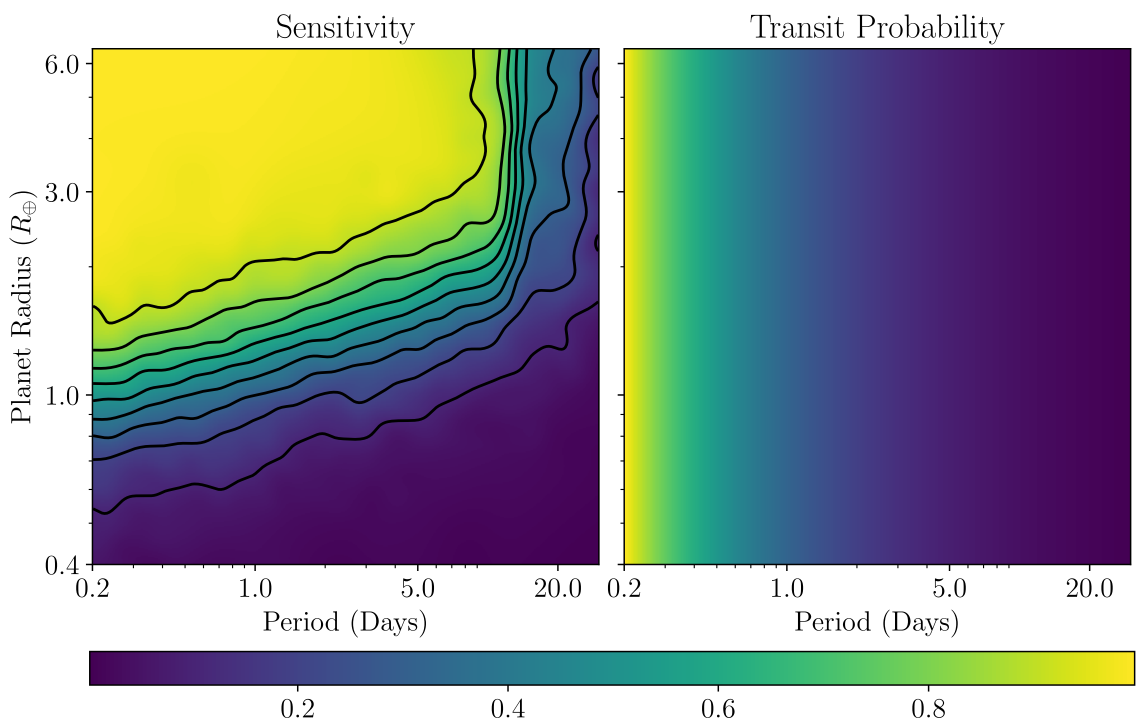

Figure 4.8: Left: Smoothed sensitivity of my pipeline to transiting planets in Radius/Period space. Decades in the sensitivity are marked with black contour lines. Right: High resolution transit probability map.

The planet distribution is generated by taking each planet as a Gaussian in period/radius space. The width in period space is fixed at 5% of the period because the planets' periods are too tightly constrained to have any width in this space. The width in period space is given by the 2.5 times the relative radius error. This extra spread over the radius uncertainty smooths the radius occurrence rate out, and without this smoothing the distribution would be too variable to be physically interpretable. Each distribution added to the map integrates to 1 over the area in period/radius space that it spans. Following equation 4.2.3, each distribution is divided by the completeness at its center, and the total number of targets in the survey. This smooth occurrence rate is plotted in Figure 4.9.

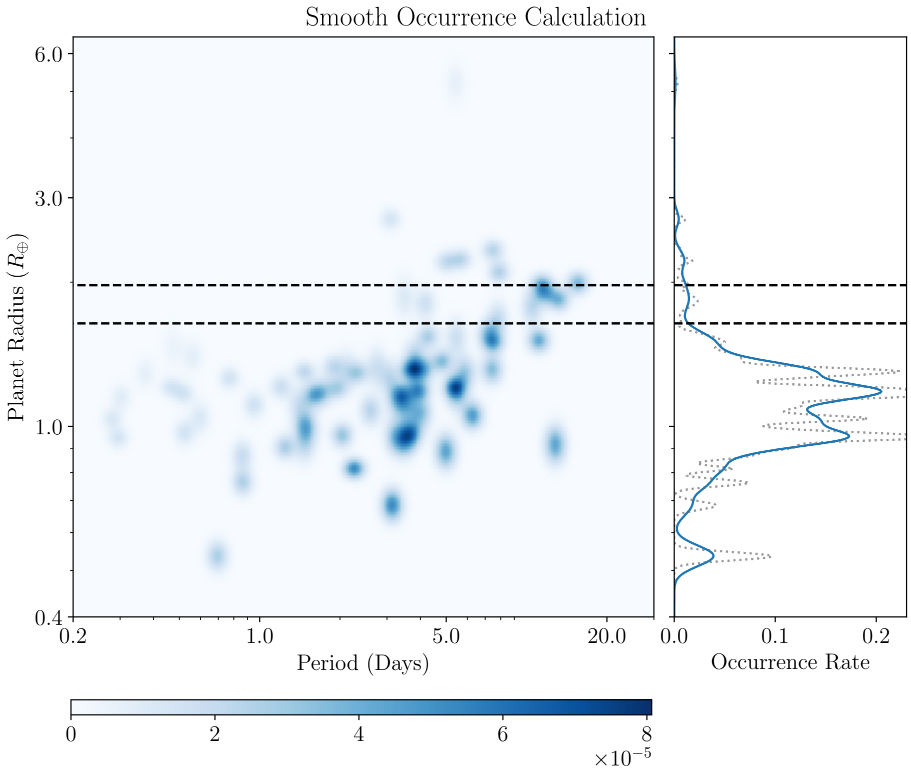

Figure 4.9: Left: Smooth occurrence rate in period/radius space. Right: Occurrence rate distribution in radius space. The primary line represents a smoothed radius occurrence with stretched uncertainties while the light-grey dashed line showing the distribution without the extra 2.5 factor spread in radius uncertainty. Fulton et al. (2017) Radius Valley highlighted between 1.64 and 1.97 $R_\oplus$.

While the smooth occurrence rate may offer a better insight into the shape of the planet distribution around mid-to-late M dwarfs compared to the binned distribution, it disagrees in magnitude with the binned occurrence rate plotted in Figure 4.7. This is somewhat expected because it does not account for the uncertainty in the occurrence rate calculation, and the tails of the binomial posterior distributions have an expectation value larger than their most likely value.

Chapter 5

Discussion

In this chapter, I will contextualize the results of this survey by comparing the occurrence rates around mid-to-late M dwarfs to previous theoretical and observational works which have published predictions and proclamations of the planet population.

5.1 Comparisons to Previous Calculations

This occurrence rate calculation follows a line of occurrence rate calculations pushing towards an understanding of the population of planets around the lowest mass stars. By appending my work to their understanding, we can understand trends in the planet population in lower-mass stars.

In their survey of early-to-mid M dwarfs, Cloutier & Menou (2020) find a cumulative occurrence rate around early-to-mid M dwarfs of 2.48 ± 0.32 and 2.26 ± 0.38 planets per star from Kepler and K2 planets, respectively. Across our sample of stars, this survey finds a cumulative occurrence rate of $1.326^{+0.21}_{-0.208}$ planets per star. As we proceed towards lower masses, planets tend to be half as plentiful. This deficit may be partially attributable to the lower maximum period of this survey with Cloutier & Menou (2020) finding a significant contribution to their occurrence from planets in the second peak of the Radius Valley from 20 to 40 days.

The most appropriate direct comparison we can make to this occurrence rate calculation Ment & Charbonneau (2023) with their survey of 363 mid-to-late M dwarfs within 15 parsecs, searching for transiting planets within 7 days. Their survey does provide constraints for the occurrence rate of super-Earths, but without any sub-Neptune detections, they could only set an upper limit on the occurrence of any second peak of the Radius Valley. Their survey also provides a direct comparison of their occurrence rate to the occurrence rate calculation from Dressing & Charbonneau (2015), an earlier survey spanning mostly early type M dwarfs looking to characterize the rate of habitable worlds. By cutting the search space of our survey, we can directly compare the results of this survey to these two previous occurrence rate calculations. In each subrange I generate 4x4 completeness grids and planet counts in these subranges following the same processes illustrated in Figures 4.2 through 4.4. An extension of the tables from Ment & Charbonneau (2023) is presented in Tables 5.1 and 5.2.

Source		DC15	MC23	This Work
Mass range		$0.08 - 0.65 M_\odot$	$0.1 - 0.3 M_\odot$	$0.1 - 0.4 M_\odot$
Median Mass		$0.5 M_\odot$	$0.17 M_\odot$	$0.22 M_\odot$
Instellation ($S_\oplus$)	Radius ($R_\oplus$)	Occurrence Rate by Source		
4-10	0.5-4.0	$0.436^{+0.140}_{-0.113}$	$0.303^{+0.120}_{-0.095}$	$0.241^{+0.056}_{-0.056}$
10-50		$0.438^{+0.096}_{-0.083}$	$0.162^{+0.064}_{-0.051}$	$0.216^{+0.061}_{-0.062}$
50-200		$0.084^{+0.024}_{-0.022}$	$0.030^{+0.012}_{-0.010}$	$0.035^{+0.013}_{-0.013}$
4-200	0.5-1.0	$0.233^{+0.122}_{-0.097}$	$0.107^{+0.098}_{-0.055}$	$0.206^{+0.058}_{-0.058}$
	1.0-1.5	$0.243^{+0.069}_{-0.068}$	$0.370^{+0.162}_{-0.118}$	$0.277^{+0.047}_{-0.047}$
	1.5-2.0	$0.198^{+0.066}_{-0.057}$	≤ 0.060	$0.033^{+0.011}_{-0.011}$
	> 2.0	$0.283^{+0.074}_{-0.067}$	≤ 0.056	$0.02^{+0.007}_{-0.007}$
	0.5-4.0	$0.99^{+0.16}_{-0.15}$	$0.49^{+0.19}_{-0.15}$	$0.405^{+0.077}_{-0.077}$

Table 5.1: Comparison of my occurrence rate calculation in planet radius and instellation to Dressing & Charbonneau (2015) (DC15) and Ment & Charbonneau (2023) (MC23). This table is an extended version of Table 7 from Ment & Charbonneau (2023).

Source		DC15	MC23	This Work
Mass range		$0.08 - 0.65 M_\odot$	$0.1 - 0.3 M_\odot$	$0.1 - 0.4 M_\odot$
Median Mass		$0.5 M_\odot$	$0.17 M_\odot$	$0.22 M_\odot$
Period (Days)	Radius ($R_\oplus$)	Occurrence Rate by Source		
0.2-1	0.5-4.0	$0.013^{+0.005}_{-0.005}$	$0.047^{+0.019}_{-0.015}$	$0.031^{+0.013}_{-0.012}$
1-2.7		$0.106^{+0.024}_{-0.024}$	$0.159^{+0.064}_{-0.050}$	$0.127^{+0.033}_{-0.033}$
2.7-7		$0.349^{+0.054}_{-0.051}$	$0.402^{+0.162}_{-0.127}$	$0.586^{+0.122}_{-0.122}$
7-30		1.54	-*	$0.461^{+0.201}_{-0.194}$
0.2-30	0.5-1.0	$0.453^{+0.047}_{-0.041}$	$0.134^{+0.123}_{-0.072}$*	$0.71^{+0.243}_{-0.233}$
	1.0-1.5	$0.547^{+0.035}_{-0.041}$	$0.446^{+0.162}_{-0.118}$*	$0.404^{+0.079}_{-0.079}$
	1.5-2.0	$0.411^{+0.022}_{-0.020}$	≤ 0.073*	$0.124^{+0.051}_{-0.051}$
	> 2.0	$0.516^{+0.025}_{-0.024}$	≤ 0.072*	$0.04^{+0.014}_{-0.014}$
	0.5-4.0	1.927	$0.47^{+0.19}_{-0.15}$*	$0.93^{+0.16}_{-0.16}$

Table 5.2: Comparison of my occurrence rate calculation in planet radius and period to Dressing & Charbonneau (2015) (DC15) and Ment & Charbonneau (2023) (MC23). This table is an extended version of Table 7 from Ment & Charbonneau (2023). *The occurrence rate calculation from Ment & Charbonneau (2023) extends from 0.2-7 days.

The results of this survey are in very good agreement with the findings of Ment & Charbonneau (2023), and the occurrence rate of sub-Neptunes lies below their lower limit. This agreement is expected, because my survey covers a very similar mass range. While the occurrence rate of sub- and super-Earths is comparable to the population presented in Dressing & Charbonneau (2015) this survey establishes definite deficit of sub-Neptunes. Our very low sub-Neptune occurrence rate means we can confidently say that the Radius Valley disappears in this regime.

Our cumulative planet rate is greater than the occurrence rate computed in Ment & Charbonneau (2023). This is possibly attributable to the larger maximum period of this survey, with an occurrence of 0.461 found for planets orbiting between 7 and 30 days. We find that a factor of two fewer planets orbit the stars in our sample compared to those surveyed in Dressing & Charbonneau (2015) which agrees with our cumulative comparison to Cloutier & Menou (2020).

Terrestrial planets within the habitable zone (Earth-like planets) present a unique opportunity in the search for life outside of the solar system. Atmospheric characterization through transmission spectroscopy determines an atmosphere's composition by observing the planet's transit depth over a range of wavelengths and determining which wavelengths the atmosphere's molecules absorb. The deeper relative transit depth of planets around M dwarfs makes the transit depth difference caused by the occulting atmosphere easier to observe. This means that atmospheric characterization of an Earth-like habitable planet will be accessible around mid-to-late M dwarfs before any Earth-like planets around larger stars. My work provides constraints on Earth-like planets around mid-to-late M dwarfs. Loosely defining an optimistic habitable zone (HZ) as 0.2 to 2 $S_{\oplus}$, my survey finds a cumulative habitable zone occurrence rate of $0.169^{+0.097}_{-0.095}$, and an occurrence rate of HZ terrestrial (0.7-1.3$R_{\oplus}$) planets of $0.141^{+0.092}_{-0.092}$. This range of habitability is very optimistic and includes Venus and Mars, which are believed to be habitable at early times. A more realistic range from Kopparapu et al. (2013) may be 0.2 to 0.9 $S_{\oplus}$ which is bounded by water loss at the inner edge and the maximum greenhouse at the outer edge, however this survey lacks the sensitivity to long-period planets to effectively probe this zone of habitability.

5.2 Comparisons to Theoretical Predictions

The TESS Guest Investigator Proposals which sought the data used to complete this survey looked to establish the dominant driver of the Radius Valley. While core-powered mass loss and photoevaporation make consistent predictions of the Radius Valley's location for Sun-like stars, their predicted location of the Radius Valley diverges towards mid-to-late M dwarfs (Cloutier & Menou (2020) comparing results from Wu (2019) and Gupta & Schlichting (2019)). Unfortunately, without a concrete emergence of the Radius Valley, this survey is not able to directly disentangle their relative effects in sculpting the Radius Valley around higher-mass stars. This being said, the non-detection of a Radius Valley in this stellar mass regime can still serve to support Radius Valley models.

Burn et al. (2021) uses the Bern NGPPS models produce populations of planets around 0.3 $M_\odot$ and 0.1 $M_\odot$ stars. These stellar masses overlap with the population of stars surveyed here. Their cumulative occurrence rate of planets with periods less than 100 days is dominated by sub-Earths and Earth-like planets $(10 - 30M_\odot)$ with 0.89 and 1.34 of these planets per star around $0.1M_\odot$ and $0.3M_\odot$ stars respectively. Their $0.1M_\odot$ star simulations produce 0.01 Neptunes $(10-30M_\odot)$ per star, and their $0.3M_\odot$ star simulations producing 0.08 Neptunes per star. Directly comparing these populations is difficult because this survey's planets do not have recorded mass. Their finding that sub- and super-Earths dominate the planet population is consistent with this survey, with their simulations possibly producing more of these planets. Their higher occurrence of sub-Neptunes is a discrepancy with our results, but our median stellar mass of 0.22 and limited sensitivity to long-period planets may explain the difference. It bears noting that the Bern model's treatment of water's contribution to the radius sub-Neptunes in has been updated in Burn et al. (2024), which is necessary to replicate the Radius Valley around Sun-like stars. However, the effect of this update on their planet population around mid-to-late M dwarfs is unknown because they have not considered these stars in their simulations.

The results of this survey strongly agree with the prediction from Venturini et al. (2024) where a Radius Valley carved out between rocky super-Earths and water-rich, sometimes steamy sub-Neptunes. This division fades towards lower mass stars where the populations overlap illustrated in Figure 1.5. Their population of planets matches the smooth occurrence rate illustrated in Figure 4.9, with a clear fall-off towards larger radii, with much fewer sub-Neptunes. Venturini et al. (2024) also propose that from Sun-like stars with a well-defined Radius Valley towards lower mass M dwarfs, the Radius Valley becomes "filled in". The Radius Valley detected in Cloutier & Menou (2020) is shallower than the valley found around Sun-like stars by Fulton et al. (2017), and occurrence rate curve from this survey completely lacks a Radius Valley. These independent surveys of different stellar types support the theoretical picture presented in Venturini et al. (2024), but deeper investigation is still necessary. If the planet population in this survey is explained by overlapping rocky and water worlds, a complete confirmation of their model's predictions is only possible with precise mass measurements which are beyond the scope of this survey.

Chapter 6

Conclusions

This work comprises the deepest search for transiting planets to date around mid-to-late M dwarfs to characterize the planet population around the smallest and most common stars in the galaxy (Winters et al., 2015).

Using a custom-built pipeline I have searched TESS light curves from 9,131 mid-to-late M dwarfs over the first 85 sectors of observation. This pipeline included robust detrending with modules removing the effects of flares and stellar rotation. Each light curve was iteratively searched for transit-like signals using Transit Least Squares (Hippke & Heller, 2019), and my pipeline combined credible results to characterize planet candidates. Each planet candidate was modeled with posterior parameter distributions derived using Monte Carlo Markov Chains. This planet search produced 73 manually vetted transiting planet signals.

Robustly testing the pipeline with injection-recovery tests characterized its sensitivity to transiting planet signals across the range of TESS magnitudes which stellar sample spans. By combining the population of detected planets with my pipeline's completeness, I calculated the occurrence rate of planets around mid-to-late M dwarfs out to 30 days. I calculate a cumulative occurrence of $1.326^{+0.210}_{-0.208}$ planets per star, with radii $< 6.5R_{\oplus}$ and orbital periods within 30 days, dominated by a population of sub- and super-Earths with very few sub-Neptunes.

This survey expands upon previous work in this stellar mass regime (Ment & Charbonneau, 2023) by expanding the stellar sample by more than an order of magnitude (25 times). My occurrence rate is consistent with previous work and provides the most refined understanding of the mid-to-late M dwarf planet population to date. The low occurrence of sub-Neptunes provides strong evidence that the Radius Valley fades in this stellar mass regime in agreement with planet population synthesis predictions (Burn et al. (2024) and Venturini et al. (2024)).

Appendix A

Visual Diagnostic Tests

In this section, I outline the visual diagnostic plots used to vet the planet candidates identified by my pipeline. These plots for a randomly selected planet candidate are illustrated below. Figure A.1 shows a best-fit model, which was useful in removing planet candidates with poor fits, which were not ruled out by the statistical vetting tests. Figure A.2 shows the highlighted transits in the detrended and non-detrended flux measurements. This plot helped identify areas where outliers and edge-effects were folded together to produce credible looking transit signals. Dips near the edges of light curves, and binary eclipses were particularly responsible in this case. Finally, Figure A.3 shows the TLS results from each sector where the planet candidate's signal was found by the TLS in the iterative transit search. These results can be cross-referenced with the light curves in figure A.2 to determine where and why the planet candidate was detected, ruling out planet candidates whose signal comes mainly from outliers in one sector.

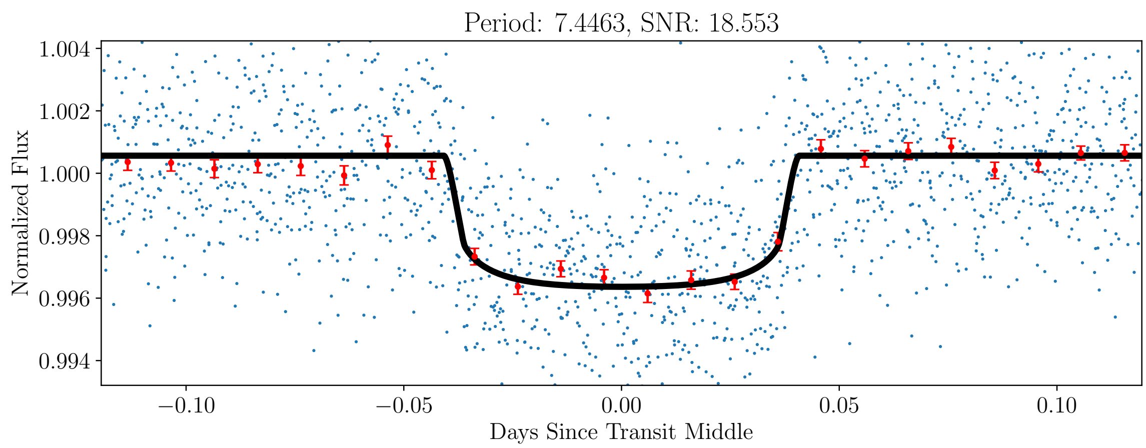

Figure A.1: Diagnostic figure showing the fit transit model used to inspect the goodness of fit.

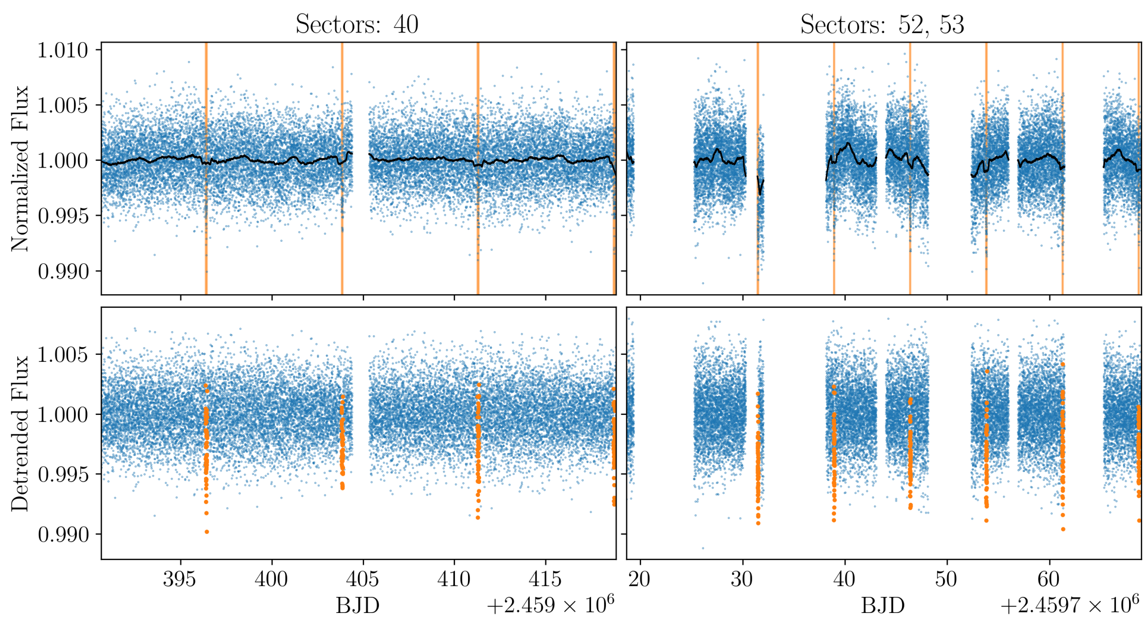

Figure A.2: Diagnostic figure showing the transits in the raw and detrended lightcurves.

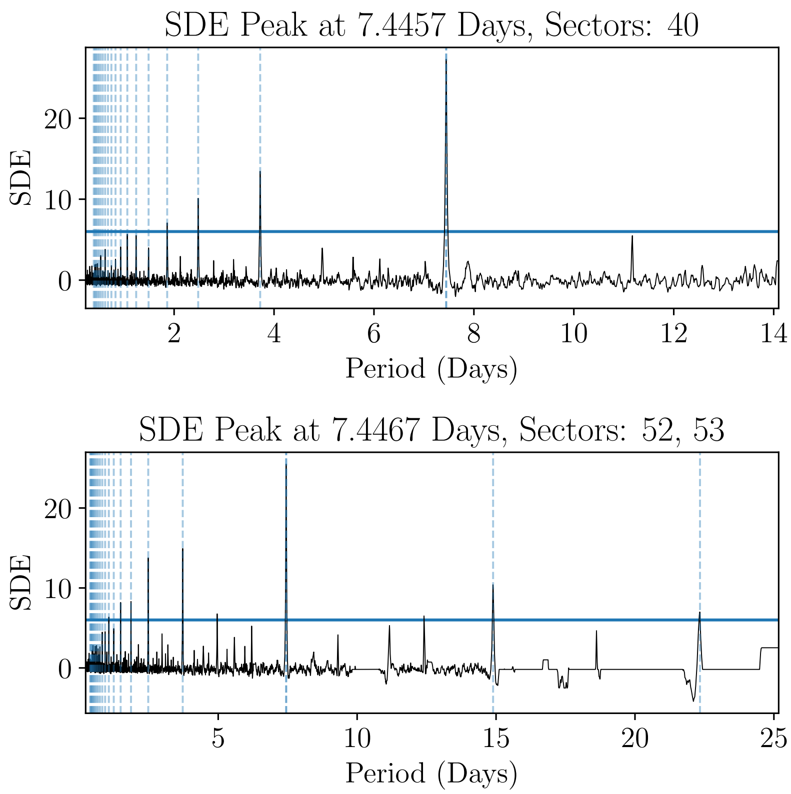

Figure A.3: Diagnostic plot illustrating the TLS results which contributed to the detection of the planet candidate.

Bibliography

Angus, R., Morton, T., Aigrain, S., Foreman-Mackey, D., & Rajpaul, V. 2018, MNRAS, 474, 2094, doi: 10.1093/mnras/stx2109

Barclay, T. 2017, tessgi/ticgen: v1.0.0, Zenodo, doi: 10.5281/zenodo.888217

Birnstiel, T., Ormel, C. W., & Dullemond, C. P. 2011, AAP, 525, A11, doi: 10.1051/0004-6361/201015228

Burn, R., Mordasini, C., Mishra, L., et al. 2024, Nature Astronomy, 8, 463, doi: 10.1038/s41550-023-02183-7

Burn, R., Mordasini, C., Mishra, L., et al. 2024, A radius valley between migrated steam worlds and evaporated rocky cores, arXiv. http://arxiv.org/abs/2401.04380

Burn, R., Schlecker, M., Mordasini, C., et al. 2021, AAP, 656, A72, doi: 10.1051/0004-6361/202140390

Chang, S. W., Byun, Y. I., & Hartman, J. D. 2015, APJ, 814, 35, doi: 10.1088/0004-637X/814/1/35

Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, ApJ, 529, L45, doi: 10.1086/312457

Claret, A. 2017, AAP, 600, A30, doi: 10.1051/0004-6361/201629705

Cloutier, R., & Menou, K. 2020, The Astronomical Journal, 159, 211, doi: 10.3847/1538-3881/ab8237

Drazkowska, J., Alibert, Y., & Moore, B. 2016, AAP, 594, A105, doi: 10.1051/0004-6361/201628983

Dressing, C. D., & Charbonneau, D. 2015, APJ, 807, 45, doi: 10.1088/0004-637X/807/1/45

Emsenhuber, A., Mordasini, C., Burn, R., et al. 2021a, AAP, 656, A69, doi: 10.1051/0004-6361/202038553

—. 2021b, AAP, 656, A70, doi: 10.1051/0004-6361/202038863

Foreman-Mackey, D. 2018, Research Notes of the American Astronomical Society, 2, 31, doi: 10.3847/2515-5172/aaaf6c

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306, doi: 10.1086/670067

Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, The Astronomical Journal, 154, 109, doi: 10.3847/1538-3881/aa80eb

Garcia, L. J., Foreman-Mackey, D., Murray, C. A., et al. 2024, Astronomy Journals, 167, 284, doi: 10.3847/1538-3881/ad3cd6

Giacalone, S., & Dressing, C. D. 2020, triceratops: Candidate exoplanet rating tool. http://ascl.net/2002.004

Giacalone, S., Dressing, C. D., Jensen, E. L. N., et al. 2021, Astronomy Journals, 161, 24, doi: 10.3847/1538-3881/abc6af

Gupta, A., & Schlichting, H. E. 2019, Monthly Notices of the Royal Astronomical Society, 487, 24, doi: 10.1093/mnras/stz1230

Herschel, W. 1783, Philosophical Transactions of the Royal Society of London, 73, 1. http://www.jstor.org/stable/106476

Hippke, M., & Heller, R. 2019, TLS: Transit Least Squares, Astrophysics Source Code Library, record ascl:1910.007

Howell, S. B., Sobeck, C., Haas, M., et al. 2014, PASP, 126, 398, doi: 10.1086/676406

Johnson, S. A., Penny, M., Gaudi, B. S., et al. 2020, The Astronomical Journal, 160, 123, doi: 10.3847/1538-3881/aba75b

Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, APJL, 713, L79, doi: 10.1088/2041-8205/713/2/L79

Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, ApJ, 765, 131, doi: 10.1088/0004-637X/765/2/131

Kovács, G., Zucker, S., & Mazeh, T. 2002, AAP, 391, 369, doi: 10.1051/0004-6361:20020802

Kreidberg, L. 2015, PASP, 127, 1161, doi: 10.1086/683602

Mayor, M., & Queloz, D. 1995, Nature, 378, 355, doi: 10.1038/378355a0

Ment, K., & Charbonneau, D. 2023, The Astronomical Journal, 165, 265, doi: 10.3847/1538-3881/acd175

Owen, J. E., & Wu, Y. 2013, The Astrophysical Journal, 775, 105, doi: 10.1088/0004-637X/775/2/105

—. 2017, The Astrophysical Journal, 847, 29, doi: 10.3847/1538-4357/aa890a

Pecaut, M. J., & Mamajek, E. E. 2013, The Astrophysical Journal Supplement, 208, 9, doi: 10.1088/0067-0049/208/1/9

Perryman, M., Hartman, J., Bakos, G. Á., & Lindegren, L. 2014, APJ, 797, 14, doi: 10.1088/0004-637X/797/1/14

Rasmussen, C. E., & Williams, C. K. I. 2006, Gaussian Processes for Machine Learning (The MIT Press)

Rauer, H., Catala, C., Aerts, C., et al. 2014, Experimental Astronomy, 38, 249, doi: 10.1007/s10686-014-9383-4

Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003, doi: 10.1117/1.JATIS.1.1.014003

Schlecker, M., Mordasini, C., Emsenhuber, A., et al. 2021, AAP, 656, A71, doi: 10.1051/0004-6361/202038554

Schröter, S., Czesla, S., & Zechmeister, M. 2019, gls.py. https://github.com/sczesla/PyAstronomy/blob/master/src/pyTiming/pyPeriod/gls.py

Smyth, W. H. 1846, Monthly Notices of the Royal Astronomical Society, 7, 121, doi: 10.1093/mnras/7.9.121

Stassun, K. G., Oelkers, R. J., Paegert, M., et al. 2019, Astronomy Journals, 158, 138, doi: 10.3847/1538-3881/ab3467

Van Eylen, V., Agentoft, C., Lundkvist, M. S., et al. 2018, Monthly Notices of the Royal Astronomical Society, 479, 4786, doi: 10.1093/mnras/sty1783

Venturini, J., Ronco, M. P., Guilera, O. M., et al. 2024, AAP, 686, L9, doi: 10.1051/0004-6361/202349088

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: 10.1038/s41592-019-0686-2

Winters, J. G., Henry, T. J., Lurie, J. C., et al. 2015, The Astronomical Hournal, 149, 5, doi: 10.1088/0004-6256/149/1/5

Wu, Y. 2019, The Astrophysical Journal, 874, 91, doi: 10.3847/1538-4357/ab06f8

Youdin, A. N. 2011, APJ, 742, 38, doi: 10.1088/0004-637X/742/1/38

(This page is intentionally left blank)