Searching the Entirety of *Kepler* Data

New Exoplanets and Occurrence Rate Estimates

by

Michelle Kunimoto

B.Sc., The University of British Columbia, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Astronomy)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a dissertation entitled:

Searching the Entirety of *Kepler* Data: New Exoplanets and Occurrence Rate Estimates

submitted by Michelle Kunimoto in partial fulfillment of the requirements for
the degree of Doctor of Philosophy
in Astronomy

Examining Committee:

Jaymie Matthews, Physics and Astronomy
Supervisor

Aaron Boley, Physics and Astronomy
Supervisory Committee Member

Brett Gladman, Physics and Astronomy
University Examiner

Ronald Clowes, Earth, Ocean and Atmospheric Sciences
University Examiner

Additional Supervisory Committee Members:

Vesna Sossi, Physics and Astronomy
Supervisory Committee Member

Catherine Johnson, Earth, Ocean and Atmospheric Sciences
Supervisory Committee Member

Abstract

First, I present the results of an independent search of all ∼200,000 stars observed over the four-year *Kepler* mission for multiplanet systems, using a three-transit minimum detection criteria to search orbital periods up to hundreds of days. My search returned 17 planet candidates in addition to thousands of known *Kepler* Objects of Interest (KOIs), with a 98.8% recovery rate of already confirmed planets. I highlight the discovery of one candidate, KIC-7340288 b, that is both rocky (radius $\leq 1.6 R_{\oplus}$) and in the habitable zone (insolation between 0.25 and 2.2 times the Earth's insolation). Another candidate is an addition to the already known KOI-4509 system. I also present adaptive optics imaging follow-up for six of my new candidates, two of which reveal a line-of-sight stellar companion within $4''$.

Using my planet catalogue, I then present exoplanet occurrence rates estimated with approximate Bayesian computation for planets with radii between 0.5 and 16 $R_{\oplus}$ and orbital periods between 0.78 and 400 days, orbiting FGK dwarf stars. I characterize the efficiency of planet recovery by both my search and vetting pipelines using injection/recovery tests, and account for both planet radius uncertainty and the estimated false positive rate due to transit-like noise signals in the data, unlike the majority of previous works. By analyzing my FGK occurrence rates as well as those computed after separating F-, G-, and K-type stars, I explore dependencies on stellar effective temperature, planet radius, and orbital period. I reveal new characteristics of the photoevaporation-driven "radius gap" between ∼1.5 and 2 $R_{\oplus}$, indicating that the significant bimodal distribution previously revealed for $P < 100$ days only exists over a much narrower range of orbital periods, above which sub-Neptunes dominate and below which super-Earths dominate. Finally, I provide several estimates of the "eta-Earth" value —

the frequency of potentially habitable, rocky planets orbiting Sun-like stars. For planets with sizes $0.75 - 1.5\ R_{\oplus}$ orbiting in a conservatively defined habitable zone (0.99 - 1.70 AU) around G-type stars, my calculations place an upper limit (84.1th percentile) of < 0.18 planets per star.

Lay Summary

The discovery of thousands of planets outside of the Solar System over the past decade has revolutionized our understanding of planet diversity. By analyzing results from missions dedicated to finding new planets, invaluable information about planet formation and evolution can be uncovered. This dissertation makes contributions on both fronts. First, I present the results of my own planet search. I found 17 new planet candidates, including a rare small planet in the habitable zone — where liquid water could potentially exist on a rocky planet's surface. Then, I analyze the search results to estimate the commonality of different kinds of planets, and explore what this can tell us about current theories. I estimate that there is likely less than one potentially habitable planet for every five Sun-like stars. Determining this is one of the major goals of astronomy, with implications for the search for life elsewhere in the Universe.

Preface

Chapter 3, which details my full independent search of *Kepler* data and its results, was adapted from a first-author paper published in *The Astronomical Journal*: Kunimoto, M., Matthews, J., & Ngo, H. 2020, Searching the Entirety of *Kepler* Data. I. 17 New Planet Candidates Including 1 Habitable Zone World, AJ, 159, 124. I wrote the paper, with co-authors providing valuable feedback on the manuscript. The development of the full pipeline from start to finish, the downloading, reduction, and search of the light curves for planets, the vetting of surviving candidates, the follow-up analysis, and the interpretation of the results, was done by me. Henry Ngo was the PI for the adaptive optics (AO) observations. He also wrote the technical justification for our telescope proposal, while I co-wrote the scientific justification. The observations themselves were obtained by staff at the Gemini North telescope in Hawaii. Dr. Ngo also helped me write my own AO data reduction scripts.

In terms of resources used to perform the research described in this chapter, all *Kepler* data was downloaded from the Mikulski Archive for Space Telescopes (`http://archive.stsci.edu/`). Some of the codes I used for the data reduction and transit search were written by Jason Rowe. These are all publicly available as part of the *Kepler* Transit Model Codebase (`https://github.com/jasonfrowe/Kepler`). Specifically, I used `kfitsread` to convert the data into a readable file format and combine each star's light curve into a single file, `detrend5` to detrend the light curves, and `transitfind2` to search through the data for transit signals. For the vetting pipeline, the code for my first uniqueness test was written by Kelsey Hoffman (private communication), while I used Jeffrey Coughlin's Model Shift package (`https://github.com/JeffLCoughlin/Model-Shift`) for the

second uniqueness test, the shape metric, significant secondary test, and second odd-even depth test. For the follow-up analysis of planet candidates, I used `emcee` (`https://github.com/dfm/emcee`) to perform Markov Chain Monte Carlo fitting of transit models to the data. The transit models themselves were adapted from code available from Ian's Astro-Python Codes (`http://www.lpl.arizona.edu/~ianc/python/`). I used `isochrones` (`https://github.com/timothydmorton/isochrones`) to produce posteriors for stellar parameters. I also analyzed each transit with `vespa` (`https://github.com/timothydmorton`), which is a probabilistic validation tool for transiting exoplanets. All other code was written by me (e.g. code for other candidacy tests not mentioned here, scripts for automating the search and vetting pipeline to run on thousands of light curves, etc.). Lastly, due to the computational time required to search each star sufficiently, I performed the majority of my work on the Orcinus and Cedar computing clusters, provided by WestGrid (`www.westgrid.ca`) and Compute Canada (`www.computecanada.ca`).

Chapter 4, which details the use of my independent search results for exoplanet occurrence rate estimates, was adapted from a first-author paper accepted for publication in *The Astronomical Journal*: Kunimoto, M. & Matthews, J. 2020, Searching the Entirety of *Kepler* Data. II. Occurrence Rate Estimates for FGK Stars. I wrote the paper, and performed the full occurrence rate estimates, analysis, and interpretation of the results. My co-author provided feedback on the manuscript.

In terms of resources, I used `cosmoabc` (`https:/github.com/COINtoolbox/CosmoABC`) to implement approximate Bayesian computation (ABC) in my occurrence rate estimates, and `emcee` to perform Markov Chain Monte Carlo fitting of power law-based functions to my occurrence rates as a function of orbital period. All other code was written by me (e.g. the exoplanet population simulator/forward model, scripts for automating the calculation of occurrence rates with different inputs, etc.). Lastly, due to the computational time required to compute ABC occurrence rates, I used the Graham Compute Canada cluster.

Chapter 2 describes the motivation behind my research, and Chapter 5

concludes my results and summarizes their significance. These were each partially adapted from the Introduction and Conclusion sections of the two papers discussed previously.

Table of Contents

Appendices

List of Tables

List of Figures

List of Symbols

$''$	arcsecond
AU	astronomical unit
$R_\oplus$	Earth radius
$M_\oplus$	Earth mass
$S_\oplus$	Earth insolation
$R_\odot$	Solar radius
$M_\odot$	Solar mass

List of Abbreviations

CNES	French Space Agency
CoRoT	Convection, Rotation and planetary Transits
ESA	European Space Agency
ESPRESSO	Echelle Spectrograph for Rocky Exoplanet- and Stable Spectroscopic Observations
EXPRES	EXtreme PREcision Spectrograph
HARPS	High Accuracy Radial velocity Planet Searcher
MAST	Mikulski Archive for Space Telescopes
NASA	National Aeronautic and Space Agency
TESS	Transiting Exoplanet Survey Satellite
WFIRST	Wide Field Infrared Survey Telescope
ABC	approximate Bayesian computation
AO	adaptive optics
EB	eclipsing binary
FP	false positive
FPP	false positive probability
FWHM	full-width half-maximum
HZ	habitable zone
PC	planet candidate
PLDF	planetary distribution function
RV	radial velocity
S/N	signal-to-noise ratio
TTV	transit timing variations

Acknowledgements

First and foremost, I would like to thank my thesis supervisor, Jaymie Matthews, for his support, guidance, and generosity over the past six years — starting with my first ever research experience, later my undergraduate honours thesis, and now, this dissertation. I would not be where I am today without him. Thank you to Aaron Boley and Henry Ngo, who both provided invaluable support, advice, patience, and numerous reference letters!

I would also like to thank Jason Rowe and Kelsey Hoffmann for fostering my interest in transiting exoplanets as an undergraduate. They were the first to show me how to hunt for planets, and were wonderful resources.

Last but certainly not least, thank you to my family and friends for always encouraging me. To my mom, for being a tireless and patient listener; to my dad, for all the pep talks and reminders to take breaks; to my brother, for always knowing how to make me smile. And to Braden, for always believing in me and providing invaluable moral support.

This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program. This research has also made use of the Exoplanet Orbit Database and the Exoplanet Data Explorer at exoplanets.org.

My adaptive optics imaging follow-up was based on observations obtained at the Gemini Observatory (Programs GN-2018B-Q-134 and GN-2019A-FT-213), which is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership: the National Science Foundation (United States), National Research Council (Canada), CONICYT (Chile), Ministerio de Ciencia, Tecnología e Innovación Productiva (Argentina), Ministério

da Ciência, Tecnologia e Inovação (Brazil), and Korea Astronomy and Space Science Institute (Republic of Korea).

This work has also made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

Chapter 1

Introduction

"Are we alone in the universe?" Playing on both our philosophical and scientific curiosities, this is one of the oldest and most fundamental questions facing humanity. We still don't know the answer, and until as recently as a few decades ago, we couldn't even say with certainty whether or not any planets existed beyond the Solar System.

That changed on January 9, 1992, when radio astronomers Alex Wolszczan and Dale Frail made history with the first definitive discovery of two "exoplanets" — planets outside of the Solar System — orbiting the pulsar PSR 1257+12 [156]. Three years later saw the first confirmed detection of an exoplanet orbiting a star similar to the Sun, the nearby 51 Pegasi [96]. By the turn of the century, several dozen worlds had been found, collectively ushering in a new era of planet-hunting. What was once science fiction has now become a mainstream and popular subfield of astronomy that spans astrophysics, planetary science, statistics, and even biology.

Today, we know of 4152 confirmed exoplanets,[1] with thousands more candidates awaiting validation. New planet detections are constantly challenging theories of planet formation, evolution, and system dynamics as we find more and more examples of planetary systems that are in stark contrast to the Solar System. For instance, many of the earliest discoveries were of "hot Jupiters", an entirely new class of planets. These giant planets orbit extremely close to their host stars, unlike our local giants that constitute all outer Solar System planets. Meanwhile, discoveries of planets between Earth and Neptune in size — fittingly known as super-Earths and sub-Neptunes — have challenged our understanding of a possible transition region between

[1] As listed on the NASA Exoplanet Archive (https://exoplanetarchive.ipac.caltech.edu), accessed April 24, 2020.

rocky and gaseous planets. Planet-hunters are also actively in pursuit of rocky exoplanets that are in the habitable zones (HZs) of their host stars, meaning they may be able to support liquid water on their surfaces, and perhaps even life. What it means for a planet to be considered "Earth-like" is constantly being refined, as are the prospects for extrasolar habitability and our understanding of our place in the Universe.

A cornerstone of this dissertation is the success of NASA's first exoplanet-finding mission, *Kepler*. *Kepler* was specifically designed to detect and characterize Earth-size planets in or near the HZs around Sun-like stars, with the goal of estimating their prevalence [13]. Since its launch in 2009, *Kepler* has confirmed 2349 planets out of 4717 announced candidates[2] from its original four-year mission, accounting for more than half of all planets known today [7, 13, 14, 21, 36, 108, 129, 143]. Furthermore, *Kepler* has confirmed 15 small ($R_p < 1.6\ R_\oplus$) planets in the HZ (flux received from its star between 0.25 and 2.5 times that received by the Earth, as defined by the NASA Exoplanet Archive). One of these planets orbits a star similar to our Sun: with a size of 1.1 $R_\oplus$ and an orbital period of 385 days, Kepler-452 b[3] is perhaps the closest Earth analogue discovered so far [73].

In this chapter, I provide an overview of topics in exoplanetary science most relevant to this thesis. In §1.1, I describe various methods for detecting exoplanets, and in §1.2 I discuss steps to go from a detected candidate to a validated exoplanet. In §1.3, I describe the general motivation and methods behind deriving population statistics from exoplanet catalogues. Lastly, I give an overview of what defines the habitable zone in §1.4.

[2]Based off confirmed and candidate *Kepler* planets listed on the NASA Exoplanet Archive, accessed April 24, 2020.

[3]A typical naming convention for exoplanets is to use the name of the host star, followed by lowercase letters (starting with 'b') by order of discovery.

1.1 Exoplanet Detection Techniques

There are several exoplanet detection techniques that have contributed to current planet catalogues. I summarize and compare the most well-known and successful here. The transit method is the technique most relevant to this thesis.

All numbers of exoplanets discovered with each method are taken from the NASA Exoplanet Archive Confirmed Exoplanet Statistics table,[4] accessed on April 24, 2020.

1.1.1 Pulsar Timing Variations

The first exoplanets to be confirmed were detected using the pulsar timing method [156]. In total, it has led to the discovery of 7 exoplanets.

A pulsar is a rapidly rotating neutron star — the dense remnant of a giant star that exploded as a supernova. Pulsars emit radio waves that are observed as regularly timed pulses when the emission is pointed toward Earth, much like how the light of a lighthouse is seen when the light shines in the direction of an observer. The time intervals between pulses range from milliseconds to seconds for an individual pulsar, and are so regular and precise that pulsars may be considered among the most accurate "clocks" in the Universe.

Like other stars, pulsars do not remain completely stationary when orbited by a companion (or companions) such as a planet. Instead, the planet and star orbit their common centre of mass. As observed on Earth, evidence for the pulsar's orbit manifests as irregularities in the timing of the pulses, where pulses arriving slightly earlier than expected indicate that the star is slightly closer to Earth in orbit, and vice-versa. The closer and more massive the orbiting planet, the larger the pulsar movement induced, and the larger the variations in pulse timings are observed. Astronomers can measure these periodic variations to deduce parameters of the orbit, such as the orbital period, as well as the mass of the planet.

[4] https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html

Due to the extremely precise nature of pulsar rotation, this method has been able to detect planets far less massive than any other. The least massive planet known so far, at only $M_p = 0.02\ M_\oplus$ (less than twice the mass of the Moon), was discovered with pulsar timing variations [155]. However, pulsar planets are rare, with fewer than 1% of pulsars having been found to host planets [93]. One explanation for this is that pulsar planet formation likely requires specific low-probability circumstances to occur, such as a favourable formation site created by the destruction of a low-mass companion star [93].

1.1.2 Radial Velocity

Early exoplanet detections were dominated by the radial velocity (RV) method. In fact, this method led to an exoplanet detection in 1988 [26], years before the first pulsar planets, though the detection was not confirmed until 2003 [58]. It also led to the discovery of 51 Pegasi b in 1995, the first exoplanet to be found around a Sun-like star [96]. To date, 801 exoplanets have been discovered using RV.

Similar to the pulsar timing method, the RV method is based on being able to detect the gravitational influence of a planet on its host star. However, rather than observing emissions from pulsars, astronomers can use spectroscopy to observe any star for signs of exoplanets. As the star orbits the system's centre of mass, the stellar spectrum is periodically shifted due to the Doppler effect. Spectral lines are shifted to shorter wavelengths (toward the blue end of the spectrum) as the star moves toward the observer along the line of sight, and longer wavelengths (toward the red end) as the star moves away. The change in wavelength $\lambda - \lambda_0$ is related to the radial (line-of-sight) velocity of the star, V_{rad}, by

$$\frac{\lambda - \lambda_0}{\lambda_0} = \frac{V_{\text{rad}}}{c}, \tag{1.1}$$

where λ and λ_0 are the observed and rest wavelengths of an emission line, c is the speed of light, and V_{rad} is positive as the star moves away from the observer. Plotting V_{rad} over time typically gives a sinusoidal curve such as the one in Fig. 1.1, where the period of the sine curve gives the orbital

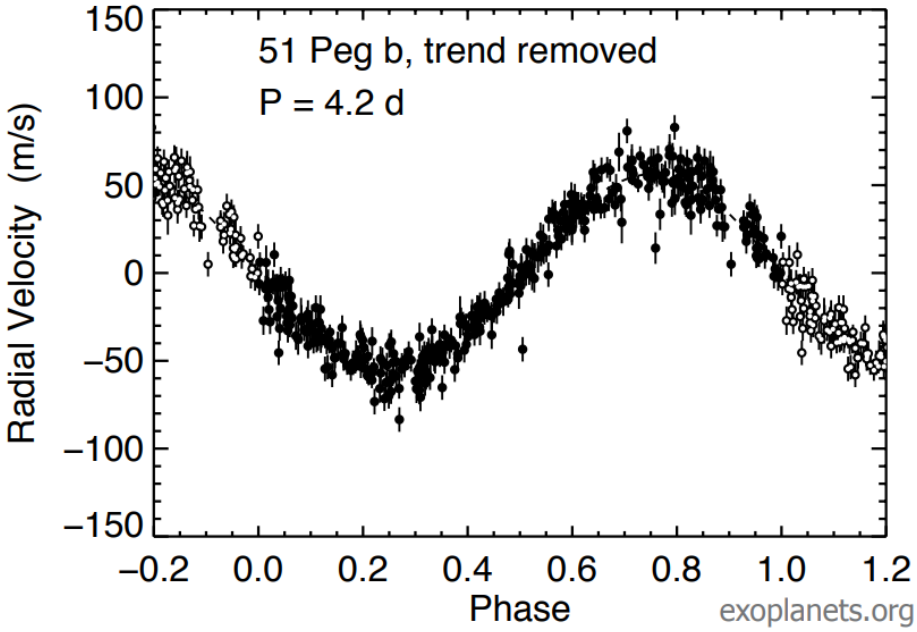

Figure 1.1: Radial velocity measurements for the star 51 Pegasi, folded over the 4.2 day orbital period in the form of a "phase diagram." The corresponding planet, 51 Pegasi b, was the first confirmed exoplanet to be found around a Sun-like star [96]. This plot was retrieved from the Exoplanet Orbit Database and the Exoplanet Data Explorer at exoplanets.org [158].

period of the star's orbit. Measurements are typically presented as a "phase diagram", for which the data are folded over the orbital period to produce a stronger signal.

RV measurements allow for several properties of the planet's orbit to be determined, including the period P (equal to the sine wave's period) and eccentricity e (where a perfect sine wave corresponds to a circular orbit, or $e = 0$, and variations from the shape of a sine wave indicate an eccentric orbit). The semi-major axis of the planet's orbit a (distance from star) can be estimated using Kepler's Third Law,

$$P^2 = \frac{4\pi^2 a^3}{GM_s}, \tag{1.2}$$

where G is the gravitational constant and M_s is the mass of the star. The mass of the planet can also be found, assuming a circular orbit, as

$$M_p = \frac{K}{\sin i}\sqrt{\frac{M_s a}{G}} \tag{1.3}$$

where K is the amplitude of the RV curve and i is the inclination angle of the orbit relative to the plane of the sky (where $i = 90^{\circ}$ means that the orbit is completely edge-on).

The RV method alone cannot be used to estimate i, leaving it unknown. Therefore, only the minimum mass of a planet, the measured product $M_p \sin i$, is typically able to be reported. This means that some RV-detected objects may actually be too massive to be planets.

The RV method is also limited by the precision of spectrographs. Early detections mostly consisted of hot Jupiters, because their large masses and close-in orbits result in the quickest and easiest-to-detect gravitational effects on their host stars. Consider, for example, 51 Pegasi b, which has a minimum mass of about half that of Jupiter (about 150 times that of the Earth) and an orbital distance of only 0.05 AU (far closer than even the innermost planet Mercury is to the Sun). It induces motion in its star of over 50 m/s, as can be read from Fig. 1.1. By comparison, the Earth is so small and so far away from the Sun that it only induces a change of $\sim$10 cm/s in the Sun's motion over the course of a year. Over the last decade, spectrographs have been able to achieve a precision of $\sim$1 m/s (e.g. HARPS [117]), so RV detections of Earth analogues have been unattainable. However, the next generation of spectrographs, such as ESPRESSO [118] and the recently built EXPRES [75], have the aim of reaching 10 cm/s RV precision for the first time.

1.1.3 The Transit Method

The transit method is the most successful detection method so far, having led to the discovery of 3158 planets overall (76% of all confirmed exoplanets). This is in large part thanks to the *Kepler* mission, which used the transit method as its chosen detection technique.

If an object passes between a star and the Earth, it will block a portion of

the star's light, and the star's observed brightness will temporarily decrease. If these events occur periodically and have a consistent duration, it is an indication that a planet could be orbiting the star and passing in front of (transiting) it once every orbital period. Examples of transits are shown in Fig. 1.2. Like with RV data, the measurements of the star's brightness (light curves) can be folded at the period of a detected transit to produce a stronger signal.

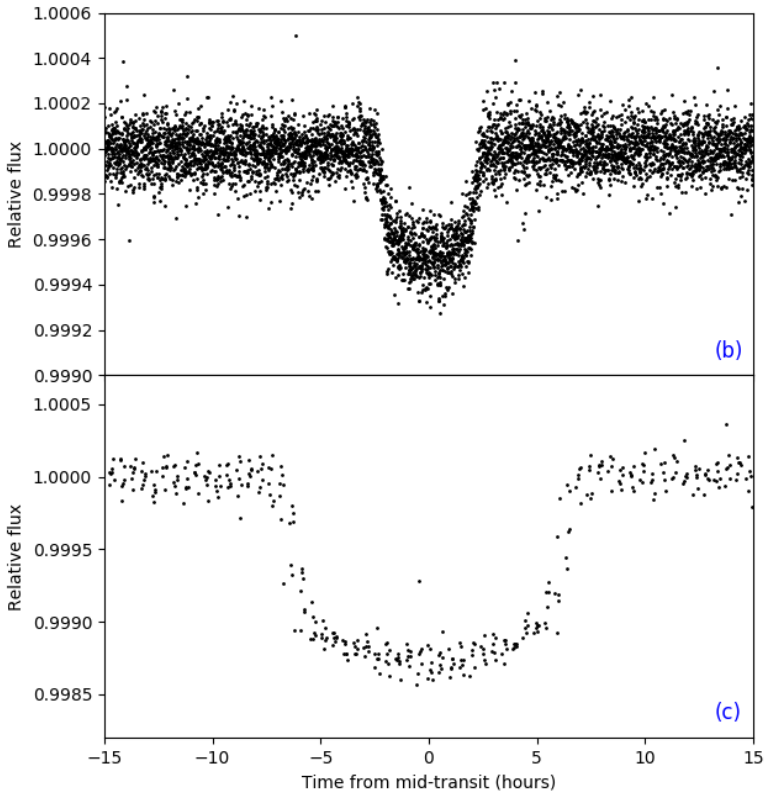

Figure 1.2: The phase-folded transits of two planets orbiting the star Kepler-103. Top: Kepler-103 b, a $3.3R_{\oplus}$ planet with a 16.0 day orbital period. Bottom: Kepler-103 c, a $5.5R_{\oplus}$ planet with a 179.6 day orbital period.

Aside from giving the orbital period of the planet (by measuring the time between each dip), the transit method is the only known method that can give an estimate of the planet's radius. To first order, the ratio of the observed change in flux during transit, ΔF, to that of the stellar flux F can be expressed as

$$\frac{\Delta F}{F} = \left(\frac{R_p}{R_s}\right)^2, \tag{1.4}$$

where R_p and R_s are the radii of the planet and star, respectively. For example, a transit with a depth of 0.25% (2500 parts per million, or ppm) means that the corresponding planet is 0.05 times the size of its host star. The shape and duration of the transit can also give an estimate of the impact parameter b (the projected distance between the planet and star centres during midtransit in units of stellar radius), the orbital inclination i (relative to an observer), and the density of the star ρ_s, even without knowledge of the star's mass or radius [136].

Furthermore, the transit method is applicable to studies of the planet's atmosphere via transit spectroscopy. While light from the star is blocked during transit by the planet itself, light may still pass through the planet's atmosphere. At some wavelengths, the light may be absorbed by certain gasses, contributing to a different observed transit depth than at other wavelengths. By observing transits at different wavelengths, astronomers can recreate the absorption spectrum and investigate the atmosphere's composition.

The transit method is also able to find small, rocky planets currently out of reach of the RV method. In particular, the CNES/ESA-led CoRoT, which operated from 2006 to 2013, was the first mission dedicated to discovering exoplanets with the transit method, as well as the first mission capable of finding planets within several times the size of Earth. It set the stage for *Kepler*, which was the first mission able to find Earth-size planets around Sun-like stars. Such a planet would cause a transit depth of only 0.0084% (84 ppm), and *Kepler* was able to achieve this with incredible precision ($\sim$10 ppm per 6 hours for a 10th magnitude star [31]).

However, the method is not without its downsides. While the RV method can detect signals from stars in orbits of all orbital inclinations (aside from $i = 0^{\circ}$, or "face-on", for which no line-of-sight motion can be detected), the transit method is restricted to orbits that are near $i = 90^{\circ}$ since the planet must pass almost directly between the observer and the star. For a circular orbit, the probability $P_{\rm tr}$ that a planet's orbital plane is sufficiently aligned can be estimated as

$$P_{\rm tr} = \frac{R_s + R_p}{a} \simeq \frac{R_s}{a}, \tag{1.5}$$

where R_p is only a consideration for small stars and is otherwise negligible. For example, the probability of observing the transit of an Earth-size planet at 1 AU around a Sun-like star is only 0.5%. Another problem is that a planet's transit usually lasts only a tiny fraction of its orbital period. While a planet might take years to orbit, the transit itself typically occurs for only a few hours (see Fig. 1.2). Additionally, astronomers need to observe multiple transits to both confirm the orbital period and establish the existence of a planet. Transit detection surveys must therefore observe thousands of stars, continuously, and for long periods of time, to even have a chance at detecting a small number of planets. For these reasons, *Kepler* observed ∼200,000 stars simultaneously, with data recorded every 30 minutes for a total of four years. Since *Kepler* adopted a three-transit minimum detection criteria for planet candidates, these four years of observations allowed the mission to be sensitive to planets with orbital periods up to $P \approx 500$ days.

Lastly, the transit method is perhaps the method most susceptible to false positives (FPs) — phenomena that mimic planet signals. I leave a discussion about transit FPs, as well as techniques to deal with them, to §1.2.

Transit Timing Variations

The transit-timing variations (TTV) method uses detections of transiting planets to find nontransiting (or missed) planets in the same system. Just as a planet will gravitationally tug on its host star as it orbits, it can also

affect the orbits of other planets in the system. Thus, if an unseen planet is significantly perturbing a transiting planet's orbit, the mid-times of each transit will vary with time.

TTVs have lead to the discovery of 21 exoplanets, and can also give upper constraints on the mass of the perturbing object.

1.1.4 Microlensing

Ranking as the third most successful detection method so far, the microlensing method has lead to the detection of 89 exoplanets.

Gravitational lensing is an astronomical phenomenon predicted by Einstein's General Theory of Relativity. As light from a distant source passes an object almost exactly aligned between it and an observer, the gravitational field of the object will act like a lens, magnifying the light of the distant source. If the lensing object is a star with a planet orbiting around it, the planet may be so aligned that its own gravitational field can contribute to the lensing effect. This causes an additional spike in magnification.

The main advantages of the microlensing method are that it can detect extremely low-mass planets, as well as planets in larger orbits than the capabilities of the RV and transit methods. For example, the upcoming WFIRST mission has anticipated sensitivity to planets as small as $\sim$0.03 $M_{\oplus}$, and planets out to hundreds of AU from their stars [10]. Furthermore, the microlensing method does not depend on the host star being bright enough to be observed; rather, its mass must simply be chance-aligned to lens a background source. This allows it to be sensitive to planets around stars much fainter and further away than other methods — perhaps even in other galaxies [38]. Microlensing is also one of the few methods that can find free-floating planets that do not orbit any star at all ("rogue" planets), as a planet itself can act as a gravitational lens even without a host star.

However, the method relies on chance events that are unique and do not repeat. Planets detected by microlensing will typically never be observed again, and both the rarity and randomness of the events makes the discovery of planets by this method difficult and unpredictable.

1.1.5 Direct Imaging

At 50 discovered exoplanets, the direct imaging method is the fourth most successful detection method.

At visible wavelengths, a star is typically billions of times brighter than any orbiting planet, and planets far enough from the star to be resolved by a telescope are lost in the glare and reflect little visible star light. However, at infrared wavelengths, the difference is on the order of only a million times, and it becomes possible to detect a planet by its thermal emission. The direct imaging method uses coronagraphs to block the light from the star, while leaving the planet visible. An example of a directly-imaged system is shown in Fig. 1.3.

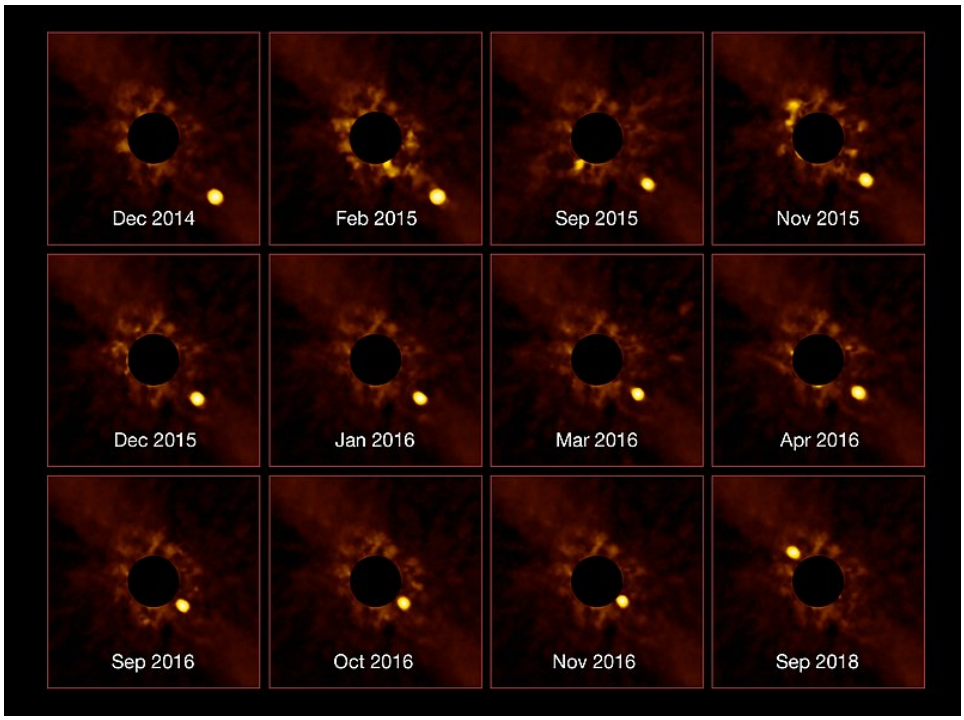

Figure 1.3: The directly imaged Beta Pictorus system, as captured by the ESO's Very Large Telescope. The central star has been masked out to increase visibility of the orbiting planet, Beta Pictorus b. Image by ESO/Lagrange/SPHERE consortium.

This method is best used for planets that are especially large (larger than Jupiter), hot (so they emit more infrared radiation), and far away from their stars (so they are easier to resolve). Due to these limitations, direct imaging is only possible on rare occasions.

1.2 Confirmation of Transiting Exoplanets

As discussed in §1.1.3, one of the main limitations to exoplanet transit surveys such as *Kepler* is that the transit method is vulnerable to false positives. These must be taken into consideration before one validates a planet candidate as a confirmed exoplanet. Of the 9564 *Kepler* Objects of Interest (KOIs; so-called because they correspond to potential transit signals) listed on the NASA Exoplanet Archive, more than half (4839) have been labeled as FPs.[5]

While other methods have their own sources and manifestations of FPs, as well as different standard steps to take toward exoplanet validation, I will focus this review on those specific to the transit method.

1.2.1 False Positives

Eclipsing Binaries

Commonly, a brightness drop is not due to a planet transiting the target star, but rather an orbiting secondary star passing in front of (eclipsing) the target that causes the decrease in brightness. Note that here a "transit" refers to a brightness dip due to a light-blocking planet, while an "eclipse" refers to a brightness dip due to a light-blocking star. 2210 KOIs have been flagged as FPs due to this phenomenon.[6]

Sometimes, these eclipsing binaries (EBs) can be identified simply by calculating the radius of the eclipsing object from the decrease in brightness, and finding that the object is too large to be consistent with a planet

[5] A "False Positive" disposition in both the "Exoplanet Archive Disposition" and "Disposition Using Kepler Data" columns.

[6] Flag of 1 in the "Stellar Eclipse False Positive" column.

(see Eqn. 1.4). However, some companions just graze the target star and cause a small depth more consistent with a planetary transit. In this scenario, the shape of the eclipse can indicate an EB. Grazing EBs have high impact parameters, corresponding to characteristically V-shaped eclipses. By comparison, planet transits are much more consistent with a U-shape (Fig. 1.4).

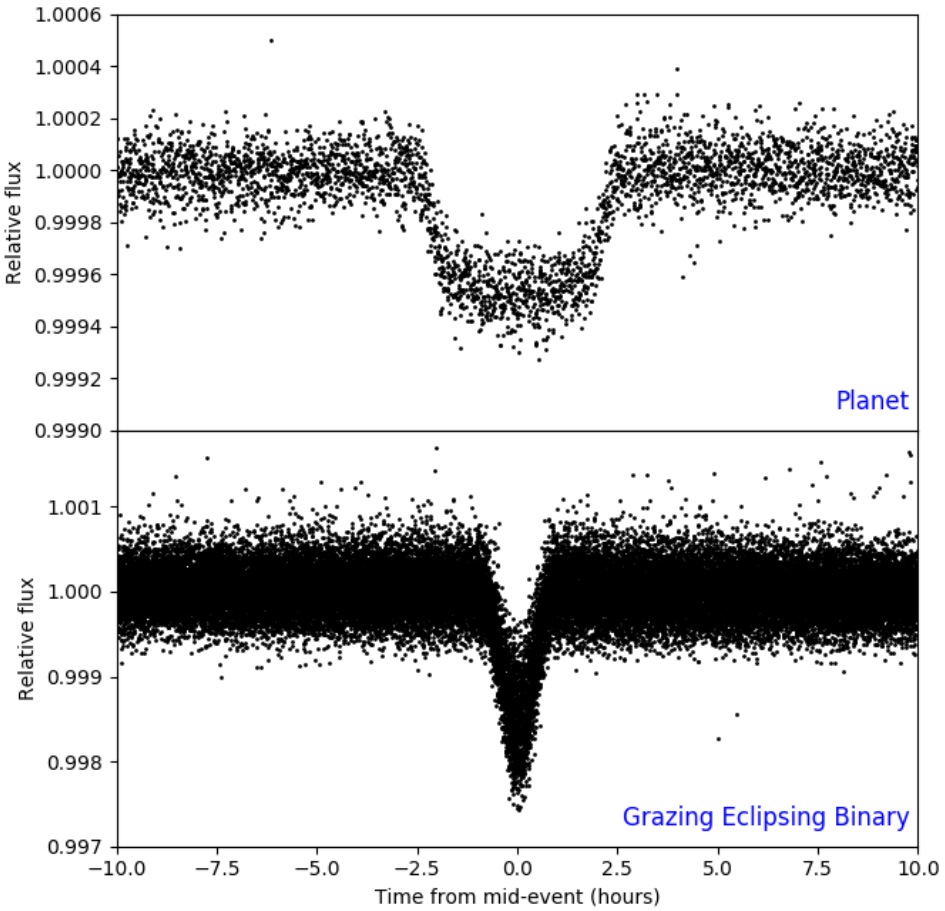

Figure 1.4: Top: Kepler-103 b, demonstrating the characteristic U-shape of a planet transit. Bottom: KOI-690.01, demonstrating the characteristic V-shape of a grazing eclipsing binary FP due to a high impact parameter.

Alternatively, one may look for the presence of a secondary eclipse in the light curve. For the case of a binary star system, the light curve consists of measurements of the combined brightness of the two stars (not just the target star), so this secondary eclipse occurs when the companion star passes behind the target and has its own light blocked. If the two stars have different brightnesses, the alternating eclipses should have different depths and the planetary nature of the companion can be excluded (Fig. 1.5).

Occasionally, hot Jupiters can be sufficiently large and bright in reflected light and thermal emission to produce a secondary eclipse as they pass behind their host star (e.g. HAT-P-7 [12]). Techniques to identify these cases include assessing the derived radius of the object, its impact parameter, and the depth of the secondary (typically much shallower for a planet than for a star).

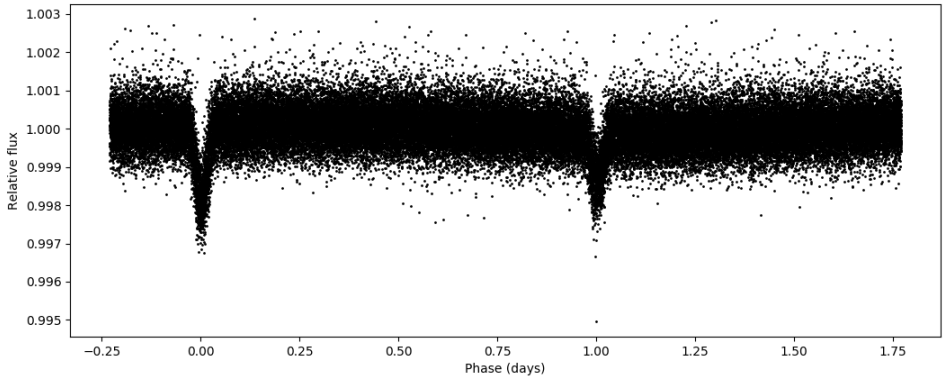

Figure 1.5: The phase diagram for KOI-888.01, an eclipsing binary FP with a 2.0 day orbital period. The primary eclipse on the left occurs when the star passes in front of the primary target star. The slightly shallower secondary eclipse on the right occurs when the star passes behind the primary target star. In this case, the secondary eclipse occurs half an orbital period (1.0 days) after the primary eclipse, which is consistent with a circular orbit.

Nearby and Background Contaminants

A transit-like signal may originate from a nearby or background source, rather than another star directly in orbit around the target (2162 KOIs).[7] For example, there may be a physically associated eclipsing binary pair (or star-planet pair) that has its light blended with the target star. This scenario is known as a hierarchical triple eclipsing binary. Similarly, there could be a blended background (or foreground) eclipsing binary system, making it look like a planet transits every time the background star is eclipsed by its companion.

This kind of contamination can occur because photometers measure all light in a specific aperture, and are not able to resolve individual stars. For example, the pixels of the *Kepler* telescope are $4'' \times 4''$ squares, and the typical photometric aperture used to observe a given target has a radius of $4-7$ pixels [17]. Thus, it is common for light from other stars to contribute to the target star's light curve and contaminate the measurements.

Bryson et al. (2013) [18] outlined several techniques for identifying contamination due to background stars in *Kepler* data. In short, pixel-level data are analyzed to measure the location (centroid) of the transit source on the sky. A transit source significantly offset from the target star would indicate an FP. Another method is known as ephemeris-matching, for which the ephemerides (period and epoch) of a transit-like signal are compared to the ephemerides of other known signals. If two signals match, then at least one of them is an FP due to contamination.

Noise and Systematics

False positives may not be due to other astrophysical objects, but rather noise or systematics in the light curve (1527 KOIs).[8] Noise can refer to statistical fluctuations, instrumental noise, intrinsic stellar variability (e.g. spots or pulsations), and more, and these sources are especially important false

[7] Flag of 1 in the "Centroid Offset False Positive" and/or "Ephemeris Match Indicates Contamination False Positive" columns.

[8] Flag of 1 in the "Not Transit-Like False Positive" column.

positive scenarios to consider when searching for planets with low signal-to-noise ratios (S/N).

Tests against these FPs can include assessing the consistency of the transit properties (on the basis that all planetary transits should have consistent depth and duration throughout the light curve), the shape of the signal (a planet transit should be a symmetric U-shape and only represent a decrease in flux), the uniqueness of the signal (there should not be other transit-like events in the phase-folded light curve with similar depth or duration), and whether the transit disappears or changes with different stellar variability detrending methods (a planet transit should be robust to detrending method). However, it is often difficult to determine the point at which a signal should be considered a planet and not noise. A somewhat arbitrary balance must be found between allowing for the passing of planets with low S/N - of which small, rocky planets with long orbital periods, which are particularly valuable additions to exoplanetary parameter space due to their difficulty to find, are examples - and avoiding the addition of too many noise FPs to a catalogue.

1.2.2 Follow-Up Observations and Validation

The above ways of identifying eclipsing binaries and nearby/background contaminants with *Kepler* data only are not always successful. For example, a stellar companion may have luminosity too low for a secondary eclipse to be visible in the light curve; a giant planet may exhibit a deep, V-shaped transit that can be mistaken for a grazing EB (e.g. Kepler-446 b [85]); or, a background eclipsing binary system may be sufficiently close to the target that pixel-level analysis does not detect a significant centroid offset. Furthermore, stellar companions and other nearby or background stars can still affect the light curve beyond being potential sources of FPs. Light from a contaminant star will dilute a planet's observed transit depth, resulting in an underestimated planet radius. Obtaining follow-up observations of a host star is thus important for confirming planets, identifying FP scenarios, and better refining the properties of planetary systems.

Spectroscopy and High-Resolution Imaging

Spectroscopy and high-resolution imaging are important tools for the follow-up of exoplanets detected with the transit method [103].

Spectroscopy is able to detect close-in binaries through radial velocity variations, and blended scenarios by cross-correlating the spectra with templates. Spectroscopic observations also provide more accurate stellar properties than those based only on photometry (e.g. [94]), which in turn provides more accurate planetary properties.

Meanwhile, high-resolution imaging is able to detect companions in wide orbits and those not physically associated with the target. If a star nearby is detected, a dilution-corrected planet radius $R_{p,\mathrm{corr}}$ can be estimated as

$$R_{p,\mathrm{corr}} = R_p \sqrt{1 + 10^{-0.4\Delta m}}, \tag{1.6}$$

where Δm is the magnitude difference between the two stars. In the worst case scenario of an equal-brightness companion ($\Delta m = 0$), correcting for this effect increases the estimated radius of the planet by $\sim$40%. For the case of multiple companions, the corrected radius becomes

$$R_{p,\mathrm{corr}} = R_p \sqrt{1 + \sum_{i=1}^{N} 10^{-0.4\Delta m_i}}, \tag{1.7}$$

where the sum is over all N companions. Furthermore, these images produce contrast curves, which place limits on the relative brightness of nearby stars as a function of separation from the target star.

Commonly, imaging follow-up is done using adaptive optics (AO), which allows for high-resolution images with quality comparable to those taken from space while being more accessible for scientists looking for follow-up time. Observations from the ground are often limited by atmospheric turbulence, but AO is able to compensate for this by employing deformable mirrors that correct for the effect. AO observes a bright reference star, or "guide star," to measure the distortion in the local atmosphere so that these mirrors can adapt in real time while taking the image.

Probabilistic Validation

Being able to detect an exoplanet with multiple methods increases confidence in planet status, and the applicability and discovery success of the RV method make it an obvious choice for follow-up. In fact, before *Kepler*, transiting exoplanet surveys required that planet candidates have RV measurements to determine their masses before they could be considered validated exoplanets [102]. The radial velocity and transit methods are also excellent complements naturally, as the determination of i from the transit allows the true mass of a planet to be known from RV, while the latter method provides the planet radius. By combining both mass and radius, astronomers can estimate the planet's density, which has important implications for assessing its composition.

However, with *Kepler* finding thousands of small planet candidates around relatively faint stars, RV follow-up has become increasingly infeasible. Instead, planet validation has become focused on probabilistic validation methods, which confirm a planet if a transiting planet scenario is far more likely to explain the candidate signal compared to an FP scenario. The *Kepler* team introduced this procedure for *Kepler* candidates with their `BLENDER` software [144], which models transits with blended scenarios informed by follow-up observations. `BLENDER` has led to the confirmation of several planets (e.g. [15, 77, 145]). `PASTIS` [41] was later introduced, focused on finding the Bayesian odds ratio between the planet and astrophysical FP scenarios. It has been used to validate both *Kepler* and CoRoT planets (e.g. [104, 131]).

Notably, Morton (2012) [99] introduced a procedure simpler than those outlined for `BLENDER` and `PASTIS`, with the goal of a fully automated, less time-intensive pipeline in mind. This was later developed into the publicly available `vespa` software [101], which calculates the astrophysical false positive probability (FPP) for a given signal without the need for follow-up observations. `vespa` was applied to the entire *Kepler* catalogue, leading to the validation of 1284 KOIs as bonafide planets using a total false positive probability threshold of FPP < 0.01 (99% confidence), and the labeling of

428 KOIs as false positives (FPP > 0.9) [102].

Some probabilistic validation methods have been based on more general arguments not focused on independent systems, in contrast to the above software. In particular, Lissauer et al. (2012) [87] presented statistical analysis that showed that planets in multiplanet systems were highly unlikely to be false positives. Between Lissauer et al. (2012) [87] and Rowe et al. (2014) [128], this approach has validated over 800 *Kepler* planets in multiplanet systems with 99% confidence.

1.3 Exploring Exoplanet Populations

With thousands of exoplanets now known, studies are no longer restricted to investigating individual objects of interest, but rather the demographics of the exoplanet population as a whole. The size of the ∼4700-large catalogue of planets and planet candidates from *Kepler* (Fig. 1.6) and its ability to find small planets beyond the capabilities of other missions makes it a particularly ideal data set for such investigations, known as occurrence rate studies. Occurrence rate measurements have improved our understanding of planet formation (e.g. [39, 57]), planet evolution (e.g. [89, 113]), and planetary system architectures (e.g. [107]).

When quantifying occurrence rates, the convention is to refer to the expected number of planets per star as a function of certain properties. For transit surveys, this is often reported as some function

$$\Gamma \equiv \frac{d^2 f}{d \log P d \log R_p} \tag{1.8}$$

giving the mean number of planets per logarithmic interval of orbital period and planet radius, where Γ is also called the occurrence rate density. Γ may then be integrated over a range of periods and radii to give the number of planets per star with those properties. Occurrence depends on other planetary parameters such as orbital eccentricity and inclination; stellar properties such as stellar radius, mass, temperature, metallicity, and age; and even the specifics of the planetary systems such as the number of planets in

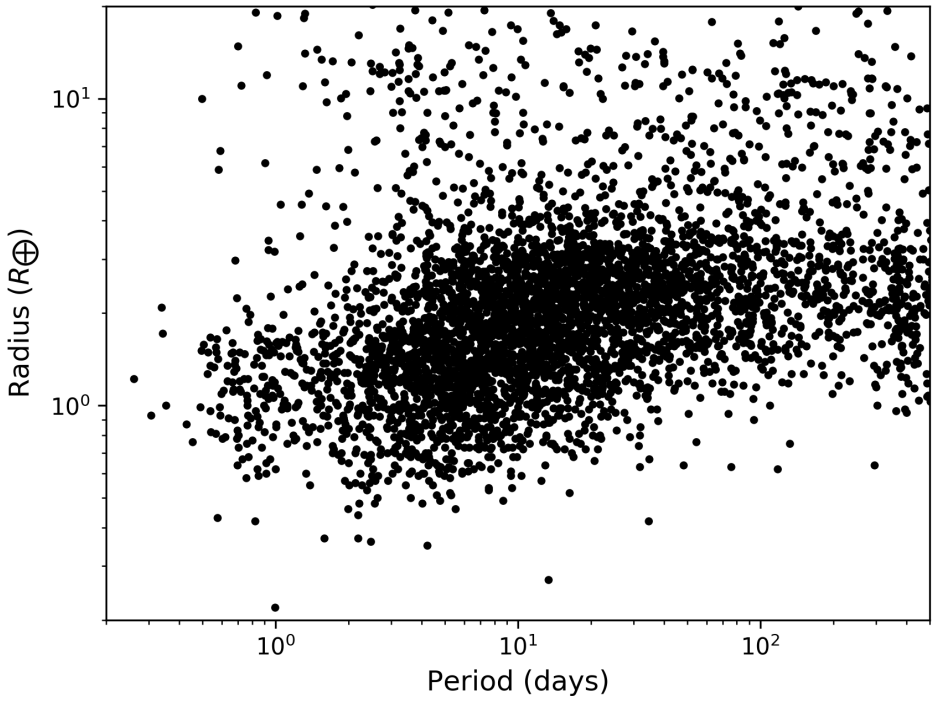

Figure 1.6: Planetary radii and orbital periods of confirmed and candidate planets found by *Kepler*, plotted up to the mission's $P \approx 500$ day sensitivity limit. Data taken from the NASA Exoplanet Archive (accessed April 24, 2020).

each system. Thus, making a single, complete, analytic form of Γ is infeasible, and techniques for calculating occurrence rates must make simplifying assumptions in both treatment of these factors and the final function.

Translating a planet catalogue into true occurrence rates is not straightforward. Ideally, one would be able to detect N_p planets around N_s stars and conclude that the integrated occurrence rate is simply $\Gamma_0 \approx N_p/N_s$, but the reality is that no transiting planet survey provides a direct measurement of the real planet population due to the presence of various selection effects. These effects must be corrected for before the true rates can be uncovered. For example, planets smaller than 0.7 $R_\oplus$ are among the rarest planets found so far, constituting only 1% of all *Kepler* confirmed planets.

Occurrence rate studies can reveal if this is indeed due to an intrinsic rarity, or if it is more a consequence of small planets simply being harder to detect.

Here, I summarize these effects, as well as some well-known approaches toward occurrence rate estimates using *Kepler* data.

1.3.1 Survey Selection Effects

First, transit planet catalogues are incomplete by nature. The majority of planets around observed stars are missed simply due to the fact that not all planets transit as observed from Earth. Furthermore, no detection pipeline is perfect, and transiting planets can be missed during the search process — particularly, planets with low S/N that are near or below the detection limit. Even following a successful detection, a catalogue's completeness can be further reduced if the vetting process mislabels a planet as an FP.

Second, planet catalogues are unreliable, in that some of the members of the catalogue may not actually be planets. Just as the vetting pipeline may incorrectly fail a planet, astrophysical or noise FPs may be incorrectly classified as planets. Again, the low-S/N regime tends to be the most affected due to the great difficulty of distinguishing such planets from noise.

It is relatively easy to correct for transit probability given the simplicity of the equation and values required. Recall Eqn. 1.5, which gives an equation for the probability of observing a planet transit, $P_{\rm tr}$. For an Earth-size planet around a Sun-like star orbiting at 1 AU, this probability is only 0.5%, indicating that 99.5% of such planets are expected to be missed by transit surveys. However, correcting for detection efficiency remains a challenge for occurrence rate studies since it is pipeline-dependent, and it can change with various factors such as the properties of individual stars and their light curves. Detection efficiency can also be affected by the order at which planets are detected in a given light curve. After finding a potential planet signal, standard detection algorithms remove the associated transits from the light curve to allow additional searches for planets. As more planets are found, less data becomes available, and transit detectability becomes more difficult.

Some occurrence rate studies have avoided the challenges by simply assuming the search pipeline is complete (e.g. [27, 146, 148]), while others have assumed a simple analytic form for the detection efficiency (e.g. [43, 62, 162]). A third method, which I use and describe in Chapter 4, involves empirically determining the detection efficiency by injecting synthetic transit signals and analysing their recovery (e.g. [29, 44, 119]). Meanwhile, imperfect vetting efficiency has only been considered in a handful of studies so far [19, 64, 65, 107, 119], and Bryson et al. (2019) [19] represents the first and so far only exoplanet occurrence rate study that corrects for reliability against astrophysical FPs and noise. Overall, how to best incorporate detection completeness, vetting completeness, and vetting reliability remains an open question.

Lastly, planet catalogues may be inaccurate, in that the properties of the planets may not be the true properties. Planet and orbital properties cannot be known exactly, due to noise in the light curve, uncertainty of stellar parameters, and potential transit dilution by nearby contaminant stars. Steps are usually made to minimize uncertainties as much as possible, such as by using adaptive optics imaging follow-up to estimate dilution, and/or using only the most up-to-date and precise stellar catalogues, but these effects have rarely been directly accounted for in occurrence rate studies so far due to the challenges involved [49, 65].

1.3.2 Occurrence Rate Methodologies

The Likelihood Function Method

One approach to characterizing occurrence rates is to model the catalogue as a Poisson realization of the rate density, taking the survey selection effects into account, and employing a maximum-likelihood method to find the model parameters. Poisson statistics are commonly used in astronomy to interpret detection surveys with various selection effects. Notably, in one of the earliest occurrence rate studies of *Kepler* data, Youdin (2011) [162] introduced a likelihood-based method that has since been adopted by members of the *Kepler* team [19, 22, 165] among others [42, 56].

In the Youdin (2011) [162] formulation, the occurrence rate density is known as a "planetary distribution function" (PLDF). Previous works have suggested that the PLDF for *Kepler* planets is well-described by a product of power laws in period and radius, such as

$$\Gamma = C R_p^{\alpha} P^{\beta} \tag{1.9}$$

where C is some constant and α and β are the power-law exponents on R_p and P respectively. Some authors also introduce a break in radius or period — in other words, they find different power-law exponents above and below some radius and/or period. For instance, Youdin (2011) [162] and Bryson et al. (2019) [19] used the smooth distribution of Eqn. 1.9, Burke et al. (2015) [22] introduced a break only in radius, and Zink et al. (2019) [165] introduced a break in both radius and period. Others have found multiple power laws only with period (or radius), one for each small range of radii (or periods) [42].

The PLDF is then used to construct a likelihood function for the outcome of a survey. Youdin (2011) [162] treated planet detection as a Poisson process, i.e. a series of independent random events, which have a total probability $\eta_{\rm tot}$ of occurring. $\eta_{\rm tot}$ is the product of various detection efficiencies that give the ratio of detections to actual planets: (i) $\eta_{\rm tr}$, the transit probability that the planet crosses our line of sight to the star (equal to $P_{\rm tr}$); (ii) $\eta_{\rm rec}$, the efficiency at which the pipeline recovers the planet (which may incorporate detection completeness, both detection and vetting completeness, or neither, depending on the author); and (iii) $\eta_{\rm fp} = 1/(1 - r_{\rm fp})$, where $r_{\rm fp}$ is the rate of FP events that are detected as planets (which is almost always assumed by authors to be negligible, i.e. $\eta_{\rm fp} = 1$). For our example Eqn. 1.9 PLDF, the likelihood function would be

$$\mathcal{L} \propto \left[C^{N_p} \prod_{i=1}^{N_p} R_{p,i}^{\alpha} P_i^{\beta} \right] \exp\left(-N_{\rm exp}\right), \tag{1.10}$$

where $N_{\rm exp}$ is the number of expected planet detections for a survey of N_s

stars within a given range of radius and period,

$$N_{\rm exp} = N_s C \int_{R_p} \int_{P} \eta_{\rm tot}(R_p, P) R_p^{\alpha} P^{\beta} d\log R_p d\log P. \tag{1.11}$$

The maximum likelihood is then calculated with analytical methods [162] or, more commonly, Bayesian estimation theory [19, 22, 165], in order to find the parameters C, α, and β that best represent the observed planet population.

An advantage of this method is that it provides a function that can be easily and flexibly integrated to find the planet occurrence rate over any desired period and radius range. It is also less sensitive to regions in period-radius space void of planet detections, given that the entire planet sample is considered at once. However, the existence of small-scale features in occurrences rates with period and radius, such as a significant gap in the distribution of planet radii between 1.5 and 2 $R_\oplus$ first explored in Fulton et al. (2017) [52], would indicate that the commonly adopted power law functions are insufficient descriptors of the data. Foreman-Mackey et al. (2014) [49] recognized these issues and broke from other works by adopting a nonparametric, piecewise-constant step function, though the popularity and near-universal use of power laws remains. Furthermore, functions fit to specific regions of period-radius space do not necessarily adequately describe *all* of period-radius space. For example, Mulders et al. (2018) [107] found that their broken power law model clearly broke down outside of $0.5 < R_p < 6.0$ $R_\oplus$ and $2 < P < 400$ days. Interpreting results based on integration outside of fitted ranges must therefore be done with great caution.

The Inverse Detection Efficiency Method

In another significant early study, Howard et al. (2012) [62] introduced the inverse detection efficiency method (IDEM), which has since been used in numerous studies [42, 43, 52, 119]. The IDEM involves making a histogram of planets where each planet is weighted by an inverse detection probability. In this formulation, the occurrence rate density is assumed to be constant

over a cell of a grid in period-radius space. In other words,

$$\Gamma_{p,r} = \frac{f_{p,r}}{\Delta \log P \Delta \log R_p} \tag{1.12}$$

is the occurrence rate density of a cell in period bin p and radius bin r, $f_{p,r}$ is the estimated occurrence rate of the cell, and $\Delta \log P \Delta \log R_p$ is the size of the cell. The cell must be small enough that this assumption is reasonable, while large enough that there are enough detected planets in the cell to place reasonable constraints on $f_{p,r}$.

Using the process outlined in Petigura et al. (2013) [119], the method starts by counting the number of detected planets in a cell, $N_{p,r}$. Then, correction factors are applied to each planet to estimate the number of planets missed due to both nontransiting orbital inclination and pipeline incompleteness. Using the same notation as in §1.3.2, the geometric transit probability of each planet corresponds to a correction of $1/\eta_{\rm tr}$, while the pipeline completeness correction is $1/\eta_{\rm rec}$. Thus, the full form of the bin's occurrence rate is

$$f_{p,r} = \frac{1}{N_s} \sum_{i=1}^{N_{p,r}} \frac{1}{\eta_{{\rm tr},i}} \frac{1}{\eta_{{\rm rec},i}} \tag{1.13}$$

where the sum is over all the planets in the bin. Poisson uncertainties in the number of planets in each bin are used to determine the uncertainty of each occurrence rate as

$$\sigma_{f_{p,r}} = \frac{f_{p,r}}{\sqrt{N_{p,r}}}. \tag{1.14}$$

The IDEM is popular because of its simplicity. Also, by finding occurrence rates on a cell-by-cell basis, the IDEM is able to uncover small-scale features that are washed out by typically assumed parametric functions, such as the Fulton et al. (2017) [52] radius gap mentioned previously. However, IDEM is not motivated probabilistically [49] and heavily relies on knowing the planet properties exactly in order to place it in the correct bin. Furthermore, the IDEM has been shown to be less accurate than other methods

and can produce artificially sharp features [49], and may be especially biased toward lower rates near the detection limit [64]. The IDEM is also unable to calculate occurrence rates over cells without planet detections. This is especially an issue for the small-planet, long-orbital-period regime where planet detections are rare or nonexistent. On a similar note, as for the likelihood function method, extrapolation and strong assumptions about the behaviour of the distribution of planets is required in order to use the IDEM to estimate occurrence rates beyond the period-radius range considered.

The Forward Modeling Method

More recently, forward modeling has been incorporated into occurrence rate calculations. Rather than starting with the data and then producing model parameters by fitting a function (as with the likelihood function method) or computing an occurrence rate cell-by-cell (as with the inverse detection efficiency method), forward modeling involves starting with guess parameters that can simulate data to be compared with the actual observations.

Mulders et al. (2018) [107] was one of the first studies to use forward modeling to estimate exoplanet occurrence rates. In their methodology, the true planet population is first drawn from an occurrence rate distribution described by a product of broken power laws for both orbital period and planet radius, with initial guess parameters. From this population, the transiting exoplanet population is simulated by excluding all planets not expected to transit according to their geometric transit probability (Eqn. 1.5), and the detectable population is simulated by further removing transiting planets not expected to be detected based on the survey detection efficiency. This final detectable population is compared to the observed population according to a summary statistic. The process is repeated by sampling the model parameter space with Markov chain Monte Carlo (MCMC) methods.

Forward modeling is also not restricted to fitting parametric functions. Hsu et al. (2018) [64] used a similar population simulator as in Mulders et al. (2018) [107], but assumed that the planet distribution could be described by a discrete period-radius grid, where the model parameters to fit were

the constant occurrence rates for individual cells. In contrast to the IDEM, their method is able to place upper limits on cells without planet detections, and occurrence rates near the detection limit are more robust [64]. Their method, which involves approximate Bayesian computation (ABC) rather than MCMC to converge on the final model parameters, will be described in full in §4.4.1.

A strength of forward modeling is that it is directly able to incorporate uncertainties in planet properties, such as uncertainties in planet radii due to both light curve noise and uncertainties in stellar properties [65]. Furthermore, Mulders et al. (2018) [107] showed how one could use their simulator to explore the population of different planetary systems, rather than simply finding the average number of planets per star. Their forward model took into account correlations in physical and orbital properties among planets in multiplanet systems, and fit parameters included the mode of the mutual inclination distribution, the orbital spacing between planets, and the frequency of planetary systems.

While forward modeling has been shown to be a powerful tool for exploring exoplanet populations from multiple perspectives, it is not without downsides. Forward modeling tends to be much more computationally expensive than other techniques, especially the IDEM, which returns occurrence rates virtually immediately. This is especially true for the grid-based studies of Hsu et al. (2018) [64] and Hsu et al. (2019) [65], for which one model parameter must be fit for every cell. Depending on cell size and the overall period-radius space considered, this could mean dozens or hundreds of model parameters, making it prohibitively expensive for convergence to be reached if fitting all cells simultaneously. As a result, grid-based exoplanet population simulators are not yet able to consider the architectures of planetary systems to the same level as power-law-based simulators. Lastly, like with all other methods, occurrence rates based on forward modeling techniques are still limited to the period-radius space covered by the planet catalogue to be compared to the simulated catalogues, and some form of extrapolation must be used for other regimes.

1.4 The Habitable Zone

One of *Kepler*'s main goals was to determine the frequency of Earth-size planets in the habitable zone of Sun-like stars [13], which is the range of orbital distances within which a rocky planet could plausibly have liquid water on its surface [76, 78]. Before we can place new planet discoveries in the context of habitability, we must define the limits of the habitable zone.

In a landmark study, Kasting et al. (1993) [76] placed estimates of the boundaries of the HZ for Earth-like planets around main-sequence stars using one-dimensional, cloud-free climate models. According to these models, the inner edge of the HZ may be determined by water loss due to photolysis and hydrogen escape. Less conservatively, the inner edge may be determined by a runaway greenhouse limit at which the oceans evaporate entirely. For the current Sun, these limits correspond to 0.95 and 0.84 AU respectively. Meanwhile, an outer edge at 1.37 AU may be determined by the maximum greenhouse limit, which is the point at which the cooling effects of condensation and scattering by CO_2 outweigh its greenhouse warming, leading to runaway glaciations.

Kopparapu et al. (2013) [78] revised the Kasting et al. (1993) [76] estimates with an updated climate model, which consequently moved the water-loss and runaway greenhouse inner edges up to 0.99 AU and 0.97 AU, and the maximum greenhouse outer edge out to 1.70 AU. Kopparapu et al. (2013) [78] also produced empirical HZ estimates based on considering the flux received by Venus and Mars when they may have last hosted liquid water on their surfaces. For Venus, this may have been as recently as 1 Gyr ago, at which point the solar flux received by Venus would have been 1.76 times that of the Earth, corresponding to an orbital distance of 0.75 AU for the present day. Similarly, a potentially water-hosting early Mars 3.8 Gyr ago would have received 0.32 times the solar flux of the Earth, corresponding to a present day orbital distance of 1.77 AU.

The above Kopparapu et al. (2013) [78] limits can be combined to give two definitions of the HZ: the wide, "optimistic" HZ, using the recent Venus and early Mars limits ($0.75 - 1.77$ AU), and the narrow, "conservative" HZ,

using the water-loss and maximum greenhouse limits (0.99−1.70 AU). These have been used by the majority of recent occurrence rate studies interested in estimating the frequency of small planets in the habitable zones of Sun-like stars (e.g. [19, 56, 65, 79, 107, 166]).

Other conventions for identifying planets in the HZ include calculating either the planet's equilibrium temperature or the stellar flux received by the planet. The equilibrium temperature $T_{\rm eq}$ can be estimated assuming thermodynamic equilibrium between the incident stellar flux and the radiated heat from the planet,

$$T_{\rm eq} = T_{\rm eff}(1-A)^{1/4}\sqrt{\frac{R_s}{2a}}, \tag{1.15}$$

where $T_{\rm eff}$ is the effective temperature of the host star, R_s is the radius of the star, and A is the Bond albedo of the planet (the fraction of total power incident upon the planet scattered back into space). Note that $T_{\rm eq}$ is not equal to the surface temperature of the planet, as it does not take into account the presence of an atmosphere. Meanwhile, the amount of stellar flux received by the planet S, or insolation, is determined relative to the solar flux received by the Earth $S_\oplus$ by

$$\frac{S}{S_\oplus} = \left(\frac{R_s}{R_\odot}\right)^2\left(\frac{a_\oplus}{a}\right)^2\left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right)^4. \tag{1.16}$$

For example, the NASA Exoplanet Archive considers an exoplanet to be in the HZ if it has an equilibrium temperature (assuming an albedo equal to the albedo of Earth, $A = 0.3$) between 180 and 310 K, or insolation between 0.25 and 2.2 $S_\oplus$.[9] Petigura et al. (2013) [119] also adopted a simple HZ definition of between 0.25 and 4 $S_\oplus$.

It is worthwhile to note that in this section, I have only considered habitable zones for planets similar to the Earth. There are several discussions of habitability that are beyond the scope of this overview, such as implications for free-floating planets [1] or desert worlds with limited surface water [2]. Furthermore, a planet in the HZ is not necessarily habitable; there exist nu-

[9] From https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html.

merous complicating properties and processes such as orbital evolution, the existence of other planets in the system, the bombardment of radiation from a host star, the presence of plate tectonics, and various atmospheric characteristics, all of which may impact a planet's ability to support surface liquid water or life. Nevertheless, the habitable zone is a useful and commonly used first order assessment of a planet's potential for habitability.

Chapter 2

Motivations for this Thesis

2.1 An Independent Search of *Kepler* Data

Kepler archival data, which is publicly available on the Mikulski Archive for Space Telescopes[10] (MAST), has been a popular focus of independent searches for exoplanets over the years. The citizen science initiative known as "Planet Hunters," for instance, was launched with the goal of involving the general public in *Kepler* data analysis and planet detection. Since their first discoveries in Fischer et al. (2012) [46], the Planet Hunters project has uncovered over 100 planet candidates with *Kepler* data [86, 133–135, 153, 154]. Meanwhile, Ofir & Dreizler (2013) [111] is one of the earliest systematic searches in the literature, restricting the subset of stars searched to known KOIs. In Q0-Q6 data,[11] they found 84 new candidates. Huang et al. (2013) [66] performed a search of 124,840 stars observed in Q1-Q6, finding 150 new candidates. Jackson et al. (2013) [70] and Sanchis-Ojeda et al. (2014) [130] searched all ∼200,000 stars for ultra-short period planets, finding four new candidates in Q0-Q11 data and 16 new candidates in Q0-Q16 data respectively.

Even following the release of the final *Kepler* catalogues based on all four years (Q1-Q17) of *Kepler* data [36, 143], independent authors have shown that there are still scientifically interesting candidates to be found. Shallue & Vanderburg (2018) [137] performed the first application of machine learning to searching *Kepler* data for exoplanets, finding two new planets among a subset of Q1-Q17 KOIs including an eighth planet in a planetary system. For my honours undergraduate thesis, I also searched a subset of KOIs, finding

[10] http://archive.stsci.edu/

[11] Each "quarter" (Q) of data refers to one quarter of a year, or ∼90 days.

four new candidates including a Neptune-sized candidate in the habitable zone [81]. More recently, Caceres et al. (2019) [25] searched 156,717 stars for planets with orbital periods between 0.2 and 100 days, finding 97 new candidates.

So far, independent *Kepler* searches have focused on only a subset of light curves (e.g. [81, 137]) or a more limited range of orbital periods than examined by the *Kepler* team (e.g. [25, 154]). This motivated me to perform the first independent, systematic search of the entirety of *Kepler* data for new exoplanets, using the same three-transit minimum detection criteria as the *Kepler* team. This independent search is the focus of Chapter 3.

While the final number of new candidates I find through my search is not expected to significantly change the overall ∼4700 *Kepler* exoplanet count, my results can still make important contributions to certain areas of exoplanet parameter space. In particular, planets with low S/N — whether because of long orbital periods resulting in few observed transits, or small radii resulting in shallow transit depths, or both — such as potentially habitable Earth-size planets in year-long orbits, are the most likely to be missed by previous searches. Discovering these kinds of planets is only the first step in the investigation of prospects for exoplanet habitability and the search for life elsewhere in the galaxy.

Searching such a large sample of stars also enables an exoplanet population analysis of my own planet catalogue. Thus, regardless of how many new candidates I find, the significance of my search will manifest through the occurrence rate investigation constituting the second half of my dissertation.

2.2 Occurrence Rate Estimates

Petigura et al. (2013) [119] and Dressing & Charbonneau (2015) [44] are two works that used independent searches of *Kepler* data as a precursor for occurrence rate statistics, focused on GK and M dwarf stars respectively. Since all other estimates so far have been based on *Kepler* catalogues, independent searches are unique contributions to the field and can test the replicability of *Kepler* results. I used these papers as inspiration to use my

own independent search for occurrence rate estimates.

Occurrence rates are fundamental observational results which have provided invaluable support and constraints for current planet formation and evolution theories, and inspired the rewriting of others. For example, early results from radial velocity surveys that showed a strong correlation between giant planet occurrence rate and host star metallicity (e.g. [47, 132]) are widely considered strong support for the core accretion theory of giant planet formation; the discovery of a gap in the radius distribution for planets with periods shorter than 100 days [52] provided invaluable support for previous theoretical predictions for the evolution of low-mass planets under the effects of photoevaporation [113]; and the findings of surprisingly high abundances of close-orbiting super-Earths and sub-Neptunes have discounted models that predicted they should be especially rare (e.g. [67]). Furthermore, determining the average number of potentially habitable Earth-size planets per Sun-like star — also known as "eta-Earth," $\eta_{\oplus}$ — is vital for the success of future missions to detect and characterize Earth-like planets and possibly signs of life. Knowledge of the rarity of these planets affects telescope design, mission lifetime, and target choice, especially for future direct imaging missions which will have sensitivity to small planets with orbits of hundreds of days, outside of the sensitivity of other missions and methods.

However, calculating occurrence rates is not straightforward. As discussed in §1.3.1, several selection effects and biases exist that complicate the process, and each have been treated differently by various authors. Some previous works have assumed an analytic function of signal-to-noise for detection efficiency (e.g. [62, 162]), some have empirically estimated detection efficiency by injecting synthetic planet transit signals into light curves and testing recovery (e.g. [29, 119]), and others have ignored it altogether by assuming that the catalogue is complete (e.g. [27]). Meanwhile, most works have ignored vetting completeness, and the reliability of a planet sample against transit-like noise was only performed for the first time in Bryson et al. (2019) [19]. Nearly all previous studies have also ignored uncertainty in planet radius, instead assuming that a detected planet's measured radius is exactly its true radius. Lastly, even when using the exact same dataset

and characterization of completeness, different methods used to calculate occurrence rates can produce inconsistent results (e.g. compare [119] and [49]).

Estimating $\eta_{\oplus}$ has even more added challenges. Finding Earth-sized planets is difficult due to their small sizes and low transit signal-to-noise ratios, meaning planet detection pipelines have greater difficulty uncovering them than larger planets, and a higher risk of confusing them with transit-like noise in the data. Finding small planets in the habitable zones of Sun-like (G-dwarf) stars is even more difficult due to their year-long orbits necessitating a bare minimum of several years of observations to observe just a few transits. Common definitions of the HZ also place potentially habitable planets hundreds of days outside even the $P \approx 500$ day sensitivity limit of *Kepler*, necessitating the extrapolation of occurrence rates based on smaller orbital periods. Lastly, numerous definitions of the limits of the HZ exist, and there is no standardized consensus yet on what size range constitutes a sufficiently "Earth-sized" planet. Reflecting these complications, $\eta_{\oplus}$ values in the literature span orders of magnitude. At one end, Catanzarite & Shao (2011) [27] found an $\eta_{\oplus}$ between 0.01 and 0.03 planets per star; at the other, Garrett et al. (2018) [56] estimated an $\eta_{\oplus}$ value of greater than 1.

New, robust estimates are invaluable in bringing the exoplanet community toward consensus. I will address these shortcomings by performing the first occurrence study to directly address all aforementioned selection effects, including the vetting completeness, reliability, and planet radius uncertainties ignored by most previous works. This work is the content of Chapter 4.

My investigations will allow me to comment on the full period-radius space accessible by *Kepler*, thanks to the wide scope of my independent search. By splitting up the stellar sample over different kinds of stars (F-, G-, and K-dwarfs), I will investigate how planet occurrence rates change with stellar effective temperature; by calculating occurrence rates over a wide range of orbital periods ($P = 0.78 - 400$ days), I will comment on dependencies with period; and by calculating occurrence rates over a wide

range of planet radii ($R_p = 0.5 - 16\ R_\oplus$), I will comment on dependencies with radius. Lastly, I will be able take into account optimistic and conservative HZ limits and different Earth-size definitions to provide multiple estimates of $\eta_\oplus$, which can be considered alongside others in the literature.

Chapter 3

An Independent Search of *Kepler* Data

This chapter has been adapted from a manuscript published in The Astronomical Journal: Kunimoto, M., Matthews, J., & Ngo, H. 2020, Searching the Entirety of Kepler Data. I. 17 New Planet Candidates Including 1 Habitable Zone World, AJ, 159, 124.

3.1 Chapter Outline

I describe my planet detection and vetting pipelines in §3.2 and §3.3. My pipeline's ability to differentiate between planets and noise-like false positives is discussed in §3.4. I compare my results to the findings across all *Kepler* catalogues in §3.5 and detail new planet candidates in §3.6. These candidates are processed through further analysis, including centroid vetting, adaptive optics imaging follow-up, and astrophysical false positive probability calculation to increase confidence in their planet status.

3.2 Planet Detection Pipeline

3.2.1 Preparing the Light Curves

Q1-Q17 DR25 long-cadence *Kepler* light curve files were downloaded from the MAST. This photometry includes systematic corrections for instrumental trends and estimates of dilution due to other stars that may contaminate the photometric aperture [141]. To initially set up the data, I used the

`kfitsread` routine from the *Kepler* Transit Model Codebase [127], which includes code previously used by the *Kepler* team for transit detection and characterization. `kfitsread` reads in each FITS file, removes data flagged as low quality, stitches all quarters of data together to create one continuous light curve, and subtracts the median flux from each data point.

I then detrended each light curve to filter out astrophysical and instrumental signatures using the `detrend5` routine [127]. Each observation was corrected by fitting a cubic polynomial to a segment W days wide centred on the time of measurement. A good choice of W is longer than the duration of a typical transit (several hours) to prevent significant transit shape distortion, while short enough to adequately filter out astrophysical signatures (several days). The ideal choice of W is star-dependent, as stars having varying levels of noise and intrinsic stellar variability. In an effort to reflect this, I detrended each light curve using $W = 1$, 1.5, and 2 days, and measured the corresponding standard deviations σ_1, $\sigma_{1.5}$, and σ_2. I used $W = 1$ if $\sigma_1/\sigma_{1.5} < 0.8$ or $\sigma_1/\sigma_2 < 0.8$, $W = 1.5$ if $\sigma_{1.5}/\sigma_2 < 0.8$, and $W = 2$ otherwise, similar to the process used by Dressing & Charbonneau (2015) [44] to flag stars with greater levels of noise. In other words, a more aggressive detrend would only be favoured if it resulted in a significant decrease in overall light curve variability.

I then 5σ-clipped outliers in the data. Only outliers in the positive flux direction were removed so as to leave deep transits untouched. Lastly, I searched for data gaps within the detrended light curves. For example, the *Kepler* telescope would execute a 90° roll every 90 days to reorient its solar panels, resulting in a break in observations of approximately one day. I defined a data gap as 0.75 or more days of missing photometry. Gaps are often accompanied by sharp increases or decreases in flux, which can interfere with the search for transits. Thus, I removed all data points within 1 day of the start and end of each gap.

3.2.2 Searching for Transiting Planets

I searched the light curves using a box least-squares (BLS) routine based on the original algorithm by Kovacs et al. (2002) [80], which was designed to identify periodic transit signals in time-series photometry. In the *Kepler* Transit Model Codebase, this is available as `transitfind2`.

Once a possible transit was identified, its period P, epoch of first transit time T_0, depth δ, and duration $T_{\rm dur}$ were estimated. I calculated the signal-to-noise ratio of the event by dividing the mean transit depth by the standard error of the mean, giving

$$\mathrm{S/N} = \sqrt{N}\frac{\delta}{\sigma}, \tag{3.1}$$

where σ is the standard deviation of the observations and N is the number of in-transit data points. Following Rowe et al. (2014) [128], I estimated σ using the standard deviation of all out-of-transit observations — defined as data outside of two transit durations of the centre of the detected signal — and used the Median Absolute Deviation (MAD) with $\sigma = 1.48\mathrm{MAD}$ assuming a normal distribution [60] to be more robust to outliers. Eqn. 3.1 is comparable to the "effective" S/N described in Kovacs et al. (2002) [80], and assumes that the depth of the transit is uniform. While this is a good approximation for small Earth-sized planets with central transits, relatively large planet-to-star radius ratios and/or large impact parameters can have significant ingress and egress durations. In these cases, the S/N will be overestimated, but this is expected to have minimal impact on the full assessment of planet candidate events [128].

For each light curve, I searched for transit signals with $\mathrm{S/N} > 6$. After identifying a transit signal, I removed its associated events from the data and searched the residuals. This enabled sensitivity to multiplanet systems. I capped this multipassthrough search at five consecutive searches and set aside light curves that reached this maximum for manual inspection. Usually this meant a particularly noisy light curves was causing the BLS algorithm to return many obviously poor signals. If this was not the case, I would continue searching until no more signals with $\mathrm{S/N} > 6$ were detected.

Choice of S/N Threshold

The significance of a detection depends primarily on its associated signal-to-noise ratio. Here, I follow the suggestion of Kovacs et al. (2002) [80] that the threshold for a significant detection with the BLS algorithm is S/N = 6.

This is a lower threshold than most previous exoplanet searches, including the MES = 7.1 threshold[12] used by the *Kepler* team [71] and some other BLS-based searches (e.g. S/N = 9 used by Vanderburg et al. (2016) [152]; S/N = 7 used by Rizzuto et al. (2017) [125]). While it is true that the rate of false alarms increases rapidly toward lower signal-to-noise, the true floor depends on characteristics of the host star, and behaviour of the instrument on timescales related to the properties of the transit candidates. Digging deeper into the noise increases the probabilities of discovering and characterizing transiting planets that are small and/or have long periods and thus very few transits. These both represent regimes of great interest in the exploration of exoplanetary parameter space. Furthermore, searching to lower S/N can increase sensitivity to signals with S/N *above* more conservative cutoffs. S/N are often underestimated, such as through the assumption of a box-shaped transit or the distortion of the transit shape from the detrending algorithm, causing a planet signal to be erroneously rejected. For these reasons, recent searches have begun to relax the noise floor, such as Shallue & Vanderburg (2018) [137] in which a BLS algorithm was used to search as low as S/N = 5, and Kunimoto et al. (2018) [81], in which I previously searched a small subset of *Kepler* light curves down to S/N = 6.

3.2.3 Initial Vetting and Identification of Transit Candidates

A total of 130,312 signals with S/N > 6 were detected with the BLS algorithm around the 198,640 stars searched. An overwhelming number of these are false alarms due to instrumental or astrophysical systematics in the time series. This is a weakness of using S/N as the only detection criteria. Thus,

[12]The Multiple Event Statistic (MES) adopted by the *Kepler* pipeline is not exactly the same as the S/N from the BLS algorithm, but may be considered comparable.

before following up each signal with the full suite of candidacy tests, I ran a first stage of vetting to discard likely false alarms, as follows. Signals that pass these tests are designated transit candidates (TCs).

Since the detrending process is destructive to the shape of a planetary transit and often results in a loss in S/N, I produced a redetrended version of each light curve before running the tests. I masked out each transit by excluding all observations within one transit duration of the central time of each transit. Then, I ran the detrending algorithm to determine the cubic polynomial fit to each segment of data as before, essentially only fitting to all out-of-transit observations. Finally, I unmasked the transit in the light curve and used an extrapolation of the fit to estimate corrections during transit. These redetrended light curves are used for the remainder of the analysis in this paper unless otherwise specified.

S/N Recalculation Test

The S/N of a planet signal in the redetrended light curve should still be strong enough compared to the noise to warrant transit candidacy. Thus, I require that both the original and recalculated S/N remain above 6.

Robust Statistic Test

A weakness of using the BLS S/N to indicate the strength of the signal is that it is unable to discriminate between a consistent set of transit events of uniform depths and durations, and a chance combination of dissimilar events. This test calculates a second signal-to-noise ratio, RS, using the median depth instead of the mean to reduce the influence of outliers on S/N. A signal passes this test with RS > 6 in both the original and redetrended light curves.

S/N Consistency Test

This test examines the signal-to-noise ratios of each transit individually and compares them to what is expected based off the full transit S/N. Since the depths of individual transits of planet candidates should be equal to

each other, the ith transit comprised of n_i observations has an expected signal-to-noise ratio of

$$\langle \mathrm{S/N}_i \rangle = \sqrt{n_i} \frac{\delta}{\sigma} \tag{3.2}$$

where δ is the overall BLS transit depth. Meanwhile, the actual signal-to-noise ratio of the ith transit with depth δ_i is

$$\mathrm{S/N}_i = \sqrt{n_i} \frac{\delta_i}{\sigma}. \tag{3.3}$$

These metrics are compared using a χ^2 statistic with N_T degrees of freedom,

$$\chi^2 = \sum_{i=1}^{N_T} (\mathrm{S/N}_i - \langle \mathrm{S/N}_i \rangle)^2, \tag{3.4}$$

where N_T is the number of transits. I define

$$\mathrm{CHI} = \mathrm{S/N} \left(\frac{\chi^2}{N_T} \right)^{-1/2} \tag{3.5}$$

for use as the false alarm discriminator, requiring a candidate to pass with CHI > 6 in both the original and redetrended light curves.

Number of Transits

Each signal must have at least three transits. A minimum of two transits are required to determine orbital period, while requiring a third improves reliability of the period estimate, reduces false detections, and increases the overall S/N. Simply dividing the total length of observations by the orbital period to get an estimate of the number of transits is insufficient as some transits may lie in gaps in the data. To avoid counting transits in gaps, I only count the number of transits that occur at epochs where data exists within 0.5 transit durations of the midpoint.

After applying all of the above cuts, I was left with 33,322 TCs out of the 130,312 signals with S/N > 6.

3.3 Vetting Pipeline

While the first stage of vetting significantly reduces the rate of false detections, some of the 33,322 TCs could still be due to noise, systematics, or astrophysical FPs. Thus, a suite of diagnostic tests must be performed to confirm (or refute) the candidacy of each signal as a bonafide transiting planet. My vetting pipeline involves automated candidacy tests followed by a round of manual inspection. TCs that pass both automated and manual triage are upgraded to planet candidates (PCs).

3.3.1 Choice of Candidacy Test Thresholds

While the thresholds chosen for the automated tests described in this section may seem arbitrary, I wish to emphasize that they were empirically chosen under several considerations.

First, my vetting pipeline was largely inspired by *Kepler*'s Robovetter, an automated vetting tool first used for *Kepler*'s DR24 catalogue [36] and again for DR25 [143] (hereafter KDR25). When possible, I used the same or similar cutoffs as used by the *Kepler* team for the *Kepler*-equivalent tests.

Second, in §3.4, I discuss the use of simulated planets and noise/systematic false positives for assessing the performance of my vetting pipeline. I also used these data products during the design of my vetting pipeline by exploring the consequences of raising and lower various thresholds. Primarily, these investigations informed the choices behind the thresholds for tests without *Kepler* equivalents.

Lastly, while I hope to make my vetting pipeline completely automated in the future, manual inspection still plays an integral role in the vetting process. The main goal of my automated candidacy tests was to reduce the number of false positives sufficiently enough that the number of transit candidates requiring manual review was feasible.

3.3.2 Transit Model Fitting

The vetting pipeline starts with obtaining a transit model fit for each TC. Several of the candidacy tests require a transit model fit, and fitting better characterizes the parameters of each candidate.

I used a Mandel & Agol (2002) [90] quadratic limb-darkening transit model assuming circular orbits,[13] fit to each transit with least-squares. Limb-darkening parameters were taken from Claret & Bloemen (2011) [33] based on the known $T_{\rm eff}$, log$g$, and [Fe/H] from the Mathur et al. (2017) [94] DR25 stellar properties catalogue. To speed up the fit process, data more than two transit durations from the centre of each transit were ignored.

Fitted Parameters

The model is parameterized by orbital period, transit epoch, ratio of planet and star radii (R_p/R_s), distance between planet and star at midtransit in units of stellar radius (a/R_s), impact parameter (b), and offset of the flux from zero near the transit, otherwise known as the zero-point flux (z).

For initial guesses, P and T_0 were taken from the BLS search results. R_p/R_s was estimated as the square root of the BLS transit depth,

$$\frac{R_p}{R_s} = \sqrt{\delta}, \tag{3.6}$$

and the initial guess for z was 0. a/R_s was estimated using Eqn. 8 in Seager & Mallen-Ornelas (2003) [136]:

$$\frac{a}{R_s} = \left[\frac{(1+\sqrt{\delta})^2 - b^2(1-\sin^2\frac{\pi T_{\rm dur}}{P})}{\sin^2\frac{\pi T_{\rm dur}}{P}}\right]^{1/2} \tag{3.7}$$

with b set to an initial guess and transit duration $T_{\rm dur}$ set to the BLS estimate. Since the model is sensitive to b, the fit was run once for each $b = 0, 0.1, 0.2, ..., 0.9$. The fit with reduced χ^2 closest to 1 was chosen to determine the best-fit parameters.

[13] Adapted from Ian's Astro-Python Codes at http://www.lpl.arizona.edu/~ianc/python/

In case the transit model fit would fail to converge, a trapezoid fit parameterized by T_0, R_p/R_s, z, width of the flat part of transit, and slope of the sides of the trapezoid was used instead. P was set fixed to the BLS-detected value.

3.3.3 Candidacy Tests Against Noise False Positives

The first candidacy tests aim to identify false positives due to noise and systematics. As discussed in §1.2.1, candidates with low S/N and/or few transits are often simply due to various types of noise or instrumental artifacts. Candidates may also be due to quasi-sinusoidal signals such as pulsating stars or star spots that were not completely removed during the detrending process.

Transit Model Fit Test

The transit model should fit the data better than a straight line, parameterized by the zero-point flux z. This tests compares each model fit's reduced chi-squared values

$$\chi^2_{\rm red} = \frac{1}{\nu} \sum_{i=1}^{N} \frac{(y_i - m_i)^2}{\sigma_i^2} \tag{3.8}$$

where y_i, m_i, and σ_i represent the flux, modeled flux, and error of the ith data point, and ν are the total degrees of freedom. Given N points fitted, the transit model with six parameters will have $N-6$ degrees of freedom while the straight line model with one parameter has $N-1$. A signal passes this test if the reduced chi-squared of the transit model is less than the reduced chi-squared of the straight line model.

Transit Model S/N Test

The signal-to-noise ratio of the model fit, MOD, should be slightly larger than the S/N of the signal due to a variety of reasons: namely, a transit model should match the shape of the TC better than a BLS square pulse, and the ephemerides should be more refined. A significantly lower MOD

than S/N calls into question the planetary origin of the signal. This test requires both MOD > 6 and MOD/S/N > 0.75.

Depth Mean-to-Median Ratio Test

The mean of all measured transit depths should be consistent with the median of all transit depths. Thus, the depth mean-to-median (DMM) ratio can be used to identify potential scenarios when a candidate is due to a systematic. If the DMM value is significantly different from 1.0, it indicates that some transits have significantly different depths from the rest, and thus the candidate is unlikely to be astrophysical in origin. A candidate fails this test if DMM > 1.5.

Chases Test

The *Kepler* team developed an individual transit metric called Chases to assess the detection strength of transit events relative to nearby signals [143]. Chases is only calculated for candidates with five or fewer transits. This test takes the median of the individual Chases metrics (see §3.3.3) and fails candidates with a value less than 0.8 — the same threshold as used by KDR25.

Uniqueness Tests

A transit must be considered unique, meaning there should be no other events in the folded light curve with a depth and duration similar to the primary signal, in either the positive or negative flux directions. My vetting pipeline employs two different tests to ensure this is the case.

First, I calculate two statistics for each TC:

$$\sigma_{\mathrm{U1}} = \frac{|d_{\mathrm{pri}} - d_{\mathrm{sec}}|}{\sqrt{\sigma_{\mathrm{pri}}^2 + \sigma_{\mathrm{sec}}^2}} \tag{3.9}$$

$$\sigma_{\mathrm{U2}} = \frac{|d_{\mathrm{pri}} - d_{\mathrm{ter}}|}{\sqrt{\sigma_{\mathrm{pri}}^2 + \sigma_{\mathrm{ter}}^2}} \tag{3.10}$$

where $d_{\rm pri}$, $d_{\rm sec}$, and $d_{\rm ter}$ are the depths of the TC event, second-largest event, and third-largest event respectively, and $\sigma_{\rm pri}$, $\sigma_{\rm sec}$, and $\sigma_{\rm ter}$ are their uncertainties. Secondary and tertiary events may be in either the positive or negative flux direction. A TC must have both $\sigma_{\rm U1} > 3.0$ and $\sigma_{\rm U2} > 3.0$ (i.e. at least 3σ significance compared to other events).

Running this analysis on the redetrended light curve is a good choice when the transit is due to a planet, as it ensures the planet transit is not distorted by detrending and the test correctly indicates strong uniqueness. However, a noise TC is essentially the only noise in the light curve that is not detrended in this version of the light curve, making an indication of uniqueness against the rest of the noise misleading. Thus, the pipeline also performs this analysis on the original light curve, requiring a slightly lower 2σ significance to pass.

Second, I use the *Kepler* team's own model-shift uniqueness test [143], which is publicly available on GitHub[14] [34]. While the statistics described above take into account the uniqueness of the TC in the form of a square pulse, this test takes into account the full transit shape as follows. After removing outliers, the best fit model of the primary transit is used to measure the best-fit depth at all other phases. The two deepest events aside from the primary event (called the secondary and tertiary events) and the most positive flux event are all identified. The significances of these events ($\sigma_{\rm pri}$, $\sigma_{\rm sec}$, $\sigma_{\rm ter}$, and $\sigma_{\rm pos}$) are computed by dividing their depths by the standard deviation of the light curve residuals outside of the primary and secondary events, assuming white noise. The amount of systematic red noise in the light curve on the timescale of the transit is also computed, as the standard deviation of the best-fit depths at phases outside of the primary and secondary events. Taking the ratio of the red noise to the white noise gives the value $F_{\rm red}$. $F_{\rm red} = 1$ means there is no red noise in the light curve.

With this test, the threshold at which an event is considered statistically

[14] https://github.com/JeffLCoughlin/Model-Shift

significant is given by

$$FA_1 = \sqrt{2}\ \mathrm{erfcinv}\left(\frac{T_{\mathrm{dur}}}{P \cdot N_{\mathrm{TCs}}}\right). \tag{3.11}$$

Here N_{TCs} is the number of transit candidates examined, the quantity P/T_{dur} represents the number of independent statistical tests for a single target, and erfcinv is the inverse complementary error function. Similarly, the threshold at which the difference in significance between two events is considered to be significant is given by

$$FA_2 = \sqrt{2}\ \mathrm{erfcinv}\left(\frac{T_{\mathrm{dur}}}{P}\right). \tag{3.12}$$

The following quantities are used as decision metrics:

$$MS_1 = FA_1 - \sigma_{\mathrm{pri}}/F_{\mathrm{red}} \tag{3.13}$$

$$MS_2 = FA_2 - (\sigma_{\mathrm{pri}} - \sigma_{\mathrm{ter}}) \tag{3.14}$$

$$MS_3 = FA_2 - (\sigma_{\mathrm{pri}} - \sigma_{\mathrm{pos}}). \tag{3.15}$$

A candidate fails the test if either $MS_1 > -3$, $MS_2 > 1$, or $MS_3 > 1$. These criteria ensure that the primary event is statistically significant when compared to the systematic noise level of the light curve, the tertiary event, and the positive event, respectively.

Transit Shape Test

The transit shape test determines if the measured depth deviates from the mean value more in the positive flux direction, negative flux direction, or are symmetrically distributed in both directions. The SHP metric, provided alongside the model-shift uniqueness test from Coughlin (2017a) [34], is

defined by

$$\mathrm{SHP} = \frac{F_{\max}}{F_{\max} - F_{\min}} \tag{3.16}$$

where $F_{\max}$ and $F_{\min}$ are the maximum and minimum measured flux amplitudes respectively. Since the light curve is normalized, $F_{\max}$ is always a positive value and $F_{\min}$ is always negative. SHP lies between 0 and 1, where 0 indicates the light curve only decreases in flux, consistent with a planet transit, and a value near 1 indicates the light curve only increases in flux, such as for a lensing event or systematic outlier. A candidate passes with SHP < 0.5.

Single Event Domination Test

Assuming all individual transits have equal signal-to-noise ratios, $\mathrm{S/N_I}$, the full transit S/N given in Eqn. 3.1 can be rewritten as

$$\mathrm{S/N} = \sqrt{N_T}\ \mathrm{S/N_I} \tag{3.17}$$

where N_T is the number of transits. It follows that if the largest individual transit's S/N value, $\mathrm{S/N}_{i,\max}$, divided by the S/N is much larger than $1/\sqrt{N_T}$, the calculation of the candidate's S/N is likely dominated by one of the individual events.

A candidate fails this test if $\mathrm{S/N}_{i,\max}/\mathrm{S/N} > 0.8$, as in the *Kepler* team's own signal event domination test from KDR25. I also make the same choice as the *Kepler* team to only test candidates with $P > 90$ days. Candidates with shorter orbital periods tend to have a larger number of individual transit events, increasing the chance of one event coinciding with a large systematic feature, which reduces the reliability of this test.

Individual Transit Metrics

This series of metrics examines individual transits, and flags those that fail. After removing flagged events, the resulting signal must still have at least three transits and S/N > 6.

Rubble Metric

As per the *Kepler* team's "Rubble" metric [143], transit events may be missing a significant amount of data, either during transit or before and/or after. For each event I count the number of datapoints within one transit duration of the centre of the transit and divide this by the number of cadences expected given 29.42 minutes per cadence. As in KDR25, an event is flagged if this value is less than 0.75.

Chases Metric

The *Kepler* team developed the Chases metric to identify non-transit-like events in long-period, low-S/N candidates by mimicking the tendency of human vetters to classify transits that "stand out" as planet candidates [143]. Chases uses the Single Event Statistic (SES) time series generated by the Transit Pipeline Search (TPS) module of the *Kepler* Pipeline [74], which measures the significance of a signal centred on every cadence. A transit produces a peak in the SES time series.

I created an analogous time series of S/N values centred on every cadence for the purpose of this test. The Chases metric is determined by first identifying the maximum S/N value for cadences in transit, $\mathrm{S/N_{max}}$. The S/N time series is searched for Δ_t, the time of the closest signal with $|\mathrm{S/N}| > 0.6$ $\mathrm{S/N_{max}}$. As in KDR25, the search range starts at 1.5 T_{dur} from midtransit, up to a maximum $\Delta_{t,\mathrm{max}} = P/10$, on either side of the transit candidate signal. The final Chase metric is determined as $C_i = \min(\Delta_t, \Delta_{t,\mathrm{max}})/\Delta_{t,\mathrm{max}}$.

A value of $C_i \approx 0$ indicates an event of comparable strength of the transit is close to the transit event, while a value of $C_i = 1$ indicates there is no comparable peak or trough, and the transit is unique.

Chases metrics are only computed for TCs with five or fewer transit events, as these events are expected to be especially significant in order to combine to have $\mathrm{S/N} > 6$. Events with $C_i < 0.01$ are flagged.

Negative Significance

A valid transit should only be comprised of events corresponding to decreases in the flux. Any individual event with $\mathrm{S/N} < 0$, indicating a flux

increase, is flagged.

3.3.4 Candidacy Tests Against Eclipsing Binary False Positives

TCs that pass the previous tests are considered transit-like. However, some may still be nonplanetary in origin. As discussed in §1.2.1, one of the most common types of astrophysical false positives are eclipsing binary stars, which could just graze the target star enough for the eclipse depth to be consistent with a planet transit.

Transit-like FPs may also be due to off-target signals, such as background eclipsing binaries or planet transit signals coming from off-target sources. These scenarios can typically be indicated by identifying significant centroid offsets. My vetting pipeline does not currently incorporate automated tests to identify these FPs. However, as described in §3.6, I performed centroid analysis as part of a more in-depth analysis of new PCs.

Significant Secondary Test

A secondary eclipse could manifest as the secondary event in the phased light curve. This test follows the same procedure as *Kepler*'s model-shift uniqueness test, but assesses the uniqueness of the secondary event rather than the primary using a new set of metrics:

$$MS_4 = FA_1 - \sigma_{\rm sec}/F_{\rm red} \tag{3.18}$$

$$MS_5 = FA_2 - (\sigma_{\rm sec} - \sigma_{\rm ter}) \tag{3.19}$$

$$MS_6 = FA_2 - (\sigma_{\rm sec} - \sigma_{\rm pos}). \tag{3.20}$$

If either $MS_4 > 2$, $MS_5 > 1$, or $MS_6 > 1$, the candidate fails due to having a significant secondary event.

Planet Candidates with Significant Secondaries

Significant secondary events are not necessarily confirmation of an eclipsing binary false positive.

Following KDR25, if the primary and secondary events have statistically indistinguishable depths and the secondary is at phase 0.5, a planet candidate may have been detected at twice its actual orbital period. Thus, a TC is allowed to pass the Significant Secondary test if $\sigma_{\rm pri} - \sigma_{\rm sec} < FA_2$ and the phase of the secondary is within $T_{\rm dur}/4$ of 0.5.

Additionally, some giant planets close to their stars such as hot Jupiters can have eclipses due to planetary occultations via reflected light and thermal emission. The depths of these eclipses are typically much smaller than those due to eclipsing binaries, while the properties of the primary events themselves should still be consistent with a planetary origin. A TC is allowed to pass the Significant Secondary test if the depth of the secondary is less than 10% of the primary, the impact parameter is less than 0.95, and the planet's radius as derived using the fitted parameter R_p/R_s is $R_p < 30R_\oplus$.

Odd-Even Depth Tests

Secondary eclipses could also be erroneously marked as half of the primary events if the eclipsing binary is detected at half its actual period and its eclipses would otherwise occur at phase 0.5, as is the case for circular orbits. These eclipsing binaries can be identified as candidates with significantly different odd and even transit depths. As with the S/N Consistency Test, the Odd-Even Depth Tests are only used for candidates with $P < 90$ days.

First, an odd-even depth statistic is calculated for each TC:

$$\sigma_{\rm OE1} = \frac{|d_{\rm odd} - d_{\rm even}|}{\sqrt{\sigma^2_{\rm odd} + \sigma^2_{\rm even}}} \tag{3.21}$$

where $d_{\rm odd}$ and $d_{\rm even}$ are the median of all points within 30 minutes of the centre of odd and even transits respectively, and $\sigma_{\rm odd}$ and $\sigma_{\rm even}$ are the standard deviations of those points. For the case of trapezoidal model fits, all points making up the flat part in-transit are also included. A TC fails if

$\sigma_{\mathrm{OE1}} > 1.0$.

A second odd-even depth statistic is also calculated as part of the model-shift uniqueness test. This method takes into account the full transit shape as well as the noise level of the full light curve. However, it is more susceptible to outliers and systematics compared to the first statistic. Thus, I use a lenient requirement of $\sigma_{\mathrm{OE}} - FA_1 < 10$.

V-Shape Test

Candidates where the ingress and egress times are a significant fraction of the total transit duration are most likely FPs. Planetary transits typically have a U-shape, while V-shaped transits are often created by grazing eclipsing binary stars (Fig. 1.4). The V-shape metric is defined as $V = b + R_p/R_s$, in order to identify eclipsing binaries both due to grazing eclipses (large impact parameter, b) and being too deep (large R_p/R_s). A candidate fails with $V > 1.05$.

3.3.5 Manual Inspection

The final round of vetting involves a visual inspection of each of the TCs that passed the automated vetting stage. I look at the full light curve, the light curve phase-folded to the transit's period, a close-up of the transit in the phase diagram, and a side-by-side comparison of odd and even transits. The latter two images include the data averaged into 30-minute bins as well as the model fit to the light curve to assess the fit. While the previous tests are able to remove the majority of FPs and attempt to mimic decisions made by human vetters, manual inspection still serves as an important "reality check" that each passing TC is convincing enough to be promoted to PC status.

A total of 5608 of the 33,322 TCs survived the automated vetting stage. Of those, 3972 passed manual vetting to become PCs. The majority of the TCs were high-S/N events that had obviously asymmetric, non-transit-like shapes, typically due to stellar variability that was not well removed by the detrending algorithm. Low-S/N events were found in flatter light curves,

but usually had inconsistent odd and even shapes that revealed a likely nonplanetary nature.

3.4 Assessing Vetting Performance

Ideally, the vetting pipeline is accurate when classifying planets as planets and false positives as false positives. Realistically, no pipeline is perfect, and sacrifices must be made to achieve balance. For example, lenient candidacy test thresholds will cause more real planets to be accepted, at the cost of more FPs incorrectly passed as planet candidates. This will call the validity of any new planet candidates coming out of the pipeline into question.

Two useful metrics used to assess vetting performance are the completeness (the fraction of true transiting planets passed as PCs) and reliability (the fraction of PCs that are actually planets). These numbers are unknown. However, they can be estimated using simulated data. Injecting fake planet transits into real *Kepler* data and vetting the resulting detections gives an estimate of the vetting completeness. Likewise, I can simulate FPs to estimate how often the vetting process mistakenly labels false positives as planets.

As in KDR25, this work only attempts to measure reliability against noise FPs. These are the largest concerns for low-S/N TCs, among which I expect most of my new planet candidates to lie.

3.4.1 Simulated Data

I injected 120,642 planet transits into the light curves and prepared, searched, and vetted the data using the same process as for the actual observed data. The only exception was that I tested the manual component on a small subset of the injected detections. The overall process is consistent with injection and recovery tests performed for completeness measurements of other independent pipelines in the literature [44, 119]. I injected signals log-uniformly distributed over the ranges $0.5 < P < 500$ days and $0.5 < R_p < 16.0\ R_\oplus$. Each transit was created using a quadratic limb-darkening Mandel & Agol

(2002) [90] model, with impact parameters uniformly distributed between 0 and 1, and circular orbits assumed.

For testing against noise FPs, the simulated data should allow realistic signals with noise properties similar to the real data, while ensuring no possibility of detecting true exoplanets still in the light curve. To achieve this, I took the original 198,640 light curves and inverted them. Essentially, this recreated the Inverted (INV) set of simulations described in Christiansen (2017) [30] for their own vetting tests. Any "transit" would actually be a positive flux increase in the observed data, and thus not a planet. The *Kepler* team also created a Scrambled (SCR) data set for testing against noise FPs, corresponding to reordering of the *Kepler* quarters by yearly chunks. Three orders were created, as described in Coughlin (2017b) [35]. I tested my pipeline on Scrambled Group 1 (SCR1).

3.4.2 Vetting Completeness

Of the 48,610 simulated TCs detected by the search pipeline, 45,676 (94.0%) passed the automated candidacy tests. More usefully, completeness is binned over period and S/N in Fig. 3.1, showing that completeness is highest at high S/N and short orbital period, and lowest at low S/N and long orbital period, as expected.

There is no region of the period-S/N grid where completeness reaches 100%; even at high S/N, planets can be incorrectly failed by the vetting pipeline due to distortion caused by noise in the light curve or the detrending process. There is also a significant drop in completeness at the lowest S/N bin (S/N $= 6 - 10$), reflecting the sacrifice of completeness that must be made to ensure that an overwhelming number of noise FPs are not passed. The $P = 300 - 500$ day bin is particularly susceptible, given that planets with such long orbital periods will have only a handful of transits in the data available to test.

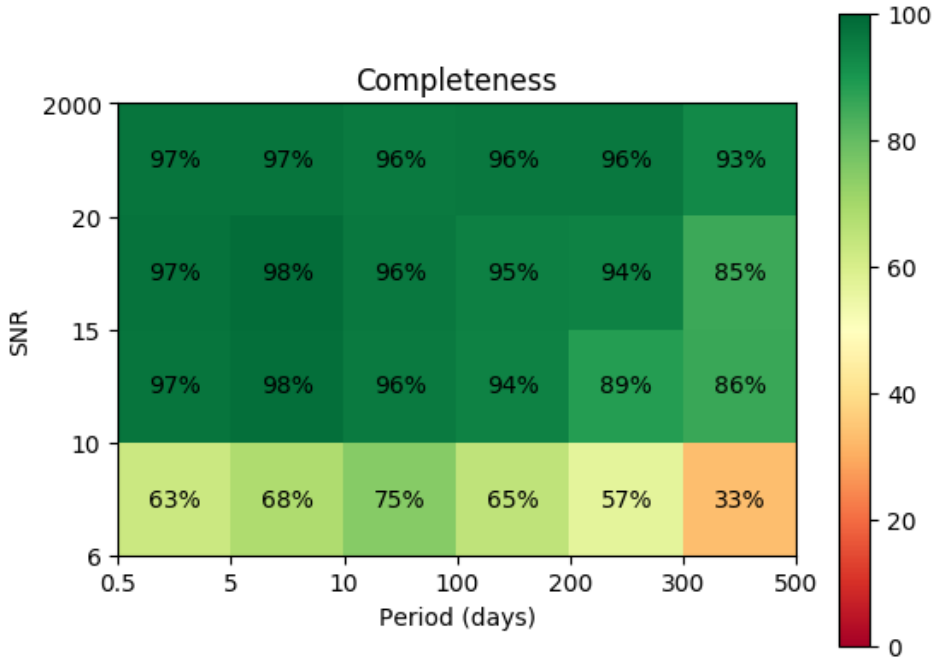

Figure 3.1: Completeness of the vetting pipeline based on running the automated tests on simulated planet TCs.

The vetting process also involves a manual component in the form of a final visual inspection. Performing a full completeness measurement that takes this stage into account is difficult due to the presence of human bias. Additionally, it is infeasible to manually review each of the 45,676 simulated TCs that passed the automated tests. Thus, I chose a random subset of 1000 of the passing TCs to review. In an attempt to remove human bias, I combined these TCs with all passing TCs from the simulated FP set, and removed any labeling that would indicate the origin set of each TC. I failed 15 of the 1000 planet TCs (1.5%) as FPs. Overall, I expect the manual vetting to reduce the overall vetting completeness by 1-2% from the 94% success rate of the automated component.

3.4.3 Vetting Reliability

My pipeline identified 15,283 TCs in the inverted set, and 12,103 in the scrambled set. The automated tests correctly labeled 14,494 (94.8%) and 11,222 (92.7%) of these as FPs, respectively. I then manually reviewed the surviving TCs, combined with the simulated planet TCs as described previously. Overall, I classified all but 8 TCs in the inverse set and 28 TCs in the scrambled set as FPs, giving a total success rate of 99.8%.

The fraction of false positives classified as FPs is also known as the effectiveness of the pipeline. This can be combined with the final vetting results to estimate the reliability. Letting E denote the effectiveness, KDR25 define reliability R as

$$R = 1 - \frac{N_{\mathrm{FP}}}{N_{\mathrm{PC}}}\left(\frac{1-E}{E}\right). \tag{3.22}$$

where N_{PC} and N_{FP} are the numbers of observed PCs and FPs identified by the vetting pipeline, respectively.

Given that I identified 3971 PCs and 29,348 FPs out of all TCs, an effectiveness of 99.8% gives an overall reliability of 98.3%. However, plotting reliability as in Fig. 3.2 reveals areas in period-S/N space where the pipeline is particularly unreliable, namely S/N < 10 and $P > 200$ days. While effectiveness in this regime was 99.6%, I only identified 10 PCs compared to 1214 FPs.

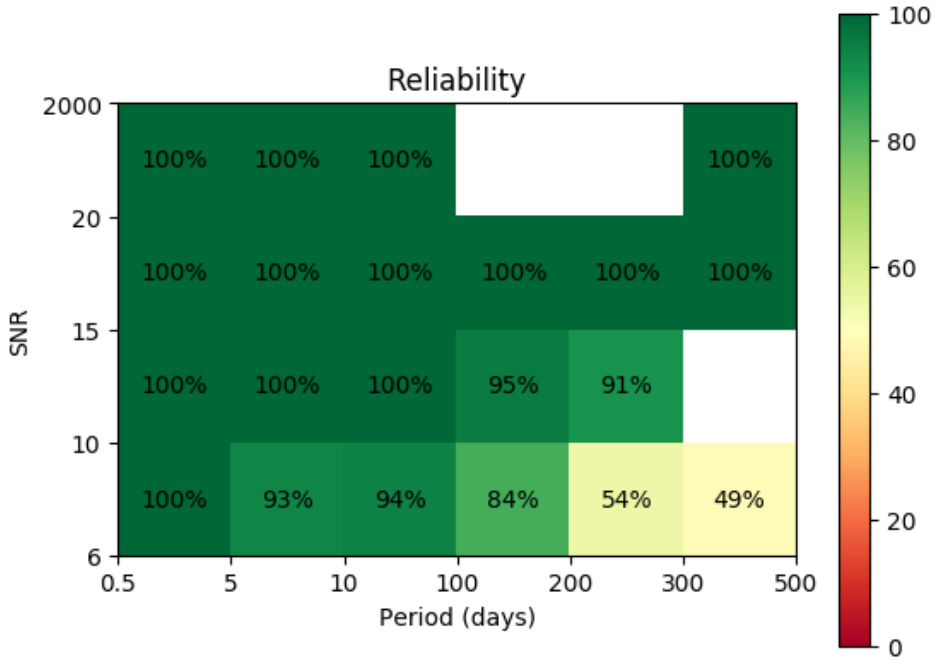

Figure 3.2: Reliability of the vetting pipeline against noise FPs based on running the automated and manual tests on simulated noise TCs (inverted + scrambled). Bins with fewer than three PCs or fewer than 20 simulated noise FPs are not shown.

3.5 Results Compared to *Kepler*

I used the federation process described in Mulally et al. (2015) [108] to match 3915 of the 3972 PCs with known KOIs identified by the *Kepler* team, accumulated over all *Kepler* catalogues. The NASA Exoplanet Archive was accessed on May 9, 2019.

3.5.1 Confirmed Planets

I successfully detected and passed 2268 of the 2295 (98.8%) planets confirmed by *Kepler*, defined as having an Exoplanet Archive Disposition of CONFIRMED and a Disposition Using Kepler Data of CANDIDATE on

the NASA Exoplanet Archive.

Another 20 were marked as TCs, but failed the candidacy tests. Upon manual inspection, it appeared that significant Transit Timing Variations (TTVs) were to blame for the failing of five confirmed planets (KOI-142.01, 227.01, 377.01, 377.02, 884.02). The other 15 planets were either very close to passing or only failed a single test (KOI-46.02, 172.02, 701.04, 1236.03, 1574.02, 2038.03, 2298.02, 2365.02, 2533.01, 3458.01, 4034.01, 4384.01, 5416.01, 5706.01, 7016.01). I found that these planets often had much lower calculated S/N than what was listed on the NASA Exoplanet Archive (for example, KOI-4384.01 had an S/N of only 6.3 according to my pipeline, but 12.2 from *Kepler*). It is likely that my use of a much simpler detrending algorithm in the search pipeline compared to the *Kepler* team's can explain these discrepancies, rather than these signals being intrinsically poor candidates. In particular, the fact that many of the TCs outputted by the search were due to incomplete removal of stellar variability would indicate not only that the pipeline is susceptible to such FPs, but also that the S/N and strength of candidacy of even valid planets could be under-estimated due to contamination from excess variability in the light curve.

Six were detected but failed to meet the initial requirements to be a TC (KOI-179.02, 245.03, 490.02, 1274.01, 1718.02, 3234.01). KOI-179.02, KOI-490.02, and KOI-1274.01 had only one or two detected transits, lower than the required three. KOI-1718.02 had a barely failing RS (5.8), KOI-3234.01 had too low of a CHI value (4.6), and KOI-245.03 failed both. Only a single confirmed planet, KOI-4846.01, was missed entirely.

3.5.2 Candidate Planets

I successfully detected and passed 1447 of the 2421 (59.8%) known *Kepler* planet candidates, defined as having both dispositions listed as CANDIDATE.

The lower rate of recovery among planet candidates is expected, and can be explained from two perspectives. First, candidates typically have lower S/N and transit shapes less clearly consistent with a planetary origin

compared to confirmed planets, resulting in a lower probability of being recovered. This is especially true near the detection limit where excess stellar variability makes identifying low-S/N signals challenging.

A second explanation is that many of the missed or failed candidates may not even be planets. For instance, 576 (around 60%) of the 974 candidates missed or failed by my pipeline were not detected by *Kepler*'s DR25 pipeline. The fact that both of our pipelines would fail to identify these signals as planet candidates would indicate that their current candidacy should be placed under additional scrutiny.

3.5.3 False Positives

193 of the PCs have both dispositions listed as FALSE POSITIVE. 109 of these were flagged by *Kepler* as FPs solely due to having a significant centroid offset, while another 71 had an ephemeris match indicating contamination. Given that I did not incorporate centroid tests or ephemeris matching between KOIs into my vetting pipeline, it is unsurprising that I would pass these as candidates. However, I address both of these issues for my new candidates in §3.6.

Six of the PCs have a Disposition Using Kepler Data of FALSE POSITIVE, but an Exoplanet Archive Disposition of CONFIRMED (KOI-125.01, 129.01, 631.01, 1416.01, 1450.01, 3032.01). Furthermore, one of the PCs is the sole KOI on the NASA Exoplanet Archive with a Disposition Using Kepler Data of CANDIDATE, but an Exoplanet Archive Disposition of FALSE POSITIVE (KOI-242.01).

3.6 New Planet Candidates

After removing all federated *Kepler* confirmed planets, candidate planets, and false positives from my PC list, I was left with 57 new planet candidates. All of my new PCs have low S/N, ranging from S/N = 7.1 to 10.7. These low S/N are unsurprising, given they are the kind of candidate most susceptible to being missed by detection pipelines.

I performed additional follow-up analysis on each of the candidates to more rigorously assess their candidacy. This involved ephemeris matching, centroid analysis, adaptive optics imaging follow-up (in select cases), and false positive probability calculation. I also performed a Markov Chain Monte Carlo (MCMC) refit to each transit, taking into account dilution effects of companions detected in the AO imaging. These fits produced the final reported planet parameters. At the end of this chapter, a table summarizing all results is given in Table 3.7, and plots of the phase diagrams of each transit with model fits and residuals are shown in Fig. 3.9. I list candidates according to their Kepler Input Catalogue (KIC) number [16].

As discussed in §3.4, I found that my pipeline has significantly lower reliability for S/N < 10 and $P > 200$ days than other regimes (54% for $200 < P < 300$ days, and 49% for $300 < P < 500$ days). Recall that the reliability is an estimate of the fraction of PCs that are planets. Since I found ten PCs across these regimes, my ∼50% reliability would indicate that only five of these are likely planets. Given that five these PCs are known KOIs and the other five were contributed by my pipeline, I made the conservative decision to downgrade the new candidates with these properties to FP status. I continue the analysis with the remaining 52 PCs.

3.6.1 Ephemeris Matching

Light that contributes to the target's light curve may not necessarily originate from the target. If this contamination is caused by a star with a variable signal, then the same signal will be observed in the target with reduced amplitude due to dilution. Thus, if two signals have the same ephemeris, then at least one of them is an FP due to contamination.

I compared the periods and epochs of each new PC to all KOIs, searching for cases where

$$|P - P_{\rm match}| \leq \min(2 \text{ hours}, 0.001P) \tag{3.23}$$

and

$$|T_0 - T_{0,\rm match}| \leq \min(4 \text{ hours}, 0.001P) \tag{3.24}$$

as in Dressing & Charbonneau (2015) [44], and found no matches.

3.6.2 Stellar Variability

False positives may also be due to stellar variability that was not fully removed during the detrending process. For example, failing to remove the rotation signal of the star can create a periodic, transit-like signal in the light curve. Finding a match between the orbital period of the PC and the rotation period of the host star could indicate an FP due to stellar variability.

I ran each undetrended light curve through the Lomb-Scargle periodogram in `Astropy` [3, 4] and searched for cases where the rotation period (or a multiple thereof) matched the detected orbital period of the PC. I determined rotation periods from the period corresponding to the highest peak in the periodogram. 14 of the host stars also had rotation periods listed in McQuillan et al. (2014) [98]. I also manually inspected each periodogram to search for smaller peaks or excess noise at the orbital periods, which could confound the search for planets. For periods that corresponded to a peak, I determined its false alarm probability (the probability of measuring a given peak height under the assumption that the noise is Gaussian with no periodic component), and flagged cases where the power had a probability less than 0.05. Following this analysis, I identified 30 of the 52 PCs as likely FPs due to stellar variability.

I also investigated whether or not the transits would change or disappear depending on different stellar variability removal methods. I used the `biweight` time-windowed slider implemented in the W$\bar{o}$tan Python package,[15] which was identified by Hippke et al. (2019) [59] as the ideal method for recovering transits from light curve data based on a comprehensive comparison of common detrending routines. Using window lengths of 0.5, 1, and 2 days, I detrended each light curve, examined the data phased at the planet period, and calculated the S/N by measuring the depth of the transit and assuming the same duration, period, and epoch as the PC. The only

[15] https://github.com/hippke/wotan

exception was that I did not use a 0.5-day width for transits with durations greater than 0.2 days, so as to avoid significantly distorting the transit itself. Four of the PCs had either S/N < 6 (KIC-6937870 b) or the transit itself was inconsistent in shape and duration with the original PC (KIC-2985262 b, KIC-6380164 b, KIC-10419787 b) using one of these alternate detrends. Then, I redetrended the remaining 18 light curves after masking out the transits, reexamined the phase diagram, recalculated the S/N, and fit a least-squares transit model. Regardless of window length used, the S/N remained above 6 for all 18 PCs, and the model best-fit parameters were within 1σ of the results using the original detrending algorithm, giving further confidence that these signals were not an artifact of stellar variability.

3.6.3 Centroid Analysis

I used the difference imaging method described in Bryson et al. (2013) [18] to identify background FPs for the remaining 18 PCs, which is summarized here. I downloaded all necessary target pixel files from the MAST. For each quarter, I combined all in-transit cadences to produce an average in-transit pixel image. I took an equal number of cadences on either side of the transit to produce an average out-of-transit pixel image. Subtracting the in-transit from the out-of-transit image gives the difference image. I fit the *Kepler* Pixel Response Function (PRF) to each of the out-of-transit and difference images. The PRF is defined as the composite of *Kepler*'s optical point spread function, integrated spacecraft pointing jitter during a nominal cadence, and other systematic effects, and is represented as a piece-wise continuous polynomial on a sub-pixel mesh [17]. I used `PyKE` [140], which provides fitting of the *Kepler* PRF as a function of flux, centre positions, width, and rotation angle to a given target pixel file. Respectively, the centre positions of the out-of-transit and difference images give the location of the target star and transit source, providing a direct measurement of the centroid offset for that quarter.

Bryson et al. (2013) [18] discuss that the difference images for low-S/N

transits are typically noise-dominated. The difference image can appear significantly different than the out-of-transit image in one quarter, and may show the transit at other locations or on the target star in others. Thus, I attain a more reliable estimate of the centroid offset and its uncertainty by averaging all quarterly offsets. I also use the bootstrapping technique described in Bryson et al. (2013) [18] to estimate the uncertainty in the result, taking the larger of the two values. Given the Q measured offsets (where Q is the number of quarters analyzed), I produce Q^2 different sets, randomly selecting from the list of offsets to fill each set. I then find the average of each set. The standard deviation of the Q^2 averages provides the bootstrap uncertainty estimate.

Following Bryson et al. (2013) [18], I classify candidates as FPs if they have a 3σ significant offset larger than $2''$, or 4σ offset larger than $1''$. One of the PCs (KIC-3336146 b) met these thresholds and was reclassified as an FP. I completed the rest of the analysis with the remaining 17 new PCs.

3.6.4 AO Observations

I obtained adaptive optics follow-up imaging for six of the host stars, prioritizing the potentially rocky candidates. As discussed in §1.2.2, the uses of AO data are twofold: first, nearby stars dilute the observed transit depth, resulting in an underestimated planet radius. This is especially of concern for small, rocky planets due to their relative rarity, and those just under the proposed $1.6R_\oplus$ "rocky limit", past which most planets are not rocky [126]. Second, a contaminant star could be the source of an FP signal, whether as a background or foreground eclipsing binary. Contrast curves derived from the AO images serve as effective constraints for unseen companions in later false positive probability calculations. I found that the AO images reduced false positive probabilities by a factor of $\sim$16 on average (see §3.6.5), emphasizing the usefulness of AO follow-up for planet validation.

Observations of three stars were collected on the Gemini North 8.1 m telescope in the K_s band with the Natural Guide Star (NGS) adaptive optics assisted Near InfraRed Imager and spectrograph (NIRI, Hodapp et al.

61). Data were taken between 2018 July and 2019 June (Program ID GN-2018B-Q-134). Another threestars were observed with the Laser Guide Star (LGS) AO system and NIRI in 2019 July as part of a Fast Turnaround program (Program ID GN-2019A-FT-213). Total exposure time for each target was between 5 and 6 minutes. Observations used the $f/32$ NIRI camera, providing a plate scale of $0.022''$/px and a $22'' \times 22''$ field of view.

Data were reduced by median-stacking each dark-subtracted and flat-divided image into a single AO image per star. I manually inspected each image for artifacts and potential companions in order to mask them out before computing 5σ contrast curves. To calculate each curve, I used the procedure outlined in Ngo et al. (2015) [110], computing the standard deviation of flux values in a series of annuli with widths equal to twice the FWHM of the central star's point spread function. Fig. 3.3 shows all 5σ contrast curves along with the median to indicate our typical sensitivity. I provide a subset of the data for the contrast curves in Table 3.1, out to a maximum separation of $\sim 8''$ (typical). I chose the maximum separation on a per-target basis based on the limit at which separations were no longer covered by all median-stacked images, due to the dither pattern used to take the observations.

Table 3.1: Contrast curve data for the six targets observed with Gemini NGS-AO and LGS-AO in the K_s band. Only a portion of this table is shown here. The full dataset is available in Kunimoto et al. (2020) [82].

KIC	Guide Star System	UT Obs. Date	Sep. ($''$)	ΔK_s
6126245	NGS-AO	31 May 2019	0.20	0.76265
			0.35	3.80569
			0.51	4.62049
			0.66	5.22778
			...	...
6224562	LGS-AO	30 June 2019	0.20	0.69001
			0.35	3.25464
			0.51	4.47640
			0.66	5.49310
			...	...

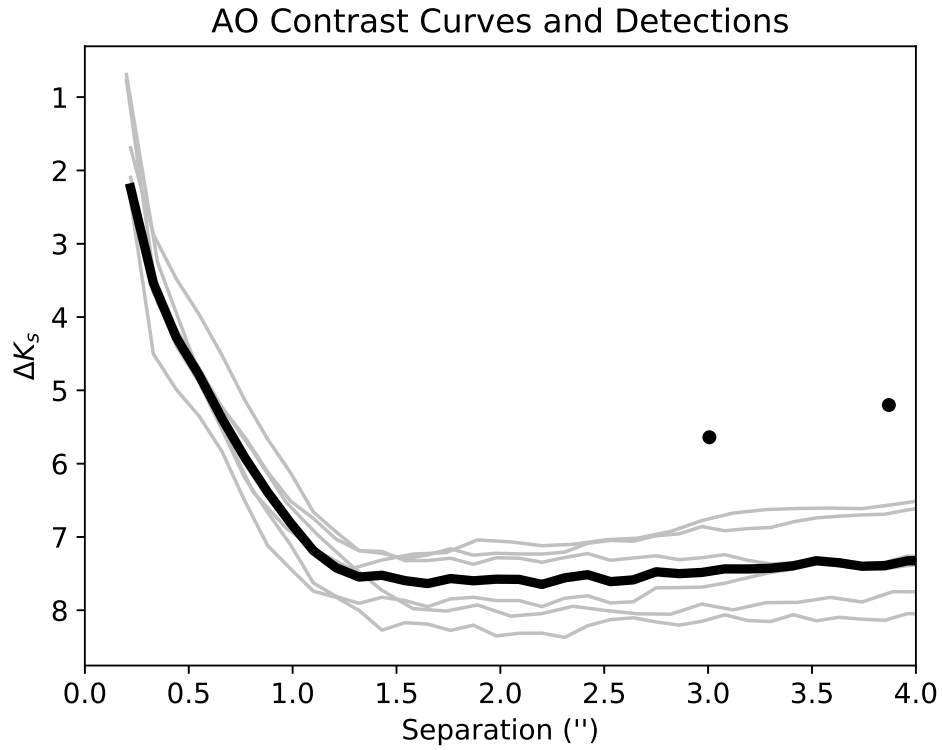

Figure 3.3: 5σ contrast curves (grey curves) for the Gemini NGS-AO and LGS-AO-observed targets. The black curve indicates the median. The black points indicate the best-fit locations of detected companions, determined from the AO images.

I manually examined each image for contaminant stars within $4''$, the size of a *Kepler* pixel. I fit a two-Gaussian model to the two targets with detected companions in order to derive the angular separation, position angle (PA), and ΔK_s. Results are shown in Table 3.2, and the AO images of targets with companions are shown in Fig. 3.4. One of the targets, KIC-7340288, has a potential companion just outside of $4''$ (at $4.2''$), that is thus excluded from the analysis but indicated in the plots by dotted circles.

One of my new PCs (KIC-11350118 c) corresponds to a known KOI already observed in the LP600 band as part of the Robo-AO KOI surveys [84]. Robo-AO did not detect any nearby stars.

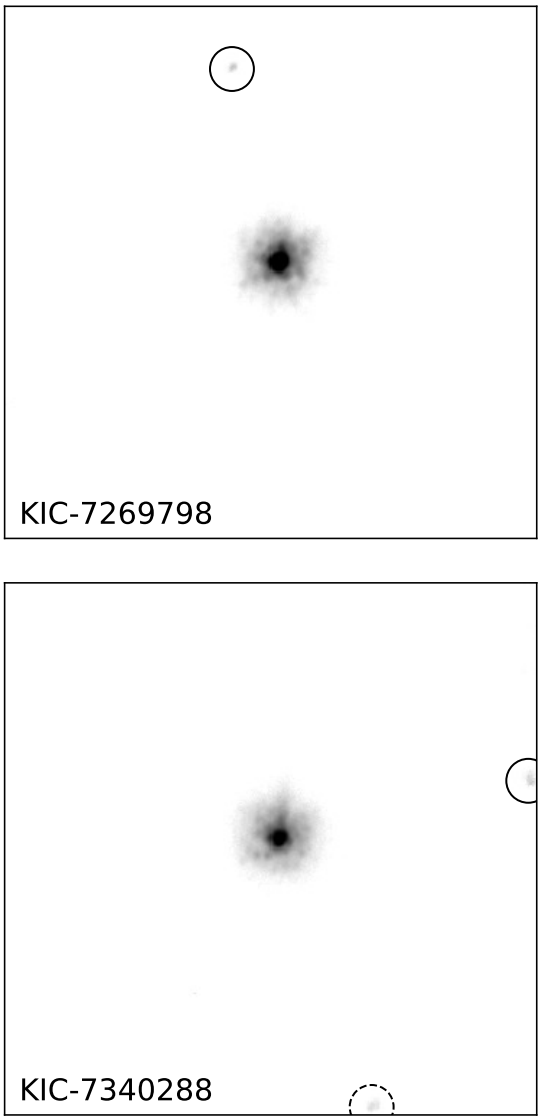

Figure 3.4: $4'' \times 4''$ AO images plotted in logscale, centred on each target with resolved companions within $4''$ indicated by a black circle. KIC-7340288 has a potential second companion just outside of the $4''$ threshold, indicated by a dotted black circle. For each image, north points up and east points left.

Table 3.2: Gemini NIRI and Robo-AO imaging searches for companions within 4″ of our target stars. Zie17 refers to Ziegler et al. (2017) [163].

KIC	Telescope	Filter	UT Obs. Date	Comp?	Sep. (″)	PA (°)	Δm	Ref.
6126245	Gemini NGS-AO	K_s	31 May 2019	N	-	-	-	This work
6224562	Gemini LGS-AO	K_s	30 June 2019	N	-	-	-	This work
6782399	Gemini NGS-AO	K_s	14 June 2019	N	-	-	-	This work
7269798	Gemini LGS-AO	K_s	30 June 2019	Y	3.006±0.002	13.034 ± 0.001	5.64±0.04	This work
7340288	Gemini LGS-AO	K_s	01 July 2019	Y	3.870±0.002	283.380±0.001	5.20±0.03	This work
7747788	Gemini NGS-AO	K_s	11 June 2019	N	-	-	-	This work
11350118	Robo-AO	LP600	01 Sept 2014	N	-	-	-	Zie17

I obtained observations for an additional 56 targets across both programs. However, during the execution of the observation program, the vetting pipeline described in §3.3 and follow-up analysis described earlier in this section were each modified and these targets are no longer planet candidates. I provide results for these targets in Tables A.1 and A.2 of Appendix A. Of note, 32 of the targets are KOIs, 12 of which do not have Robo-AO observations. These observations can be used in future follow-up analysis of all the confirmed and candidate planets associated with these KOIs.

3.6.5 Astrophysical False Positive Probabilities

I tested each of my candidates against astrophysical false positive hypotheses using `vespa`, a Python package built to enable astrophysical false positive probability analysis of transiting signals [99, 101].

`vespa` uses stellar posteriors calculated with `isochrones`, a Python package that provides MCMC fitting of single-, binary-, and triple-star model stellar properties to MIST stellar model grids [100], as an input. I provided `isochrones` with RA/Dec coordinates and $grizJHK$ photometry from the KIC, with $griz$ bands corrected to the Sloan Digital Sky Survey (SDSS) according to Pinsonneault et al. (2012) [122]. $T_{\rm eff}$, log$g$, and [Fe/H] from Mathur et al. (2017) [94] were used if the provenance of these values is from spectroscopy or asteroseismology. I provided parallaxes from *Gaia* DR2 [54, 55] when available.

As constraints, I followed the convention of Morton et al. (2016) [102] by setting the allowed exclusion radius for a blend scenario as 3 times the uncertainty in the fitted centroid position, floored at $0.5''$. I also set the maximum secondary eclipse depth allowed by the *Kepler* photometry as

$$\delta_{\rm max} = \delta_{\rm sec} + 3\sigma_{\rm sec}, \tag{3.25}$$

where $\delta_{\rm sec}$ and $\sigma_{\rm sec}$ are the fitted depth and uncertainty of the secondary event in the light curve, as calculated by the model-shift uniqueness test in the vetting pipeline. Lastly, I inputted contrast curves derived from the AO imaging when available. In `vespa`, these eliminate the possibility of bound

or background stars above a certain brightness at a given projected distance.

Considering all of these inputs, `vespa` assigns probabilities to different hypotheses that might describe a transiting planet candidate signal: unblended eclipsing binary, hierarchical-triple eclipsing binary, chance-aligned background/foreground eclipsing binary, and transiting planet. Following the convention of Morton et al. (2016) [102], I consider candidates with total astrophysical false positive probabilities FPP > 0.9 as FPs. All other candidates, including those that have `vespa` fail to return a false positive probability (typically due to a nonconverging MCMC fit), remain planet candidates. None of the 17 remaining PCs were classified as FPs as a result of the `vespa` results.

12 of the candidates have FPP < 0.01 (confidence at the 99% level). `vespa` has been used to validate over a thousand KOIs as confirmed planets using this threshold [102]. However, given that all of these PCs have low signal-to-noise ratios (S/N < 10), a noise or systematic explanation for the signals cannot be ignored. Burke et al. (2019) [23] indicated that statistical validation methods based only on astrophysical scenarios, such as `vespa`, are insufficient for such a low-S/N regime. Thus, I chose to retain their candidate disposition. A more comprehensive study of my pipeline's reliability against noise (to more careful detail than the coarse period-S/N grid from §3.4.3), or the detection of these candidates with a different detection method, will be required before I can more confidently confirm these candidates as bonafide planets.

3.6.6 MCMC Fit

I refit each transit using `emcee`, a Python implementation of an affine invariant MCMC ensemble sampler [48], with walkers initialized in a tight Gaussian ball near the best-fit parameters from the least-squares fit discussed in §3.3.2. I set P and T_0 fixed to their least-squares values to aid in convergence. Fit results are shown in Table 3.3. These fit results are mixed with `isochrones` stellar parameter posteriors, given in Table 3.4, to produce the derived planet parameters shown in Table 3.7. The reported

values use the median, with uncertainties given by 15.9% and 84.1% percentiles, corresponding to a 68.3% confidence region.

Derived Parameters from MCMC

The planet radius R_p is determined from the fitted parameter R_p/R_s using

$$R_p = \left(\frac{R_p}{R_s}\right) R_s, \tag{3.26}$$

where R_s is the known stellar radius.

The semi-major axis of the planet's orbit a is determined from the fitted P and known stellar parameters, rather than the fitted a/R_s,

$$a = \left(\frac{GM_s P^2}{4\pi^2}\right)^{1/3}, \tag{3.27}$$

where G is the gravitational constant and M_s is the stellar mass.

The planet equilibrium temperature $T_{\rm eq}$ is calculated assuming thermodynamic equilibrium between the incident stellar flux and the radiated heat from the planet,

$$T_{\rm eq} = T_{\rm eff}(1-A)^{1/4}\sqrt{\frac{R_s}{2a}}, \tag{3.28}$$

where A is the albedo of the planet. As for the *Kepler* planets listed on the NASA Exoplanet Archive, I assumed Earth's albedo, $A = 0.3$, for all cases.

The planet's stellar insolation S, defined as the stellar flux received by the planet, is determined relative to solar flux received by the Earth $S_\oplus$ by

$$\frac{S}{S_\oplus} = \left(\frac{R_s}{R_\odot}\right)^2 \left(\frac{a_\oplus}{a}\right)^2 \left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right)^4. \tag{3.29}$$

Table 3.3: MCMC fit results for select fitted planet parameters (R_p/R_s, a/R_s, and b). P and T_0 were set to their least-squares best-fit values. Note that BKJD indicates Kepler Barycentric Julian Date (BJD - 2,454,833.0). The number of significant figures in each column has been chosen to match the convention used by the NASA Exoplanet Archive.

KIC	P (days)	T_0 (BKJD)	R_p/R_s	a/R_s	b
1570311 b	$23.44253108 \pm 0.00046135$	148.98683 ± 0.01301	$0.017478^{+0.005396}_{-0.002075}$	$8.487^{+4.907}_{-2.065}$	$0.977^{+0.016}_{-0.036}$
2696784 b	$82.30223397 \pm 0.00196956$	130.31551 ± 0.01966	$0.009613^{+0.000652}_{-0.000561}$	$22.104^{+5.043}_{-6.402}$	$0.923^{+0.038}_{-0.042}$
2861140 b	$36.87848594 \pm 0.00033789$	364.07192 ± 0.00612	$0.016318^{+0.00173}_{-0.001597}$	$63.344^{+10.242}_{-19.783}$	$0.43^{+0.351}_{-0.296}$
2985262 b	$13.0351506 \pm 5.443 \times 10^{-5}$	140.06886 ± 0.00392	$0.009084^{+0.000517}_{-0.000449}$	$29.901^{+2.269}_{-6.193}$	$0.383^{+0.3}_{-0.266}$
3336146 b	$3.27626622 \pm 1.652 \times 10^{-5}$	134.62799 ± 0.00298	$0.007098^{+0.000596}_{-0.000508}$	$14.447^{+1.248}_{-3.703}$	$0.41^{+0.33}_{-0.279}$
3345775 b	$6.22112577 \pm 2.924 \times 10^{-5}$	122.33384 ± 0.00401	$0.004294^{+0.000232}_{-0.000189}$	$13.329^{+1.97}_{-2.986}$	$0.847^{+0.064}_{-0.055}$
3347135 b	$226.52674578 \pm 0.00125028$	160.17927 ± 0.0051	$0.017495^{+0.000529}_{-0.000459}$	$242.158^{+8.295}_{-23.965}$	$0.271^{+0.229}_{-0.185}$
3662290 b	$288.23951462 \pm 0.01314029$	322.24425 ± 0.04069	$0.011739^{+0.000882}_{-0.000658}$	$101.68^{+17.014}_{-29.933}$	$0.815^{+0.101}_{-0.081}$
3728762 b	$6.73928932 \pm 6.924 \times 10^{-5}$	122.20544 ± 0.00867	$0.007275^{+0.000553}_{-0.000501}$	$5.557^{+1.023}_{-1.476}$	$0.914^{+0.042}_{-0.04}$
3967744 b	$57.88535219 \pm 0.00040897$	143.26298 ± 0.0065	$0.011696^{+0.000888}_{-0.000693}$	$47.663^{+6.252}_{-15.352}$	$0.439^{+0.358}_{-0.304}$
4346258 b	$4.90776291 \pm 1.936 \times 10^{-5}$	354.03734 ± 0.00291	$0.013645^{+0.001287}_{-0.00114}$	$19.513^{+1.919}_{-5.171}$	$0.418^{+0.329}_{-0.291}$
4551429 b	$35.37625461 \pm 0.00023684$	147.08621 ± 0.00434	$0.012458^{+0.00096}_{-0.000774}$	$75.028^{+6.129}_{-17.846}$	$0.386^{+0.325}_{-0.262}$
4556565 b	$5.54559321 \pm 3.186 \times 10^{-5}$	353.80389 ± 0.004	$0.011497^{+0.001035}_{-0.000861}$	$16.36^{+1.875}_{-4.241}$	$0.427^{+0.32}_{-0.289}$
5095499 b	$4.29501271 \pm 1.861 \times 10^{-5}$	134.0872 ± 0.00411	$0.008775^{+0.000575}_{-0.000546}$	$9.371^{+0.698}_{-1.944}$	$0.379^{+0.305}_{-0.274}$

KIC	P (days)	T_0 (BKJD)	R_p/R_s	a/R_s	b
5184017 b	$6.25985244 \pm 2.664 \times 10^{-5}$	135.91422 ± 0.00367	$0.006305^{+0.000557}_{-0.000386}$	$11.235^{+1.007}_{-3.042}$	$0.409^{+0.345}_{-0.279}$
5342061 c	$11.49044213 \pm 7.721 \times 10^{-5}$	137.65928 ± 0.00558	$0.013311^{+0.001104}_{-0.000975}$	$23.445^{+2.578}_{-5.977}$	$0.413^{+0.327}_{-0.29}$
5628770 b	$11.42952942 \pm 6.504 \times 10^{-5}$	132.35509 ± 0.00511	$0.008349^{+0.000662}_{-0.000623}$	$39.42^{+3.527}_{-8.723}$	$0.4^{+0.303}_{-0.278}$
5649129 b	$2.82857392 \pm 1.138 \times 10^{-5}$	133.09613 ± 0.00321	$0.007052^{+0.000701}_{-0.000446}$	$8.999^{+1.073}_{-2.605}$	$0.469^{+0.314}_{-0.324}$
5794479 b	$5.92912541 \pm 3.014 \times 10^{-5}$	125.34438 ± 0.00419	$0.007263^{+0.000445}_{-0.000324}$	$14.551^{+1.261}_{-2.751}$	$0.412^{+0.263}_{-0.277}$
5893807 b	$7.66465733 \pm 5.173 \times 10^{-5}$	133.13869 ± 0.00444	$0.009196^{+0.000885}_{-0.000853}$	$18.378^{+1.785}_{-4.85}$	$0.41^{+0.338}_{-0.287}$
6021193 e	$26.48588826 \pm 0.0001843$	126.89598 ± 0.00597	$0.009068^{+0.000736}_{-0.000482}$	$27.39^{+3.054}_{-7.436}$	$0.401^{+0.351}_{-0.275}$
6126245 b	$3.48546885 \pm 1.049 \times 10^{-5}$	134.83373 ± 0.0022	$0.004019^{+0.000346}_{-0.000334}$	$11.424^{+1.283}_{-2.809}$	$0.4^{+0.332}_{-0.276}$
6139884 b	$4.80084532 \pm 2.32 \times 10^{-5}$	122.04373 ± 0.00401	$0.005274^{+0.000447}_{-0.000372}$	$10.93^{+1.501}_{-2.984}$	$0.611^{+0.21}_{-0.184}$
6224562 b	$2.32907482 \pm 4.46 \times 10^{-6}$	133.31436 ± 0.00164	$0.012356^{+0.001302}_{-0.000948}$	$18.11^{+1.936}_{-4.558}$	$0.42^{+0.316}_{-0.287}$
6347299 b	$38.64138416 \pm 0.00025459$	148.42833 ± 0.006	$0.009747^{+0.000678}_{-0.000596}$	$43.08^{+3.242}_{-9.797}$	$0.374^{+0.326}_{-0.26}$
6380164 d	$167.78839179 \pm 0.00333756$	206.04216 ± 0.01562	$0.014345^{+0.000522}_{-0.000478}$	$189.128^{+9.523}_{-26.707}$	$0.326^{+0.26}_{-0.225}$
6440915 b	$365.41156475 \pm 0.01311087$	317.94313 ± 0.01688	$0.023918^{+0.00191}_{-0.001569}$	$105.273^{+35.13}_{-30.741}$	$0.829^{+0.087}_{-0.163}$
6782399 b	$34.20150223 \pm 0.00018815$	134.63965 ± 0.00479	$0.008004^{+0.000517}_{-0.000335}$	$43.655^{+5.013}_{-11.713}$	$0.387^{+0.359}_{-0.274}$
6837899 b	$8.99702134 \pm 5.882 \times 10^{-5}$	137.73126 ± 0.00441	$0.010686^{+0.00098}_{-0.000889}$	$22.957^{+2.935}_{-5.947}$	$0.428^{+0.32}_{-0.293}$
6888194 b	$46.04031439 \pm 0.00077561$	163.79766 ± 0.01181	$0.009233^{+0.000687}_{-0.000632}$	$96.009^{+8.866}_{-21.171}$	$0.382^{+0.317}_{-0.267}$
6929071 b	$61.85364904 \pm 0.00031287$	183.29319 ± 0.00792	$0.012057^{+0.000992}_{-0.000935}$	$126.476^{+13.322}_{-33.009}$	$0.433^{+0.315}_{-0.295}$
6937870 b	$27.46007046 \pm 0.00027629$	140.54046 ± 0.00769	$0.010037^{+0.001033}_{-0.000705}$	$45.134^{+4.254}_{-12.159}$	$0.423^{+0.33}_{-0.299}$
7020834 b	$369.4781786 \pm 0.01213654$	187.4524 ± 0.01568	$0.018399^{+0.003931}_{-0.00126}$	$43.816^{+4.82}_{-4.313}$	$0.985^{+0.009}_{-0.006}$
7119412 b	$10.54126725 \pm 0.00012629$	136.69996 ± 0.01003	$0.011053^{+0.001328}_{-0.00088}$	$8.323^{+1.625}_{-2.616}$	$0.919^{+0.047}_{-0.04}$
7186892 b	$17.23935628 \pm 7.05 \times 10^{-5}$	131.71803 ± 0.00421	$0.006875^{+0.000666}_{-0.000397}$	$38.689^{+5.189}_{-9.623}$	$0.416^{+0.329}_{-0.284}$

KIC	P (days)	T_0 (BKJD)	R_p/R_s	a/R_s	b
7187389 b	$23.77032399 \pm 0.00027154$	359.57155 ± 0.00879	$0.01145^{+0.000984}_{-0.00086}$	$27.557^{+2.576}_{-6.466}$	$0.411^{+0.308}_{-0.272}$
7269798 b	$21.44308742 \pm 0.00011838$	152.55136 ± 0.00472	$0.014886^{+0.00144}_{-0.001118}$	$59.932^{+7.183}_{-16.839}$	$0.44^{+0.326}_{-0.296}$
7340288 b	$142.53244069 \pm 0.00335958$	204.71041 ± 0.01799	$0.025258^{+0.00201}_{-0.001766}$	$156.563^{+12.21}_{-32.38}$	$0.369^{+0.311}_{-0.255}$
7747788 b	$133.09439782 \pm 0.00221722$	215.34785 ± 0.01357	$0.00893^{+0.000627}_{-0.000531}$	$161.086^{+15.814}_{-42.123}$	$0.409^{+0.328}_{-0.286}$
7974496 b	$3.96943045 \pm 2.652 \times 10^{-5}$	133.95697 ± 0.0056	$0.008433^{+0.000843}_{-0.000747}$	$8.43^{+1.201}_{-2.343}$	$0.632^{+0.203}_{-0.188}$
8172679 b	$194.05437841 \pm 0.00399057$	197.40431 ± 0.00878	$0.017765^{+0.000949}_{-0.000449}$	$182.339^{+11.466}_{-32.662}$	$0.359^{+0.287}_{-0.244}$
9274173 b	$4.43040627 \pm 1.401 \times 10^{-5}$	134.02566 ± 0.00261	$0.011594^{+0.000968}_{-0.000835}$	$17.616^{+3.045}_{-4.748}$	$0.804^{+0.097}_{-0.089}$
9716483 b	$209.40859648 \pm 0.00225065$	166.46013 ± 0.00829	$0.012104^{+0.000696}_{-0.000497}$	$128.227^{+10.877}_{-32.405}$	$0.4^{+0.33}_{-0.281}$
9777962 b	$367.20909928 \pm 0.00969414$	359.4191 ± 0.02022	$0.02824^{+0.004365}_{-0.00282}$	$80.972^{+40.269}_{-19.615}$	$0.949^{+0.027}_{-0.08}$
10018357 b	$133.78748404 \pm 0.00230495$	253.99492 ± 0.01191	$0.01555^{+0.001482}_{-0.000618}$	$97.298^{+12.202}_{-31.568}$	$0.481^{+0.328}_{-0.341}$
10083396 b	$113.46453674 \pm 0.0020158$	210.81344 ± 0.01041	$0.006687^{+0.000366}_{-0.000304}$	$70.969^{+4.505}_{-13.633}$	$0.347^{+0.309}_{-0.243}$
10419787 b	$122.71394705 \pm 0.00096645$	208.46146 ± 0.00552	$0.015731^{+0.001187}_{-0.001016}$	$140.29^{+12.715}_{-34.704}$	$0.413^{+0.319}_{-0.283}$
10598829 b	$67.52966257 \pm 0.00045439$	191.28971 ± 0.00578	$0.011558^{+0.000907}_{-0.000819}$	$76.452^{+6.627}_{-18.646}$	$0.396^{+0.329}_{-0.275}$
10879314 b	$49.19380825 \pm 0.00037771$	164.61498 ± 0.0069	$0.014659^{+0.001378}_{-0.001153}$	$70.276^{+9.683}_{-18.504}$	$0.418^{+0.338}_{-0.289}$
11092463 b	$6.87343628 \pm 8.866 \times 10^{-5}$	135.83495 ± 0.01066	$0.013895^{+0.002293}_{-0.001421}$	$7.309^{+1.561}_{-2.8}$	$0.923^{+0.054}_{-0.043}$
11139863 b	$7.22517263 \pm 3.616 \times 10^{-5}$	120.75217 ± 0.00397	$0.004123^{+0.00029}_{-0.000177}$	$19.012^{+3.395}_{-5.91}$	$0.558^{+0.264}_{-0.356}$
11350118 c	$2.65550668 \pm 1.532 \times 10^{-5}$	133.27041 ± 0.00483	$0.009019^{+0.000877}_{-0.0007}$	$6.396^{+0.612}_{-1.649}$	$0.419^{+0.33}_{-0.293}$
11565976 b	$24.24399123 \pm 0.00016113$	143.36071 ± 0.0042	$0.006635^{+0.000506}_{-0.00048}$	$52.545^{+4.43}_{-12.444}$	$0.397^{+0.318}_{-0.272}$
11805835 b	$23.52676998 \pm 0.00024171$	142.63053 ± 0.00886	$0.013617^{+0.001429}_{-0.001029}$	$46.944^{+5.132}_{-12.626}$	$0.427^{+0.33}_{-0.296}$
12023559 b	$84.55709677 \pm 0.00127186$	206.60834 ± 0.01139	$0.016528^{+0.001056}_{-0.000877}$	$77.011^{+6.172}_{-16.798}$	$0.39^{+0.305}_{-0.271}$
12216301 b	$116.53116276 \pm 0.00220776$	160.14551 ± 0.01344	$0.012935^{+0.000896}_{-0.000673}$	$83.959^{+11.893}_{-23.552}$	$0.676^{+0.173}_{-0.148}$

KIC	P (days)	T_0 (BKJD)	R_p/R_s	a/R_s	b
12505309 b	$2.89755848 \pm 3.03 \times 10^{-6}$	122.99708 ± 0.00109	$0.003888^{+0.00036}_{-0.000306}$	$16.478^{+2.215}_{-4.924}$	$0.447^{+0.337}_{-0.305}$

Table 3.4: `isochrones` fit results for select fitted stellar parameters (R_s, M_s, $T_{\rm eff}$, $\log g$, [Fe/H], and distance d).

KIC	R_s ($R_\odot$)	M_s ($M_\odot$)	$T_{\rm eff}$ (K)	$\log g$ (cm/s^2)	[Fe/H] (dex)	d (kpc)
1570311	$5.26^{+0.26}_{-0.34}$	$2.11^{+0.27}_{-0.2}$	5062^{+39}_{-45}	$3.347^{+0.035}_{-0.072}$	$-0.181^{+0.086}_{-0.171}$	$2.17^{+0.11}_{-0.13}$
2696784	$1.42^{+0.04}_{-0.03}$	$1.45^{+0.04}_{-0.06}$	7106^{+419}_{-291}	$4.292^{+0.028}_{-0.026}$	$-0.021^{+0.133}_{-0.126}$	$0.62^{+0.01}_{-0.01}$
2861140	$1.28^{+0.1}_{-0.09}$	$1.2^{+0.09}_{-0.08}$	6391^{+221}_{-203}	$4.302^{+0.053}_{-0.06}$	$-0.026^{+0.149}_{-0.139}$	$1.78^{+0.14}_{-0.13}$
2985262	$0.95^{+0.01}_{-0.01}$	$1.02^{+0.03}_{-0.04}$	5902^{+158}_{-167}	$4.494^{+0.011}_{-0.017}$	$-0.059^{+0.153}_{-0.176}$	$0.53^{+0.0}_{-0.0}$
3336146	$1.11^{+0.02}_{-0.02}$	$1.13^{+0.04}_{-0.05}$	6281^{+176}_{-195}	$4.407^{+0.017}_{-0.029}$	$-0.08^{+0.172}_{-0.157}$	$0.69^{+0.01}_{-0.01}$
3345775	$1.87^{+0.03}_{-0.04}$	$2.58^{+0.07}_{-0.13}$	11580^{+214}_{-193}	$4.306^{+0.018}_{-0.022}$	$-0.215^{+0.114}_{-0.199}$	$0.28^{+0.0}_{-0.0}$
3347135	$1.13^{+0.05}_{-0.04}$	$1.16^{+0.05}_{-0.05}$	5932^{+65}_{-89}	$4.395^{+0.029}_{-0.029}$	$0.272^{+0.091}_{-0.095}$	$0.38^{+0.02}_{-0.02}$
3662290	$1.56^{+0.05}_{-0.04}$	$1.48^{+0.09}_{-0.1}$	6986^{+265}_{-322}	$4.221^{+0.039}_{-0.05}$	$0.078^{+0.132}_{-0.133}$	$0.58^{+0.01}_{-0.01}$
3728762	$1.7^{+0.06}_{-0.07}$	$1.43^{+0.16}_{-0.07}$	6706^{+269}_{-234}	$4.129^{+0.084}_{-0.042}$	$0.111^{+0.151}_{-0.196}$	$1.02^{+0.03}_{-0.03}$
3967744	$2.43^{+0.12}_{-0.1}$	$1.85^{+0.08}_{-0.07}$	7004^{+266}_{-260}	$3.932^{+0.041}_{-0.039}$	$0.273^{+0.106}_{-0.121}$	$1.86^{+0.08}_{-0.08}$
4346258	$0.8^{+0.03}_{-0.03}$	$0.86^{+0.04}_{-0.04}$	5279^{+163}_{-146}	$4.567^{+0.021}_{-0.025}$	$-0.038^{+0.145}_{-0.125}$	$0.97^{+0.04}_{-0.03}$
4551429	$0.55^{+0.0}_{-0.0}$	$0.58^{+0.01}_{-0.01}$	3786^{+52}_{-32}	$4.727^{+0.007}_{-0.01}$	$0.334^{+0.086}_{-0.115}$	$0.15^{+0.0}_{-0.0}$
4556565	$1.37^{+0.19}_{-0.1}$	$1.29^{+0.05}_{-0.04}$	6063^{+120}_{-139}	$4.272^{+0.078}_{-0.113}$	$0.345^{+0.096}_{-0.049}$	$1.51^{+0.19}_{-0.09}$
5095499	$1.22^{+0.05}_{-0.04}$	$1.14^{+0.07}_{-0.06}$	6286^{+194}_{-185}	$4.323^{+0.039}_{-0.041}$	$-0.044^{+0.153}_{-0.167}$	$1.51^{+0.07}_{-0.05}$
5184017	$2.61^{+0.16}_{-0.5}$	$1.73^{+0.04}_{-0.1}$	6187^{+293}_{-81}	$3.841^{+0.159}_{-0.041}$	$0.388^{+0.05}_{-0.082}$	$1.27^{+0.05}_{-0.14}$

KIC	R_s ($R_\odot$)	M_s $M_\odot$	$T_{\rm eff}$ (K)	$\log g$ (cm/s^2)	[Fe/H] (dex)	d (kpc)
5342061	$1.09^{+0.05}_{-0.04}$	$1.09^{+0.06}_{-0.08}$	6120^{+214}_{-193}	$4.395^{+0.037}_{-0.038}$	$-0.047^{+0.147}_{-0.154}$	$1.29^{+0.06}_{-0.05}$
5628770	$1.2^{+0.03}_{-0.02}$	$1.22^{+0.04}_{-0.05}$	6510^{+195}_{-193}	$4.363^{+0.017}_{-0.025}$	$-0.057^{+0.148}_{-0.169}$	$0.84^{+0.01}_{-0.01}$
5649129	$4.31^{+0.19}_{-0.35}$	$1.6^{+0.26}_{-0.26}$	4839^{+173}_{-108}	$3.37^{+0.072}_{-0.058}$	$0.004^{+0.111}_{-0.438}$	$1.47^{+0.06}_{-0.08}$
5794479	$2.56^{+0.16}_{-0.12}$	$3.65^{+0.19}_{-0.85}$	12631^{+961}_{-2685}	$4.189^{+0.055}_{-0.14}$	$0.257^{+0.113}_{-0.174}$	$1.44^{+0.04}_{-0.06}$
5893807	$1.71^{+0.16}_{-0.11}$	$1.51^{+0.09}_{-0.11}$	6692^{+260}_{-283}	$4.147^{+0.076}_{-0.083}$	$0.212^{+0.119}_{-0.156}$	$2.1^{+0.16}_{-0.13}$
6021193	$1.62^{+0.03}_{-0.02}$	$1.29^{+0.03}_{-0.03}$	5919^{+131}_{-85}	$4.128^{+0.019}_{-0.015}$	$0.302^{+0.077}_{-0.079}$	$0.78^{+0.01}_{-0.01}$
6126245	$1.54^{+0.04}_{-0.04}$	$1.5^{+0.07}_{-0.1}$	7190^{+346}_{-330}	$4.235^{+0.035}_{-0.044}$	$0.013^{+0.137}_{-0.172}$	$0.76^{+0.02}_{-0.02}$
6139884	$0.89^{+0.01}_{-0.01}$	$0.96^{+0.03}_{-0.04}$	5931^{+194}_{-166}	$4.523^{+0.012}_{-0.018}$	$-0.263^{+0.178}_{-0.196}$	$0.37^{+0.0}_{-0.0}$
6224562	$0.8^{+0.03}_{-0.02}$	$0.86^{+0.04}_{-0.03}$	4977^{+106}_{-121}	$4.571^{+0.018}_{-0.024}$	$0.234^{+0.108}_{-0.125}$	$0.68^{+0.03}_{-0.03}$
6347299	$1.01^{+0.01}_{-0.01}$	$1.07^{+0.03}_{-0.04}$	5965^{+80}_{-79}	$4.455^{+0.014}_{-0.024}$	$0.01^{+0.085}_{-0.091}$	$0.7^{+0.01}_{-0.01}$
6380164	$2.19^{+0.1}_{-0.13}$	$1.67^{+0.05}_{-0.06}$	6717^{+135}_{-120}	$3.977^{+0.054}_{-0.037}$	$0.21^{+0.113}_{-0.119}$	$1.03^{+0.04}_{-0.05}$
6440915	$2.14^{+0.13}_{-0.15}$	$1.64^{+0.07}_{-0.08}$	6577^{+220}_{-184}	$3.986^{+0.07}_{-0.05}$	$0.265^{+0.12}_{-0.145}$	$1.89^{+0.11}_{-0.11}$
6782399	$1.89^{+0.06}_{-0.09}$	$1.51^{+0.08}_{-0.08}$	6659^{+101}_{-110}	$4.06^{+0.062}_{0.031}$	$0.133^{+0.176}_{-0.154}$	$0.84^{+0.02}_{-0.03}$
6837899	$1.09^{+0.03}_{-0.02}$	$1.12^{+0.04}_{-0.05}$	6228^{+191}_{-188}	$4.415^{+0.018}_{-0.031}$	$-0.078^{+0.168}_{-0.153}$	$1.1^{+0.02}_{-0.02}$
6888194	$2.74^{+0.25}_{-0.35}$	$1.82^{+0.08}_{-0.07}$	6482^{+282}_{-149}	$3.812^{+0.118}_{-0.058}$	$0.33^{+0.077}_{-0.128}$	$1.36^{+0.1}_{-0.12}$
6929071	$1.98^{+0.08}_{-0.07}$	$1.54^{+0.09}_{-0.07}$	6633^{+254}_{-251}	$4.03^{+0.041}_{-0.035}$	$0.148^{+0.153}_{-0.153}$	$1.51^{+0.05}_{-0.04}$
6937870	$0.6^{+0.01}_{-0.01}$	$0.62^{+0.02}_{-0.02}$	4209^{+89}_{-89}	$4.681^{+0.008}_{-0.014}$	$-0.089^{+0.146}_{-0.144}$	$0.21^{+0.0}_{-0.0}$
7020834	$2.14^{+0.09}_{-0.08}$	$1.76^{+0.17}_{-0.11}$	7166^{+589}_{-356}	$4.018^{+0.077}_{-0.049}$	$0.203^{+0.146}_{-0.164}$	$0.76^{+0.02}_{-0.02}$
7119412	$0.74^{+0.01}_{-0.01}$	$0.79^{+0.02}_{-0.03}$	4880^{+102}_{-83}	$4.6^{+0.011}_{-0.017}$	$0.021^{+0.094}_{-0.1}$	$0.4^{+0.0}_{-0.0}$
7186892	$0.74^{+0.03}_{-0.02}$	$0.81^{+0.03}_{-0.03}$	4979^{+77}_{-72}	$4.603^{+0.017}_{-0.016}$	$-0.027^{+0.117}_{-0.095}$	$0.19^{+0.01}_{-0.01}$
7187389	$0.88^{+0.02}_{-0.02}$	$0.95^{+0.03}_{-0.04}$	5658^{+176}_{-165}	$4.529^{+0.015}_{-0.02}$	$-0.044^{+0.148}_{-0.182}$	$0.86^{+0.02}_{-0.02}$

KIC	R_s ($R_\odot$)	M_s $M_\odot$	$T_{\rm eff}$ (K)	$\log g$ (cm/s^2)	[Fe/H] (dex)	d (kpc)
7269798	$0.54^{+0.01}_{-0.01}$	$0.58^{+0.01}_{-0.01}$	3758^{+28}_{-21}	$4.73^{+0.009}_{-0.008}$	$0.377^{+0.055}_{-0.075}$	$0.22^{+0.0}_{-0.0}$
7340288	$0.55^{+0.01}_{-0.01}$	$0.57^{+0.02}_{-0.01}$	3949^{+79}_{-52}	$4.722^{+0.008}_{-0.012}$	$0.029^{+0.114}_{-0.149}$	$0.33^{+0.0}_{-0.0}$
7747788	$1.71^{+0.05}_{-0.05}$	$1.63^{+0.07}_{-0.1}$	7146^{+389}_{-314}	$4.186^{+0.031}_{-0.042}$	$0.174^{+0.115}_{-0.124}$	$0.75^{+0.02}_{-0.01}$
7974496	$1.62^{+0.09}_{-0.1}$	$1.33^{+0.08}_{-0.07}$	6448^{+218}_{-284}	$4.14^{+0.065}_{-0.043}$	$0.079^{+0.156}_{-0.126}$	$1.91^{+0.1}_{-0.1}$
8172679	$4.89^{+0.55}_{-0.48}$	$1.7^{+0.19}_{-0.17}$	5040^{+64}_{-72}	$3.284^{+0.084}_{-0.078}$	$-0.452^{+0.181}_{-0.18}$	$1.37^{+0.17}_{-0.12}$
9274173	$1.12^{+0.04}_{-0.03}$	$1.12^{+0.05}_{-0.06}$	6006^{+97}_{-106}	$4.385^{+0.035}_{-0.038}$	$0.122^{+0.122}_{-0.112}$	$1.17^{+0.04}_{-0.03}$
9716483	$1.57^{+0.05}_{-0.05}$	$1.54^{+0.09}_{-0.1}$	7327^{+410}_{-417}	$4.229^{+0.041}_{-0.043}$	$0.008^{+0.143}_{-0.128}$	$0.98^{+0.02}_{-0.02}$
9777962	$2.42^{+0.14}_{-0.15}$	$1.77^{+0.08}_{-0.08}$	6764^{+282}_{-241}	$3.916^{+0.05}_{-0.045}$	$0.24^{+0.126}_{-0.124}$	$2.6^{+0.15}_{-0.14}$
10018357	$4.69^{+0.14}_{-1.02}$	$1.99^{+0.78}_{-0.21}$	5251^{+667}_{-178}	$3.406^{+0.252}_{-0.056}$	$-0.435^{+0.382}_{-0.491}$	$1.83^{+0.04}_{-0.11}$
10083396	$1.56^{+0.03}_{-0.03}$	$1.33^{+0.05}_{-0.06}$	6387^{+116}_{-137}	$4.176^{+0.026}_{-0.026}$	$0.109^{+0.14}_{-0.087}$	$0.46^{+0.01}_{-0.01}$
10419787	$1.2^{+0.03}_{-0.03}$	$1.21^{+0.05}_{-0.06}$	6361^{+208}_{-195}	$4.364^{+0.022}_{-0.032}$	$0.012^{+0.141}_{-0.163}$	$1.01^{+0.03}_{-0.02}$
10598829	$1.55^{+0.2}_{-0.13}$	$1.47^{+0.1}_{-0.08}$	6743^{+166}_{-235}	$4.235^{+0.054}_{-0.109}$	$0.206^{+0.137}_{-0.15}$	$1.35^{+0.17}_{-0.11}$
10879314	$2.47^{+0.21}_{-0.3}$	$1.69^{+0.08}_{-0.08}$	6235^{+152}_{-120}	$3.877^{+0.092}_{-0.054}$	$0.378^{+0.062}_{-0.087}$	$2.18^{+0.19}_{-0.24}$
11092463	$0.88^{+0.03}_{-0.03}$	$0.95^{+0.03}_{-0.04}$	5601^{+148}_{-168}	$4.533^{+0.02}_{-0.024}$	$-0.004^{+0.154}_{-0.144}$	$1.1^{+0.04}_{-0.04}$
11139863	$1.8^{+0.05}_{-0.06}$	$1.63^{+0.14}_{-0.13}$	7185^{+330}_{-353}	$4.14^{+0.057}_{-0.058}$	$0.119^{+0.196}_{-0.277}$	$0.26^{+0.0}_{-0.0}$
11350118	$0.67^{+0.01}_{-0.01}$	$0.72^{+0.02}_{-0.02}$	4751^{+154}_{-153}	$4.646^{+0.011}_{-0.014}$	$-0.171^{+0.15}_{-0.153}$	$0.52^{+0.01}_{-0.01}$
11565976	$1.93^{+0.05}_{-0.05}$	$1.51^{+0.1}_{-0.07}$	6668^{+239}_{-211}	$4.044^{+0.04}_{-0.031}$	$0.102^{+0.167}_{-0.132}$	$0.81^{+0.02}_{-0.02}$
11805835	$0.63^{+0.01}_{-0.01}$	$0.67^{+0.02}_{-0.02}$	4723^{+184}_{-141}	$4.663^{+0.012}_{-0.014}$	$-0.364^{+0.151}_{-0.191}$	$0.38^{+0.0}_{-0.01}$
12023559	$1.03^{+0.02}_{-0.02}$	$1.06^{+0.04}_{-0.05}$	6136^{+172}_{-201}	$4.438^{+0.019}_{-0.026}$	$-0.126^{+0.194}_{-0.175}$	$1.02^{+0.02}_{-0.02}$
12216301	$2.57^{+0.18}_{-0.1}$	$3.54^{+0.16}_{-0.33}$	12170^{+930}_{-1230}	$4.165^{+0.054}_{-0.087}$	$0.293^{+0.092}_{-0.159}$	$1.63^{+0.04}_{-0.04}$
12505309	$1.63^{+0.04}_{-0.03}$	$1.52^{+0.09}_{-0.09}$	6931^{+289}_{-270}	$4.196^{+0.034}_{-0.041}$	$0.138^{+0.13}_{-0.134}$	$0.48^{+0.01}_{-0.01}$

3.6.7 Dilution

As discussed in §1.2.2, when a nearby star is resolved in AO images, the effects of dilution on the estimated planet radius must be considered. I do not take into account changes to the measured values of the primary star's properties due to the presence of a companion. Since most of the targets have properties inferred from photometry only, light from a companion could cause the stellar type to be misidentified. However, this is likely negligible for companions with large contrast ratios.

I consider two cases: that the planet candidate is transiting the brighter primary star (pri), or the fainter companion (sec). Using Eqns. 3 and 4 in Law et al. (2014) [84], the corresponding radius corrections are

$$R_{p,\mathrm{pri}} = R_p \sqrt{\frac{F_{\mathrm{tot}}}{F_{\mathrm{pri}}}} \tag{3.30}$$

and

$$R_{p,\mathrm{sec}} = R_p \frac{R_{\mathrm{sec}}}{R_{\mathrm{pri}}} \sqrt{\frac{F_{\mathrm{tot}}}{F_{\mathrm{sec}}}} \tag{3.31}$$

where F_i/F_{tot} is the fraction of total light contributed by star i in the aperture and R_p is the planet radius without dilution corrections. Assuming that the total flux is provided by the two stars, $F_{\mathrm{tot}} = F_{\mathrm{pri}} + F_{\mathrm{sec}}$, I can rewrite these equations in terms of magnitudes as

$$R_{p,\mathrm{pri}} = R_p \sqrt{1 + 10^{-0.4\Delta m}} \tag{3.32}$$

and

$$R_{p,\mathrm{sec}} = R_p \frac{R_{\mathrm{sec}}}{R_{\mathrm{pri}}} \sqrt{1 + 10^{0.4\Delta m}}. \tag{3.33}$$

Since the Δm values in Eqns. 3.32 and 3.33 are in the Kepler band (Kp), they must be converted from our K_s band AO results. Howell et al. (2012)

[63] derived the following conversion between Kp and K_s magnitudes:

$$\begin{aligned} Kp - K_s = &-643.05169 + 246.00603K_s - 37.136501K_s^2 \\ &+ 2.7802622K_s^3 - 0.10349091K_s^4 \\ &+ 0.0015364343K_s^5 \end{aligned} \tag{3.34}$$

for $10 < K_s < 15.4$ mag, and

$$Kp - K_s = -2.7284 + 0.3311K_s \tag{3.35}$$

for $K_s > 15.4$ mag, allowing me to convert ΔK_s to ΔKp.

Furthermore, for the case that the planet candidate transits the secondary, an estimate of the ratio of stellar radii, $R_{\rm sec}/R_{\rm pri}$ is required. If the companion is in the background, the single-band photometry is not able to constrain $R_{\rm sec}$. However, if I assume the primary and secondary stars are bound, I can use knowledge of the primary star to estimate the properties of the secondary.

I followed the general strategy outlined in Furlan et al. (2017) [53], summarized here. First, I assumed that K_p magnitudes are roughly equivalent to R magnitudes. I used the primary star's known $T_{\rm eff,pri}$ (from the `isochrones` fit) with a table of colours and effective temperature[16] [116] to derive $(V-R)_{\rm pri}$ colours and absolute V magnitudes ($M_{V,\rm pri}$). Then, I assumed that the bound stars are the same distance to the Sun to let $M_{V,\rm sec} = m_{V,\rm sec} - m_{V,\rm pri} + M_{V,\rm pri}$, or $M_{V,\rm sec} = K_{p,\rm sec} + (V-R)_{\rm sec} - K_{p,\rm pri} - (V-R)_{\rm pri} + M_{V,\rm pri}$. I found the $(V-R)_{\rm sec}$ colour that yielded a self-consistent $M_{V,\rm sec}$ value, which in turn gives an estimate of the radius of secondary from the table.

I determined correction factors only for companions which could physically account for the observed transit depth. If the planet needed to fully obscure the companion in order to explain the transit, I ruled out this scenario for the planet host star. For example, a 1% transit depth would rule

[16] http://www.pas.rochester.edu/∼emamajek/EEM_dwarf_UBVIJHK_colors_Teff.txt

out any companion fainter than 5 mags or more as a potential planet host.

Table 3.5 shows the results of each case for the candidates with resolved stars within $4''$. For both cases, the secondary was too faint to account for the transit depth or cause significant dilution.

Table 3.5: Revised planetary radii for the two candidates with AO-resolved stars within $4''$, considering whether the planet transits the primary or secondary.

KIC	R_p ($R_\oplus$)	$R_{p,\mathrm{pri}}$ ($R_\oplus$)	$R_{p,\mathrm{sec}}$ ($R_\oplus$)
7269798 b	$0.88^{+0.09}_{-0.07}$	$0.88^{+0.09}_{-0.07}$	-
7340288 b	$1.51^{+0.13}_{-0.11}$	$1.51^{+0.13}_{-0.11}$	-

3.6.8 Highlighted Discoveries

One of the candidates, KIC-7340288 b, is both likely rocky and in the habitable zone of its star, defined here according to the NASA Exoplanet Archive criteria as having an insolation between 0.25 and $2.2S_\oplus$. Finding Earth-sized planets in the HZ was one of the original goals of the *Kepler* mission [13]. Furthermore, KIC-11350118 c is a new candidate associated with a known KOI system.

KIC-7340288 b: A Candidate Super-Earth in the Habitable Zone

KIC-7340288 b is a $1.51R_\oplus$ planet candidate orbiting a K-dwarf ($T_{\mathrm{eff}} = 3959K$, $R_s = 0.547R_\odot$, $M_s = 0.574M_\odot$) with an orbital period of 142.5 days. This candidate is in the HZ with an insolation of $0.33S_\oplus$, and is also likely rocky given its $< 1.6R_\oplus$ radius.

Phase diagrams are plotted in Fig. 3.5, showing the full transit as well as indicating data corresponding to odd and even transits. The odd-even plot shows consistency between their depths and durations, compared to each other as well as the full transit's best fit model.

I also review the results from the stellar variability analysis in §3.6.2. Fig. 3.6 shows the Lomb-Scargle periodogram, with the 142.5 day orbital period indicated by a dotted line. There is a strong rotation period at ∼13.4

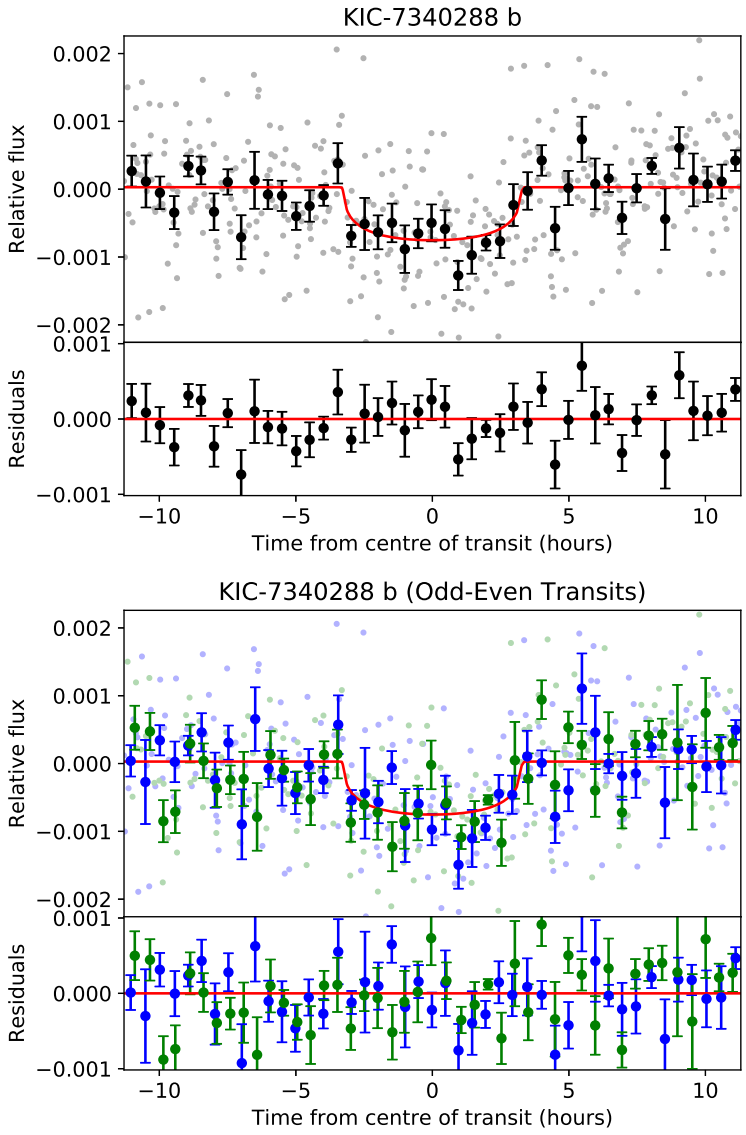

Figure 3.5: Phase diagrams of the `vespa`-confirmed $1.51 R_\oplus$ habitable zone planet KIC-7340288 b, plotting all data together (top) and indicating every odd and even transit (bottom). Odd transits are in blue and even transits are in green. Points plotted faintly in the background represents the actual data, while data binned into 30-minute bins are darker. Error bars represent the standard error of each bin. The full transit model MCMC fit to the data is plotted in red.

days (half of the 26.711 ± 0.231 day rotation period reported in McQuillan et al. (2014) [98]), but no multiples of this rotation period correspond to the orbital period. Meanwhile, Fig. 3.7 confirms that the transit remains consistent regardless of the choice of detrending algorithm used to remove the stellar variability. Plotted are phase diagrams of KIC 7340288, created by detrending the MAST light curve with the original algorithm as well as the Hippke et al. (2019) [59] time-windowed slider with 1- and 2-day window lengths. In each case, the light curve was folded at the BLS-detected period of the planet ($P = 142.5282$ days) and centred at the epoch ($T_0 = 204.7231$ BKJD). Assuming the BLS-detected duration of 0.2355 days, the transits have S/N of 7.4, 7.5, and 7.2 respectively. I also fit least-squares and MCMC transit models to each light curve after masking the transits and redetrending as in the standard pipeline. The best-fit parameters are given in Table 3.6, indicating good agreement within 1σ.

Table 3.6: LS + MCMC fit results for KIC-7340288 b using three different detrends.

Detrend	P (days)	T_0 (BKJD)	R_p/R_s
Original	142.5324 ± 0.0034	204.7104 ± 0.0180	$0.02526^{+0.00201}_{-0.00177}$
Biweight (1 day)	142.5319 ± 0.0039	204.7125 ± 0.0167	$0.02417^{+0.00183}_{-0.00198}$
Biweight (2 day)	142.5319 ± 0.0039	204.7125 ± 0.0170	$0.02232^{+0.00198}_{-0.00206}$

Detrend	a/R_s	b
Original	$156.56^{+12.21}_{-32.38}$	$0.369^{+0.311}_{-0.255}$
Biweight (1 day)	$159.98^{+13.52}_{-40.96}$	$0.385^{+0.347}_{-0.286}$
Biweight (2 day)	$159.58^{+15.68}_{-44.28}$	$0.402^{+0.351}_{-0.293}$

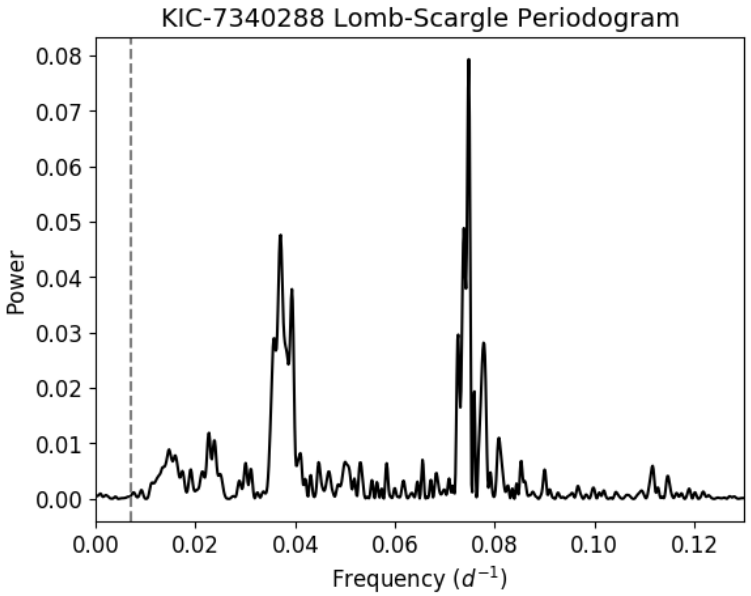

Figure 3.6: Lomb-Scargle periodogram for the (un-detrended) KIC-7340288 light curve, indicating the 142.5 day planet orbital period with a dotted line against the peaks of the periodogram. The strong peak on the right corresponds to the detected ∼13.4 day rotation period, while the strong peak on the left corresponds to its harmonic. No additional peaks were seen at higher frequencies (not plotted).

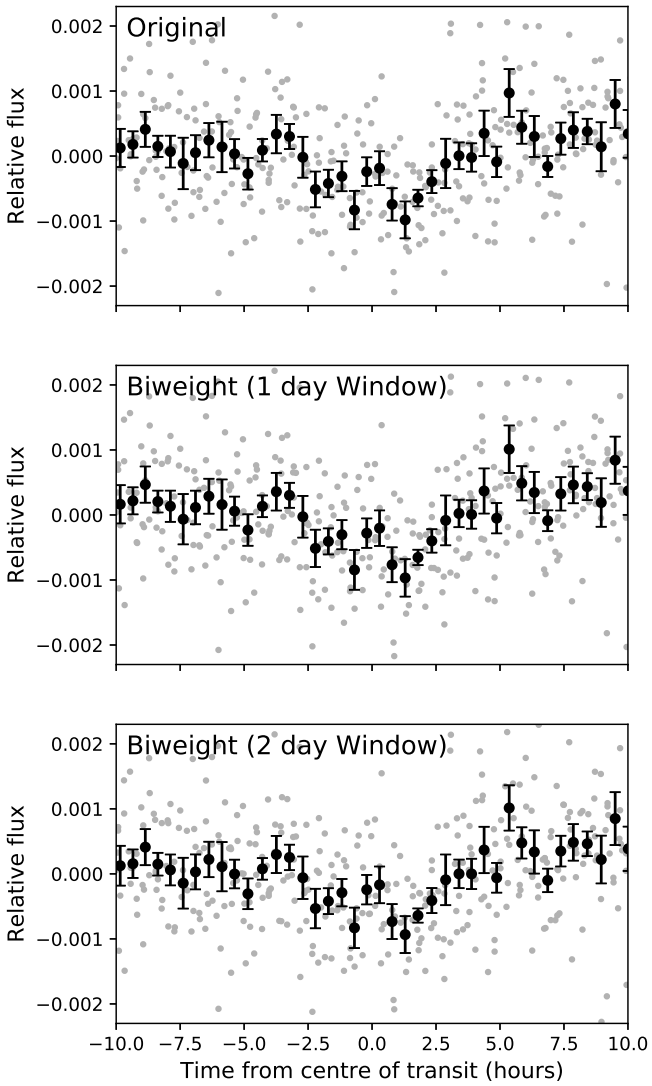

Figure 3.7: Phase diagrams of KIC-7340288 b, plotting original datapoints in grey and data binned into 30-minute bins are in black. Error bars represent the standard error of each bin. Comparison can be made between the following detrending algorithms: the original detrend (top), the time-windowed slider described in Hippke et al. (2019) [59] with a 1-day window length (middle), and the slider with a 2-day window length (bottom).

AO imaging of KIC-7340288 revealed one stellar companion within $4''$, with $\Delta K = 5.20$ and an angular separation of $3.9''$. Assuming the planet orbits the primary star, its radius is unchanged by dilution due to the negligible contribution of light from this companion, and it remains below the rocky limit. I am also able to rule out the scenario that the planet orbits the companion, since the faintness of the star implies it would need to be fully obscured by the planet to explain the observed 0.06% transit depth. After incorporating the AO results into `vespa`, I found an astrophysical false positive probability of 7.91×10^{-4}, which confidently rejects astrophysical explanations. As discussed in §3.6.5, this is a sufficiently low FPP for the standard FPP < 0.01 threshold for *Kepler* exoplanet validation, though I maintain its candidate disposition out of caution for noise explanations due to its low S/N.

KIC-11350118 b: A Small Candidate in a KOI System

The KIC-11350118 system, also known as KOI-4509, already has a single known candidate, with $P = 12.0$ days and $R_p = 0.97 R_\oplus$. I detect an additional, smaller candidate with $P = 2.7$ days and $R_p = 0.66 R_\oplus$. The full transit and odd-even phase diagrams are plotted in Fig. 3.8.

The Robo-AO survey observed the host star and did not detect any companions within $4''$ [163]. Furthermore, the candidate's membership in a multiplanet system lowers its false positive probability [87]. For systems containing two planets in the *Kepler* field, Lissauer et al. (2012) [87] estimated a "multiplicity-boost" factor of 25 to the planet prior probability. After incorporating this into the `vespa` calculation, I found an astrophysical false positive probability of 9.62×10^{-4}.

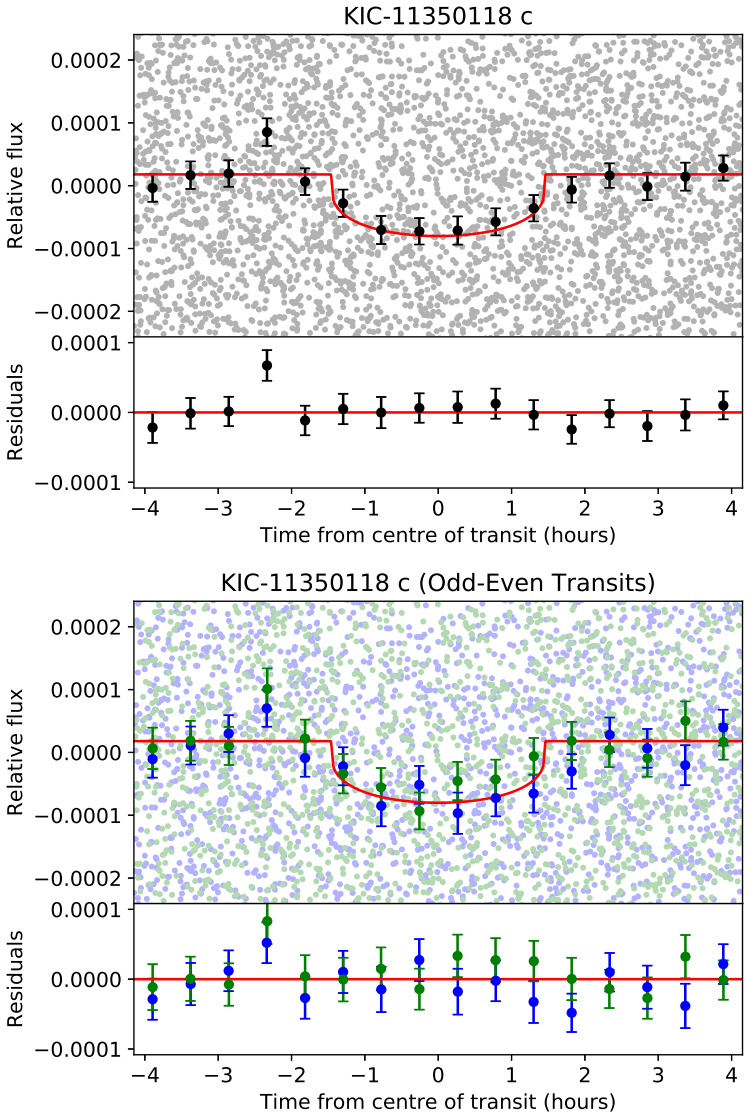

Figure 3.8: Phase diagrams of KIC-11350118 c, otherwise known as KOI-4509.02, plotting all data together (top) and indicating every odd and even transit (bottom) (see Fig. 3.5).

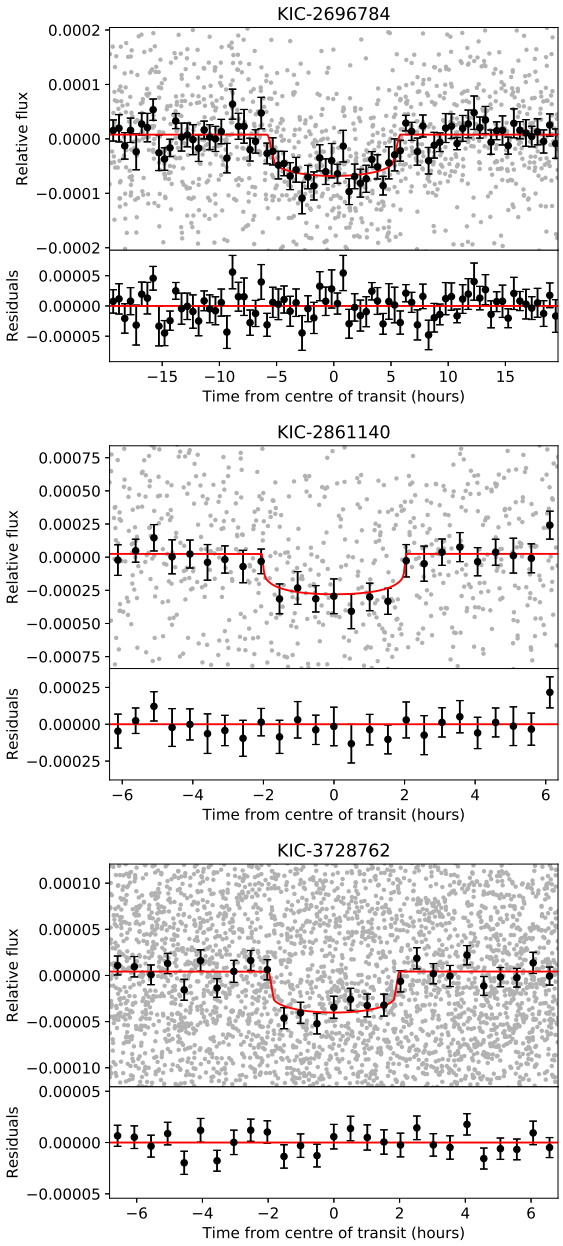

Figure 3.9: Binned phase diagrams of the remainder of the 17 new planet candidates, showing data and model fit with residuals. Original datapoints are plotted in grey, while data binned into 30-minute bins are in black. Error bars represent the standard error of each bin. The transit model MCMC fit to the data is plotted in red

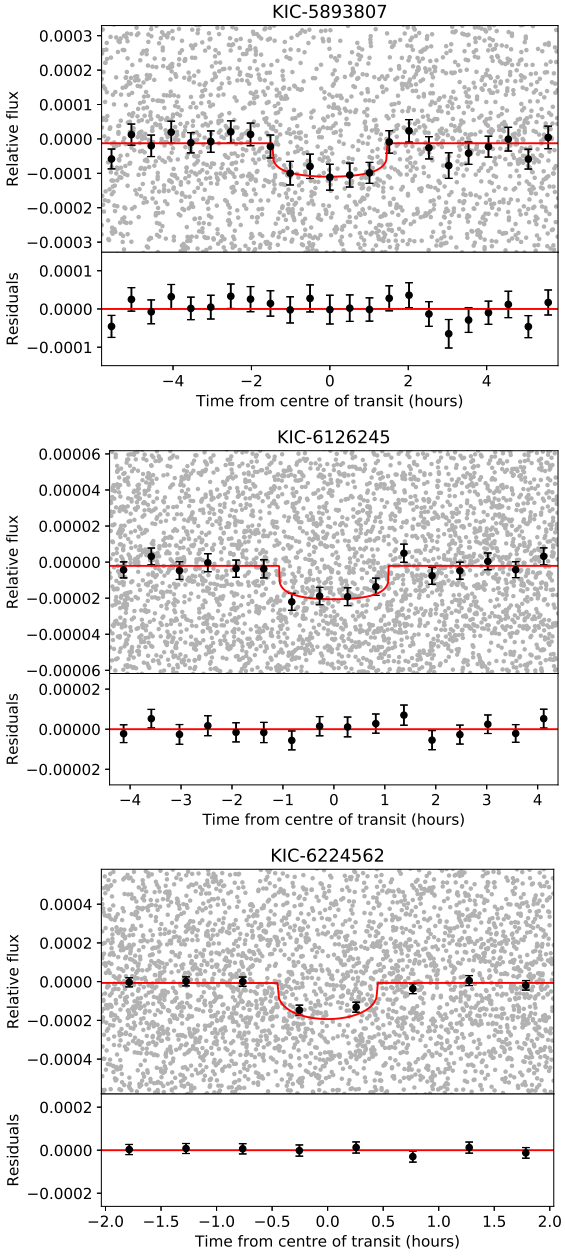

Figure 3.9: (cont.)

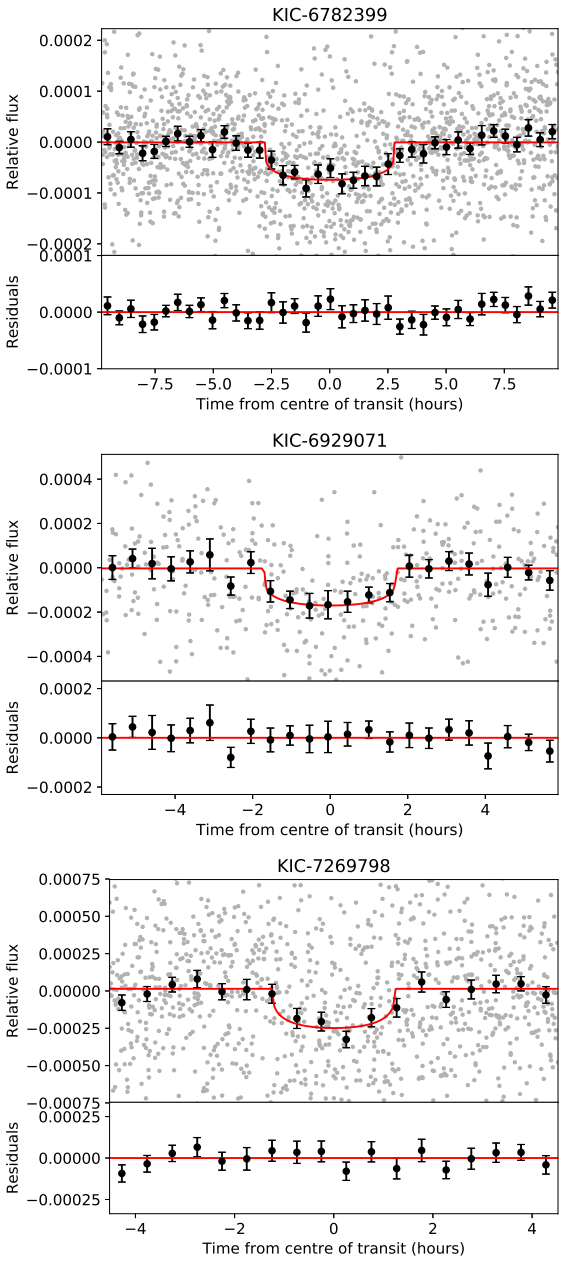

Figure 3.9: (cont.)

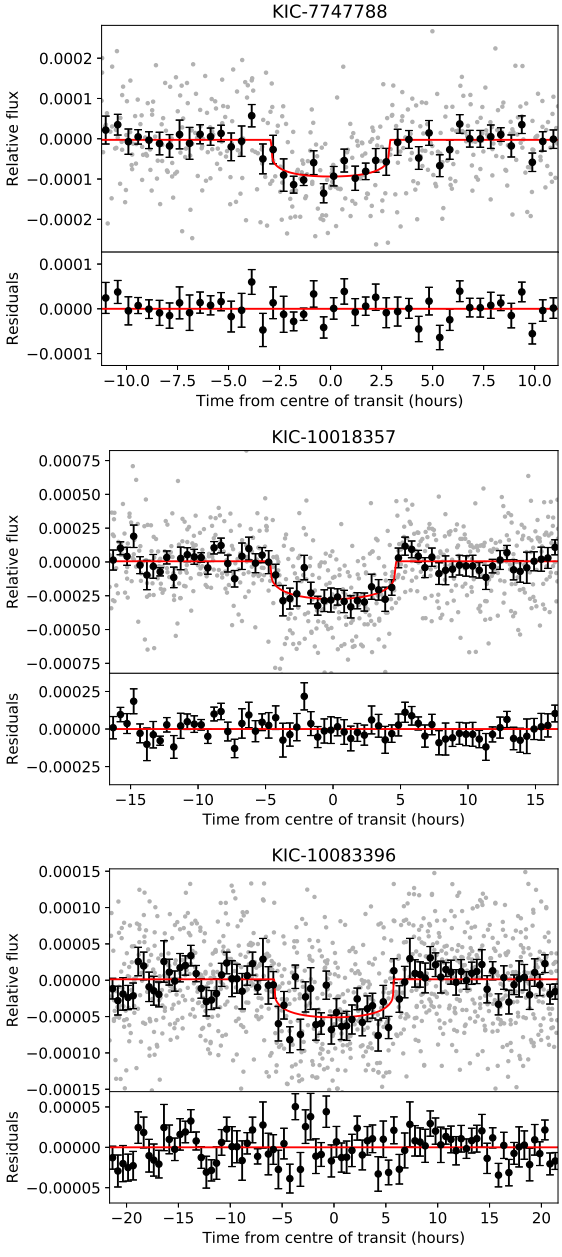

Figure 3.9: (cont.)

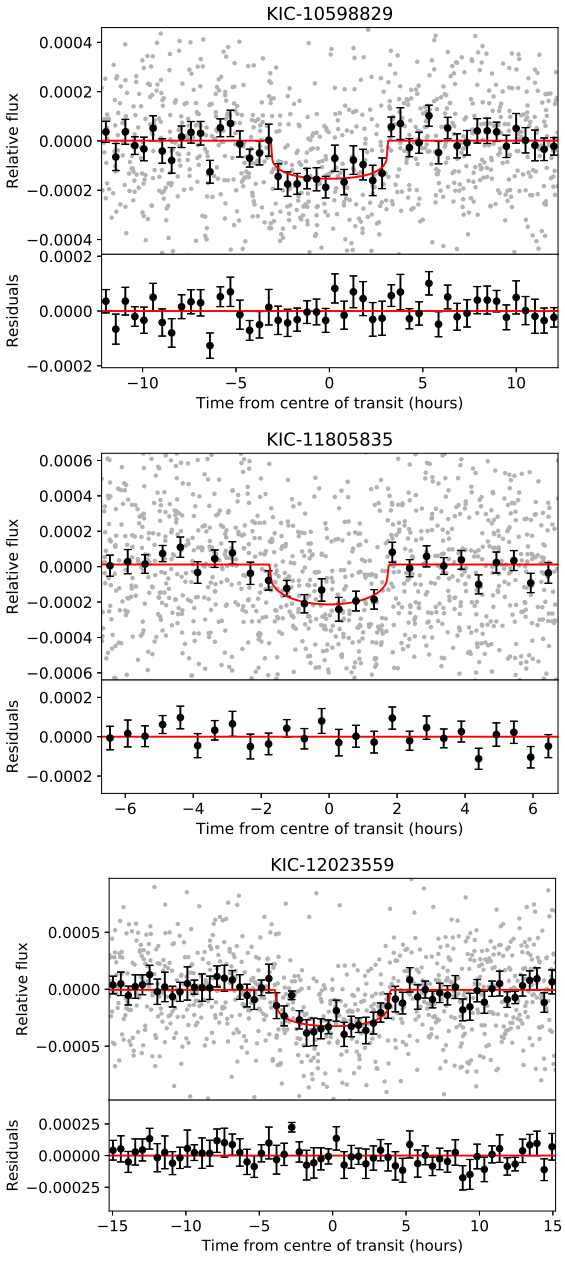

Figure 3.9: (cont.)

Table 3.7: Summary of results for all planet candidates (PC; FPP < 0.9) and false positives (FP; due to low reliability, stellar variability, centroid offset, or FPP > 0.9). Planetary radii do not take into account dilution; refer to Table 3.5.

KIC	P (days)	R_p ($R_\oplus$)	a (AU)	$T_{\rm eq}$ (K)	S ($S_\oplus$)	S/N	FPP	Status	Notes
1570311 b	23.4	$8.40^{+2.49}_{-1.77}$	$0.197^{+0.005}_{-0.008}$	1033^{+88}_{-88}	$270.45^{+104.82}_{-81.37}$	8.9	-	FP	stellar variability
2696784 b	82.3	$1.50^{+0.11}_{-0.10}$	$0.418^{+0.005}_{-0.007}$	579^{+30}_{-25}	$26.68^{+6.06}_{-4.27}$	7.2	-	PC	`vespa` failed
2861140 b	36.9	$2.28^{+0.32}_{-0.27}$	$0.230^{+0.006}_{-0.005}$	666^{+37}_{-33}	$46.52^{+11.15}_{-8.50}$	7.2	0.0526	PC	
2985262 b	13.0	$0.94^{+0.06}_{-0.05}$	$0.109^{+0.001}_{-0.001}$	767^{+23}_{-22}	$81.92^{+10.26}_{-9.15}$	10.7	-	FP	stellar variability
3336146 b	3.3	$0.86^{+0.08}_{-0.07}$	$0.045^{+0.001}_{-0.001}$	1374^{+45}_{-44}	$845.59^{+115.09}_{-102.32}$	9.2	-	FP	centroid offset
3345775 b	6.2	$0.87^{+0.05}_{-0.04}$	$0.091^{+0.001}_{-0.002}$	2320^{+55}_{-50}	$6867.32^{+678.38}_{-568.34}$	7.8	-	FP	stellar variability
3347135 b	226.5	$2.16^{+0.12}_{-0.10}$	$0.765^{+0.010}_{-0.011}$	318^{+8}_{-8}	$2.42^{+0.27}_{-0.24}$	10.0	-	FP	stellar variability
3662290 b	288.2	$2.01^{+0.17}_{-0.13}$	$0.974^{+0.018}_{-0.023}$	391^{+17}_{-19}	$5.53^{+1.05}_{-0.99}$	6.5	-	FP	likely noise
3728762 b	6.7	$1.35^{+0.12}_{-0.11}$	$0.079^{+0.003}_{-0.001}$	1372^{+63}_{-59}	$845.85^{+159.92}_{-140.52}$	8.2	1.45×10^{-5}	PC	
3967744 b	57.9	$3.10^{+0.30}_{-0.25}$	$0.360^{+0.005}_{-0.004}$	801^{+40}_{-34}	$97.73^{+21.24}_{-15.76}$	7.8	-	FP	stellar variability
4346258 b	4.9	$1.19^{+0.12}_{-0.11}$	$0.054^{+0.001}_{-0.001}$	899^{+31}_{-29}	$155.10^{+22.71}_{-19.27}$	8.3	-	FP	stellar variability
4551429 b	35.4	$0.74^{+0.06}_{-0.05}$	$0.176^{+0.001}_{-0.001}$	294^{+4}_{-3}	$1.78^{+0.10}_{-0.07}$	9.0	-	FP	stellar variability
4556565 b	5.5	$1.74^{+0.26}_{-0.20}$	$0.067^{+0.001}_{-0.001}$	1213^{+82}_{-54}	$513.70^{+153.38}_{-85.37}$	9.2	-	FP	stellar variability
5095499 b	4.3	$1.17^{+0.09}_{-0.08}$	$0.054^{+0.001}_{-0.001}$	1317^{+50}_{-48}	$714.34^{+114.32}_{-98.96}$	8.9	-	FP	stellar variability
5184017 b	6.3	$1.72^{+0.23}_{-0.30}$	$0.080^{+0.001}_{-0.002}$	1555^{+93}_{-132}	$1387.16^{+364.15}_{-415.38}$	7.7	-	FP	stellar variability
5342061 c	11.5	$1.59^{+0.15}_{-0.13}$	$0.103^{+0.002}_{-0.002}$	884^{+35}_{-34}	$144.89^{+24.70}_{-21.26}$	7.8	-	FP	stellar variability

KIC	P (days)	R_p ($R_\oplus$)	a (AU)	$T_{\rm eq}$ (K)	S ($S_\oplus$)	S/N	FPP	Status	Notes
5628770 b	11.4	$1.10^{+0.09}_{-0.08}$	$0.106^{+0.001}_{-0.002}$	969^{+32}_{-31}	$208.86^{+28.60}_{-25.80}$	8.1	-	FP	stellar variability
5649129 b	2.8	$3.30^{+0.37}_{-0.32}$	$0.046^{+0.002}_{-0.003}$	2068^{+104}_{-106}	$4338.53^{+941.03}_{-824.17}$	8.2	-	FP	stellar variability
5794479 b	5.9	$2.03^{+0.27}_{-0.14}$	$0.099^{+0.002}_{-0.006}$	2865^{+309}_{-511}	$15980.38^{+8086.37}_{-8692.91}$	9.1	-	FP	stellar variability
5893807 b	7.7	$0.85^{+0.12}_{-0.09}$	$0.087^{+0.002}_{-0.002}$	1309^{+74}_{-71}	$695.49^{+172.15}_{-138.60}$	8.0	1.70×10^{-6}	PC	
6021193 e	26.5	$1.60^{+0.14}_{-0.09}$	$0.189^{+0.002}_{-0.001}$	763^{+19}_{-13}	$80.41^{+8.14}_{-5.5}$	8.3	-	FP	stellar variability
6126245 b	3.5	$0.68^{+0.06}_{-0.06}$	$0.052^{+0.001}_{-0.001}$	1739^{+91}_{-86}	$2166.86^{+488.70}_{-398.49}$	7.3	6.65×10^{-3}	PC	
6139884 b	4.8	$0.51^{+0.04}_{-0.04}$	$0.055^{+0.001}_{-0.001}$	1053^{+36}_{-31}	$291.76^{+42.43}_{-33.13}$	8.7	-	FP	stellar variability
6224562 b	2.3	$1.08^{+0.12}_{-0.09}$	$0.033^{+0.001}_{-0.001}$	1083^{+32}_{-32}	$326.34^{+40.05}_{-36.61}$	9.3	0.110	PC	
6347299 d	38.6	$1.08^{+0.08}_{-0.07}$	$0.229^{+0.002}_{-0.003}$	444^{+10}_{-9}	$22.48^{+1.58}_{-1.37}$	8.2	-	FP	stellar variability
6380164 b	167.8	$3.42^{+0.21}_{-0.23}$	$0.707^{+0.007}_{-0.009}$	521^{+17}_{-19}	$17.51^{+2.34}_{-2.36}$	9.2	-	FP	stellar variability
6440915 b	365.4	$5.56^{+0.60}_{-0.52}$	$1.179^{+0.017}_{-0.019}$	391^{+18}_{-18}	$5.54^{+1.11}_{-0.96}$	8.8	-	FP	likely noise
6782399 b	34.2	$1.65^{+0.12}_{-0.10}$	$0.237^{+0.004}_{-0.004}$	828^{+21}_{-25}	$111.59^{+11.86}_{-12.95}$	8.2	1.29×10^{-4}	PC	
6837899 b	9.0	$1.27^{+0.12}_{-0.11}$	$0.088^{+0.001}_{-0.001}$	968^{+33}_{-33}	$207.91^{+29.44}_{-26.61}$	7.7	-	FP	stellar variability
6888194 b	46.0	$2.76^{+0.33}_{-0.40}$	$0.307^{+0.004}_{-0.004}$	858^{+50}_{-59}	$128.74^{+33.07}_{-32.03}$	7.4	-	FP	stellar variability
6929071 b	61.9	$2.45^{+0.24}_{-0.21}$	$0.363^{+0.005}_{-0.006}$	692^{+27}_{-28}	$54.38^{+8.86}_{-8.25}$	7.7	-	PC	`vespa` failed
6937870 b	27.5	$0.65^{+0.07}_{-0.05}$	$0.152^{+0.001}_{-0.002}$	368^{+9}_{-8}	$4.33^{+0.42}_{-0.36}$	7.7	-	FP	stellar variability
7020834 b	369.5	$4.34^{+0.92}_{-0.39}$	$1.218^{+0.039}_{-0.026}$	420^{+35}_{-25}	$7.34^{+2.80}_{-1.60}$	9.6	-	FP	likely noise
7119412 b	10.5	$0.89^{+0.10}_{-0.07}$	$0.087^{+0.001}_{-0.001}$	623^{+14}_{-12}	$36.84^{+3.34}_{-2.65}$	9.3	-	FP	stellar variability
7186892 b	17.2	$0.56^{+0.06}_{-0.04}$	$0.122^{+0.001}_{-0.002}$	543^{+13}_{-13}	$20.61^{+2.09}_{-1.88}$	10.1	-	FP	stellar variability
7187389 b	23.8	$1.10^{+0.10}_{-0.09}$	$0.159^{+0.002}_{-0.002}$	488^{+21}_{-20}	$28.28^{+4.36}_{-3.57}$	6.8	-	FP	stellar variability
7269798 b	21.4	$0.88^{+0.09}_{-0.07}$	$0.126^{+0.001}_{-0.001}$	344^{+4}_{-3}	$3.34^{+0.14}_{-0.12}$	8.2	0.0113	PC	

KIC	P (days)	R_p ($R_\oplus$)	a (AU)	$T_{\rm eq}$ (K)	S ($S_\oplus$)	S/N	FPP	Status	Notes
7340288 b	142.5	$1.51^{+0.13}_{-0.11}$	$0.444^{+0.004}_{-0.004}$	194^{+4}_{-3}	$0.33^{+0.03}_{-0.02}$	7.4	7.91×10^{-4}	PC	rocky HZ
7747788 b	133.1	$1.67^{+0.13}_{-0.11}$	$0.601^{+0.009}_{-0.012}$	533^{+29}_{-25}	$19.19^{+4.61}_{-3.33}$	7.6	4.92×10^{-4}	PC	
7974496 b	4.0	$1.48^{+0.18}_{-0.16}$	$0.054^{+0.001}_{-0.001}$	1549^{+75}_{-81}	$1364.17^{+283.65}_{-262.54}$	7.1	-	FP	stellar variability
8172679 b	194.0	$9.61^{+1.23}_{-1.09}$	$0.784^{+0.029}_{-0.027}$	557^{+34}_{-33}	$22.80^{+6.11}_{-5.00}$	10.5	-	FP	stellar variability
9274173 b	4.4	$1.42^{+0.13}_{-0.11}$	$0.055^{+0.001}_{-0.001}$	1199^{+30}_{-29}	$490.71^{+51.21}_{-46.03}$	8.7	-	FP	stellar variability
9716483 b	209.4	$2.08^{+0.14}_{-0.11}$	$0.797^{+0.015}_{-0.017}$	454^{+28}_{-26}	$10.10^{+2.68}_{-2.15}$	8.6	-	FP	likely noise
9777962 b	367.2	$7.46^{+1.35}_{-0.84}$	$1.215^{+0.017}_{-0.019}$	422^{+21}_{-21}	$7.51^{+1.61}_{-1.37}$	8.9	-	FP	likely noise
10018357 b	133.8	$7.87^{+0.71}_{-1.72}$	$0.644^{+0.075}_{-0.023}$	616^{+89}_{-74}	$34.05^{+24.49}_{-13.69}$	9.1	1.13×10^{-8}	PC	
10083396 b	113.5	$1.14^{+0.07}_{-0.06}$	$0.504^{+0.006}_{-0.007}$	495^{+12}_{-12}	$14.19^{+1.39}_{-1.28}$	7.4	5.86×10^{-5}	PC	
10419787 b	122.7	$2.06^{+0.17}_{-0.14}$	$0.515^{+0.007}_{-0.009}$	429^{+15}_{-15}	$8.03^{+1.19}_{-1.05}$	7.6	-	FP	stellar variability
10598829 b	67.5	$1.96^{+0.30}_{-0.22}$	$0.369^{+0.009}_{-0.007}$	607^{+42}_{-32}	$32.29^{+9.81}_{-6.37}$	7.4	3.59×10^{-3}	PC	
10879314 b	49.2	$3.92^{+0.57}_{-0.54}$	$0.313^{+0.005}_{-0.005}$	773^{+43}_{-51}	$84.68^{+20.52}_{-20.40}$	7.2	-	FP	stellar variability
11092463 b	6.9	$1.33^{+0.24}_{-0.14}$	$0.070^{+0.001}_{-0.001}$	877^{+30}_{-31}	$140.06^{+20.17}_{-18.65}$	7.8	-	FP	stellar variability
11139863 b	7.2	$0.81^{+0.07}_{-0.05}$	$0.086^{+0.002}_{-0.002}$	1444^{+87}_{-77}	$1031.85^{+272.09}_{-202.39}$	10.6	-	FP	stellar variability
11350118 c	2.7	$0.66^{+0.07}_{-0.05}$	$0.034^{+0.001}_{-0.001}$	935^{+31}_{-31}	$181.58^{+25.48}_{-22.96}$	8.2	9.62×10^{-4}	PC	KOI-4509.02
11565976 b	24.2	$1.40^{+0.12}_{-0.11}$	$0.188^{+0.004}_{-0.003}$	941^{+39}_{-37}	$186.19^{+32.67}_{-27.77}$	7.4	-	FP	stellar variability
11805835 b	23.5	$0.94^{+0.10}_{-0.07}$	$0.141^{+0.001}_{-0.001}$	442^{+17}_{-14}	$9.01^{+1.49}_{-1.08}$	7.2	2.30×10^{-3}	PC	
12023559 b	84.6	$1.86^{+0.13}_{-0.11}$	$0.385^{+0.005}_{-0.006}$	444^{+14}_{-16}	$9.19^{+1.24}_{-1.25}$	8.1	4.54×10^{-4}	PC	
12216301 b	116.5	$3.66^{+0.41}_{-0.26}$	$0.712^{+0.010}_{-0.023}$	1029^{+104}_{-108}	$265.44^{+124.64}_{-95.33}$	7.6	-	FP	stellar variability
12505309 b	2.9	$1.20^{+0.10}_{-0.08}$	$0.046^{+0.001}_{-0.002}$	1823^{+81}_{-62}	$2617.39^{+497.97}_{-336.756}$	7.4	-	FP	stellar variability

Chapter 4

Occurrence Rate Estimates

This chapter has been adapted from a manuscript accepted for publication in The Astronomical Journal: Kunimoto, M. & Matthews, J. 2020, Searching the Entirety of Kepler Data. II. Occurrence Rate Estimates for FGK Stars.

4.1 Chapter Outline

The results of the previous chapter allow for the determination of occurrence rates. I describe cuts made to produce my input stellar and planet catalogues, which are subsets of those from the independent search, in §4.2. In §4.3, I describe the determination of both search and vetting completeness using injection/recovery tests. In §4.4, I give an overview of the specific occurrence rate methodology used, which incorporates forward modeling via approximate Bayesian computation, and discuss the application of ABC to exoplanet occurrence rates in §4.5.

My overall results are presented in §4.6, in which I discuss the dependence of exoplanet occurrence rates on stellar effective temperature (§4.6.2), planet radius (§4.6.3), and orbital period (§4.6.4). I also describe the incorporation of the coarse catalogue reliability model from §3.4.3 to assess the impact of a nonzero FP rate and better constrain estimates (§4.6.5). In §4.7, I present both baseline and reliability-incorporated results over several definitions of the potentially habitable, rocky exoplanet parameter space, and give a final recommended $\eta_{\oplus}$ estimate. Lastly, in §4.8, I review the limitations of my methodology to indicate primary areas of improvement for the future.

4.2 Input Catalogues

4.2.1 Stellar Sample

I started with the 197,096 *Kepler* stars in the Q1-Q17 DR25 stellar catalogue [94], and calculated limb-darkening coefficients using $T_{\rm eff}$, log$g$, and [Fe/H] from Claret & Bloemen (2011) [33]. With the arrival of stellar parallaxes in *Gaia* Data Release 2 (DR2), Berger et al. (2018) [11] produced improved radii for 177,911 targets, yielding an average radius precision of less than 10% for most *Kepler* stars, compared to the ≈40% precision from photometry alone. Given that a fully updated set of stellar properties has not yet been released, I used these radii in tandem with other properties from Mathur et al. (2017) [94] and only kept stars present in both catalogues.

To clean this sample, I removed stars flagged in Berger et al. (2018) [11] as likely binary stars (BIN flag = 1 or 3; 174,769 stars remain). I did not remove those flagged as binaries due to companions revealed with high-resolution imaging (BIN flag = 2), as these observations were only available for a subset of stars. Because the focus of my study is FGK dwarfs, I also removed stars with Evol flag > 0 (116,637 stars remained), which indicate that they are unlikely to be on the main sequence.

To ensure each star's light curve had enough data to allow for the discovery of long-orbit planets, I required that the time length of the data ($T_{\rm obs}$) is at least 2 yr, and the duty cycle ($f_{\rm duty}$) was at least 0.6; in other words, at least 60% of the observations must be filled. After these cuts, 100,823 stars remained.

Lastly, I retained only FGK stars by using suggested $T_{\rm eff}$ limits from Pecaut & Mamajek (2013) [116]. This left 40,010 F- ($6000 \leq T_{\rm eff} < 7300K$), 39,173 G- ($5300 \leq T_{\rm eff} < 6000K$), and 17,097 K-type ($3900 \leq T_{\rm eff} < 5300K$) stars, for a total of 96,280 stars in the sample.

Some stars in this sample may have been chosen as targets for reasons other than the *Kepler* exoplanet search program, such as for asteroseismology. These stars would be expected to exhibit different noise and variability properties than typical main-sequence stars, which could introduce a systematic bias in the results relative to studies that focus on only exoplanet

search targets. I checked the investigation ID of each star in the sample using the *Kepler* Data Search & Retrieval form on the MAST, and found that 290 did not have an "EX*" ID. In other words, 99.7% of the 96,280 FGK stars were selected for the exoplanet search program, and I do not expect a significant bias to be present.

4.2.2 Planet Sample

My full search and vetting pipeline was described in Chapter 3. To review, I used a Kovacs et al. (2002) [80] box least-squares (BLS) algorithm to search for potential transits around all ∼200,000 stars observed by *Kepler*, defining a transit candidate (TC) as a signal with signal-to-noise ratio S/N > 6, at least three transits, and the meeting of other requirements in an initial vetting stage to reject false alarms caused by instrumental and astrophysical systematics. Each TC was passed through a vetting pipeline, involving both machine and manual triage. Automated candidacy tests were used to flag both noise and astrophysical FPs, while visual inspection was used as a "reality check" to confirm each surviving TC as a planet candidate (PC).

Around the 96,280 FGK stars considered, I identified 2623 PCs that matched with already known planet candidates as listed on the NASA Exoplanet Archive,[17] defined as *Kepler* Objects of Interest (KOIs) with either a CONFIRMED or CANDIDATE disposition. Additionally, I introduce eight previously unknown candidates from my independent search, for a total of 2631 planets in the full catalogue. By comparison, *Kepler*'s Q1-Q17 DR25 pipeline identified 2829 planet candidate KOIs corresponding to this stellar sample.

Confirmed and Candidate KOIs Missed

I had a 98.9% recovery rate for all confirmed FGK KOIs, finding and passing 1655 of 1673. Nine of the KOIs (KOI-172.02, 701.04, 1236.03, 2038.03, 2365.02, 4034.01, 4384.01, 5706.01, and 7016.01) were either very close to

[17] https://exoplanetarchive.ipac.caltech.edu/, accessed 09 May 2019

passing the vetting pipeline, or failed only one of my tests, while four (KOI-245.03, 490.02, 1274.01, 3234.01) were detected but failed to meet the requirements to become a TC. KOI-490.02 and KOI-1274.01 were strong signals, but had less than the required three transits. The only planet completely missed was KOI-245.04, though despite its Confirmed Exoplanet Archive Disposition, it is also flagged as a Not Transit-Like FP.

Another four of the failed confirmed KOIs (KOI-142.01, 377.01, 377.02, and 884.02) displayed significant transit timing variations (TTVs). Because my vetting pipeline did not correct for TTVs, it is unsurprising that these failed despite their high S/N. Given the unique nature of these planets and considering that neither my search nor vetting completeness models take into account TTVs, I decided to include these in the catalogue.

I summarize all confirmed planets not included in the catalogue in Table 4.1. Note that four were also missed by the Q1-Q17 DR25 pipeline, and six that were detected may not necessarily be considered "high-quality" candidates (e.g. requiring Disposition Score > 0.9 [107]). Meanwhile, I had a lower recovery rate of candidate KOIs, finding 961 of 1487 (64.6%). A lower rate is to be expected considering that confirmed planets typically have higher S/N and transit shapes more clearly consistent with a planetary origin. Furthermore, 299 (around 60%) of the candidates missed or failed by the pipeline were not detected by the DR25 pipeline.

My goal was to produce an independent pipeline that could both search for planets and be used for completeness modeling conducive to occurrence rate statistics. Thus, with the exception of the confirmed KOIs failed due to exhibiting TTVs, I do not include any of the KOIs missed or failed in my determination of occurrence rates.

New Planet Candidates

I added eight new candidates to the FGK planet catalogue, listed in Table 4.2. As discussed in §3.6, these candidates passed the full vetting pipeline, and underwent additional analysis including astrophysical false positive probability calculation.

Table 4.1: Confirmed planet KOIs corresponding to the FGK stars in the sample missed or failed by my pipeline. Table entries are taken from the NASA Exoplanet Archive.

KOI	P (days)	R_p ($R_\oplus$)	S/N	Disposition Score	TCE Delivery
172.02	242.5	1.73	23.20	0.6930	Q1-Q17 DR25
245.03	13.4	0.27	7.40	-	-
245.04	51.2	-	-	-	-
490.02	1071.2	9.27	544.20	0.0000	Q1-Q17 DR25
701.04	267.3	1.43	19.30	0.0000	Q1-Q17 DR25
1236.03	54.4	3.20	44.90	-	Q1-Q17 DR24
1274.01	705.0	4.53	96.10	-	-
2038.03	17.9	1.39	11.40	0.8890	Q1-Q17 DR25
3234.01	2.4	0.85	13.40	0.9930	Q1-Q17 DR25
4034.01	7.0	6.14	18.60	0.1000	Q1-Q17 DR25
4384.01	122.4	2.15	12.20	0.9970	Q1-Q17 DR25
5706.01	425.5	3.20	19.60	0.9040	Q1-Q17 DR25
7016.01	384.8	1.09	12.30	0.7710	Q1-Q17 DR25

Table 4.2: New planet candidates added to the FGK planet catalogue from Chapter 3. Planet candidates are listed according to their *Kepler* Input Catalogue (KIC) ID.

KIC	KOI	P (days)	R_p ($R_\oplus$)
2696784 b	-	82.3	1.50
2861140 b	-	36.9	2.28
6126245 b	-	3.5	0.68
6782399 b	-	34.2	1.65
7747788 b	-	133.1	1.67
11350118 c	4509.02	2.7	0.66
11805835 b	-	23.5	0.94
12023559 b	-	84.6	1.86

Planet Properties

As part of the vetting pipeline, I found a least-squares best fit of each planet transit with a Mandel & Agol (2002) [90] quadratic limb-darkening transit model assuming circular orbits. The model is parameterized by orbital period (P), transit epoch (T_0), ratio of the planet and star radii (R_p/R_s),

distance between planet and star at midtransit in units of stellar radius (a/R_s), impact parameter (b), and zero-point flux (z). Following the vetting pipeline, I refit each transit using `emcee`, with each walker initialized near the best-fit parameters from the least-squares fit. I set P and T_0 fixed to their least-squares values to aid in convergence.

For the four planets with TTVs, I used the MCMC fit results listed on the NASA Exoplanet Archive.

Dilution

The planet radius R_p can be determined from the fitted parameter R_p/R_s by multiplying by the known stellar radius. However, as discussed in §3.6.4, there may be one or more nearby stars that contribute light to the *Kepler* aperture, causing the measured transit depth to be diluted and the radius to be underestimated.

For the purposes of this occurrence rate study, I made the assumption that the planet orbits the brighter primary star. I used the high-resolution imaging results from the *Kepler* Follow-Up Observation Program [53] to correct the radii of planets around stars with a potential companion within $4''$, the size of a *Kepler* pixel. Furlan et al. (2017) [53] compiled observations for a total of 3557 KOIs, including those observed in the first three Robo-AO surveys [5, 84, 163], and provided a weighted average of correction factors across a variety of bands for 1891 KOIs with companions.

Ziegler et al. (2018) [164] presented a fourth Robo-AO survey for 532 KOIs published after Furlan et al. (2017) [53]. Their results were provided as Δm in the LP600 band, which I approximate to be equal to the *Kepler* band for use in Eqns. 1.6 and 1.7.

Three of my FGK PCs (KIC 6126245 b, 6782399 b, and 7747788 b) had AO imaging follow-up from §3.6.4, and none had a nearby stellar companion. In total, 2578 of the 2631 PCs (98.0%) in my FGK sample had high-resolution imaging observations, and I applied correction factors to 679.

Final Planet Catalogue

My focus for this work is on the occurrence rates of planets in a period-radius grid spanning orbital periods $0.78 < P < 400$ days and radii $0.5 < R_p < 16.0$ $R_\oplus$. Lower and upper limits on these properties were chosen so as to split the grid into logarithmically spaced bins comparable to bins used in previous grid-based works [51, 52, 62, 105, 119]. After applying the radius correction factors and including only candidates that fit these criteria, my final planet catalogue involves 557 candidates around F-type stars, 1,276 around G-type stars, and 700 around K-type stars, for a total of 2533 planet candidates. Fig. 4.1 shows the distribution of planets based on orbital period and radius, while Table 4.3 summarizes the sizes of each star and planet sample.

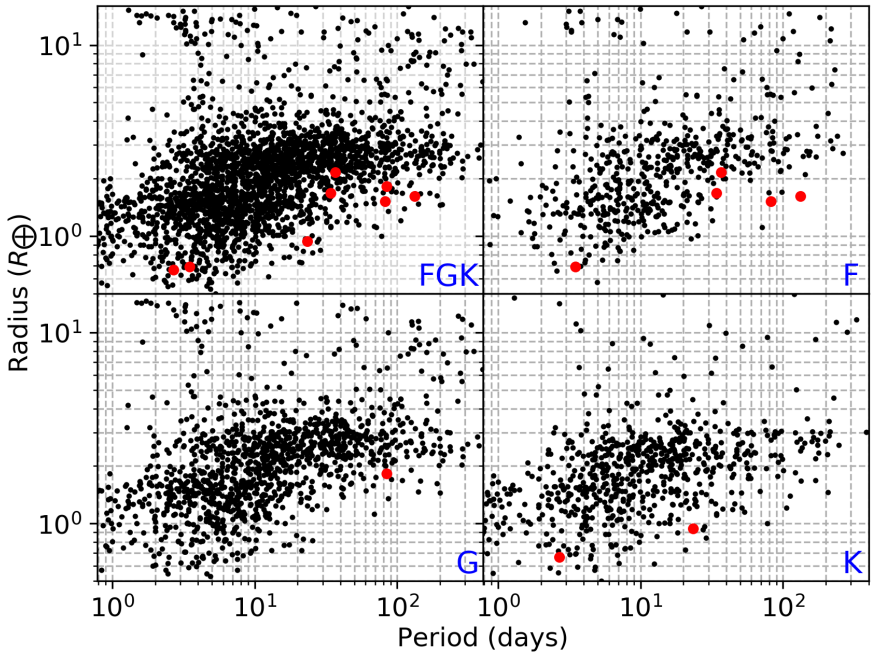

Figure 4.1: Planets in the final catalogue, plotted according to orbital period and radius. Plots are organized by host star stellar type, including F- ($6000 \leq T_{\rm eff} < 7300K$), G- ($5300 \leq T_{\rm eff} < 6000K$), and K-type ($3900 \leq T_{\rm eff} < 5300K$) stars. The eight new candidates from my independent search are plotted in red.

Table 4.3: Number of stars and planets by stellar type. N_{DR25} gives the number of planets found in DR25 around the same sample of stars for comparison.

Type	T_{min} (K)	T_{max} (K)	N_{stars}	N_{planets}	N_{DR25}
FGK	3900	7300	96,280	2,533	2,700
F	6000	7300	40,010	557	639
G	5300	6000	39,173	1,276	1,338
K	3900	5300	17,097	700	723

4.3 Completeness Model

This planet sample is not expected to be complete. As discussed in §1.3.1, particularly near the detection limit, transiting planets are often missed or even mislabeled as false positives. Thus, it is important to quantify the completeness corrections for both the transit detection pipeline and vetting pipeline to derive accurate occurrence rates. Here, search completeness refers to the fraction of transiting planets that are detected, while vetting completeness refers to the fraction of detected planets that are correctly classified as planet candidates.

Search completeness is a common feature of occurrence rate studies. Most notably, using injection/recovery tests, Christiansen et al. (2015) [29] showed that the *Kepler* detection efficiency is well modeled by a gamma cumulative distribution function, of the form

$$P_{\mathrm{det}}(\mathrm{S/N}) = \frac{c}{b^a(a-1)!}\int_0^{\mathrm{S/N}} x^{a-1}e^{-x/b}dx, \tag{4.1}$$

giving the probability of detecting a transit with a given signal-to-noise ratio S/N.

However, vetting completeness has often been ignored, with most previous studies assuming perfect efficiency at classifying planet transit signals as planets. In a comparison between *Kepler* DR25 occurrence rates derived under this assumption and various vetting models, Hsu et al. (2019) [65] found that taking into account imperfect vetting was important for small

planets ($R_p < 2R_\oplus$) and planets with orbital periods longer than a month ($P > 32$ days). They also found that their occurrence rates were robust to the choice of vetting model, as differences between the two models tested were still significantly smaller than the uncertainty due to the *Kepler* sample size.

With these considerations, I adopt the Hsu et al. (2019) [65] combined detection and vetting efficiency model described in §2.2.2 of their paper, using injection/recovery tests to determine the fraction of planets both successfully detected and vetted by the automated pipeline. These results are fit to the Christiansen et al. (2015) [29] gamma cumulative distribution function, and a direct dependence on the number of transits $N_{\rm tr}$ is introduced by fitting separate functions for injections with 3, 4, 5, 6, 7-9, 10-18, 19-36, and ≥ 37 transits.

Similar to Petigura et al. (2013) [119], I injected 96,280 planet transits, one for each FGK star in the sample, into Q1-Q17 light curves downloaded from MAST. Half the signals were log-uniformly distributed over $0.78 < P < 100$ days and $0.5 < R_p < 16.0$ $R_\oplus$, with the other half log-uniformly distributed over $100 < P < 500$ days so as to improve the determination of completeness for planets with low numbers of transits. Each transit was created using a quadratic limb-darkening Mandel & Agol (2002) [90] model, with impact parameters (b) uniformly distributed between 0 and 1 and circular orbits assumed.

I prepared, searched, and vetted the simulated data with the same process as for the actual observed data, using the federation process described in Mulally et al. (2015) [108] to match detections with the injected planets. The only exception was that I did not perform the manual vetting stage given that it would be infeasible to review the tens of thousands of simulated PCs that were passed by the automated stage. Using similar injection/recovery tests in Chapter 3, I estimated that the manual inspection would lower overall vetting completeness by only ∼1-2%, which would indicate that neglecting to account for these planets failed by the manual stage should not significantly impact occurrence rates.

Eqn. 4.1 requires a calculation of each injected transit's expected S/N

— in other words, an estimate of the S/N that the BLS algorithm would calculate (Eqn. 3.1) if it detected the planet. This can be found using the injected planet's known radius, period, and impact parameter, and basic properties known about the star and corresponding light curve.

First, I estimate the number of transits from the length of observations in the light curve and the planet's orbital period, taking into account loss of data with $f_{\rm duty}$,

$$N_{\rm tr} = \frac{T_{\rm obs} f_{\rm duty}}{P}. \tag{4.2}$$

The transit duration $T_{\rm dur}$ can be estimated as

$$T_{\rm dur} = \frac{R_s}{a}\frac{P}{\pi}\sqrt{1-b^2}, \tag{4.3}$$

where b is the injected transit's impact parameter and a is the semi-major axis of its orbit, from

$$a = \left(\frac{GM_s P^2}{4\pi^2}\right)^{1/3} \tag{4.4}$$

with stellar mass M_s. Combined with $N_{\rm tr}$ and a rate of one observation every 29.42 minutes (one *Kepler* long cadence), the total number of data points during transit can be approximated as

$$N = \frac{N_{\rm tr} T_{\rm dur}}{29.42 \text{ min}}. \tag{4.5}$$

Lastly, I calculate the expected depth of the transit δ from the ratio of planet to star radii, $k = R_p/R_s$, taking into account quadratic limb-darkening

coefficients u_1 and u_2. Zink et al. (2019) [165] estimated this as

$$\begin{aligned} A &= 1 - (u_1 + u_2) \\ B &= \frac{A}{4} + \frac{u_1 + 2u_2}{6} - \frac{u_2}{8} \\ \delta &= 1 - \frac{1}{B}\left(\frac{A}{4} + \frac{(u_1 + 2u_2)(1 - k^2)^{3/2}}{6}\right. \\ &\quad \left. - \frac{u_2(1 - k^2)}{8}\right). \end{aligned} \tag{4.6}$$

Putting everything together, an injected transit's expected S/N is

$$\mathrm{S/N} = \sqrt{N}\frac{\delta}{\sigma} \tag{4.7}$$

where σ is the noise level of the star, estimated using the MAD of the light curve with $\sigma = 1.48\mathrm{MAD}$ [60].

Fig. 4.2 shows the fraction of successful detections as a function of expected S/N for $N_{\mathrm{tr}} = 7 - 9$ and $N_{\mathrm{tr}} \geq 37$ as examples. The recovery fractions based on the search pipeline alone and the combined search and vetting pipeline are shown for comparison. As expected, the vetting process affects recovery at lower S/N (≤ 15) significantly more than at higher S/N, and overall recovery is improved for planets with more transits. The full fit results are shown in Table 4.4.

Fig. 4.2 and Table 4.4 also give the corresponding combined search and vetting completeness models from Hsu et al. (2019) [65]. As a reminder, these were based on the *Kepler* DR25 pipeline's injection/recovery tests [30], and were fit as a function of expected MES. The DR25 pipeline is significantly better at recovering low-S/N events and those with few transits. This is expected given my more simplistic pre-search data reduction. At higher S/Ns, especially for events with more transits, pipeline performance is more similar.

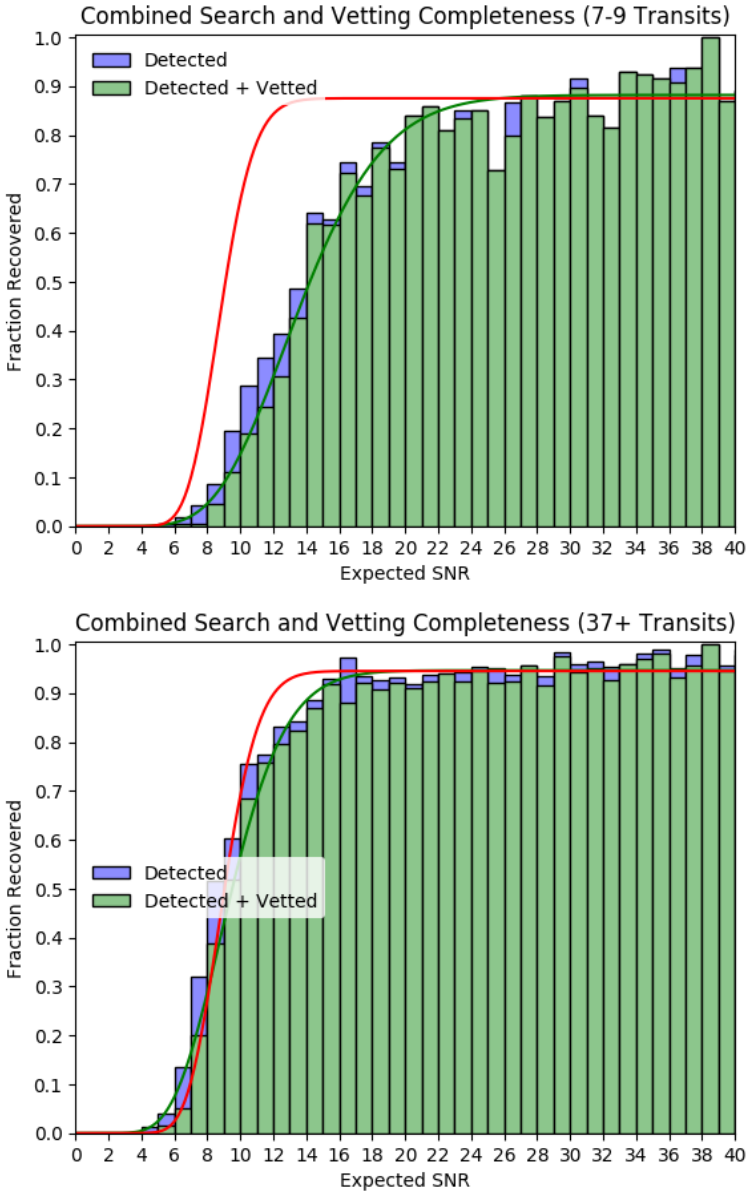

Figure 4.2: Combined search and vetting completeness of my pipeline, showing the fraction of injected transits recovered based only on the search (blue) and both search and vetting (green). A gamma cumulative distribution function (Eqn. 4.1 is fit to the combined recovery fraction (green line). These examples correspond to $N_{\rm tr} = 7 - 9$ (top) and $N_{\rm tr} \geq 37$ (bottom). For comparison, the Hsu et al. (2019) [65] best-fit gamma CDF fits as a function of expected MES statistic are also shown (red lines).

Table 4.4: Best-fit parameters for $P_{\rm det}$, the combined search and vetting model, with comparisons to the DR25 model results of Hsu et al. (2019) [65].

	This work			Hsu et al. (2019) [65]		
$N_{\rm tr}$	a	b	c	a	b	c
3	12.0239	1.3892	0.4653	33.3884	0.2645	0.6991
4	17.4744	0.9059	0.6651	32.8860	0.2696	0.7684
5	13.5488	1.0900	0.7704	31.5196	0.2827	0.8337
6	11.4812	1.2763	0.8369	30.9919	0.2870	0.8599
7-9	11.5413	1.2063	0.8817	30.1906	0.2947	0.8750
10-18	11.4538	1.0725	0.9118	31.6432	0.2794	0.8861
19-36	14.8651	0.7292	0.9164	32.6448	0.2689	0.8897
≥ 37	12.2332	0.7820	0.9465	27.8185	0.3243	0.9451

4.4 Occurrence Rate Methodology

4.4.1 Approximate Bayesian Computation

Bayesian inference is an increasingly popular approach of statistical inference on unknown parameters. Bayes' theorem is used to estimate the posterior probability distribution $P(\boldsymbol{\theta}|D)$ of a model with parameters $\boldsymbol{\theta}$ given the data D,

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}, \tag{4.8}$$

where $P(D|\boldsymbol{\theta})$ is the likelihood function, indicating the compatibility of the data given the model; $P(\boldsymbol{\theta})$ is the prior probability, representing initial beliefs toward the model; and $P(D)$ is a normalization constant. The best-fit model parameters can be estimated from $P(\boldsymbol{\theta}|D)$ such as by finding the posterior mode (most probable values of $\boldsymbol{\theta}$) or posterior median (50th percentile), with credible intervals representing our uncertainty about the model parameters.

For simple models, the likelihood function can typically be derived analytically. However, for more complex models, the likelihood may be unknown or too computationally expensive to evaluate. It is in these cases that the

"likelihood-free" method of ABC steps in as an effective and rigorous way of performing an approximate Bayesian analysis.

ABC circumvents the need for a likelihood function by using our prior information along with an ability to simulate, or "forward model," the observed data under investigation. By simulating a large number of datasets and quantifying the "distance" between each data set and the observed data set, the distribution of model parameters that provides the best matches can be determined. This distribution serves as an approximation to the posterior probability distribution.

4.4.2 Population Monte Carlo ABC

The specific form of ABC used here is the Population Monte Carlo (ABC-PMC) algorithm proposed by Beaumont et al. (2009) [9], wherein multiple generations of simulated data are created and an adaptive importance sampling scheme is used to evolve the ABC posterior. I use the ABC-PMC algorithm implemented in `cosmoabc`, a Python ABC Sampler [69], which is summarized here.

To initialize the ABC-PMC algorithm, a set of M values are drawn from the prior distribution, called "particles," $\{\boldsymbol{\theta}^i\}$ with $i \in [1, M]$. M is chosen to be much larger than N, the number of samples needed to characterize the prior. For each particle, a simulated dataset D_S^i is generated, a distance function ρ is used to calculate the distance between the simulated and real dataset, $\rho^i = \rho(D, D_S^i)$. From the whole set of M particles, only the N particles with the smallest ρ^i are kept. These constitute the zeroth "generation" ($S_{t=0}$), and the 75% quantile of all $\rho \in S_{t=0}$ gives the distance threshold for the next iteration ($\epsilon_{t=1}$). Each particle is assigned an equal weight, $W_{t=0}^j = 1/N$, for $j \in [1, N]$.

An importance sampling technique is used to produce subsequent generations ($t > 0$). A trial particle ($\boldsymbol{\theta}_{\rm try}$) is drawn from the previous generation S_{t-1} with weights W_{t-1}, and used to simulate a catalogue and find its associated distance, $\rho_{\rm try}$. $\boldsymbol{\theta}_{\rm try}$ is stored to the current generation S_t if $\rho_{\rm try} \leq \epsilon_t$. This process is repeated until S_t is filled with N accepted particles. The

weights of each particle are then calculated as

$$W_t^j = \frac{P(\boldsymbol{\theta}_t^j)}{\sum_{i=1}^{N} W_{t-1}^i N(\boldsymbol{\theta}_t^j; \boldsymbol{\theta}_{t-1}^i, C_{t-1})} \tag{4.9}$$

where $P(\boldsymbol{\theta}_t^j)$ is the prior probability distribution calculated at $\boldsymbol{\theta}_t^j$, and $N(\boldsymbol{\theta}_t^j; \boldsymbol{\theta}_{t-1}^i, C_{t-1})$ represents a Gaussian probability density function (PDF) centred at $\boldsymbol{\theta}_{t-1}^i$ with covariance matrix built from S_{t-1} and calculated at $\boldsymbol{\theta}^j$.

Following the determination of the new weights, the algorithm repeatedly produces new generations until subsequent iterations no longer significantly change the ABC posterior. In `cosmoabc`, this convergence occurs when the number of draws necessary to construct a generation is much larger than N.

4.5 ABC Applied to Exoplanet Occurrence Rates

Planet surveys have a variety of complexities that make the determination of the correct likelihood impractical, such as the existence of selection effects that are pipeline-dependent, the choice of targets, and the measurement uncertainties in the planet properties. Thus, ABC is well suited to the inference of occurrence rates based on *Kepler* planet catalogues and my independent catalogue outlined in Chapter 3.

As discussed in §4.4, ABC depends on the following elements:

- A prior probability distribution over the model parameters,
- A forward model, to simulate the data given the model parameters, and
- A distance function, to assess the agreement between the simulated data and the observed data.

Because I calculate occurrence rates over a 2D grid of orbital period and planet size in this work, the model parameters of interest are $f_{p,r}$, the average number of planets per star in period bin p and radius bin r. I assume that each $f_{p,r}$ is constant over the relevant range of periods and radii.

Meanwhile, the forward model must simulate the planet population around the considered stellar sample using each bin's guess occurrence rate and take into account selection effects and biases such as catalogue completeness and planet radius uncertainty to produce a simulated catalogue. The distance function must then compare the simulated catalogue to the actual observed catalogue to indicate which occurrence rates most closely describe the distribution.

4.5.1 Prior Probability

As in the baseline occurrence rates of Hsu et al. (2019) [65], I assign independent uniform priors for each occurrence rate over $[0, f_{\max,p,r})$. The upper limit for each bin is

$$f_{\max,p,r} = C \times \log_2\left(\frac{P_{\max,p}}{P_{\min,p}}\right) \times \log_2\left(\frac{R_{p,\max,r}}{R_{p,\min,r}}\right) \tag{4.10}$$

with $C = 2$, small enough that proposals with more than three planets per factor of 2 in period are rare [65]. This is consistent with expectations based on long-term orbital stability.

4.5.2 Forward Model

It is within the exoplanet population simulator that many of the complexities that make a likelihood function infeasible to compute are able to be incorporated into the determination of occurrence rates.

One such complexity is the existence of selection effects. To review the notation of §1.3.1, Youdin (2011) [162] outlines three main selection effects to be accounted for as part of robust exoplanet population analysis. These are quantified as detection efficiencies, η, which give the ratio of detections to actual planets: (i) $\eta_{\rm tr}$, the transit probability that the planet crosses our line of sight to the star; (ii) $\eta_{\rm rec}$, the efficiency at which the detection pipeline recovers the planet; and (iii) $\eta_{\rm fp} = 1/(1 - r_{\rm fp})$, where $r_{\rm fp}$ is the rate of false positive events that are detected as planets. The net detection efficiency of a given planet is found by multiplying all of the above efficiencies together.

For the baseline results, I assume the FP rate is low enough that it can be ignored for simplicity ($\eta_{\rm fp} = 1$). However, I discuss potential implications of this assumption in §4.6.5.

Importantly, these selection effects change on a per-star basis. For instance, the completeness model depends on both the physical properties of a star and the characteristics of its associated *Kepler* light curve. My forward model allows me to take these into account and find a specific completeness for a planet around a specific star, with little sacrifice of computational efficiency. By comparison, studies that have used likelihood functions in occurrence rate statistics have had to utilize star-averaged detection efficiencies that depend only on P and R_p, as incorporating information about individual stars would be too computationally expensive. In these cases, two planets with the same period and radius but host stars with vastly different properties would still be assigned the same completeness.

Furthermore, given that I am focused on specific period and radius bins, occurrence rates may be sensitive to the accuracy of a planet's membership in its correct bin. While orbital period is typically known to an accuracy of minutes or better, uncertainties in planet radius are significantly larger. First, measurement errors caused by fitting a transit model to a noisy light curve can cause the fitted ratio R_p/R_s to differ from its true value. Second, and more significantly, uncertainty in the star's radius used to derive R_p from R_p/R_s directly leads to uncertainty in the planet's radius, even if R_p/R_s is known exactly. As a result, a planet's "observed" radius bin may differ from its true radius bin, particularly if it is near the boundary between two bins. My forward model is able to take into account these measurement uncertainties by simulating both true and observed stellar and planetary radii, while most other studies assume that a planet's properties are known exactly.

Step 1: Generate Planets

I start by determining the number of planets to be simulated in my population. Given that the occurrence rate $f_{p,r}$ represents the average number

of planets per star in period bin p and radius bin r, and considering there are N_s stars in the sample, the number of planets in each bin can be drawn from a Poisson distribution with rate $\lambda = f_{p,r} N_s$.

Then, I assign each planet a star at random, and draw physical and orbital properties from model distributions. I draw each planet's precise orbital period and radius uniformly in log period and log radius, constrained to be within the assigned bin. I assume circular orbits ($e = 0$), and assume the orbital inclinations (i) are uniformly distributed across the sky, drawing from $\cos i \sim U(0, 1)$.

Note that I do not take into account correlations in planet properties in multiplanet systems, and only assign a star to each planet for the purpose of attaining a stellar radius, mass, and other relevant parameters. In other words, planets are drawn completely independently of one another. My assumption of circular orbits, while consistent with previous works, is also simplistic, and systems with a single transiting planet have been shown to have a different eccentricity distribution than systems with multiple planets (e.g. a mean of $e \approx 0.3$ compared to 0.04; [161]). However, these choices are primarily due to the computational expensiveness of running the ABC forward model, restricting the fitting to only a select number of bins at a time and thus preventing me from simulating full system architectures. Burke et al. (2015) [22] also showed that incorporating nonzero eccentricity (e.g. assuming all planets have $e = 0.4$) had only a modest impact on occurrence rates, comparable to statistical errors.

Step 2: Calculate Selection Effects

Transit Probability

Many planets will be undetected simply because they do not cross our line of sight to the star. I use the planet's semi-major axis $a = (GM_s P^2/4\pi^2)^{1/3}$ and inclination i drawn previously to determine the planet's impact parameter,

$$b = \frac{a \cos i}{R_s}, \tag{4.11}$$

requiring that $b \leq 1$. In other words, the planet transits if the centre of the planet passes inside the disk of the star. As in Hsu et al. (2019) [65], I ignore the small number of transiting planets with $b > 1$, as large impact parameters are often associated with grazing eclipsing binaries and these planets are likely to be flagged as FPs. Thus, I set

$$\eta_{\rm tr} = \begin{cases} 1 & b \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.12}$$

Recovery Efficiency

I estimate the recoverability of each planet by taking into account pipeline search completeness, vetting completeness, and the probability that at least three transits occur in the *Kepler* window.

For search and vetting completeness, I use the combined search and vetting model as outlined in §4.3. I follow the same process outlined in §4.3 to estimate each simulated planet's transit S/N and $N_{\rm tr}$ to determine the corresponding $P_{\rm det}$.

For the window probability, I use the binomial probability function described in Burke et al. (2015) [22]

$$\begin{aligned} P_{\rm win,\geq 3} = 1 - (1 - f_{\rm duty})^M - M f_{\rm duty}(1 - f_{\rm duty})^{M-1} \\ - \frac{M(M-1)}{2} f_{\rm duty}^2 (1 - f_{\rm duty})^{M-2} \end{aligned} \tag{4.13}$$

where $M = T_{\rm obs}/P$. Thus, I find the total recovery efficiency for a given planet as

$$\eta_{\rm rec} = P_{\rm det} P_{\rm win,\geq 3}. \tag{4.14}$$

Step 3: Simulate Detected Exoplanet Population

Following Hsu et al. (2019) [65], I determine if a planet is detected by drawing from a Bernoulli distribution with probability

$$\eta_{\rm tot} = \eta_{\rm tr}\eta_{\rm rec}. \tag{4.15}$$

At this point, I remove all planets flagged as undetected from the simulation, and the forward model focuses only on the recovered population.

Step 4: Incorporate Planet Radius Uncertainty

I cannot assume that once a planet is detected, I also recover its true radius exactly. I take into account measurement errors caused by fitting a transit model as well as uncertainty in a host star's radius following Hsu et al. (2019) [65].

First, I compute a planet's true planet-to-star radius ratio $k = R_p/R_s$ using the true planet and stellar radius, and draw an observed $k_{\rm obs}$ centred on k based on the transit's S/N and the diagonal noise model of Price & Rogers (2014) [123]. Then, I draw an observed stellar radius $R_{s,\rm obs}$ from two half-normal distributions, with median equal to the Berger et al. (2018) [11] radius and widths equal to the upper and lower radius uncertainties. Finally, I compute the observed planet radius as $R_{p,\rm obs} = k_{\rm obs}R_{s,\rm obs}$.

I place each simulated planet into a new radius bin depending on the results of this process. In doing so, I simulate an observed exoplanet population, to be compared with the catalogue produced from the *Kepler* search.

Step 5: Compare to Observed Population

I generate summary statistics for each bin in both observed and simulated catalogues,

$$s_k = \frac{N_k}{N_s}, \tag{4.16}$$

where N_k is the number of planets in the kth bin. I use the fraction of planets per star rather than the absolute number of planets so as to allow

for differences in the choice of N_s between catalogues. For instance, I could choose to run a quick inference by comparing my search results from all 96,280 FGK stars with a catalogue that simulates planets around only 10,000 FGK stars.

It is at this point that I apply a distance function to quantify the distance between summary statistics and thus assess the agreement between the simulated and observed planet catalogues.

4.5.3 Distance Function

When modeling only a single period-radius bin at a time, such as in Hsu et al. (2018) [64], the summary statistic for each catalogue is scalar. The choice of distance function may be simply

$$\rho(s_{\rm obs}, s_{\rm sim}) = (s_{\rm obs} - s_{\rm sim})^2, \tag{4.17}$$

where $s = N/N_s$ is calculated for only the single bin of interest, and obs and sim refer to the observed and simulated catalogues respectively.

When fitting multiple bins simultaneously (as done here; see next section), I use the distance suggestion of Hsu et al. (2019) [65],

$$\rho(s_{{\rm obs},k}, s_{{\rm sim},k}) = \sum_k \frac{|s_{{\rm obs},k} - s_{{\rm sim},k}|}{\sqrt{s_{{\rm obs},k} + s_{{\rm sim},k}}}, \tag{4.18}$$

inspired by the Canberra distance [83]. Hsu et al. (2019) [65] found that this distance allowed ABC to converge more rapidly than other tested functions. This function also weights the absolute value of the differences in s_k by the square root of the sum, resulting in a similar fractional error in occurrence rates for all bins rather than a similar absolute error.

4.5.4 Model Verification

With the ABC framework set, I can justify the number of bins to fit at once and verify that the algorithm is able to recover occurrence rates accurately with the appropriate choices.

Had I not incorporated planet radius uncertainty into the forward model, the placement of each simulated planet into a specific period-radius bin would be without ambiguity. The occurrence rates of each bin would not affect those of others, and thus fitting only one bin at a time would be an obvious choice due to computational efficiency.

Because my simulator takes into account measurement error, it may place a planet into a radius bin different from its true bin. If two neighbouring bins have different occurrence rates, the number of planets exchanged across the radius boundary may be asymmetric. Furthermore, the edge bins being fit will display noticeable bias. As the simulator does not simulate planets with true radii above the upper limit of the top bin and below the lower limit of the bottom bin, the exchange of planets over these radii limits will be strictly one sided, and their occurrence rates will be overestimated.

These considerations necessitate the fitting of multiple bins simultaneously. However, fitting more parameters comes at the cost of the performance of the ABC-PMC algorithm, as it becomes less likely that the proposed values for all parameters will result in good agreement between the observed and simulated catalogues. Both the width of the ABC posterior and the computational time required for the algorithm to achieve convergence will increase significantly.

To explore these issues, I used my forward model to simulate 10 "true" planet catalogues in the $6.25-12.5$ day period range using the full 96,280 star sample. I used occurrence rates of $f = \{0.06, 0.05, 0.04, 0.03, 0.02, 0.01, 0.005\}$ for the bins with boundaries $R_p = \{0.35, 0.5, 0.71, 1, 1.41,\ 2, 2.83, 4\}\ R_\oplus$. I fit a total of one, three, five, and seven bins simultaneously, centred on the $1.25-1.5R_\oplus$ bin ($f = 0.03$, or 3%), for each simulated catalogue. As inputs to `cosmoabc`, I set the number of particles for the initial generation at 500 and all subsequent generations at 200. I considered the system converged when at least 2000 draws (10 times the size of each generation) were required to construct the next generation.

Fig. 4.3 shows the final ABC posterior for the $6.25-12.5$ day, $1-1.41R_\oplus$ occurrence rate after each run. As expected, the one-bin fit consistently overestimates the true occurrence rate due to the fact that simulated plan-

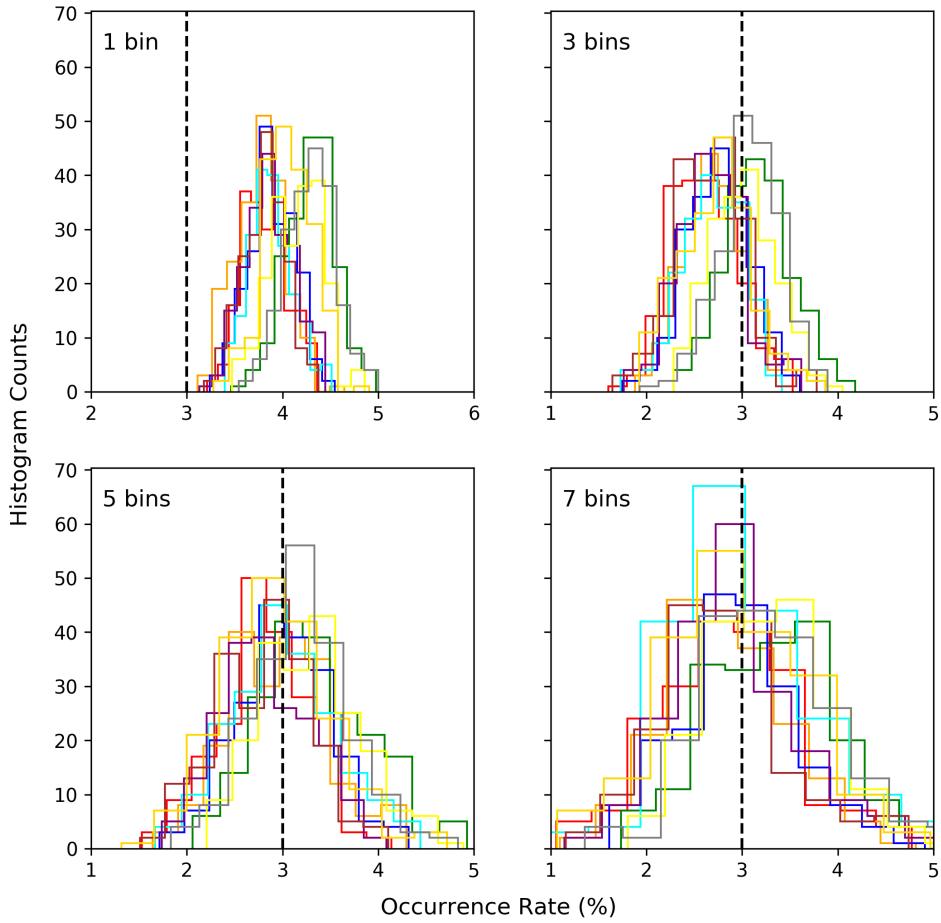

Figure 4.3: Results from testing how the recovery of the $6.25 - 12.5$ day, $1-1.414\ R_{\oplus}$ simulated occurrence rate ($f = 0.03$, or 3%) changes depending on the number of bins fit simultaneously. Each colour corresponds to 1 of 10 simulated planet catalogues.

ets can only leak out of the bin. The average absolute difference between the ABC posterior median and the true 3% occurrence rate was 0.98%. The three-, five-, and seven-bin fits all show significant improvement, with ABC posteriors well clustered around the true occurrence rate. Average absolute differences were 0.24%, 0.15%, and 0.17% respectively. Also note the expected widening of the ABC posterior with more bins.

When examining the results for the edge bins in each multibin-fit run, I confirmed that they tended to be overestimated compared to the interior bins. This was especially apparent for the bottom-edge bins, likely due to the fact that they were assigned the highest occurrence rates and thus had more outward leakage of planets than inward. These considerations prompted me to exclude the results for the two edge bins when performing multibin fits, and only report the results for the interior bins.

Overall, I agree with the conclusions of Hsu et al. (2019) [65] that five to seven radius bins are the optimal choice, and that one should be careful when considering the results of edge bins.

4.6 Occurrence Rate Results

My baseline exoplanet occurrence rates are defined using the combined FGK sample without reliability, as well as F, G, and K stars separately, using a period-radius grid with logarithmically spaced bin edges of $P = \{0.78, 1.63, 3.13, 6.25, 12.5, 25, 50, 100, 200, 400\}$ days and $R_p = \{0.5, 0.71, 1, 1.41, 2, 2.83, 4, 5.66, 8, 11.31, 16\}R_\oplus$. For the FGK sample, which is expected to be the best constrained due to having the largest number of bins populated with planets, I produce additional occurrence rates after taking into account the reliability of the pipeline. I also follow the recommendations of the Study Analysis Group (SAG) 13 of the NASA Exoplanet Exploration Program Analysis Group (ExoPAG)[18] and estimate F, G, and K occurrence rates on a grid with bin edges of $P = \{10, 20, 40, 80, 160, 320, 640\}$ days and $R_p = \{0.67, 1, 1.5, 2.25, 3.38, 5.06, 7.59, 11.39, 17.09\}R_\oplus$. These results are presented in Table B.1 of Appendix B.

I used my investigations of the multibin fits to determine the final setup for my methodology. Because my interest is in planets with radii down to 0.5 $R_\oplus$, the final results involve fitting additional 0.35 - $0.5R_\oplus$ bin for the sole purpose of acting as an edge bin to ensure accuracy for the 0.5 - $0.71R_\oplus$ bin. I am also interested in planets with radii up to 16

[18] https://exoplanets.nasa.gov/system/presentations/files/67_Belikov_SAG13_ExoPAG16_draft_v4.pdf

$R_\oplus$, but given that planets with $R_p > 16\ R_\oplus$ are rare, I do not expect the same bias to be present and keep my results for the 11.31 - 16 $R_\oplus$ bins as is. Therefore, I report the results using five-bin fits with radius boundaries $R_p = \{0.35, 0.5, 0.71, 1, 1.41, 2\}$, $\{1, 1.41, 2, 2.83, 4, 5.66\}$, and $\{2.83, 4, 5.66, 8, 11.31, 16\}$ $R_\oplus$ for each period range. Five-bin fits were chosen as a balance between minimizing edge-bin bias while avoiding the unnecessary broadening of the ABC posterior. After removing the edge bins (with the exception of the 11.31 - 16 $R_\oplus$ bin) from each subset, the entire $0.5 < R_p < 16\ R_\oplus$ radius range of interest is covered.

My final FGK, F, G, and K results are given in Table 4.5. I report the occurrence rate as the median of the ABC posterior for each $f_{p,r}$, with the difference between the median and 15.9th and 84.1th percentiles as the lower and upper uncertainties, respectively. For bins with zero detected planets, I report only the upper limit (84.1th percentile). I plot baseline FGK occurrence rates in Fig. 4.4, followed by occurrence rates for F-, G-, and K-type stars in Figs. 4.5, 4.6, and 4.7 respectively. Uncertainties are represented by the larger of the lower and upper uncertainties for each bin for easier readability. I set the colour scale to be the same in all four plots for more direct visual comparison.

Table 4.5: Occurrence rate results for FGK-, F-, G-, and K-type stars over the whole period-radius grid. Results are given in percent (i.e. 10^{-2}).

Period (days)	Radius ($R_\oplus$)	FGK (%)	F (%)	G (%)	K (%)
$0.78 - 1.56$	$0.5 - 0.71$	$0.02^{+0.02}_{-0.01}$	< 0.03	$0.03^{+0.04}_{-0.02}$	$0.1^{+0.09}_{-0.06}$
$0.78 - 1.56$	$0.71 - 1.0$	$0.1^{+0.04}_{-0.03}$	$0.02^{+0.02}_{-0.01}$	$0.14^{+0.07}_{-0.06}$	$0.3^{+0.17}_{-0.14}$
$0.78 - 1.56$	$1.0 - 1.41$	$0.18^{+0.05}_{-0.05}$	$0.04^{+0.03}_{-0.02}$	$0.23^{+0.1}_{-0.08}$	$0.61^{+0.23}_{-0.21}$
$0.78 - 1.56$	$1.41 - 2.0$	$0.12^{+0.04}_{-0.04}$	$0.03^{+0.03}_{-0.02}$	$0.13^{+0.07}_{-0.06}$	$0.4^{+0.19}_{-0.18}$
$0.78 - 1.56$	$2.0 - 2.83$	$0.01^{+0.01}_{-0.01}$	$0.02^{+0.02}_{-0.01}$	< 0.05	< 0.17
$0.78 - 1.56$	$2.83 - 4.0$	$0.01^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$	$0.02^{+0.02}_{-0.02}$	< 0.11
$0.78 - 1.56$	$4.0 - 5.66$	$0.01^{+0.01}_{-0.01}$	< 0.03	$0.02^{+0.03}_{-0.02}$	< 0.11
$0.78 - 1.56$	$5.66 - 8.0$	< 0.02	< 0.03	< 0.04	< 0.1

Period (days)	Radius ($R_\oplus$)	FGK (%)	F (%)	G (%)	K (%)
$0.78 - 1.56$	$8.0 - 11.31$	< 0.02	< 0.03	< 0.03	< 0.12
$0.78 - 1.56$	$11.31 - 16.0$	$0.02^{+0.02}_{-0.01}$	$0.02^{+0.02}_{-0.01}$	$0.02^{+0.03}_{-0.02}$	$0.08^{+0.08}_{-0.05}$
$1.56 - 3.13$	$0.5 - 0.71$	$0.1^{+0.06}_{-0.05}$	$0.03^{+0.04}_{-0.02}$	$0.17^{+0.12}_{-0.1}$	$0.22^{+0.22}_{-0.14}$
$1.56 - 3.13$	$0.71 - 1.0$	$0.25^{+0.09}_{-0.08}$	$0.08^{+0.07}_{-0.05}$	$0.33^{+0.16}_{-0.14}$	$0.66^{+0.35}_{-0.3}$
$1.56 - 3.13$	$1.0 - 1.41$	$0.46^{+0.09}_{-0.1}$	$0.25^{+0.11}_{-0.09}$	$0.54^{+0.19}_{-0.17}$	$1.04^{+0.41}_{-0.37}$
$1.56 - 3.13$	$1.41 - 2.0$	$0.52^{+0.1}_{-0.09}$	$0.14^{+0.08}_{-0.07}$	$0.78^{+0.2}_{-0.16}$	$1.2^{+0.4}_{-0.36}$
$1.56 - 3.13$	$2.0 - 2.83$	$0.1^{+0.05}_{-0.05}$	$0.04^{+0.04}_{-0.03}$	$0.13^{+0.1}_{-0.07}$	$0.35^{+0.25}_{-0.18}$
$1.56 - 3.13$	$2.83 - 4.0$	$0.05^{+0.03}_{-0.03}$	$0.03^{+0.03}_{-0.02}$	$0.07^{+0.06}_{-0.05}$	$0.18^{+0.15}_{-0.11}$
$1.56 - 3.13$	$4.0 - 5.66$	$0.02^{+0.02}_{-0.01}$	< 0.04	$0.05^{+0.05}_{-0.03}$	< 0.17
$1.56 - 3.13$	$5.66 - 8.0$	< 0.02	< 0.04	< 0.05	< 0.13
$1.56 - 3.13$	$8.0 - 11.31$	$0.01^{+0.02}_{-0.01}$	< 0.05	$0.03^{+0.03}_{-0.02}$	< 0.16
$1.56 - 3.13$	$11.31 - 16.0$	$0.09^{+0.04}_{-0.03}$	$0.06^{+0.05}_{-0.03}$	$0.16^{+0.09}_{-0.07}$	$0.12^{+0.11}_{-0.07}$
$3.13 - 6.25$	$0.5 - 0.71$	$0.34^{+0.18}_{-0.18}$	$0.11^{+0.1}_{-0.07}$	$0.49^{+0.33}_{-0.26}$	$0.72^{+0.51}_{-0.39}$
$3.13 - 6.25$	$0.71 - 1.0$	$0.71^{+0.21}_{-0.21}$	$0.27^{+0.16}_{-0.14}$	$1.04^{+0.38}_{-0.34}$	$1.19^{+0.67}_{-0.55}$
$3.13 - 6.25$	$1.0 - 1.41$	$1.37^{+0.23}_{-0.23}$	$0.86^{+0.26}_{-0.23}$	$1.58^{+0.47}_{-0.38}$	$2.67^{+0.92}_{-0.72}$
$3.13 - 6.25$	$1.41 - 2.0$	$1.79^{+0.27}_{-0.26}$	$0.54^{+0.2}_{-0.18}$	$2.56^{+0.46}_{-0.43}$	$4.36^{+1.0}_{-0.94}$
$3.13 - 6.25$	$2.0 - 2.83$	$0.99^{+0.18}_{-0.18}$	$0.4^{+0.18}_{-0.15}$	$1.14^{+0.35}_{-0.3}$	$2.8^{+0.81}_{-0.7}$
$3.13 - 6.25$	$2.83 - 4.0$	$0.27^{+0.11}_{-0.1}$	$0.19^{+0.13}_{-0.1}$	$0.3^{+0.21}_{-0.16}$	$0.59^{+0.35}_{-0.34}$
$3.13 - 6.25$	$4.0 - 5.66$	$0.18^{+0.09}_{-0.07}$	$0.13^{+0.1}_{-0.07}$	$0.23^{+0.13}_{-0.1}$	$0.29^{+0.25}_{-0.19}$
$3.13 - 6.25$	$5.66 - 8.0$	$0.08^{+0.06}_{-0.05}$	$0.06^{+0.05}_{-0.04}$	$0.1^{+0.09}_{-0.06}$	$0.24^{+0.23}_{-0.15}$
$3.13 - 6.25$	$8.0 - 11.31$	$0.09^{+0.06}_{-0.04}$	< 0.08	$0.18^{+0.11}_{-0.09}$	$0.22^{+0.19}_{-0.12}$
$3.13 - 6.25$	$11.31 - 16.0$	$0.19^{+0.08}_{-0.06}$	$0.13^{+0.08}_{-0.06}$	$0.34^{+0.15}_{-0.13}$	< 0.22
$6.25 - 12.5$	$0.5 - 0.71$	$0.18^{+0.17}_{-0.12}$	< 0.33	$0.39^{+0.37}_{-0.26}$	$0.47^{+0.49}_{-0.33}$
$6.25 - 12.5$	$0.71 - 1.0$	$0.51^{+0.26}_{-0.22}$	$0.2^{+0.18}_{-0.14}$	$0.57^{+0.4}_{-0.33}$	$1.38^{+0.77}_{-0.75}$
$6.25 - 12.5$	$1.0 - 1.41$	$2.34^{+0.39}_{-0.39}$	$1.25^{+0.4}_{-0.38}$	$3.12^{+0.68}_{-0.68}$	$4.26^{+1.36}_{-1.22}$
$6.25 - 12.5$	$1.41 - 2.0$	$2.66^{+0.4}_{-0.38}$	$1.58^{+0.49}_{-0.44}$	$3.44^{+0.68}_{-0.69}$	$4.95^{+1.45}_{-1.43}$
$6.25 - 12.5$	$2.0 - 2.83$	$3.37^{+0.43}_{-0.39}$	$1.08^{+0.39}_{-0.34}$	$4.12^{+0.69}_{-0.69}$	$9.25^{+1.65}_{-1.52}$
$6.25 - 12.5$	$2.83 - 4.0$	$0.61^{+0.22}_{-0.21}$	$0.49^{+0.26}_{-0.2}$	$1.05^{+0.42}_{-0.36}$	$0.41^{+0.45}_{-0.28}$
$6.25 - 12.5$	$4.0 - 5.66$	$0.17^{+0.1}_{-0.09}$	$0.08^{+0.08}_{-0.05}$	$0.34^{+0.22}_{-0.18}$	$0.42^{+0.36}_{-0.25}$
$6.25 - 12.5$	$5.66 - 8.0$	$0.07^{+0.06}_{-0.05}$	$0.09^{+0.08}_{-0.06}$	$0.12^{+0.13}_{-0.08}$	< 0.54
$6.25 - 12.5$	$8.0 - 11.31$	$0.16^{+0.09}_{-0.07}$	$0.1^{+0.1}_{-0.07}$	$0.16^{+0.15}_{-0.09}$	$0.46^{+0.36}_{-0.25}$

Period (days)	Radius ($R_\oplus$)	FGK (%)	F (%)	G (%)	K (%)
$6.25 - 12.5$	$11.31 - 16.0$	$0.18^{+0.08}_{-0.07}$	$0.2^{+0.12}_{-0.1}$	$0.26^{+0.16}_{-0.13}$	< 0.49
$12.5 - 25.0$	$0.5 - 0.71$	< 0.28	< 0.5	< 0.78	< 0.98
$12.5 - 25.0$	$0.71 - 1.0$	$0.26^{+0.22}_{-0.16}$	$0.19^{+0.29}_{-0.13}$	$0.66^{+0.51}_{-0.4}$	$0.94^{+0.86}_{-0.61}$
$12.5 - 25.0$	$1.0 - 1.41$	$1.59^{+0.44}_{-0.37}$	$1.0^{+0.46}_{-0.48}$	$1.15^{+0.72}_{-0.58}$	$3.84^{+1.68}_{-1.56}$
$12.5 - 25.0$	$1.41 - 2.0$	$2.41^{+0.59}_{-0.51}$	$1.08^{+0.51}_{-0.46}$	$3.18^{+0.99}_{-0.89}$	$5.52^{+2.07}_{-1.92}$
$12.5 - 25.0$	$2.0 - 2.83$	$6.56^{+0.69}_{-0.73}$	$2.82^{+0.63}_{-0.6}$	$8.43^{+1.16}_{-1.23}$	$15.39^{+2.57}_{-2.28}$
$12.5 - 25.0$	$2.83 - 4.0$	$2.13^{+0.46}_{-0.46}$	$1.27^{+0.49}_{-0.45}$	$2.89^{+0.95}_{-0.74}$	$2.87^{+1.54}_{-1.28}$
$12.5 - 25.0$	$4.0 - 5.66$	$0.19^{+0.17}_{-0.11}$	$0.26^{+0.22}_{-0.16}$	$0.23^{+0.28}_{-0.16}$	$0.76^{+0.7}_{-0.49}$
$12.5 - 25.0$	$5.66 - 8.0$	$0.24^{+0.13}_{-0.11}$	$0.1^{+0.11}_{-0.06}$	$0.34^{+0.23}_{-0.19}$	$0.81^{+0.54}_{-0.47}$
$12.5 - 25.0$	$8.0 - 11.31$	$0.06^{+0.07}_{-0.04}$	< 0.2	$0.16^{+0.18}_{-0.11}$	< 0.63
$12.5 - 25.0$	$11.31 - 16.0$	$0.18^{+0.1}_{-0.09}$	$0.17^{+0.13}_{-0.09}$	$0.26^{+0.19}_{-0.15}$	$0.45^{+0.41}_{-0.27}$
$25.0 - 50.0$	$0.5 - 0.71$	< 0.57	< 0.93	< 1.25	< 2.49
$25.0 - 50.0$	$0.71 - 1.0$	$0.25^{+0.26}_{-0.16}$	< 0.52	$0.34^{+0.42}_{-0.24}$	$1.37^{+1.28}_{-0.94}$
$25.0 - 50.0$	$1.0 - 1.41$	$0.43^{+0.38}_{-0.25}$	$0.34^{+0.35}_{-0.22}$	$0.41^{+0.53}_{-0.29}$	$2.34^{+1.86}_{-1.37}$
$25.0 - 50.0$	$1.41 - 2.0$	$2.19^{+0.64}_{-0.55}$	$0.73^{+0.54}_{-0.4}$	$2.47^{+1.1}_{-0.98}$	$6.48^{+2.69}_{-2.36}$
$25.0 - 50.0$	$2.0 - 2.83$	$7.57^{+0.98}_{-0.88}$	$4.25^{+0.91}_{-0.84}$	$10.48^{+1.82}_{-1.82}$	$12.55^{+3.1}_{-3.01}$
$25.0 - 50.0$	$2.83 - 4.0$	$3.45^{+0.65}_{-0.63}$	$1.4^{+0.62}_{-0.56}$	$5.94^{+1.51}_{-1.36}$	$3.31^{+2.04}_{-1.53}$
$25.0 - 50.0$	$4.0 - 5.66$	$0.26^{+0.22}_{-0.17}$	$0.52^{+0.38}_{-0.29}$	$0.39^{+0.42}_{-0.26}$	$0.74^{+0.82}_{-0.53}$
$25.0 - 50.0$	$5.66 - 8.0$	$0.27^{+0.16}_{-0.13}$	$0.26^{+0.24}_{-0.16}$	$0.22^{+0.25}_{-0.15}$	$1.02^{+0.86}_{-0.64}$
$25.0 - 50.0$	$8.0 - 11.31$	$0.2^{+0.14}_{-0.12}$	< 0.32	$0.34^{+0.31}_{-0.21}$	$1.01^{+1.18}_{-0.66}$
$25.0 - 50.0$	$11.31 - 16.0$	$0.3^{+0.18}_{-0.13}$	$0.27^{+0.21}_{-0.15}$	$0.49^{+0.35}_{-0.24}$	< 1.21
$50.0 - 100.0$	$0.5 - 0.71$	< 1.11	< 2.93	< 2.87	< 5.3
$50.0 - 100.0$	$0.71 - 1.0$	$0.28^{+0.33}_{-0.2}$	< 1.05	< 1.32	$1.39^{+1.66}_{-0.95}$
$50.0 - 100.0$	$1.0 - 1.41$	$0.33^{+0.36}_{-0.23}$	$0.4^{+0.45}_{-0.27}$	$0.57^{+0.67}_{-0.4}$	$1.61^{+1.97}_{-1.12}$
$50.0 - 100.0$	$1.41 - 2.0$	$1.5^{+0.72}_{-0.64}$	$0.6^{+0.51}_{-0.39}$	$1.2^{+0.89}_{-0.77}$	$6.45^{+3.34}_{-2.61}$
$50.0 - 100.0$	$2.0 - 2.83$	$6.52^{+1.18}_{-1.03}$	$2.96^{+1.06}_{-0.89}$	$10.2^{+1.94}_{-2.04}$	$10.09^{+4.83}_{-3.36}$
$50.0 - 100.0$	$2.83 - 4.0$	$3.62^{+0.89}_{-0.81}$	$2.33^{+1.08}_{-0.84}$	$4.56^{+1.65}_{-1.55}$	$5.28^{+2.94}_{-2.45}$
$50.0 - 100.0$	$4.0 - 5.66$	$0.92^{+0.49}_{-0.41}$	$1.04^{+0.68}_{-0.51}$	$1.47^{+0.9}_{-0.73}$	< 2.79
$50.0 - 100.0$	$5.66 - 8.0$	$0.53^{+0.33}_{-0.26}$	$0.56^{+0.54}_{-0.32}$	$0.67^{+0.55}_{-0.41}$	$1.18^{+1.28}_{-0.79}$
$50.0 - 100.0$	$8.0 - 11.31$	$0.58^{+0.33}_{-0.26}$	$0.54^{+0.41}_{-0.3}$	$0.53^{+0.48}_{-0.33}$	$1.5^{+1.23}_{-0.91}$
$50.0 - 100.0$	$11.31 - 16.0$	$0.37^{+0.26}_{-0.21}$	$0.46^{+0.36}_{-0.25}$	$0.58^{+0.49}_{-0.33}$	< 2.12

Period (days)	Radius ($R_\oplus$)	FGK (%)	F (%)	G (%)	K (%)
$100.0 - 200.0$	$0.5 - 0.71$	< 3.46	< 9.11	< 8.1	< 10.23
$100.0 - 200.0$	$0.71 - 1.0$	< 1.2	< 2.19	< 2.72	< 4.38
$100.0 - 200.0$	$1.0 - 1.41$	$0.41^{+0.48}_{-0.28}$	< 1.25	< 1.79	$2.65^{+2.55}_{-1.71}$
$100.0 - 200.0$	$1.41 - 2.0$	$1.04^{+0.71}_{-0.61}$	$0.62^{+0.79}_{-0.43}$	$1.48^{+1.22}_{-0.88}$	$4.07^{+3.46}_{-2.48}$
$100.0 - 200.0$	$2.0 - 2.83$	$6.01^{+1.32}_{-1.23}$	$2.96^{+1.56}_{-1.2}$	$7.23^{+2.67}_{-2.14}$	$13.12^{+5.26}_{-4.89}$
$100.0 - 200.0$	$2.83 - 4.0$	$2.82^{+1.07}_{-0.97}$	$1.21^{+1.1}_{-0.71}$	$4.83^{+2.16}_{-1.92}$	$3.96^{+2.73}_{-2.4}$
$100.0 - 200.0$	$4.0 - 5.66$	$0.67^{+0.56}_{-0.39}$	< 0.89	$1.79^{+1.35}_{-1.04}$	$1.7^{+2.0}_{-1.2}$
$100.0 - 200.0$	$5.66 - 8.0$	$0.86^{+0.51}_{-0.45}$	$0.63^{+0.54}_{-0.38}$	$1.68^{+1.11}_{-0.89}$	< 3.09
$100.0 - 200.0$	$8.0 - 11.31$	$1.55^{+0.69}_{-0.59}$	$0.71^{+0.8}_{-0.44}$	$2.98^{+1.49}_{-1.11}$	$1.64^{+1.78}_{-1.1}$
$100.0 - 200.0$	$11.31 - 16.0$	$0.81^{+0.47}_{-0.38}$	$1.08^{+0.72}_{-0.58}$	$0.88^{+0.83}_{-0.55}$	$2.25^{+1.94}_{-1.45}$
$200.0 - 400.0$	$0.5 - 0.71$	< 14.11	< 42.06	< 33.44	< 41.64
$200.0 - 400.0$	$0.71 - 1.0$	< 3.37	< 8.52	< 8.3	< 11.45
$200.0 - 400.0$	$1.0 - 1.41$	< 1.43	< 2.81	< 3.58	< 6.8
$200.0 - 400.0$	$1.41 - 2.0$	$0.57^{+0.68}_{-0.42}$	< 1.75	< 3.3	$3.74^{+3.88}_{-2.44}$
$200.0 - 400.0$	$2.0 - 2.83$	$4.78^{+1.74}_{-1.68}$	$1.43^{+1.42}_{-0.91}$	$8.48^{+4.08}_{-3.19}$	$5.19^{+4.97}_{-3.21}$
$200.0 - 400.0$	$2.83 - 4.0$	$2.21^{+1.41}_{-1.14}$	$1.6^{+1.63}_{-0.97}$	$1.95^{+2.24}_{-1.31}$	$10.01^{+6.86}_{-5.47}$
$200.0 - 400.0$	$4.0 - 5.66$	$1.31^{+0.95}_{-0.79}$	$1.05^{+1.22}_{-0.69}$	$2.42^{+2.08}_{-1.44}$	$3.64^{+3.77}_{-2.41}$
$200.0 - 400.0$	$5.66 - 8.0$	$2.75^{+1.24}_{-1.05}$	$1.48^{+1.3}_{-0.9}$	$5.63^{+2.56}_{-2.34}$	< 5.33
$200.0 - 400.0$	$8.0 - 11.31$	$1.09^{+0.87}_{-0.63}$	$0.99^{+1.1}_{-0.65}$	$1.79^{+1.59}_{-1.24}$	$3.9^{+3.98}_{-2.67}$
$200.0 - 400.0$	$11.31 - 16.0$	$1.66^{+0.95}_{-0.78}$	$1.08^{+1.14}_{-0.67}$	$2.02^{+1.79}_{-1.22}$	$6.52^{+5.68}_{-3.6}$

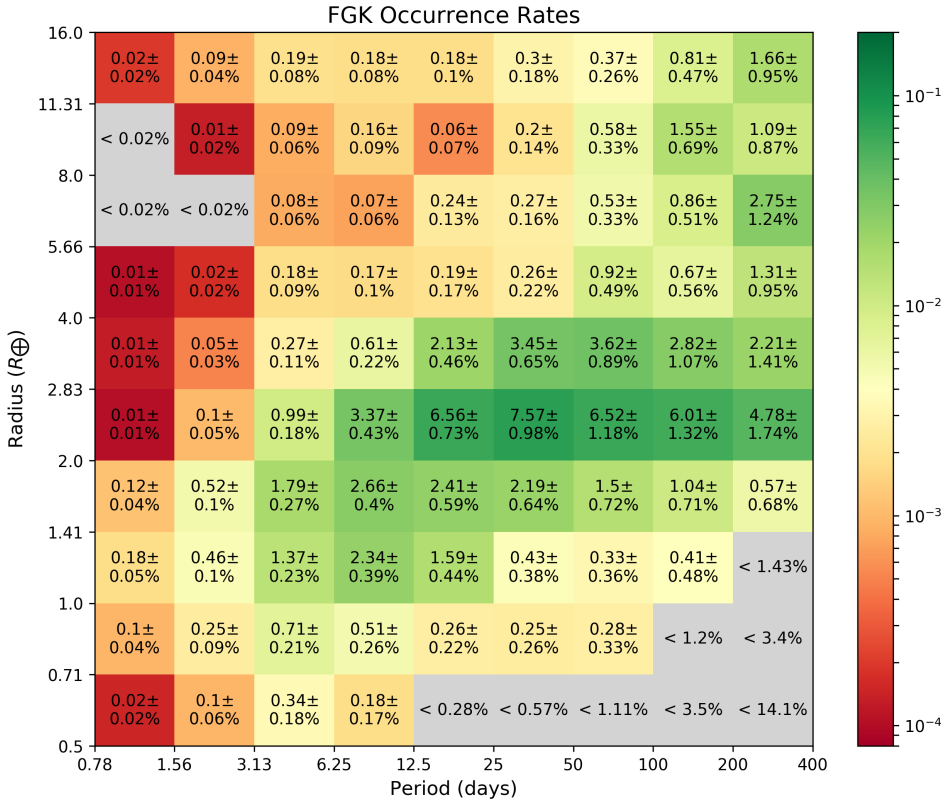

Figure 4.4: Occurrence rate estimates for FGK stars. The number of planets per star is given in percent (i.e. 10^{-2}) and as the median of the ABC posterior. Uncertainties are the larger of the lower and upper uncertainties, calculated as the difference between the median and 15.9th and 84.1th percentiles, respectively. Bins with no detected planets are in grey, with only the upper limit (84.1th percentile) shown.

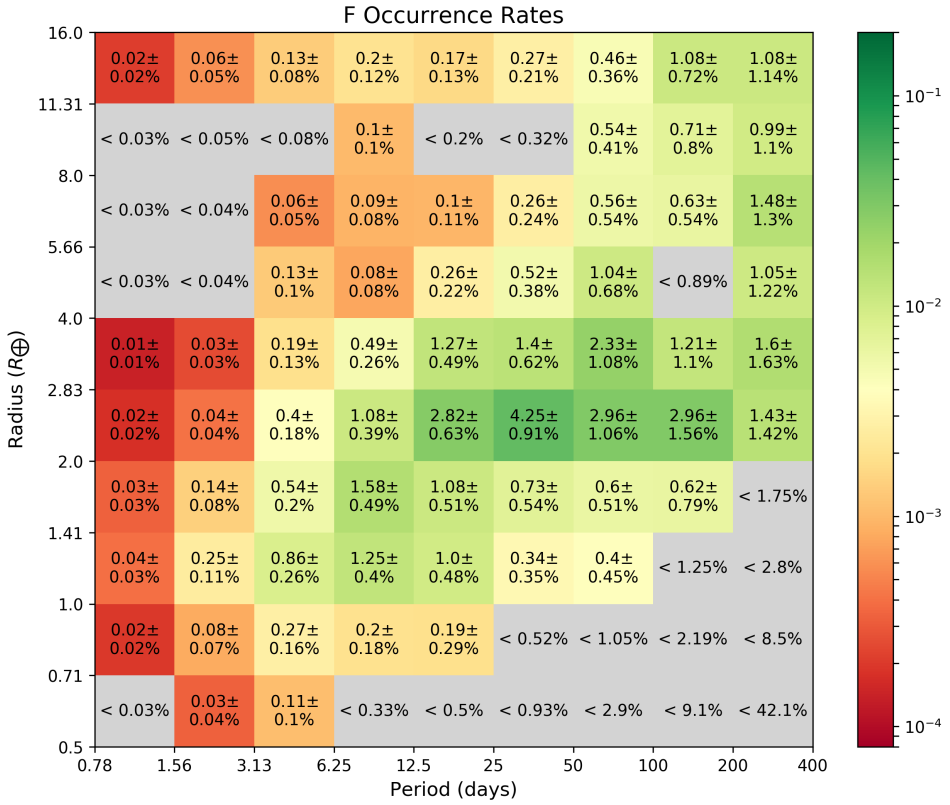

Figure 4.5: Same as Fig. 4.4, but for F-type stars only.

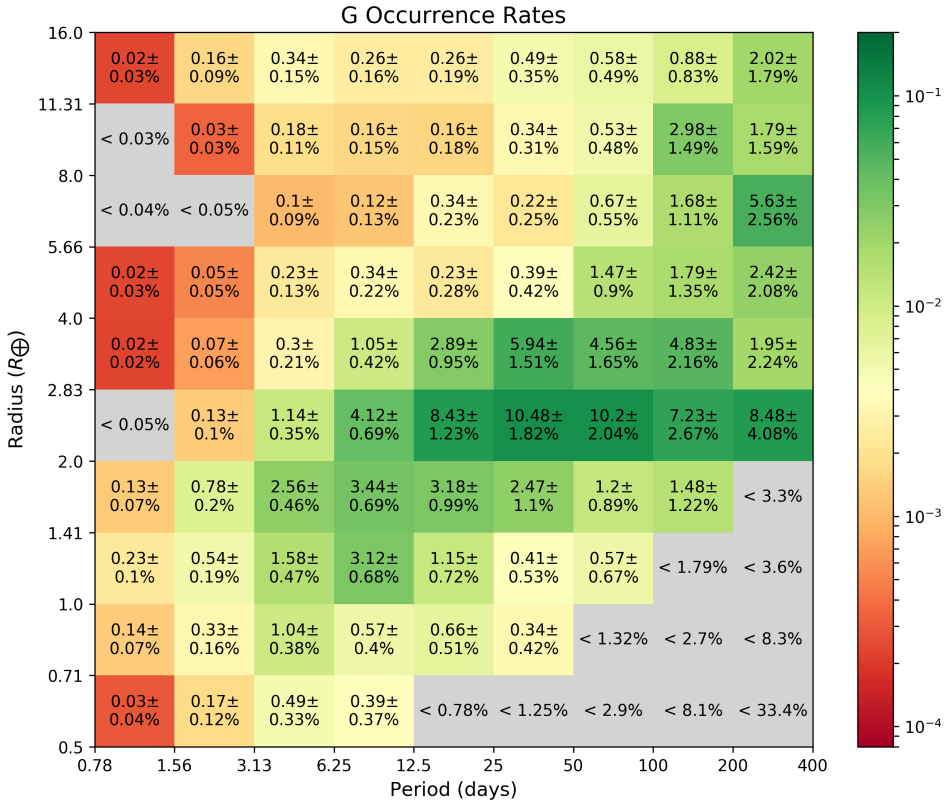

Figure 4.6: Same as Fig. 4.4, but for G-type stars only.

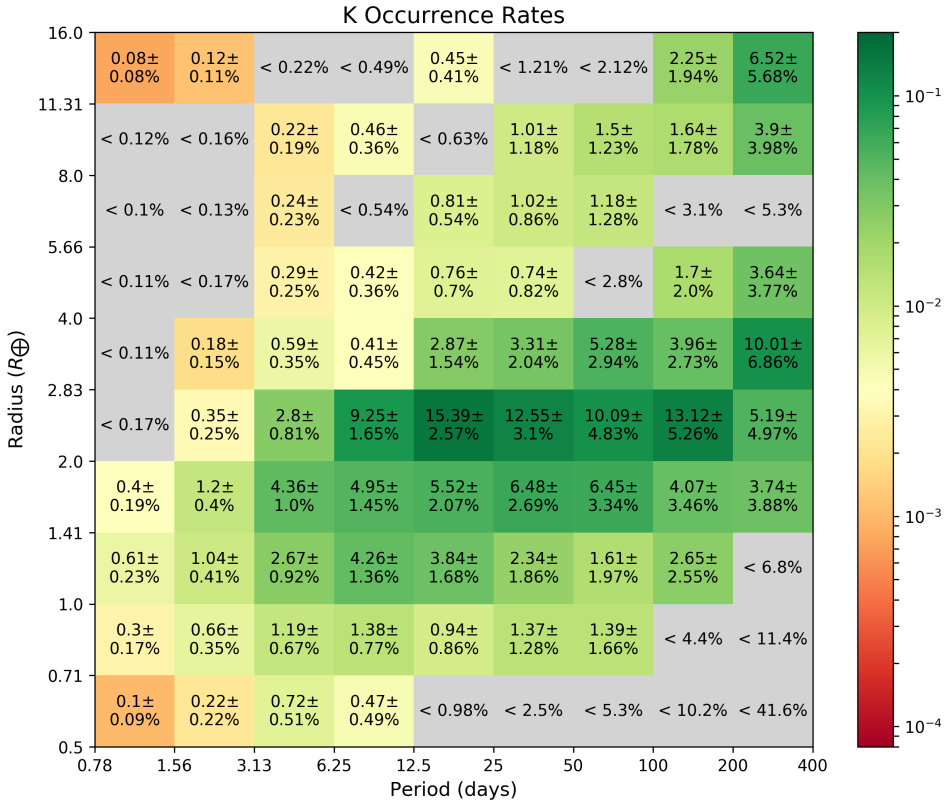

Figure 4.7: Same as Fig. 4.4, but for K-type stars only.

4.6.1 General Comparison to Previous Works

Here, I compare my occurrence rates with those reported by other studies, choosing Fulton et al. (2017) [52], Mulders et al. (2015a) [105], Petigura et al. (2013) [119], and Fressin et al. (2013) [51] as targets for comparison on the basis of having the most similar period and radius ranges. Given that studies calculate occurrence rates with different methods, account for completeness in different ways, use catalogues based on different amounts of available *Kepler* photometry, and more, direct comparison is difficult. Additionally, Petigura et al. (2013) [119] inferred occurrence rates using only the first planet found in each system. Consequently, their occurrence rates are not estimates of the average number of planets per star as in the other works. Recognizing these challenges, Fulton et al. (2017) [52] compared the *ratios* of occurrence rates between bins, which are less sensitive to these issues, rather than the absolute occurrence rates of individual bins. I adopt the same reasoning here.

Table 4.6 shows my FGK star results compared to Fressin et al. (2013) [51] (FGK), Mulders et al. (2015a) [105] (FGKM), and Fulton et al. (2017) [52] (FGK) while Table 4.7 shows my G star results compared to Petigura et al. (2013) [119] (GK) and Mulders et al. (2015a) [105] (G). I find that the main discrepancy is that the ratios of occurrence rates for $2-4\ R_\oplus$ planets to $1-2\ R_\oplus$ planets, across all period ranges and for both FGK- and G-type stars, are higher than all previous works. This is most noticeable when comparing to the older studies of Petigura et al. (2013) [119] and Fressin et al. (2013) [51]. Petigura et al. (2013) [119] found that $1-2\ R_\oplus$ planets within 50 days are 1.2 times more common than those with radii $2-4\ R_\oplus$. Fressin et al. (2013) [51] found a similar ratio of 1.1. Meanwhile, Mulders et al. (2015a) [105] found that planets with $1-2\ R_\oplus$ are less common, with ratios of 0.9 (FGKM stars) and 0.6 (GK). My results are even more favoured toward the larger radius bin, with ratios of 0.6 (FGK) and 0.4 (G).

Fulton et al. (2017) [52] also found a lower fraction of planets below $2\ R_\oplus$ than older works. In particular, they found a $P < 100$ day, $\{1.41, 2\}/\{2, 2.83\}\ R_\oplus$ ratio of 0.6, whereas Petigura et al. (2013) [119]

Table 4.6: Rough comparisons for FGK-type stars. My results are FGK occurrence rates marginalized over periods down to 0.78 days. F17 results are taken from Table 5 in Fulton et al. (2017) [52]. M15 results are taken from Table 6 in Mulders et al. (2015a) [105], summing bins down to 0.68 days. F13 results are taken from Table 3 in Fressin et al. (2013) [51], reported down to 0.8 days. For all results that involved summing multiple bins, I used the propagation of error to estimate uncertainty.

R_p ($R_\oplus$)	P (days)	**This work** (%)	F17 (%)	M15 (%)	F13 (%)
$1-2$	<50	$16.2^{+1.2}_{-1.1}$	-	16.3 ± 0.7	19.4 ± 2.0[1]
$2-2.8$	<50	$18.6^{+1.3}_{-1.2}$	19.4 ± 1.4	12.7 ± 0.5	-
$2-4$	<50	$25.2^{+1.5}_{-1.5}$	25.4 ± 1.6	18.7 ± 0.6	18.3 ± 1.3
$4-8$	<50	$1.6^{+0.4}_{-0.3}$	-	3.1 ± 0.2	-
$8-16$	<50	$1.6^{+0.3}_{-0.3}$	-	2.0 ± 0.2	-
$1-2$	<100	$18.1^{+1.4}_{-1.4}$	-	16.3 ± 0.7[2]	23.0 ± 2.4[1]
$1.4-2.8$	<100	$36.5^{+2.1}_{-2.0}$	43.1 ± 2.2	26.7 ± 0.8[2]	-
$2-4$	<100	$35.4^{+2.1}_{-2.1}$	36.6 ± 2.2	23.0 ± 0.8[2]	23.5 ± 1.6[2]
$4-8$	<100	$3.1^{+6.7}_{-0.6}$	-	4.4 ± 0.3[2]	-
$8-16$	<100	$2.6^{+0.5}_{-0.4}$	-	2.6 ± 0.2[2]	-

[1] $1.25-2$ $R_\oplus$
[2] $P < 85$ days

Table 4.7: Rough comparisons for G-type stars. My results are G occurrence rates marginalized over periods down to 6.25 days (lower bound chosen to match the lower bound of Petigura et al. (2013) [119] results). M15 results are taken from Table 7 in Mulders et al. (2015a) [105], summing bins down to 5.8 days. P13 results are taken from Fig. 2 in Petigura et al. (2013) [119], summing bins down to 6.25 days. For all results that involved summing multiple bins, I used the propagation of error to estimate uncertainty.

R_p ($R_\oplus$)	P (days)	**This work** (%)	M15 (%)	P13[1](%)
$1-2$	< 50	$14.1^{+2.0}_{-1.9}$	12.9 ± 0.9	19.2 ± 1.7
$2-4$	< 50	$33.2^{+3.0}_{-3.0}$	20.8 ± 1.0	16.4 ± 1.2
$4-8$	< 50	$1.8^{+0.6}_{-0.5}$	3.3 ± 0.4	1.5 ± 0.3
$8-16$	< 50	$1.8^{+0.5}_{-0.5}$	1.8 ± 0.3	1.0 ± 0.4
$1-2$	< 100	$16.0^{+2.2}_{-2.0}$	16.0 ± 1.4[2]	25.0 ± 2.3
$2-4$	< 100	$48.0^{+4.0}_{-3.8}$	25.2 ± 1.2[2]	24.1 ± 1.8
$4-8$	< 100	$4.1^{+1.2}_{-1.0}$	4.8 ± 0.6[2]	2.8 ± 0.7
$8-16$	< 100	$3.1^{+2.3}_{-1.8}$	2.5 ± 0.4[2]	1.6 ± 0.5

[1] Fraction of stars with planets instead of number of planets per star

[2] $P < 85$ days

found 1.3. They explained this difference using their knowledge of a gap in the radius distribution between 1.5 and 2 $R_\oplus$ and a peak near $\sim$2.5 $R_\oplus$, which were revealed in their study. Because these features were recovered in part due to their use of more precise stellar radii from spectroscopy, they suggested that the large ($\approx$40%) radius uncertainties from photometry alone would scatter planets with true sizes between 2 and 2.83 $R_\oplus$ to the $1.41-2$ $R_\oplus$ bin, both filling the gap and reducing the peak. Given that I used updated stellar radii from the Berger et al. (2018) [11] catalogue, which brought typical radius uncertainties down to $\approx$8% and found low ratio similar to Fulton et al. (2017) [52] (0.4), the same explanation likely applies here. Thus, in combination with the results of Fulton et al. (2017) [52], my results emphasize the sensitivity of planet occurrence rates to accurate stellar radii. I also argue that my results are more robust than previous works, given my use of updated stellar radii in combination with direct accounting for planet radius uncertainties.

In the following sections, I discuss further comparisons to previous works in the context of interesting and informative features previously uncovered from exoplanet population analysis and occurrence rate estimates.

4.6.2 Dependence on Stellar Effective Temperature

Marginalizing over the entire period-radius grid, I find occurrence rates of $0.89^{+0.23}_{-0.16}$ planets per F-type star, $1.67^{+0.21}_{-0.16}$ planets per G-type star, and $2.56^{+0.29}_{-0.24}$ per K-type star, compared to $1.06^{+0.09}_{-0.07}$ for the combined FGK sample.[19] For planets within 200 days (i.e. omitting the $200-400$ day bins which have low completeness and little to no planet detections below 2 $R_\oplus$), I find occurrence rates of $0.53^{+0.06}_{-0.05}$ (F), $1.17^{+0.08}_{-0.07}$ (G), $1.84^{+0.15}_{-0.13}$ (K), and $0.81^{+0.04}_{-0.04}$ (FGK), indicating a statistically significant trend of increasing overall planet occurrence toward cooler stars.

[19]The FGK rate is slightly lower than what would be found if one were to combine each F, G, and K occurrence rate (after weighting by stellar sample size). This is because the larger sample size of the FGK catalogue produces more constrained occurrence rates, especially in the low-completeness and low-planet-detection regime where only upper limits can be reported.

Fig. 4.8 shows the entire radius distribution marginalized over $P < 200$ days for each stellar type. I indicate any estimates that involved marginalizing over a bin with no planet detections with a downward pointing arrow, representing an upper limit. For small planets, my results are in strong agreement with Howard et al. (2012) [62] and Mulders et al. (2015a) [105] that occurrence rates increase substantially toward cooler stars. Overall, for $R_p = 1 - 2.83\ R_\oplus$ and $P < 200$ days, I find occurrence rates of $0.26^{+0.03}_{-0.02}$ (F), $0.67^{+0.05}_{-0.05}$ (G), and $1.20^{+0.11}_{-0.10}$ (K). In other words, small planets around K-type stars are about twice as abundant than around G-type stars, and five times as abundant than around F-type stars. I also agree with Mulders et al. (2015b) [106] that each distribution shows a clear drop-off in planets beyond 2.83 $R_\oplus$. Note that some measurements on both sides of the transition are only upper limits, but this is due to a lack of planets in the $0.78 - 1.56$ day

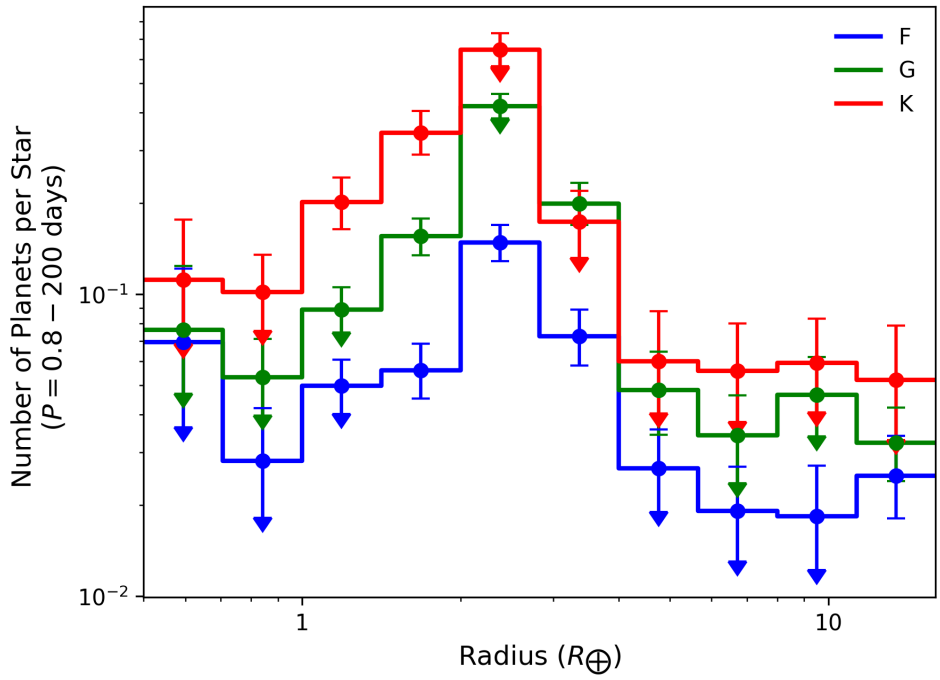

Figure 4.8: Occurrence rates for planets within 200 days as a function of radius for F-, G-, and K-type stars.

period range, which have upper limits of less than 0.002 planets per star. The existence of the drop-off is not dependent on their contribution.

For planets beyond 2.83 $R_\oplus$, I find that the trend becomes less clear. The $2.83 - 4$ $R_\oplus$ bin demonstrates no statistically significant difference between G and K occurrence rates, with a G-K (G minus K) difference of $0.02^{+0.05}_{-0.05}$ planets per star yet a K-F difference of $0.10^{+0.05}_{-0.04}$. Over the same radius bin, Mulders et al. (2015b) [106] also found that G and K occurrence rates were indistinguishable while still significantly more common than around F-type stars. At larger radii where the distributions flatten out for all stellar types, I do not attempt to interpret trends, as the majority of occurrence rate estimates involve summing bins without planet detections and thus only represent upper limits.

4.6.3 Dependence on Planet Radius

My FGK occurrence rate results marginalized over different period ranges are shown in Fig. 4.9.

First, I note the recovery of the "Neptune" or "sub-Jupiter desert," which is a dearth in Neptune- to sub-Jupiter-sized exoplanets in close-in orbits ($P \lesssim 3$ days) that has been noted and studied in many previous works (e.g. [97, 112, 142]). On the lower-mass end of the desert, several studies have emphasized the role of photoevaporation on the mass loss of highly irradiated planets, while the dearth at the higher-mass end likely requires a different explanation [68] such as tidal disruption that results from high-eccentricity migration [112]. In particular, I observe a significant decrease in occurrence rates for planets between 4 and 11.31 $R_\oplus$ and the shortest orbital periods ($0.78-3.13$ days), but not larger planets, in line with expectations. Note that I did not detect any planets in the $5.66-11.31$ $R_\oplus$, $0.78 - 1.56$ day range, nor in the $5.66 - 8.0$ $R_\oplus$, $1.56 - 3.13$ day bins, so the dip is likely even lower than indicated in the plot. Meanwhile, at higher orbital periods (especially beyond 6.3 days), the distribution is flat over the same radii.

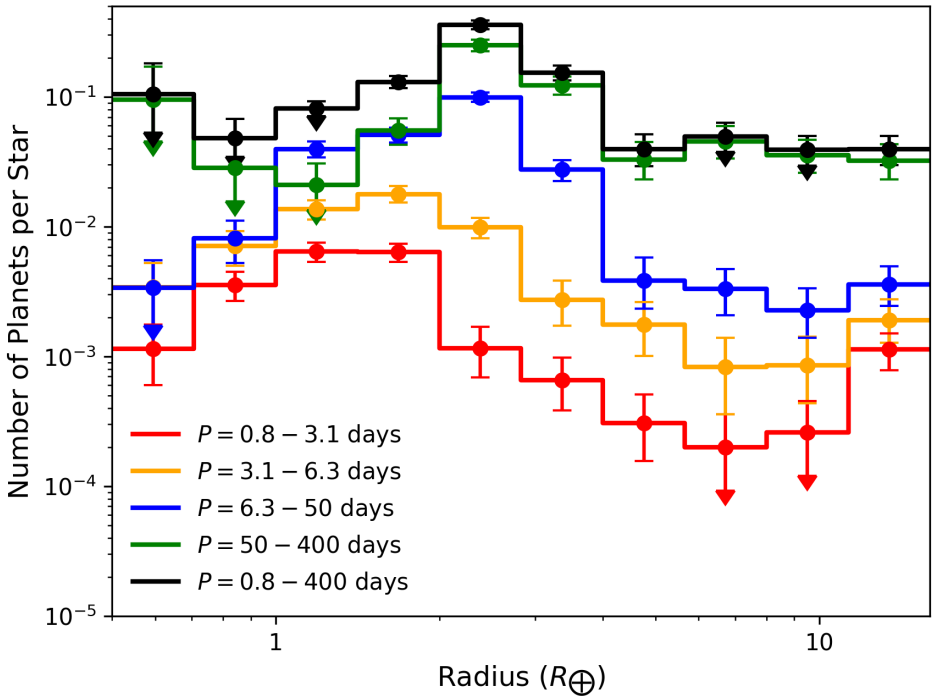

Figure 4.9: FGK occurrence rates as a function of radius, marginalized over different period ranges.

I now turn to smaller planets. An informative feature recently noted in exoplanet radius distributions is a gap between ~ 1.5 and 2 $R_\oplus$ for planets within $P < 100$ days, also known as the "radius valley" [52, 151]. The gap is accompanied by two peaks in radius, near $\sim$1.3 $R_\oplus$ (super-Earths) and $\sim$2.4 $R_\oplus$ (sub-Neptunes). This bimodal distribution had been predicted years earlier by numerical models involving the atmospheric erosion of highly irradiated low-mass planets [89, 113], while its statistical significance was not established by completeness-corrected observations until Fulton et al. (2017) [52], hereafter F17. F17 attributed the revelation of the radius gap to their use of precise stellar radius measurements from the California-Kepler Survey (CKS) [120].

My bin sizes were not small enough to be able to resolve this feature. This was a consequence of choosing logarithmically spaced bins large

enough to be appropriate for the entire grid down to 0.5 $R_\oplus$ and out to 400 days. Thus, I recomputed $0.78 - 100$ day occurrence rates using the same $P = \{0.78, 1.56, 3.13, 6.25, 12.5, 25, 50, 100\}$ days period bins, but much finer bins in radius space over the regime of interest: $R_p = \{1, 1.10, 1.21, 1.35, 1.49, 1.64, 1.81, 2, 2.21, 2.44, 2.69, 2.97, 3.28, 3.62, 4\}$ $R_\oplus$. The resulting distribution marginalized over the entire period range is compared with occurrence rates from Table 3 of F17, shown in the top panel of Fig. 4.10.

While I do find some evidence for a first peak near $\sim$1.4 $R_\oplus$, a minimum at $\sim$1.7 $R_\oplus$, and a second peak at $\sim$2.6 $R_\oplus$, differences between my distribution and that of F17 are obvious. Planets smaller than 1.5 $R_\oplus$ are considerably less abundant according to my sample, and I cannot confirm that the radius valley is statistically significant from these results alone. Meanwhile, the second peak is shifted to a higher radius than in F17, and it is twice as tall as the first peak rather than being comparable in height.

It should be noted that F17 only considered planets with hosts fainter than *Kepler* magnitude $Kp = 14.2$, given that the core sample of the CKS is magnitude limited, and the distribution of CKS planet radii above and below $Kp = 14.2$ is statistically different [52]. In order to see if this could explain the differences between our results, I recalculated my occurrence rates using only FGK stars with $Kp < 14.2$, which reduced the stellar sample from 96,280 to 30,688, and the $P < 100$ day planet population from 2377 to 833. This is shown in the bottom panel of Fig. 4.10. While interpreting the results is difficult given that the occurrence rates are less well constrained by the data, I still do not recover the shape of their radius valley.

Potential explanations include are our difference in occurrence rate methodology, where F17 used the simpler inverse detection efficiency method and did not take into account uncertainty in planet radius. They estimated that the underlying radius distribution after removing the smear due to this uncertainty would cause the gap to become slightly deeper, but the sub-Neptune peak would be increased (see Fig. 7 of F17). Furthermore, while the CKS sample represented a significant improvement in precise stellar radii over previous works, it did not yet incorporate *Gaia* DR2 parallaxes

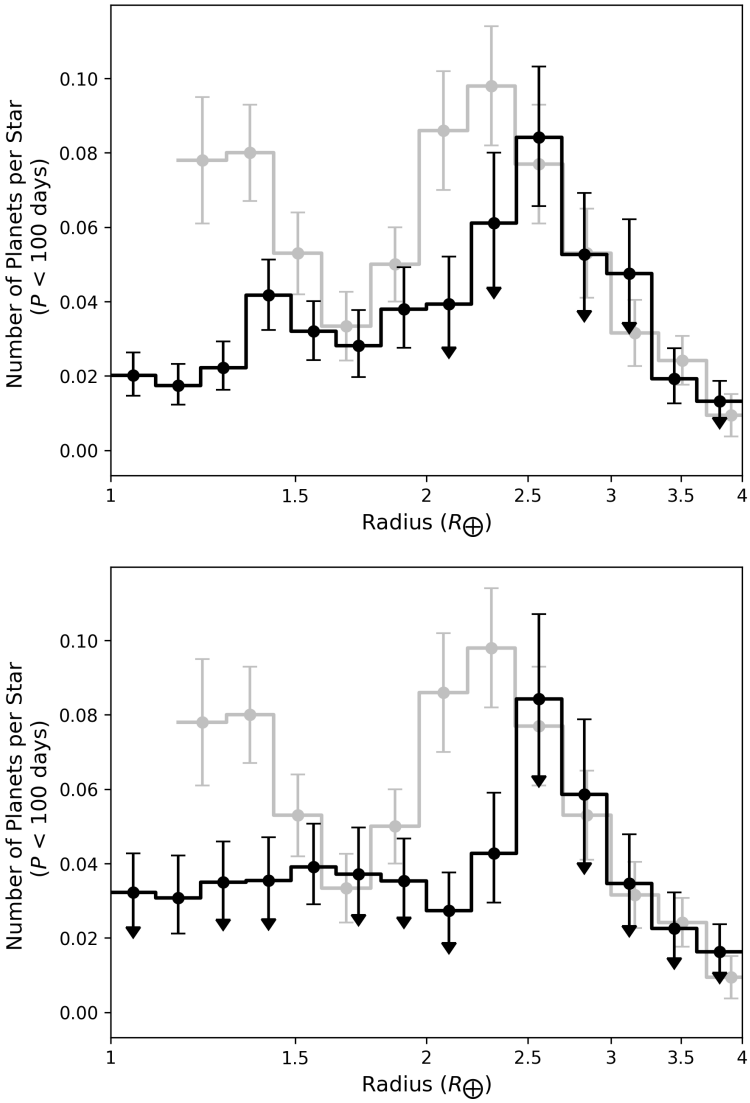

Figure 4.10: Top: recalculated $P < 100$ day occurrence rates using radius bins smaller than my baseline study in order to compare to the Fulton et al. (2017) [52] radius valley. The occurrence rates from Fulton et al. (2017) [52] are shown in light grey down to 1.16 $R_\oplus$, beyond which results were not reported due to low completeness. Bottom: the same as above, but after applying a cut of $Kp < 14.2$ to the FGK sample. The latter results are poorly constrained due to the significantly smaller stellar and planet sample.

like the Berger et al. (2018) [11] catalogue used here. Berger et al. (2018) [11] compared histograms of planet radii (uncorrected for biases) using their catalogue and CKS-derived radii. Even under the same cuts as utilized by F17, they found a higher number of sub-Neptunes in their sample and a radius valley shifted further to the right (see Fig. 8 of Berger et al. (2018) [11]).

I continue my analysis with these caveats and maintain my focus on overall planet occurrence rates for FGK stars (using the full 96,280 star sample, without the magnitude cut). An advantage over the Fulton et al. (2017) [52] paper is that I have the means to investigate the radius valley further, as a function of period, due to finding separate occurrence rates over specific period bins. These results are shown in Fig. 4.11.

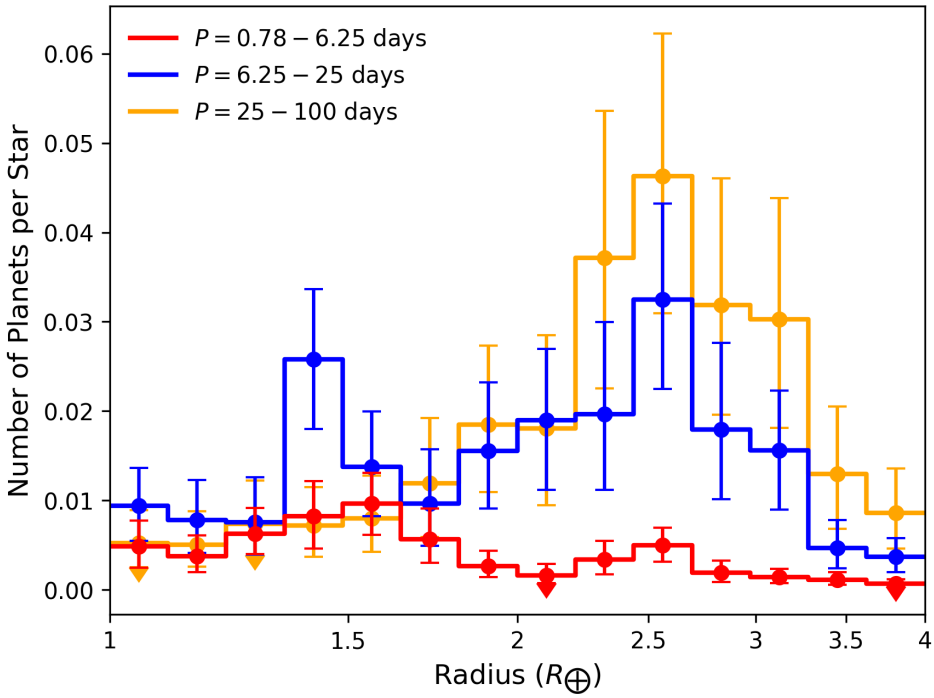

Figure 4.11: My recalculated $P < 100$ day occurrence rates split over smaller period ranges to compare to the Owen & Wu (2017) [114] evolutionary model results.

Owen & Wu (2017) [114] revisited the radius valley following the results of F17, and developed an analytical model to demonstrate that photoevaporation should separate planets into bare cores ($\sim$1.3 $R_\oplus$) and those with double the core's radius ($\sim$2.6 $R_\oplus$). Starting with a primordial *Kepler* planet population and evolving the population under the effects of cooling contraction and mass loss by evaporation, Owen & Wu (2017) [114] produced a prediction for the final radius distribution across different period ranges. Within $P < 10$ days, they found that super-Earths dominate, with only a small peak past 2 $R_\oplus$. This is qualitatively similar to my $P < 6.25$ day occurrence rates. Not shown are my results only within $P < 3.13$ days, for which I no longer find any sub-Neptune peak. The $10-20$ day bins in Owen & Wu (2017) [114] demonstrated the most significant radius valley, with peaks at $\sim$1.3 $R_\oplus$ and $\sim$2.6 $R_\oplus$ exhibiting similar heights. Importantly, I clearly recover a similar bimodal distribution with strong peaks near $\sim$1.3 $R_\oplus$ and $\sim$2.6 $R_\oplus$ over roughly the same periods ($6.25-25$ days). Lastly, Owen & Wu (2017) [114] found that the distribution became dominated by sub-Neptunes at larger orbital periods, with only a small super-Earth peak over $20-40$ days and no such peak over $40-100$ days. I find that the super-Earth peak disappears earlier, with no such peak over $25-100$ days. Nevertheless, my occurrence rates provide strong observational evidence in support of the Owen & Wu (2017) [114] model and can inform future studies on theoretical explanations for the radius valley.

4.6.4 Dependence on Orbital Period

My FGK occurrence rate results marginalized over different radius ranges are shown in Fig. 4.12.

The $1-2$ $R_\oplus$ and $2-4$ $R_\oplus$ distributions show clear increases in $df/d\log P$ with period up to a transition period P_0, followed by a decreasing trend for $1-2$ $R_\oplus$ and a flatter trend for $2-4$ $R_\oplus$. P_0 could indicate an important orbital distance down to which migration deposits planets, or an orbital distance at which most planets form, and differences between these distributions could indicate such mechanisms depend on planet size. I fit

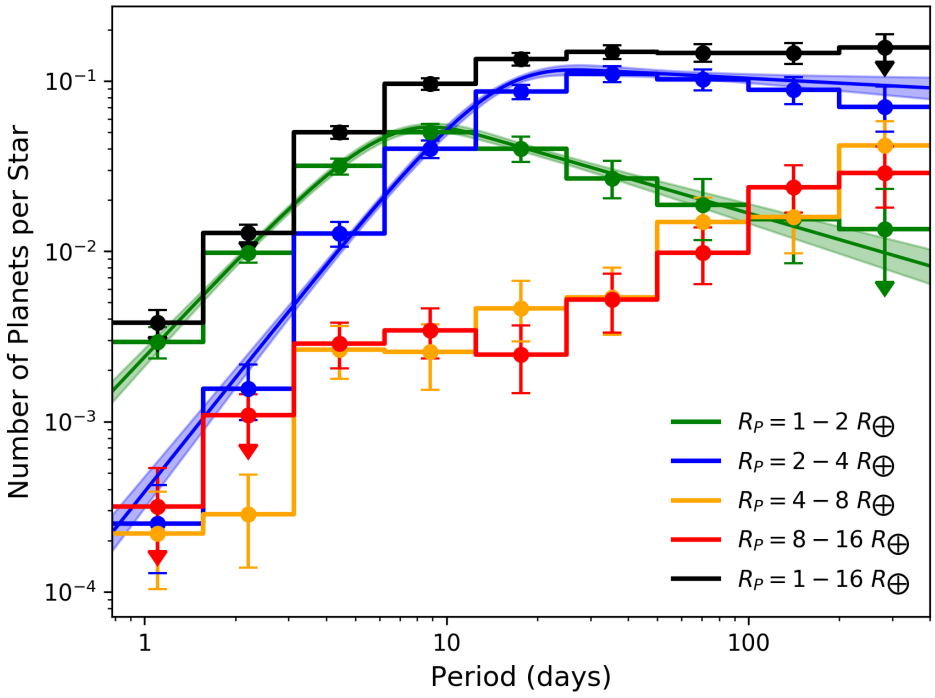

Figure 4.12: FGK occurrence rates as a function of period, marginalized over different radius ranges. Also shown are fits of Eqn. 4.19 to the $1-2$ $R_\oplus$ and $2-4$ $R_\oplus$ distributions.

the function

$$F = \frac{df}{d\log P} = CP^{\beta}\left(1 - e^{-(P/P_0)^{\gamma}}\right) \tag{4.19}$$

from Howard et al. (2012) [62] to these distributions using the extension of their maximum-likelihood method outlined in Petigura et al. (2018) [121]. For the ith bin, the log-likelihood of the model is

$$\ln L_i = n_{\mathrm{pl},i} \ln F\Delta\boldsymbol{x} + n_{\mathrm{nd},i} \ln(1 - F\Delta\boldsymbol{x}) \tag{4.20}$$

where $\Delta\boldsymbol{x} = \Delta\log P \Delta\log R_p$ is the size of the bin, $n_{\mathrm{pl},i}$ is the number of planets detected in the bin, and $n_{\mathrm{nd},i} = n_{\mathrm{pl},i}/f_i - n_{\mathrm{pl},i}$ is the effective number of nondetections as estimated using the bin's occurrence rate f_i.

The maximum-likelihood solution is obtained by maximizing the combined log-likelihood over all bins:

$$\ln L = \sum_{i=1}^{n_{\rm bin}} L_i. \tag{4.21}$$

I used MCMC sampling with `emcee` in Python [48] to explore the parameter space. The median and 68.3% credible interval for each distribution is shown in Eqn. 4.19, with associated parameters in Table 4.8.

Table 4.8: Median and 68.3% credible interval parameters for Eqn. 4.19, describing the shape of the occurrence rate distribution with orbital period over different size ranges.

R_p ($R_\oplus$)	C	β	γ	P_0 (days)
$1-2$	$0.36^{+0.07}_{-0.10}$	$-0.5^{+0.1}_{-0.1}$	$2.42^{+0.1}_{-0.1}$	$5.9^{+0.5}_{-0.5}$
$2-4$	$0.33^{+0.08}_{-0.12}$	$-0.1^{+0.1}_{-0.1}$	$2.3^{+0.1}_{-0.1}$	$13.3^{+1.4}_{-1.5}$

This function simplifies to two power laws far from P_0:

$$\frac{df}{d\log P} \propto \begin{cases} P^\alpha & \text{if } P \ll P_0, \text{ where } \alpha = \beta + \gamma \\ P^\beta & \text{if } P \gg P_0. \end{cases} \tag{4.22}$$

The $1-2$ $R_\oplus$ planet occurrence rate rises with P, with $df \propto P^\alpha d\log P$ where $\alpha = 1.9^{+0.1}_{-0.1}$, up to a transition period $P_0 = 5.9^{+0.5}_{-0.5}$ days. Beyond this, $df \propto P^\beta d\log P$ where $\beta = -0.5^{+0.1}_{-0.1}$. Comparatively, Petigura et al. (2018) [121] looked at $1-1.7$ $R_\oplus$ bins and found a slightly higher initial increase ($\alpha = 2.4^{+0.4}_{-0.3}$), a slightly higher transition period ($P_0 = 6.5^{+1.6}_{-1.2}$ days), and a slightly shallower decrease at longer orbital periods ($\beta = -0.3^{+0.2}_{-0.2}$), though the 68.3% credible intervals of the latter two parameters overlap with ours. Meanwhile, the long-period distribution found by Dong & Zhu (2013) [42] was flat, with $\beta = -0.10 \pm 0.12$.

The transition for $2-4$ $R_\oplus$ occurrence rates occurs farther out, at $13.3^{+1.4}_{-1.5}$ days, with a similar rapidly rising distribution at shorter orbital periods ($\alpha = 2.2^{+0.1}_{-0.1}$) and a nearly flat distribution at longer orbital periods ($\beta = -0.1^{+0.1}_{-0.1}$). These are consistent with a $P_0 = 11.9^{+1.7}_{-1.5}$ day transition

period, $\alpha = 2.3^{+0.2}_{-0.2}$, and $\beta = -0.1^{+0.1}_{-0.1}$ for 1.7−4 $R_\oplus$ planets from Petigura et al. (2018) [121]. Dong & Zhu (2013) [42] also found a nearly flat distribution for 2−4 $R_\oplus$ with $\beta = 0.11 \pm 0.05$. However, while all three studies agree that $\sim 1-2$ $R_\oplus$ planets are more common than $\sim 2-4$ $R_\oplus$ planets before the small-planet transition, our results and those of Petigura et al. (2018) [121] would indicate the opposite is true past the transition while Dong & Zhu (2013) [42] found similar occurrence rates for both distributions. I believe this is due to our use of more up-to-date and precise stellar radii, causing some $1-2$ $R_\oplus$ planets to be pushed into the $2-4$ $R_\oplus$ bin.

All three studies indicate that the distributions of larger planets (here, $4-8$ $R_\oplus$ and $8-16$ $R_\oplus$) are inconsistent with this power-law cut-off model for the period range in question, with occurrence rates only gradually increasing. My $4-8$ $R_\oplus$ occurrence rates do jump suddenly at 3.1 days, though this is likely another look at the Neptune Desert described in §4.6.3.

For $8-16$ $R_\oplus$, I do not confirm the three-day "pile-up" of hot Jupiters clear from RV surveys (e.g. [37, 149, 157]). This pile-up features strongly in various high-eccentricity migration scenarios (e.g. [45, 159, 160]). However, other *Kepler*-based studies have called into question the pile-up [51, 62], and differences in overall hot Jupiter ($P \lesssim 10$ days) occurrence rates between *Kepler* and RV surveys have been previously noted. In particular, *Kepler* hot Jupiter occurrence rates typically lie at around $0.4-0.6\%$ (e.g. [51, 62, 105, 121]), while RV occurrence rates are at around $0.9-1.2\%$ (e.g. [91, 95]). My own hot Jupiter estimate should be intermediate between $0.43^{+0.10}_{-0.09}\%$ ($0.78-6.25$ days) and $0.77^{+0.16}_{-0.14}\%$ ($0.78-12.5$ days), consistent with other *Kepler* results. Dawson & Murray-Clay (2013) [40] suggested that the pile-up was a stronger feature of metal-rich stars ([Fe/H]≥ 0), while the *Kepler* sample has systematically lower metallicity than RV samples. They somewhat recovered the pile-up in the *Kepler* sample when considering only stars with super-solar metallicity. Giant planet occurrence has also been shown to correlate strongly with host-star metallicity [47, 121, 132]. An assessment of the presence of the pile-up in my planet catalogue under these conditions can be left to a future investigation focused on planet occurrence and its dependence on stellar metallicity.

4.6.5 Impact of Catalogue Reliability

While my baseline results incorporate catalogue completeness due to both the detection and vetting process, my planet sample is also not expected to be completely reliable (see §3.4.3). Here, I focus on my catalogue's reliability against noise specifically, which is the largest concern for candidates near the detection limit, including small, rocky planets in orbits with long orbital periods.

The incorporation of reliability against noise into occurrence rate estimates remains an open question and was only first directly tackled in Bryson et al. (2019) [19]. Bryson et al. (2019) [19] took a probabilistic approach to reliability models, fitting components of the reliability with functions over finely spaced bins in period-S/N space (where S/N is represented by the MES employed by the *Kepler* team). Rather than assume a functional form of the reliability, I take a simplified approach, described here

In §3.4.3, I had already determined reliability over a coarse grid in period-S/N space using the Inverted and Scrambled datasets. I remade this reliability grid, this time over the period bin edges equal to those used for my occurrence rates. While ideally I would find the reliability across only the FGK stars in the sample, a concern is that small number statistics would be more significant than for the full, better characterized ∼200,000-star sample. Thus, while I recognize that noise properties between the two samples should be different, I elected to use my results across all stars (top panel of Fig. 4.13) to improve the signal-to-noise ratio of the estimate. The FGK-only results are included in the bottom panel of Fig. 4.13 for comparison.

To incorporate reliability into the ABC methodology itself, recall the discussion of selection effects outlined in §4.5.2 for my forward model. The FP-related selection effect was $\eta_{\rm fp} = 1/(1-r_{\rm fp})$, where $r_{\rm fp}$ is the rate of false positive events that are detected as planets. This is equal to 1 minus the reliability, giving $\eta_{\rm fp} = 1/R$. Thus, I replace Eqn. 4.15 in my forward model with

$$\eta_{\rm tot} = \eta_{\rm tr}\eta_{\rm rec}/R, \tag{4.23}$$

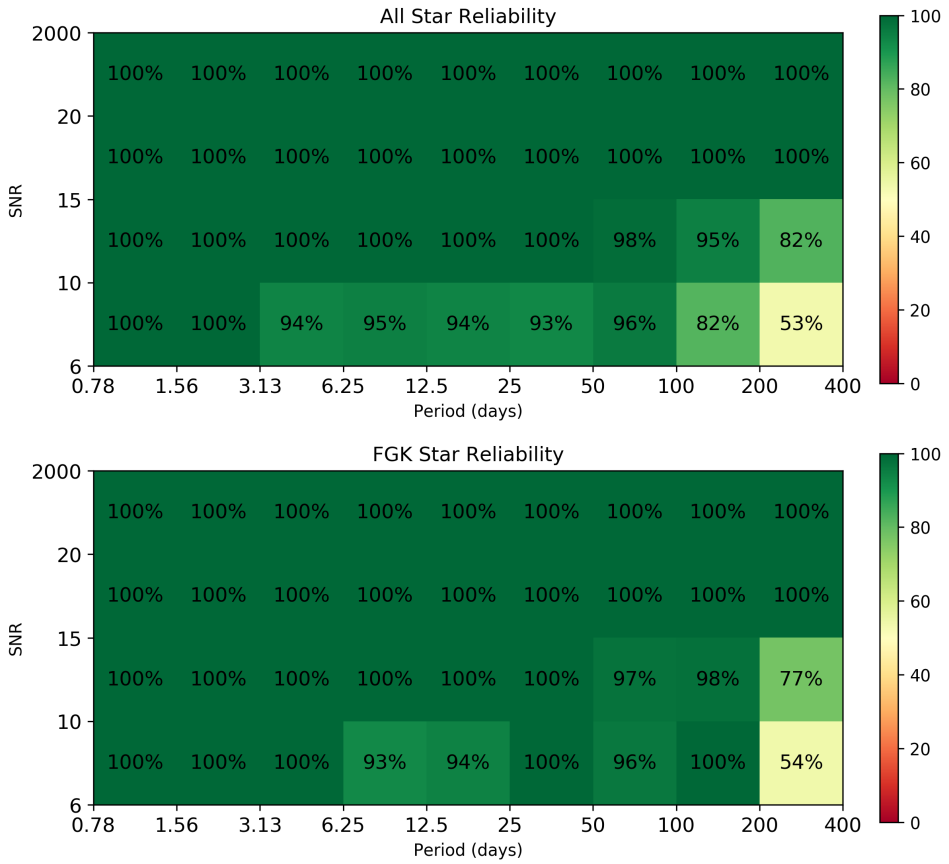

Figure 4.13: Estimate of the reliability of my catalogue, where reliability refers to the fraction of PCs within a given period-S/N bin that are actually planets. Top: considering all stars searched in the *Kepler* sample. Bottom: only including the FGK stars used in this occurrence rate study.

where R is found for a given planet according to its period and S/N and using the coarse grid.

I recalculated my FGK occurrence rates under these changes, with median and 68.3% credible interval results included in Table 4.9 alongside my baseline FGK estimates. Reliability did not significantly impact my results, with posteriors for every cell overlapping significantly. This is not unexpected, given that the reliability is high ($> 90\%$) outside only the long-period, low-S/N corner in period-S/N space. For the corresponding period-radius cells, it is difficult to assess the full impact of reliability given that the same areas have very low completeness and no planet detections, meaning estimates are already not as well constrained as other areas. However, I do tend to reduce upper limits in the most affected areas. In particular, the $0.5 - 0.71$ $R_\oplus$, $200 - 400$ day cell had its upper limit reduced from 14.4% to 13.1%, and the $0.71 - 1$ $R_\oplus$, $200 - 400$ day cell had its upper limit reduced from 3.4% to 2.7%.

Given that I focused in my previous sections on high-reliability regimes (i.e. mainly $R_p > 1$ $R_\oplus$, $P < 200$ days), I do not expect the lack of reliability incorporation to affect my previous analysis, and leave it as a comparison to other studies that also did not incorporate reliability. However, for my upcoming terrestrial habitable zone planet frequency discussion, which will specifically depend on calculations over regions of low reliability, I will discuss both versions of occurrence rate results.

Table 4.9: Occurrence rate results for FGK stars over whole period-radius grid. The "FGK w/ R" column refers to FGK results after incorporating reliability against transit-like noise. Results are given in % (10^{-2}).

Period (days)	Radius ($R_\oplus$)	FGK (%)	FGK w/ R (%)
$0.78 - 1.56$	$0.5 - 0.71$	$0.02^{+0.02}_{-0.01}$	$0.02^{+0.02}_{-0.01}$
$0.78 - 1.56$	$0.71 - 1.0$	$0.1^{+0.04}_{-0.03}$	$0.09^{+0.04}_{-0.04}$
$0.78 - 1.56$	$1.0 - 1.41$	$0.18^{+0.05}_{-0.05}$	$0.18^{+0.05}_{-0.05}$
$0.78 - 1.56$	$1.41 - 2.0$	$0.12^{+0.04}_{-0.04}$	$0.11^{+0.04}_{-0.03}$

Period (days)	Radius ($R_\oplus$)	FGK (%)	FGK w/ R (%)
$0.78 - 1.56$	$2.0 - 2.83$	$0.01^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$
$0.78 - 1.56$	$2.83 - 4.0$	$0.01^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$
$0.78 - 1.56$	$4.0 - 5.66$	$0.01^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$
$0.78 - 1.56$	$5.66 - 8.0$	< 0.02	< 0.02
$0.78 - 1.56$	$8.0 - 11.31$	< 0.02	< 0.03
$0.78 - 1.56$	$11.31 - 16.0$	$0.02^{+0.02}_{-0.01}$	$0.02^{+0.02}_{-0.01}$
$1.56 - 3.13$	$0.5 - 0.71$	$0.1^{+0.06}_{-0.05}$	$0.1^{+0.06}_{-0.06}$
$1.56 - 3.13$	$0.71 - 1.0$	$0.25^{+0.09}_{-0.08}$	$0.27^{+0.08}_{-0.09}$
$1.56 - 3.13$	$1.0 - 1.41$	$0.46^{+0.09}_{-0.1}$	$0.46^{+0.1}_{-0.09}$
$1.56 - 3.13$	$1.41 - 2.0$	$0.52^{+0.1}_{-0.09}$	$0.52^{+0.1}_{-0.1}$
$1.56 - 3.13$	$2.0 - 2.83$	$0.1^{+0.05}_{-0.05}$	$0.1^{+0.06}_{-0.05}$
$1.56 - 3.13$	$2.83 - 4.0$	$0.05^{+0.03}_{-0.03}$	$0.05^{+0.03}_{-0.03}$
$1.56 - 3.13$	$4.0 - 5.66$	$0.02^{+0.02}_{-0.01}$	$0.02^{+0.02}_{-0.01}$
$1.56 - 3.13$	$5.66 - 8.0$	< 0.02	< 0.02
$1.56 - 3.13$	$8.0 - 11.31$	$0.01^{+0.02}_{-0.01}$	$0.01^{+0.01}_{-0.01}$
$1.56 - 3.13$	$11.31 - 16.0$	$0.09^{+0.04}_{-0.03}$	$0.09^{+0.04}_{-0.03}$
$3.13 - 6.25$	$0.5 - 0.71$	$0.34^{+0.18}_{-0.18}$	$0.34^{+0.19}_{-0.16}$
$3.13 - 6.25$	$0.71 - 1.0$	$0.71^{+0.21}_{-0.21}$	$0.69^{+0.2}_{-0.19}$
$3.13 - 6.25$	$1.0 - 1.41$	$1.37^{+0.23}_{-0.23}$	$1.39^{+0.23}_{-0.23}$
$3.13 - 6.25$	$1.41 - 2.0$	$1.79^{+0.27}_{-0.26}$	$1.82^{+0.25}_{-0.24}$
$3.13 - 6.25$	$2.0 - 2.83$	$0.99^{+0.18}_{-0.18}$	$0.99^{+0.18}_{-0.17}$
$3.13 - 6.25$	$2.83 - 4.0$	$0.27^{+0.11}_{-0.1}$	$0.27^{+0.13}_{-0.1}$
$3.13 - 6.25$	$4.0 - 5.66$	$0.18^{+0.09}_{-0.07}$	$0.17^{+0.08}_{-0.06}$
$3.13 - 6.25$	$5.66 - 8.0$	$0.08^{+0.06}_{-0.05}$	$0.08^{+0.06}_{-0.04}$
$3.13 - 6.25$	$8.0 - 11.31$	$0.09^{+0.06}_{-0.04}$	$0.09^{+0.05}_{-0.05}$
$3.13 - 6.25$	$11.31 - 16.0$	$0.19^{+0.08}_{-0.06}$	$0.19^{+0.07}_{-0.06}$
$6.25 - 12.5$	$0.5 - 0.71$	$0.18^{+0.17}_{-0.12}$	$0.19^{+0.16}_{-0.13}$
$6.25 - 12.5$	$0.71 - 1.0$	$0.51^{+0.26}_{-0.22}$	$0.51^{+0.24}_{-0.21}$
$6.25 - 12.5$	$1.0 - 1.41$	$2.34^{+0.39}_{-0.39}$	$2.32^{+0.38}_{-0.35}$
$6.25 - 12.5$	$1.41 - 2.0$	$2.66^{+0.4}_{-0.38}$	$2.65^{+0.37}_{-0.39}$
$6.25 - 12.5$	$2.0 - 2.83$	$3.37^{+0.43}_{-0.39}$	$3.37^{+0.45}_{-0.39}$
$6.25 - 12.5$	$2.83 - 4.0$	$0.61^{+0.22}_{-0.21}$	$0.61^{+0.21}_{-0.19}$

Period (days)	Radius ($R_\oplus$)	FGK (%)	FGK w/ R (%)
$6.25 - 12.5$	$4.0 - 5.66$	$0.17^{+0.1}_{-0.09}$	$0.17^{+0.11}_{-0.08}$
$6.25 - 12.5$	$5.66 - 8.0$	$0.07^{+0.06}_{-0.05}$	$0.07^{+0.06}_{-0.04}$
$6.25 - 12.5$	$8.0 - 11.31$	$0.16^{+0.09}_{-0.07}$	$0.15^{+0.08}_{-0.06}$
$6.25 - 12.5$	$11.31 - 16.0$	$0.18^{+0.08}_{-0.07}$	$0.18^{+0.09}_{-0.08}$
$12.5 - 25.0$	$0.5 - 0.71$	< 0.28	< 0.32
$12.5 - 25.0$	$0.71 - 1.0$	$0.26^{+0.22}_{-0.16}$	$0.28^{+0.22}_{-0.18}$
$12.5 - 25.0$	$1.0 - 1.41$	$1.59^{+0.44}_{-0.37}$	$1.6^{+0.49}_{-0.4}$
$12.5 - 25.0$	$1.41 - 2.0$	$2.41^{+0.59}_{-0.51}$	$2.42^{+0.53}_{-0.48}$
$12.5 - 25.0$	$2.0 - 2.83$	$6.56^{+0.69}_{-0.73}$	$6.59^{+0.71}_{-0.71}$
$12.5 - 25.0$	$2.83 - 4.0$	$2.13^{+0.46}_{-0.46}$	$2.1^{+0.51}_{-0.39}$
$12.5 - 25.0$	$4.0 - 5.66$	$0.19^{+0.17}_{-0.11}$	$0.2^{+0.16}_{-0.12}$
$12.5 - 25.0$	$5.66 - 8.0$	$0.24^{+0.13}_{-0.11}$	$0.24^{+0.14}_{-0.11}$
$12.5 - 25.0$	$8.0 - 11.31$	$0.06^{+0.07}_{-0.04}$	$0.05^{+0.06}_{-0.04}$
$12.5 - 25.0$	$11.31 - 16.0$	$0.18^{+0.1}_{-0.09}$	$0.17^{+0.11}_{-0.08}$
$25.0 - 50.0$	$0.5 - 0.71$	< 0.57	< 0.61
$25.0 - 50.0$	$0.71 - 1.0$	$0.25^{+0.26}_{-0.16}$	$0.27^{+0.24}_{-0.17}$
$25.0 - 50.0$	$1.0 - 1.41$	$0.43^{+0.38}_{-0.25}$	$0.42^{+0.39}_{-0.26}$
$25.0 - 50.0$	$1.41 - 2.0$	$2.19^{+0.64}_{-0.55}$	$2.18^{+0.61}_{-0.55}$
$25.0 - 50.0$	$2.0 - 2.83$	$7.57^{+0.98}_{-0.88}$	$7.55^{+0.87}_{-0.81}$
$25.0 - 50.0$	$2.83 - 4.0$	$3.45^{+0.65}_{-0.63}$	$3.41^{+0.66}_{-0.64}$
$25.0 - 50.0$	$4.0 - 5.66$	$0.26^{+0.22}_{-0.17}$	$0.26^{+0.21}_{-0.17}$
$25.0 - 50.0$	$5.66 - 8.0$	$0.27^{+0.16}_{-0.13}$	$0.28^{+0.16}_{-0.14}$
$25.0 - 50.0$	$8.0 - 11.31$	$0.2^{+0.14}_{-0.12}$	$0.2^{+0.15}_{-0.12}$
$25.0 - 50.0$	$11.31 - 16.0$	$0.3^{+0.18}_{-0.13}$	$0.29^{+0.15}_{-0.12}$
$50.0 - 100.0$	$0.5 - 0.71$	< 1.11	< 1.05
$50.0 - 100.0$	$0.71 - 1.0$	$0.28^{+0.33}_{-0.2}$	$0.24^{+0.28}_{-0.17}$
$50.0 - 100.0$	$1.0 - 1.41$	$0.33^{+0.36}_{-0.23}$	$0.36^{+0.36}_{-0.25}$
$50.0 - 100.0$	$1.41 - 2.0$	$1.5^{+0.72}_{-0.64}$	$1.53^{+0.58}_{-0.59}$
$50.0 - 100.0$	$2.0 - 2.83$	$6.52^{+1.18}_{-1.03}$	$6.6^{+1.22}_{-1.08}$
$50.0 - 100.0$	$2.83 - 4.0$	$3.62^{+0.89}_{-0.81}$	$3.69^{+0.94}_{-0.86}$
$50.0 - 100.0$	$4.0 - 5.66$	$0.92^{+0.49}_{-0.41}$	$0.9^{+0.43}_{-0.41}$
$50.0 - 100.0$	$5.66 - 8.0$	$0.53^{+0.33}_{-0.26}$	$0.51^{+0.34}_{-0.25}$

Period (days)	Radius ($R_\oplus$)	FGK (%)	FGK w/ R (%)
$50.0 - 100.0$	$8.0 - 11.31$	$0.58^{+0.33}_{-0.26}$	$0.6^{+0.31}_{-0.28}$
$50.0 - 100.0$	$11.31 - 16.0$	$0.37^{+0.26}_{-0.21}$	$0.37^{+0.27}_{-0.19}$
$100.0 - 200.0$	$0.5 - 0.71$	< 3.46	< 4.34
$100.0 - 200.0$	$0.71 - 1.0$	< 1.2	< 1.59
$100.0 - 200.0$	$1.0 - 1.41$	$0.41^{+0.48}_{-0.28}$	$0.52^{+0.64}_{-0.36}$
$100.0 - 200.0$	$1.41 - 2.0$	$1.04^{+0.71}_{-0.61}$	$1.03^{+0.71}_{-0.58}$
$100.0 - 200.0$	$2.0 - 2.83$	$6.01^{+1.32}_{-1.23}$	$5.86^{+1.39}_{-1.26}$
$100.0 - 200.0$	$2.83 - 4.0$	$2.82^{+1.07}_{-0.97}$	$2.81^{+1.05}_{-0.87}$
$100.0 - 200.0$	$4.0 - 5.66$	$0.67^{+0.56}_{-0.39}$	$0.66^{+0.54}_{-0.43}$
$100.0 - 200.0$	$5.66 - 8.0$	$0.86^{+0.51}_{-0.45}$	$0.91^{+0.54}_{-0.45}$
$100.0 - 200.0$	$8.0 - 11.31$	$1.55^{+0.69}_{-0.59}$	$1.49^{+0.74}_{-0.53}$
$100.0 - 200.0$	$11.31 - 16.0$	$0.81^{+0.47}_{-0.38}$	$0.82^{+0.45}_{-0.41}$
$200.0 - 400.0$	$0.5 - 0.71$	< 14.11	< 13.07
$200.0 - 400.0$	$0.71 - 1.0$	< 3.37	< 2.72
$200.0 - 400.0$	$1.0 - 1.41$	< 1.43	< 1.44
$200.0 - 400.0$	$1.41 - 2.0$	$0.57^{+0.68}_{-0.42}$	$0.55^{+0.73}_{-0.4}$
$200.0 - 400.0$	$2.0 - 2.83$	$4.78^{+1.74}_{-1.68}$	$4.81^{+1.9}_{-1.63}$
$200.0 - 400.0$	$2.83 - 4.0$	$2.21^{+1.41}_{-1.14}$	$2.21^{+1.33}_{-1.19}$
$200.0 - 400.0$	$4.0 - 5.66$	$1.31^{+0.95}_{-0.79}$	$1.24^{+1.05}_{-0.73}$
$200.0 - 400.0$	$5.66 - 8.0$	$2.75^{+1.24}_{-1.05}$	$2.79^{+1.27}_{-1.14}$
$200.0 - 400.0$	$8.0 - 11.31$	$1.09^{+0.87}_{-0.63}$	$1.09^{+0.84}_{-0.65}$
$200.0 - 400.0$	$11.31 - 16.0$	$1.66^{+0.95}_{-0.78}$	$1.66^{+0.87}_{-0.78}$

4.7 Terrestrial Habitable Zone Planet Frequency

A partial explanation for the lack of consistency between literature $\eta_\oplus$ values lies in how authors define the habitable zone. As discussed in §1.4, while most recent occurrence rate studies refer to Kopparapu et al. (2013) [78], the use of optimistic vs. conservative limits will affect a final $\eta_\oplus$ estimate. Another complicating factor is how authors define the size of a potentially habitable, rocky planet. Too small, and a planet will not be able to retain

an atmosphere or support plate tectonics [76]. Raymond et al. (2007) [124], for instance, considered 0.3 $M_\oplus$ as the lower-mass limit for planetary habitability. Using the mass-radius relation for $R_p < 1.23\ R_\oplus$ from Chen & Kipping (2017) [28],

$$M_p = 0.972\left(\frac{R_p}{R_\oplus}\right)^{3.584} M_\oplus, \tag{4.24}$$

this corresponds to a radius of 0.72 $R_\oplus$. This lower limit was used by Zink & Hansen (2019) [166], while other studies have somewhat arbitrarily used lower bounds anywhere between 0.5 and 1 $R_\oplus$. As for an upper radius limit, a potential transition between rocky super-Earths and volatile-shrouded sub-Neptunes must be considered. It is difficult to simplify this to a single radius since the composition of a planet is much more informative about its potentially rocky nature. However, Rogers (2015) [126] took a statistical approach to a sample of small planets with both masses and radii, finding that most planets above 1.6 $R_\oplus$ are not expected to be rocky, and a best-fit transition occurs at $R_p = 1.48^{+0.08}_{-0.04}$. Furthermore, as found in Fulton et al. (2017) [52] and further explored here, $\sim$1.5 $R_\oplus$ precedes a gap in the exoplanet size distribution, which would support this prediction. Previous papers have typically chosen upper radius limits between 1.5 and 2 $R_\oplus$.

In an effort to standardize $\eta_\oplus$ determination, the ExoPAG SAG13 report recommended that authors produce estimates using both $1 - 1.5\ R_\oplus$ and $0.5 - 1.5\ R_\oplus$ radius ranges. They also defined their G-type star $\eta_{\rm habSol,SAG13}$ value as lying between 237 and 860 days, corresponding to the Kopparapu et al. (2013) [78] optimistic HZ for a Sun-like star

Following Hsu et al. (2019) [65], I start with occurrence rates using bin edges of $R_p = \{0.5, 0.75, 1, 1.25, 1.5, 1.75, 2\}\ R_\oplus$ and $P = 237 - 500$ days. These are the same radius bin edges as in Hsu et al. (2019) [65], but with an additional $0.5 - 0.75\ R_\oplus$ bin to meet the radius range recommendation of SAG13. The 237 day lower bound on the period bin corresponds to the inner edge of the optimistic HZ as defined by Kopparapu et al. (2013) [78] for a Sun-like star, while the 500 day upper bound corresponds to the limit of *Kepler*'s (and my pipeline's) sensitivity. As in the SAG13 report,

I considered G-type stars to represent "Sun-like" stars for my calculations. SAG13 defined G-type stars using the same temperature limits as I use here ($5300 - 6000$ K).

Since I am also interested in incorporating reliability, I needed to estimate my catalogue's reliability specific over the $237 - 500$ day period range. Using my all-star results from §4.6.5, I found $R = 0.39$, 0.80, and 1.0 for S/N < 10, $10 \leq$ S/N < 15, and S/N ≥ 15 respectively.

4.7.1 Optimistic Habitable Zone Estimate

My direct calculation over the $237 - 500$ day bin represents a subset of the $237 - 860$ day optimistic HZ. For the $1 - 1.5$ $R_\oplus$ range, I find an occurrence rate of $0.05^{+0.04}_{-0.03}$ planets per star. When incorporating reliability, I find an occurrence rate of $0.05^{+0.03}_{-0.02}$ planets per star. For the $0.75 - 1.5$ $R_\oplus$ range considered by Hsu et al. (2019) [65], I find an occurrence rate of $0.12^{+0.08}_{-0.06}$ ($0.10^{+0.07}_{-0.04}$ with reliability), which overlaps well with the $0.16^{+0.11}_{-0.06}$ estimate from Hsu et al. (2019) [65]. Lastly, I find a considerably increased estimate with substantial uncertainties when adopting the $0.5 - 1.5$ $R_\oplus$ size range, giving $0.38^{+0.29}_{-0.19}$ ($0.31^{+0.28}_{-0.15}$), on account of the low completeness and poor constraints provided by the data. Note that neither my work nor Hsu et al. (2019) [65] had planet detections over any of these radius ranges, so these occurrence rates should be interpreted as upper limits.

Considering the entire $237 - 860$ day HZ range requires extrapolating these results to longer periods. Interpreting results obtained via extrapolation should be done with added caution, considering I have demonstrated substantial uncertainties in sub-$\eta_\oplus$ occurrence rate estimates and will necessarily have to make an assumption about the nature of planet distributions beyond the limit of *Kepler*'s sensitivity. However, the ABC methodology requiring such extrapolation is not unique. All studies are limited by the *Kepler* mission duration to planets within ≈500 days, and small planets typically require more observed transits than larger planets in order to produce signals with sufficient S/N for detection. Other grid-based occurrence rates must make similar assumptions to my work. Meanwhile, studies that find

a function to describe planet distributions with period and/or radius may integrate over a desired $\eta_\oplus$ range to produce an estimate, but the function itself will have been based on planets with larger sizes and/or shorter orbital periods. In these cases, the employed assumption is that the same parametric model can also explain the $\eta_\oplus$ regime, which is not necessarily true. Furthermore, I have shown that there are numerous period- and size-dependent small-scale variations (such as the clear radius valley for orbital periods within 25 days), indicating that a parametric occurrence rate model is not necessarily the best descriptor of the data even over shorter orbital periods and larger planet sizes considered by these works.

With these caveats, I adopt the method of extrapolation used by Hsu et al. (2019) [65] and assume that the differential occurrence rate derived over $237 - 500$ days and a given radius range is the same over longer periods. Under this assumption, I estimate (upper limit) optimistic HZ occurrence rates of $\eta_\oplus = 0.08^{+0.07}_{-0.04}$ ($0.08^{+0.06}_{-0.04}$) planets per star for $1 - 1.5$ $R_\oplus$, $\eta_\oplus = 0.21^{+0.14}_{-0.10}$ ($0.17^{+0.11}_{-0.08}$) for $0.75 - 1.5$ $R_\oplus$, and $\eta_\oplus = 0.66^{+0.51}_{-0.32}$ ($0.53^{+0.48}_{-0.26}$) for $0.5 - 1.5$ $R_\oplus$.

4.7.2 Conservative Habitable Zone Estimate

The $0.99-1.70$ AU conservative HZ from Kopparapu et al. (2013) [78] corresponds to orbital periods of $360-809$ days for a Sun-like star. Extrapolating my $237-500$ day results over these periods, I find (upper limit) occurrence rates of $\eta_\oplus = 0.05^{+0.04}_{-0.03}$ ($0.05^{+0.04}_{-0.02}$) planets per star for $1 - 1.5$ $R_\oplus$, $\eta_\oplus = 0.13^{+0.09}_{-0.06}$ ($0.11^{+0.07}_{-0.05}$) for $0.75 - 1.5$ $R_\oplus$, and $\eta_\oplus = 0.42^{+0.32}_{-0.20}$ ($0.34^{+0.30}_{-0.16}$) for $0.5 - 1.5$ $R_\oplus$.

4.7.3 Comparison to Previous Works

The challenges involved with defining the bounds of $\eta_\oplus$ have motivated recent studies to instead report and compare $\Gamma_\oplus$, the differential occurrence rate near the HZ [22, 49, 162]. I follow Hsu et al. (2019) [65] in defining $\Gamma_\oplus$ using the $237 - 500$ day, $0.75 - 1.5$ $R_\oplus$ results, giving $\Gamma_\oplus \equiv (d^2 f)/[d(\ln P)d(\ln R_p)] = 0.23^{+0.16}_{-0.11}(0.19^{+0.13}_{-0.08})$.

Comparisons with other $\Gamma_\oplus$ estimates in the literature are shown in Fig. 4.14. The values from Pascucci et al. (2019) [115] correspond to their Model #4 and #6 results given in their Table 2, which address the impact of photoevaporated cores by excluding planets within 12 and 25 days from their analysis, respectively. The ExoPAG SAG13 estimate, obtained via Kopparapu et al. (2018) [79], is based on a meta-analysis of community-submitted G-type star occurrence rate studies, for which a broken power law was fit to a combined period-radius grid. The plotted central values were found by plugging in the Earth's radius and period into their baseline power law, while the lower and upper uncertainties correspond to their pessimistic and optimistic power laws, respectively. The values from Burke et al. (2015) [22] represent the allowable range from their sensitivity analysis. The Petigura et al. (2013) [119] result is calculated by converting their $200 - 400$ day, $1 - 2\ R_\oplus$ extrapolated occurrence rate into a differential rate. The Dong & Zhu (2013) [42] value is an extrapolation of their $1 - 2\ R_\oplus$ best-fit function, evaluated at the orbital period of Earth and with error bars given by the propagation of uncertainty. All other values are the $\Gamma_\oplus$ results explicitly reported in their respective papers.

My results compare most favourably with those from Pascucci et al. (2019) [115], Hsu et al. (2019) [65], Bryson et al. (2019) [19], Kopparapu et al. (2018) [79], and Dong & Zhu (2013) [42], and are well within the allowable range of Burke et al. (2015) [22]. Meanwhile, the values from Youdin (2011) [162], Garrett et al. (2018) [56], and Zink et al. (2019) [165] all lie above my upper limits. Lack of consistency with the early Youdin (2011) [162] result can readily be explained by the fact that it is based on a much older *Kepler* catalogue, having included only planets with $P < 50$ days. Part of the disagreement with Zink & Hansen (2019) [166] can be explained by their incorporation of transit multiplicity — in other words, they took into account a reduction in detection efficiency with planet detection order, which would result in an increased occurrence rate — though I cannot confirm whether or not this would explain the whole discrepancy. Nevertheless, I argue that my uncertainty estimates are more realistic than those of Garrett et al. (2018) [56] and Zink & Hansen (2019) [166], which are considerably

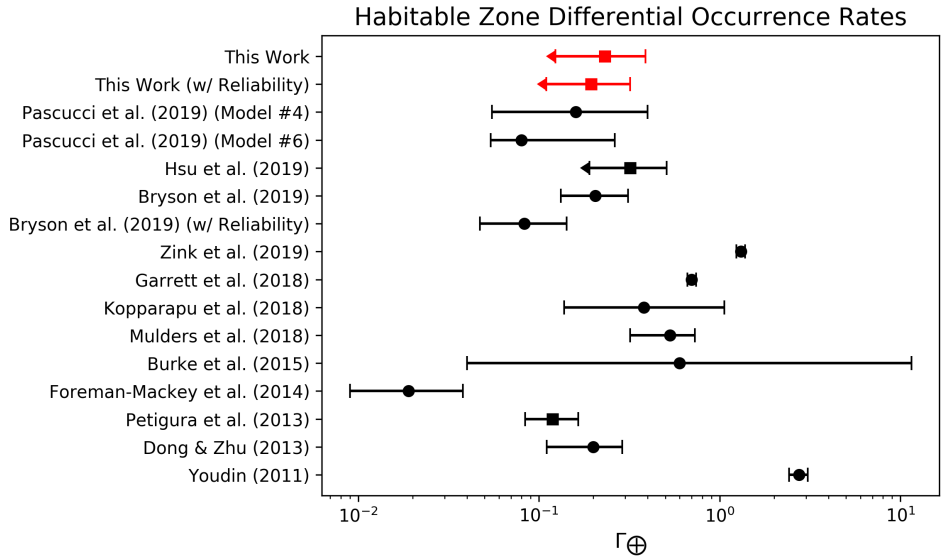

Figure 4.14: A collection of $\Gamma_\oplus$ values from the literature: Pascucci et al. (2019) [115], Hsu et al. (2019) [65], Bryson et al. (2019) [19], Zink et al. (2019) [165], Garrett et al. (2018) [56], ExoPAG SAG13 via Kopparapu et al. (2018) [79], Mulders et al. (2018) [107], Burke et al. (2015) [22], Foreman-Mackey et al. (2014) [49], Petigura et al. (2013) [119], Dong & Zhu (2013) [42], and Youdin (2011) [162]. Squares indicate that grid-based occurrence rates were explored (my work, Hsu et al. (2019) [65], and Petigura et al. (2013) [119]), while circles indicate a functional form for the occurrence rate was assumed (all others). Left-pointing arrows indicate that the result is meant to be interpreted as an upper limit.

more optimistic than all other works.

Turning to the impact of reliability, I did not find significantly different occurrence rates as noted previously, even over this low-reliability regime. This is contrasted to the results of Bryson et al. (2019) [19] who found that incorporating reliability dropped $\Gamma_\oplus$ from $0.205^{+0.106}_{-0.073}$ down to $0.083^{+0.050}_{-0.036}$, whereas mine dropped from $0.23^{+0.16}_{-0.11}$ to only $0.19^{+0.13}_{-0.08}$. I note three things: first, I reemphasize that given the total lack of planet detections in the relevant bins, my results are poorly constrained by the data and should be interpreted as upper limits. Had I been able to place better constraints

on these occurrence rates, a clearer picture of the full impact of reliability might appear. Second, I only took into account reliability against noise and systematics, whereas Bryson et al. (2019) [19] additionally considered reliability against astrophysical FPs. Their differential rate using noise-FP-only reliability was $\Gamma_{\oplus} = 0.128^{+0.077}_{-0.049}$. Lastly, my grid-based occurrence rate method means that the reliability of a given bin will only directly affect the occurrence rates of that bin, and indirectly affect the occurrence rates of neighbouring, simultaneously fit bins due to the leaking of planets between bins. Bryson et al. (2019) [19] fit a joint power-law model across all planets with $R_p = 0.75 - 2.5\ R_{\oplus}$ and $P = 50 - 400$ days, meaning that the reliability of any planet would affect the fit results across the entire space considered. I would expect this to cause the reliability to have a greater affect on occurrence rates than in my work.

A subset of $\eta_{\oplus}$ not yet mentioned is the concept of $\zeta_{\oplus}$: the number of planets per Sun-like star with radii and orbital periods within 20% of Earth's values, as introduced by Burke et al. (2015) [22]. Combining both an extrapolated analysis (assuming their $0.75-2.5\ R_{\oplus}$, $50-300$ day broken power law) and a direct analysis (recomputing a broken power law over $300-700$ days) of the *Kepler* Q1-Q16 planet sample around GK dwarf stars, Burke et al. (2015) [22] reported $\zeta_{\oplus} = 0.10$ with an allowable range of $0.01-2$. To compare to this value, I recalculated my G-type star occurrence rates with bin edges of $R_p = \{0.8, 1.0, 1.2\}\ R_{\oplus}$ and $P = 292 - 438$ days. I recovered their central value almost exactly, finding $\zeta_{\oplus} = 0.12^{+0.09}_{-0.06}$ ($0.10^{+0.07}_{-0.06}$).

4.7.4 Final $\eta_{\oplus}$ Recommendation

For the definition of the habitable zone, Kopparapu et al. (2013) [78] recommended the use of their conservative ($0.99 - 1.70$ AU) boundaries for informing future missions with $\eta_{\oplus}$ values. For the lower radius limit, I am inclined to consider potentially habitable rocky planets as those down to $0.75\ R_{\oplus}$ rather than $0.5\ R_{\oplus}$, as this is near the scientifically motivated $0.72\ R_{\oplus}$ limit used by Zink et al. (2019) [166]. For the upper radius limit, the $1.5\ R_{\oplus}$ radius considered throughout this section is already well mo-

tivated by both the characteristics of the radius valley and the findings of Rogers (2015) [126]. Meanwhile, incorporating reliability allows for better constraints on $\eta_{\oplus}$ upper limits.

For my reliability-incorporated, $0.75 - 1.5\ R_{\oplus}$ conservative HZ, my $\{5, 15.9, 50, 84.1, 95\}$th percentiles are $\eta_{\oplus} = \{0.04, 0.06, 0.11, 0.18, 0.24\}$ planets per star. Thus, I suggest future exoplanet characterization and habitability-related missions to consider an upper limit (84.1th percentile) of < 0.18 terrestrial HZ planets per Sun-like star.

4.8 Limitations of Occurrence Rates

Here, I summarize the main assumptions and design choices made in my study, each of which can motivate future improvements to my occurrence rate model and methodology.

Astrophysical FPs: I assumed that the false positive rate due to astrophysical FPs (such as background, grazing, or hierarchical triple eclipsing binaries) was negligible, and only took into account the FP rate due to noise and systematics. In order to minimize the effect of contamination from astrophysical FPs on my occurrence rates, I vetted against them in my vetting pipeline, and only included candidates in the final catalogue that were also passed as planet candidates by the *Kepler* Robovetter. However, in the future I could characterize my catalogue's reliability against astrophysical FPs similarly to how I explored reliability against noise and systematics.

Behaviour of simulated noise: I used the Inverted and Scrambled datasets to estimate my catalogue reliability against noise and systematics, under the assumption that they accurately simulate the noise and systematics present in the actual observed *Kepler* data. However, this is not the case for all periods and all types of false alarms. For example, the use of the Inverted set relies on the assumption that the false alarms are symmetric upon data inversion. This means it will not reproduce the drops in flux caused by cosmic-ray-induced sudden pixel sensitivity dropout [72]. Meanwhile, the Scrambled dataset leaves each *Kepler* quarter untouched. Significant signals with periods less than a month may appear in both the original and

"simulated" noise sets. A more careful accounting of reliability in my occurrence rate estimates would require alternative methods for characterizing reliability.

Reliability Model: I used a simplistic estimate of my catalogue's reliability against noise, in the form of a coarse grid in period-S/N space, largely due to small-number statistics. Bryson et al. (2019) [19] found a functional for reliability, which I could adopt after a more thorough investigation into my catalogue's reliability against both astrophysical and noise FPs.

Choice of prior: I assumed independent, uniform priors for all bins throughout this study. While this was the same choice used by Hsu et al. (2019) [65] for their baseline study, they cautioned that this was equivalent to assuming a prior on the total rate that was peaked toward high occurrence rates. After using a Dirichlet prior over radius bins and a uniform prior on the total rate, they found significantly lower rates for bins poorly constrained by the data, such as the $\eta_{\oplus}$ regime. It is unclear whether or not my methodology is as sensitive to choice of prior, but I am unable to implement multivariate priors in `cosmoabc`, though this would be a line of inquiry I would be interested in exploring in the future.

Eccentricity: In my forward model, I simulated all planets in perfectly circular orbits ($e = 0$) for simplicity. Nonzero eccentricity would affect both the transit duration and detection probability of a given planet. Burke et al. (2015) [22] found that introducing eccentricity had only a minor effect on lowering occurrence rates, comparable to systematic errors. However, in principle I could draw from a distribution for the eccentricity (e.g. [64, 65]), or take into account some dependence on planet properties in the future.

Planetary system architectures: I have only attempted to characterize the average occurrence rates of individual planets, rather than the orbital architectures of multiplanet systems. Notably, Mulders et al. (2018) [107] introduced a forward model that simulates a planetary system for each star and takes into account correlations in the properties of each planet in the system (i.e. orbital inclination, period, and radius). However, their simulations were made possible by assuming a parametric power-law based function for exoplanet occurrence rates. Grid-based occurrence rates as in my study

are more computationally limited, and I can only simulate $\sim$5 bins (covering a small section of period-radius space) at a time. Nevertheless, future improvements to the efficiency of ABC algorithms may make this possible.

Window function: I adopted the binomial approximation of the window function probability of detecting at least three transits in *Kepler* data from Burke et al. (2015) [22]. This form of the window probability was also used in Hsu et al. (2018) [64] and Zink et al. (2019) [165]. An alternative would be to use the *Kepler* DR25 target-by-target window functions from Burke & Catanzarite (2017) [24], which have been shown to result in reduced occurrence rates due to better accounting for the detection probability for planets with few transits [65]. I chose not to use these data products given they are unique to the *Kepler* DR25 pipeline and TPS algorithm, though given I use the same three-transit minimum detection criteria, the differences between our pipelines may be minor enough that it would be worth incorporating these data products in the future.

Transit multiplicity: I did not take into account reductions in detection efficiency due to transit multiplicity. Similar to the *Kepler* pipeline, my search was a multipassthrough process in which the strongest S/N signal in the light curve would be removed after detection in order to facilitate another search for more planets. With more potential signals removed from the data, subsequent searches would be based on less available data, and the detection probability of finding a surviving candidate would be reduced. Zink et al. (2019) [165] found that the *Kepler* pipeline experiences an additional 5.5% and 15.9% efficiency loss for planets with $P < 200$ and $P > 200$ days respectively, after finding the strongest transit signal in a multiple-planet system.

Chapter 5

Conclusions and Future Work

5.1 Summary

In Chapter 3, I described my independent transiting exoplanet search and vetting pipeline and its application to all ∼200,000 stars observed over the four years of the original *Kepler* mission — the largest independent *Kepler* search performed to date. Through my search, I rediscovered thousands of known planets and planet candidates, with a 98.8% recovery rate of confirmed *Kepler* planets. I also reported the discovery of 17 candidates new to my work, 12 of which have astrophysical false positive probabilities less than 1% which has important implications for future steps toward validating these candidates as bonafide planets. Several of my new detections represent particularly valuable additions to exoplanetary parameter space. KIC-7340288 b is both rocky (1.51 $R_\oplus$) and in the habitable zone (0.33 $S_\oplus$) of its star — a part of phase space currently occupied by only 15 smaller confirmed KOIs according to the NASA Exoplanet Archive. The abundance of such planets is of great interest to the exoplanet community, but is also poorly constrained. Furthermore, less than 1% of all *Kepler* planets and planet candidates are smaller than 0.7 $R_\oplus$. At 0.66 $R_\oplus$, KIC-11350118 c is larger than only 15 confirmed KOIs, and also transforms the KOI-4509 system into a multiplanet system.

I also obtained adaptive optics imaging follow-up for six my new planet candidates. My results emphasize the role that high-resolution imaging can play in validating planets. The input of contrast curves to `vespa` reduced

astrophysical false positive probabilities by 16 times on average for the targets for which I obtained observations. The confirmation of no nearby stars bright enough to significantly dilute the transit depth of my rocky HZ planet KIC-7340288 b is also value, considering how close this planet is to the suggested rocky limit of $R_p < 1.6\ R_\oplus$.

In Chapter 4, I described the computation and analysis of occurrence rates based on my independent planet catalogue for F-, G-, and K-type stars. I used approximate Bayesian computation — a promising methodology that was only recently applied to exoplanet occurrence for the first time [64]. I directly incorporated search completeness, vetting completeness, and planet radius uncertainty, the latter two of which have been absent from most previous studies. I also described incorporating catalogue reliability into grid-based occurrence rate measurements for the first time, and suggest that the impact of low reliability is less severe than for methods that assume parametric planet distribution functions. My occurrence rate study is the first to directly take into account all of the above biases, which is an essential step toward more accurate occurrence rates. I also used the most up-to-date, *Gaia*-incorporated stellar radius values. With all of these considerations, my occurrence rate should be more robust than previous works.

In my investigation of the dependence of planet occurrence on host star effective temperature, I confirmed the findings of Howard et al. (2012) [62] and Mulders et al. (2015a, 2015b) [105, 106] that small planets are significantly more abundant around cooler stars than hotter stars. In my investigation of the overall distribution of planets with orbital period, I found that power laws well describe $1-2\ R_\oplus$ and $2-4\ R_\oplus$ occurrence rates, separated by transition periods of $5.9^{+0.5}_{-0.5}$ days and $13.3^{+1.4}_{-1.5}$ days respectively. These transition periods may inform future studies on potential stopping mechanisms for planet migration, or orbital distances at which most small planets form. Meanwhile, larger planets demonstrate a consistent rise in occurrence rates with orbital period across the whole $P < 400$ days examined.

A particularly significant result came out of my investigation of the gap in the distribution of planets with radius. I provided the first bias-corrected observational evidence for this gap since the original Fulton et al. (2017) [52]

paper, and was also able to take an even deeper look into its dependence on orbital period for the first time. In particular, my results indicate that the bimodal distribution does not exist over the full $P < 100$ day range; rather, it is significant from $\sim$6 days up to $\sim$25 days, while sub-Neptunes dominate at longer orbital periods and super-Earths dominate at shorter orbital periods. This agreement with recent model predictions by Owen & Wu (2017) [114] is perhaps the strongest support thus far for a photoevaporation-driven model of small-planet evolution.

Lastly, determining the frequency of potentially habitable planets was a primary motivation of my work. Recognizing the sensitivity of $\eta_\oplus$ to the assumed definitions of the habitable zone and the size limits of potentially habitable, rocky planets, I reported a wide variety of upper limits that can be compared to other literature values, both before and after taking into account catalogue reliability. Robustly calculated $\eta_\oplus$ values are essential for informing the design of future missions dedicated to the search and characterization of potentially habitable exoplanets. In conclusion, I recommend an upper limit (84.1th percentile) of < 0.18 potentially habitable, rocky planets per Sun-like star.

5.2 Future Work

Both my independent search and vetting pipeline, and occurrence rate methodology, can be extended into numerous future studies. I summarize potential opportunities for follow-up here.

5.2.1 Applications to Other Missions

The search and vetting pipeline I developed for this thesis was designed to work on *Kepler* light curves, but it can be adapted to any missions that use the transit method. Likewise, the occurrence rate method I have described and implemented would work on any transiting planet catalogue, as long as the completeness (and ideally reliability) of the catalogue is characterized.

One such mission is K2. K2 was the continuation of the original four-

year *Kepler* mission ($2009 - 2013$), following the failure of two of *Kepler*'s reaction wheels that were required to point the telescope at the same section of the sky throughout its observations. Before the telescope finally retired in 2018, K2 observed hundreds of thousands more stars, with observations lasting ∼80 days per star. There are a variety of publicly available codes provided by the exoplanet community for analyzing K2 light curves (e.g. `lightkurve`, `PyKE`), which would be invaluable resources to both improve my code and help with the transition, such as removing K2 motion systematics not present in *Kepler* data.

A specific avenue I would be interested in exploring with K2 would be occurrence rates for M dwarf stars. Due to its primary goal of finding Earth-size planets around Sun-like stars, *Kepler* observed only a few thousand M dwarfs out of all ∼200,000 stars. Utilizing data from K2, which observed tens of thousands more (∼30,000), could represent a significant improvement for M dwarf occurrence rates. Furthermore, K2 remains a largely untapped resource for occurrence rate studies. This is due to the fact that the published planet candidate lists are an amalgamation of many different community-led planet discoveries (e.g. [50, 152]), the vast majority of which did not perform the analyses necessary for such statistics. K2 studies also tend to be biased toward planets and stars most conducive to follow-up, further lowering the completeness of a combined catalogue. Thus, a homogeneous catalogue based on a uniform search and vetting pipeline, which I would be able to provide, is needed.

On a related note, M dwarf stars are becoming increasingly more appealing for the hunt for habitable worlds, motivating the determination of an improved $\eta_\oplus$ estimate for M dwarfs specifically. M dwarfs are the most abundant stars by number, comprising roughly 75% of the nearby stellar population. Their smaller sizes mean Earth-size planets are easier to find than around other stars, while their lower temperatures mean that any planets in the HZ must orbit closer in, resulting in orbital periods of only dozens of days — short enough that the ∼80 day observations from K2 would be able to place meaningful constraints.

Another example mission is TESS, NASA's follow-up to *Kepler*. TESS

is nearing the end of its originally planned two-year mission (2018 − 2020), though it will soon be continuing operations in an extended mission. Unlike *Kepler* and K2, TESS is an all-sky survey focused on nearby, bright stars, and observes different sections of the sky for different amounts of time (with the majority of observations lasting ∼27 days, and others up to one year). TESS has already provided a wealth of publicly available archival data on MAST upon which an independent search can be run. Furthermore, while TESS is not expected to be as amenable to occurrence rates studies as *Kepler* and K2 since its unique mission footprint makes it challenging to characterize its completeness, the ABC methodology may be the most viable method to attempt this. A strength of the ABC method is that it allows for the properties and corresponding detection efficiencies of individual stars to be considered during the occurrence rate calculation, rather than assuming a star-averaged completeness function. This will be essential for incorporating differences in observation timescales between stars.

5.2.2 Other *Kepler* Occurrence Rate Explorations

Since I searched the entire *Kepler* stellar sample, a wide range of dependencies of occurrence rates on stellar properties can be explored. I focused my dissertation on occurrence rates for F, G, and K dwarf stars, defining each stellar type according to effective temperature. A future study could look at other dimensions, make the appropriate cuts on the stellar sample, and observe how the distributions of orbital period and planet radius change accordingly.

One particular dimension that has featured prominently in exoplanet population analysis, and was briefly discussed earlier in this dissertation, is stellar metallicity. Giant planet occurrence has been shown to increase with host star metallicity (e.g. [47, 132]), while correlation is not as strong for smaller planets (e.g. [20, 139]). In §4.6.4, I discussed how the lack of hot Jupiter pile-up in my results could be because it is a characteristic of metal-rich stars specifically, as indicated by Dawson & Murray-Clay (2013) [40]. Recalculating the distribution with period after only considering metal-rich

stars could lend support to this theory.

5.2.3 Combining Transit Surveys with Other Methods

As more small planets are found by other missions following the *Kepler* era, being able to combine the results of multiple surveys, especially those based on different detection methods, will become more important. For example, the next generation of spectrographs will have the ability to detect Earth-size planets in the HZs of M dwarf stars with the radial velocity method for the first time, and will not be limited by the specific orbital geometry required by the transit method. Combining their discoveries with those of *Kepler* and/or K2 could significantly improve $\eta_{\oplus}$ estimates for M dwarfs even further. In a more general sense, no single method is able to fill all of exoplanetary parameter space alone due to differences in detection sensitivity, meaning that occurrence rate studies have so far been limited to specific regions. Combining results from several methods into a single occurrence rate calculation can give a more complete picture of planet diversity than possible by looking at surveys individually.

However, combining results is challenging due to various differences in systematics — such as the behaviour of noise in the data, the completeness of the search and vetting pipelines used, and selection effects specific to each detect method — as well as differences in the planet properties recovered — such as radius with the transit method, compared to minimum mass with the radius velocity method. A small number of papers have begun addressing these challenges (e.g. Clanton & Gaudi (2016) [32]), but more work is still to be done, and transit surveys have not yet appeared in such studies. Therefore, a specific procedure is an open question which I would be excited to investigate, such as starting with the occurrence rate method I used for this dissertation.

Bibliography

[1] Abbot, D. & Switzer, E. 2011, The Steppenwolf: A proposal for a habitable planet in interstellar space, ApJL, 735, L27

[2] Abe, Y., Abe-Ouchi, A., Sleep, N., & Zahnle, K. 2011, Habitable zone limits for dry planets, Astrobiology, 11, 443

[3] Astropy Collaboration, Roabitaille, T., Tollerud, E., et al. 2013, Astropy: A community Python package for astronomy, A&A, 558, A33

[4] Astropy Collaboration, Price-Whelan, A., Sipocz, B., et al. 2018, The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package, AJ, 156, 123

[5] Baranec, C., Ziegler, C., Law, N., et al. 2016, Robo-AO Kepler Planetary Candidate Survey II: Adaptive Optics Imaging of 969 Kepler Exoplanet Candidate Host Stars, AJ, 152, 18

[6] Barclay, T., Rowe, J., Lissauer, J., et al. 2013, A sub-Mercury-sized exoplanet, Nature, 494, 452

[7] Batalha, N., Rowe, J., Bryson, S., et al. 2013, Planetary Candidates Observed by Kepler. III. Analysis of the First 16 Months of Data, ApJS, 204, 24

[8] Beaulieu, J.-P., Bennett, D., Fouque, P., et al. 2006, Discovery of a cool planet of 5.5 Earth masses through gravitational microlensing, Nature, 439, 437

[9] Beaumont, M., Cornuet, J-M., Marin, J-M., & Robert, C. 2009, Adaptive approximate Bayesian computation, Biometrika, 96, 983

[10] Bennett, D., Akeson, R., Anderson, J., et al. 2018, The WFIRST Exoplanet Microlensing Survey, arXiv:1803.08564

[11] Berger, T., Huber, D., Gaidos, E., & van Saders, J. 2018, Revised Radii of Kepler Stars and Planets Using Gaia Data Release 2, ApJ, 866, 99

[12] Borucki, W., Koch, D., Jenkins, J., et al. 2009, Kepler's optical phase curve of the exoplanet HAT-P-7b, Science, 325, 709

[13] Borucki, W., Koch, D., Basri, G., et al. 2010, Kepler Planet-Detection Mission: Introduction and First Results, Science, 327, 977

[14] Borucki, W., Koch, D., Basri, G., et al. 2011, Characteristics of Planetary Candidates Observed by Kepler. II. Analysis of the First Four Months of Data, ApJ, 736, 19

[15] Borucki, W., Koch, D., Batalha, N., et al. 2012, Kepler-22b: A 2.4 Earth-radius Planet in the Habitable Zone of a Sun-like Star, ApJ, 745, 120

[16] Brown, T., Latham, D., Everett, M., & Esquerdo, G. 2011, Kepler Input Catalog: Photometric Calibration and Stellar Classification, AJ, 142, 112

[17] Bryson, S., Tenenbaum, P., Jenkins, J., et al. 2010, The Kepler Pixel Response Function, ApJL, 713, L97

[18] Bryson, S., Jenkins, J., Gilliland, R., et al. 2013, Identification of Background False Positives from Kepler Data, PASP, 125, 930

[19] Bryson, S., Coughlin, J., Batalha, N., et al. 2019, A Probabilistic Approach to Kepler Completeness and Reliability for Exoplanet Occurrence Rates, arXiv:1906.03575

[20] Buchhave, L., Latham, D., Johansen, A., et al. 2012, An abundance of small exoplanets around stars with a wide range of metallicities, Nature, 486, 375

[21] Burke, C., Bryson, S., Mulally, F., et al. 2014, Planetary Candidates Observed by Kepler IV: Planet Sample from Q1-Q8 (22 Months), ApJS, 210, 19

[22] Burke, C., Christiansen, J., Mulally, F., et al. 2015, Terrestrial Planet Occurrence Rates for the Kepler GK Dwarf Sample, ApJ, 809, 8

[23] Burke, C., Mulally, F., Thompson, S., Coughlin, J., & Rowe, J. 2019, Re-evaluating Small Long-period Confirmed Planets from Kepler, AJ, 157, 143

[24] Burke, C. & Catanzarite, J. 2017, Planet Detection Metrics: Window and One-Sigma Depth Functions for Data Release 25, KSCI-19101-002

[25] Caceres, G., Feigelson, E., Babu, G., et al. 2019, Autoregressive Planet Search: Application to the Kepler Mission, AJ, 158, 58

[26] Campbell, B., Walker, G., & Yang, S. 1988, A Search for Substellar Companions to Solar-type Stars, ApJ, 331, 902

[27] Catanzarite, J. & Shao, M. 2011, The Occurrence Rate of Earth Analog Planets Orbiting Sun-like Stars, ApJ, 738, 151

[28] Chen, J. & Kipping, D. 2017, Probabilistic Forecasting of the Masses and Radii of Other Worlds, ApJ, 834, 17

[29] Christiansen, J. 2015, Planet Detection Metrics: Pipeline Detection Efficiency, KSCI-19094-001

[30] Christiansen, J. 2017, Planet Detection Metrics: Pixel-Level Transit Injection Tests of Pipeline Detection Efficiency for Data Release 25, KSCI-19110-001

[31] Christiansen, J., Jenkins, J., Caldwell, D., et al. 2012, The Derivation, Properties, and Value of Kepler's Combined Differential Photometric Precision, PASP, 124, 1279

[32] Clanton, C. & Gaudi, B. 2016, Synthesizing Exoplanet Demographics: A Single Population of Long-Period Planetary Companions to M Dwarfs Consistent with Microlensing, Radial Velocity, and Direct Imaging Surveys, ApJ, 819, 125

[33] Claret, A. & Bloeman, S. 2011, Gravity and limb-darkening coefficients for the Kepler, CoRoT, Spitzer, uvby, UBVRIJHK, and Sloan photometric systems, A&A, 529, 75

[34] Coughlin, J. 2017A, Description of the TCERT Vetting Reports for Data Release 25, KSCI-19105-002

[35] Coughlin, J. 2017B, Planet Detection Metrics: Robovetter Completeness and Effectiveness for Data Release 25, KSCI-19114-002

[36] Coughlin, J., Mullally, F., Thompson, S., et al. 2016, Planetary Candidates Observed by Kepler. VII. The First Fully Uniform Catalog Based on the Entire 48-month Data Set (Q1-Q17 DR24), ApJS, 224, 12

[37] Cumming, A., Marcy, G., & Butler, R. 1999, The Lick Planet Search: Detectability and Mass Thresholds, ApJ, 526, 890

[38] Dai, X. & Guerras, E. 2017, Probing Extragalactic Planets Using Quasar Microlensing, ApJL, 853, L27

[39] Dawson, R., Lee, E., & Chiang, E. 2016, Correlations Between Compositions and Orbits Established by the Giant Impact Era of Planet Formation, ApJ, 822, 54

[40] Dawson, R. & Murray-Clay, R. 2013, Giant Planets Orbiting Metal-rich Stars Show Signatures of Planet-Planet Interactions, ApJL, 767, L24

[41] Diaz, R., Almenara, J., Santerne, A., et al. 2014, PASTIS: Bayesian extrasolar planet validation – I. General framework, models, and performance, MNRAS, 441, 983

[42] Dong, S. & Zhu, Z. 2013, Fast Rise of "Neptune-size" Planets (4-8 $R_{\oplus}$) from P $\sim$ 10 to $\sim$250 Days—Statistics of Kepler Planet Candidates up to $\sim$0.75 AU, ApJ, 778, 53

[43] Dressing, C. & Charbonneau, D. 2013, The Occurrence Rate of Small Planets around Small Stars, ApJ, 767, 95

[44] Dressing, C. & Charbonneau, D. 2015, The Occurrence of Potentially Habitable Planets Orbiting M Dwarfs Estimated from the Full Kepler Dataset and an Empirical Measurement of the Detection Sensitivity, ApJ, 807, 45

[45] Fabrycky, D. & Tremaine, S. 2007, Shrinking Binary and Planetary Orbits by Kozai Cycles with Tidal Friction, ApJ, 669, 1293

[46] Fischer, D., Schwamb, M., Schawinski, K., et al. 2012, Planet Hunters: the first two planet candidates identified by the public using the Kepler public archive data, MNRAS, 419, 2900

[47] Fischer, D. & Valenti, J. 2005, The Planet-Metallicity Correlation, ApJ, 622, 1102

[48] Foreman-Mackey, D., Hogg, D., Lang, D., & Goodman, J. 2013, emcee: The MCMC Hammer, PASP, 125, 306

[49] Foreman-Mackey, D., Hogg, D., & Morton, T. 2014, Exoplanet population inference and the abundance of Earth analogs from noisy, incomplete catalogs, ApJ, 795, 6

[50] Foreman-Mackey, D., Montet, B., Hogg, D., et al. 2015, A Systematic Search for Transiting Planets in the K2 Data, ApJ, 806, 215

[51] Fressin, F., Torres, G., Charbonneau, D., et al. 2013, The False Positive Rate of Kepler and the Occurrence of Planets, ApJ, 766, 81

[52] Fulton, B., Petigura, B., Howard, A., et al. 2017, The California-Kepler Survey. III. A Gap in the Radius Distribution of Small Planets, AJ, 154, 109

[53] Furlan, E., Ciardi, D., Everett, M., et al. 2017, The Kepler Follow-Up Observation Program. I. A Catalog of Companions to Kepler Stars from High-Resolution Imaging, AJ, 153, 71

[54] *Gaia* Collaboration. 2016, The Gaia mission, arXiv:1609.04153

[55] *Gaia* Collaboration. 2018, Gaia Data Release 2. Summary of the contents and survey properties, arXiv:1804.09365

[56] Garrett, D., Savransky, D., & Belikov, R. 2018, Planet Occurrence Rate Density Models Including Stellar Effective Temperature, PASP, 130, 114403

[57] Hansen, B. & Murray, N. 2013, Testing in Situ Assembly with the Kepler Planet Candidate Sample, ApJ, 775, 53

[58] Hatzes, A., Cochran, W., Endl, M., et al. 2003, A Planetary Companion to γ Cephei A, ApJ, 599, 1383

[59] Hippke, M., David, T., Mulders, G., & Heller, R. 2019, Wōtan: Comprehensive Time-series Detrending in Python, AJ, 158, 143

[60] Hoaglin, D., Mosteller, F., & Tukey, J. 1983, Understanding Robust and Exploratory Data Analysis (New York: Wiley)

[61] Hodapp, K., Jensen, J., Irwin, E., et al. 2003, The Gemini Near-Infrared Imager (NIRI), PASP, 115, 814

[62] Howard, A., Marcy, G., Bryson, S., et al. 2012, Planet Occurrence within 0.25 AU of Solar-type Stars from Kepler, ApJS, 201, 15

[63] Howell, S., Rowe, J., Bryson, S., et al. 2012, Kepler-21b: A 1.6 R Earth Planet Transiting the Bright Oscillating F Subgiant Star HD 179070, ApJ, 746, 123

[64] Hsu, D., Ford, E., Ragozzine, D., & Morehead, R. 2018, Improving the Accuracy of Planet Occurrence Rates from Kepler Using Approximate Bayesian Computation, AJ, 155, 205

[65] Hsu, D., Ford, E., Ragozzine, D., & Ashby, K. 2019, Occurrence Rates of Planets orbiting FGK Stars: Combining Kepler DR25, Gaia DR2 and Bayesian Inference, AJ, 158, 109

[66] Huang, X., Bakos, G., & Hartman, J. 2013, 150 new transiting planet candidates from Kepler Q1-Q6 data, MNRAS, 429, 2001

[67] Ida, S. & Lin, D. 2008, Toward a Deterministic Model of Planetary Formation. V. Accumulation Near the Ice Line and Super-Earths, ApJ, 685, 584

[68] Ionov, D., Pavlyuchenkov, Y., & Shematovich, V. 2018, Survival of a planet in short-period Neptunian desert under effect of photoevaporation, MNRAS, 476, 5639

[69] Ishida, E., Vitenti, S., Penna-Lima, M., et al. 2015, cosmoabc: Likelihood-free inference via Population Monte Carlo Approximate Bayesian Computation, Astronomy & Computing, 13, 1

[70] Jackson, B., Stark, C., Adams, E., Chambers, J., & Deming, D. 2013, A Survey for Very Short-Period Planets in the Kepler Data, ApJ, 779, 165

[71] Jenkins, J., Caldwell, D., & Borucki, W. 2002, Some Tests to Establish Confidence in Planets Discovered by Transit Photometry, ApJ, 564, 495

[72] Jenkins, J., Caldwell, D., Chandrasekaran, H., et al. 2010, Overview of the Kepler Science Processing Pipeline, ApJL, 713, L87

[73] Jenkins, J., Twicken, J., Batalha, N., et al. 2015, Discovery and Validation of Kepler-452b: A 1.6 $R_{\oplus}$ Super Earth Exoplanet in the Habitable Zone of a G2 Star, AJ, 150, 56

[74] Jenkins, J., (ed.). 2017, Kepler Data Processing Handbook, KSCI-19081-002

[75] Jurgenson, C., Fischer, D., McCracken, T., et al. 2016, EXPRES: a next generation RV spectrograph in the search for earth-like worlds, Proc. SPIE, 9908

[76] Kasting, J., Whitmire, D., & Reynolds, R. 1993, Habitable zones around main sequence stars, Icarus, 101, 108

[77] Kipping, D., Torres, G., Buchhave, L., et al. 2014, Discovery of a Transiting Planet near the Snow-line, ApJ, 795, 25

[78] Kopparapu, R., Ramirez, R., Kasting, J., et al. 2013, Habitable Zones around Main-sequence Stars: New Estimates, ApJ, 765, 131

[79] Kopparapu, R., Hebrard, E., Belikov, R., et al. 2018, Exoplanet Classification and Yield Estimates for Direct Imaging Missions, ApJ, 856, 122

[80] Kovács, G; Zucker, S; Mazeh, T. 2002, A box-fitting algorithm in the search for periodic transits, A&A, 391, 369

[81] Kunimoto, M., Matthews, J., Rowe, J., & Hoffman, K. 2018, Lifting Transit Signals from the Kepler Noise Floor. I. Discovery of a Warm Neptune, AJ, 155, 43

[82] Kunimoto, M., Matthews, J., & Ngo, H. 2020, Searching the Entirety of Kepler Data. I. 17 New Planet Candidates Including One Habitable Zone World, AJ, 159, 124

[83] Lance, G. & Williams, W. 1967, Mixed-Data Classificatory Programs I - Agglomerative Systems, Australian Computer Journal, 1, 15

[84] Law, N., Morton, T., Baranec, C., et al. 2014, Robotic Laser Adaptive Optics Imaging of 715 Kepler Exoplanet Candidates Using Robo-AO, ApJ, 791, 35

[85] Lillo-Box, J., Barrado, D., Santos, N., et al. 2015, Kepler-447b: a hot-Jupiter with an extremely grazing transit, A&A, 577, A105,

[86] Lintott, C., Schwamb, M., Barclay, T., et al. 2013, Planet Hunters: New Kepler Planet Candidates from Analysis of Quarter 2, AJ, 145, 151

[87] Lissauer, J., Marcy, G., Rowe, J., et al. 2012, Almost All of Kepler's Multiple-planet Candidates Are Planets, ApJ, 750, 112

[88] Lissauer, J., Marcy, G., Bryson, S., et al. 2014, Validation of Kepler's Multiple Planet Candidates. II. Refined Statistical Framework and Descriptions of Systems of Special Interest, ApJ, 784, 44

[89] Lopez, E. & Fortney, J. 2013, The Role of Core Mass in Controlling Evaporation: The Kepler Radius Distribution and the Kepler-36 Density Dichotomy, ApJ, 776, 214

[90] Mandel, K. & Agol, E. 2002, Analytic Lightcurves for Planetary Transit Searches, ApJL, 580, L171

[91] Marcy, G., Butler, R., Fischer, D., et al. 2005, Observed Properties of Exoplanets: Masses, Orbits, and Metallicities, Progress of Theoretical Physics Supplement, 158, 24

[92] Marois, C., Macintosh, B., Barman, T., et al. 2008, Direct imaging of multiple planets orbiting the star HR 8799, Science, 322, 1348

[93] Martin, R., Livio, M., & Palaniswamy, D. 2016, Why are Pulsar Planets Rare?, ApJ, 832, 122

[94] Mathur, S., Huber, D., Batalha, N., et al. 2017, Revised Stellar Properties of Kepler Targets for the Q1-17 (DR25) Transit Detection Run, ApJS, 229, 30

[95] Mayor, M., Marmier, M., Lovis, C., et al. 2011, The HARPS search for southern extra-solar planets XXXIV. Occurrence, mass distribution and orbital properties of super-Earths and Neptune-mass planets, arXiv:1109.2497

[96] Mayor, M. & Queloz, D. 1995, A Jupiter-mass companion to a solar-type star, Nature, 378, 355

[97] Mazeh, T., Holczer, T., & Faigler, S. 2016, Dearth of short-period Neptunian exoplanets: A desert in period-mass and period-radius planes, A&A, 589, 75

[98] McQuillan, A., Mazeh, T., & Aigrain, S. 2014, Rotation Periods of 34,030 Kepler Main-sequence Stars: The Full Autocorrelation Sample, ApJ, 211, 24

[99] Morton, T. 2012, An Efficiency Automated Validation Procedure for Exoplanet Transit Candidates, ApJ, 761, 6

[100] Morton, T. 2015, isochrones: Stellar model grid package, ascl:1503.010

[101] Morton, T. 2015, VESPA: False positive probabilities calculator, ascl:1503.011

[102] Morton, T., Bryson, S., Coughlin, J., et al. 2016, False Positive Probabilities for All Kepler Objects of Interest: 1284 Newly Validated Planets and 428 Likely False Positives, ApJ, 822, 86

[103] Morton, T. & Johnson, J. 2011, On the Low False Positive Probabilities of Kepler Planet Candidates, ApJ, 738, 170

[104] Moutou, C., Almenara, J., Diaz, R., et al. 2014, CoRoT-22 b: a validated 4.9 $R_{\oplus}$ exoplanet in 10-d orbit, MNRAS, 444, 2783

[105] Mulders, G., Pascucci, I., & Apai, D. 2015a, A Stellar-mass-dependent Drop in Planet Occurrence Rates, ApJ, 798, 112

[106] Mulders, G., Pascucci, I., & Apai, D. 2015b, An Increase in the Mass of Planetary Systems Around Lower-Mass Stars, ApJ, 814, 130

[107] Mulders, G., Pascucci, I., Apai, D., & Ciesla, F. 2018, The Exoplanet Population Observation Simulator. I. The Inner Edges of Planetary Systems, AJ, 156, 24

[108] Mulally, F., Coughlin, J., Thompson, J., et al. 2015, Planetary Candidates Observed by Kepler. VI. Planet Sample from Q1–Q16 (47 Months), ApJS, 217, 31

[109] Nesvorny, D., Kipping, D., Buchhave, L., et al. 2012, The Detection and Characterization of a Nontransiting Planet by Transit Timing Variations, Science, 336, 6085

[110] Ngo, H., Knutson, H., Hinkley, S., et al. 2015, Friends of Hot Jupiters. II. No Correspondence Between Hot-Jupiter Spin-Orbit Misalignment and the Incidence of Directly Imaged Stellar Companions, ApJ, 800, 138

[111] Ofir, A. & Dreizler, S. 2013, An Independent Planet Search In The Kepler Dataset. I. A hundred new candidates and revised KOIs, A&A, 555, A58

[112] Owen, J. & Lai, D. 2018, Photoevaporation and high-eccentricity migration created the sub-Jovian desert, MNRAS, 479, 5012

[113] Owen, J. & Wu, Y. 2013, Kepler Planets: A Tale of Evaporation, ApJ, 775, 105

[114] Owen, J. & Wu, Y. 2017, The Evaporation Valley in the Kepler Planets, ApJ, 847, 29

[115] Pascucci, I., Mulders, G., & Lopez, E. 2019, The Impact of Stripped Cores on the Frequency of Earth-size Planets in the Habitable Zone, ApJL, 883, L15

[116] Pecaut, M. & Mamajek, E. 2013, Intrinsic Colors, Temperatures, and Bolometric Corrections of Pre-main-sequence Stars, ApJS, 208, 9

[117] Pepe, F., Mayor, M., Delabre, B., et al. 2000, HARPS: a new high-resolution spectrograph for the search of extrasolar planets, Proc. SPIE, 4008

[118] Pepe, F., Cristiani, S., Lopez, R., et al. 2010, ESPRESSO: the Echelle spectrograph for rocky exoplanets and stable spectroscopic observations, Proc. SPIE, 7735

[119] Petigura, E., Howard, A., & Marcy, G. 2013, Prevalence of Earth-size planets orbiting Sun-like stars, PNAS, 110, 48

[120] Petigura, E., Howard, A., Marcy, G., et al. 2017, The California-Kepler Survey. I. High-resolution Spectroscopy of 1305 Stars Hosting Kepler Transiting Planets, AJ, 154, 107

[121] Petigura, E., Marcy, G., Winn, J., et al. 2018, The California-Kepler Survey. IV. Metal-rich Stars Host a Greater Diversity of Planets, AJ, 155, 89

[122] Pinsonneault, M., Deokkeun, A., Molenda-Zakowicz, J., et al. 2012, A Revised Effective Temperature Scale for the Kepler Input Catalog, ApJS, 199, 30

[123] Price, E. & Rogers, L. 2014, Transit Light Curves with Finite Integration Time: Fisher Information Analysis, ApJ, 794, 92

[124] Raymond, S., Scalo, J., & Meadows, V. 2007, A Decreased Probability of Habitable Planet formation Around Low-Mass Stars, ApJ, 669, 606

[125] Rizzuto, A., Mann, A., Vanderburg, A., Kraus, A., & Covey, K. 2017, Zodiacal Exoplanets in Time (ZEIT). V. A Uniform Search for Transiting Planets in Young Clusters Observed by K2, AJ, 154, 224

[126] Rogers, L. 2015, Most 1.6 Earth-Radius Planets are not Rocky, ApJ, 801, 41

[127] Rowe, J. 2016, Kepler Transit Model Codebase, http://doi.org/10.5281/zenodo.60297

[128] Rowe, J., Bryson, S., Marcy, G., et al. 2014, Validation of Kepler's Multiple Planet Candidates. III. Light Curve Analysis and Announcement of Hundreds of New Multi-planet Systems, ApJ, 784, 45

[129] Rowe, J., Coughlin, J., Antoci, V., et al. 2015, Planetary Candidates Observed by Kepler. V. Planet Sample from Q1-Q12 (36 Months), ApJ, 217, 16

[130] Sanchis-Ojeda, R., Rappaport, S., Winn, J., et al. 2014, A Study of the Shortest-Period Planets Found with Kepler, AJ, 787, 47

[131] Santerne, A., Hebrard, G., Deleuil, M., et al. 2014, SOPHIE velocimetry of Kepler transit candidates XII. KOI-1257 b: a highly eccentric three-month period transiting exoplanet, A&A, 571, A37

[132] Santos, N., Udry, S., Mayor, M., et al. 2003, The CORALIE survey for southern extra-solar planets XI. The return of the giant planet orbiting HD 192263, A&A, 406, 373

[133] Schmitt, J., Wang, J., Fischer, D., et al. 2014A, Planet Hunters. VI. An Independent Characterization of KOI-351 and Several Long Period Planet Candidates from the Kepler Archival Data, AJ, 148, 28

[134] Schmitt, J., Agol, E., Deck, K., et al. 2014B, Planet Hunters. VII. Discovery of a New Low-Mass, Low-Density Planet (PH3 C) Orbiting Kepler-289 with Mass Measurements of Two Additional Planets (PH3 B and D), ApJ, 795, 167

[135] Schwamb, M., Orosz, J., Carter, J., et al. 2013, Planet Hunters: A Transiting Circumbinary Planet in a Quadruple Star System, ApJ, 768, 127

[136] Seager, S. & Mallen-Ornelas, G. 2002, On the Unique Solution of Planet and Star Parameters from an Extrasolar Planet Transit Light Curve, ApJ, 585, 1038

[137] Shallue, C. & Vanderburg, A. 2018, Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90, AJ, 155, 94

[138] Silburt, A., Gaidos, E. & Yanqin, W. 2015, A Statistical Reconstruction of the Planet Population Around Kepler Solar-Type Stars, ApJ, 799, 180

[139] Sousa, S., Santos, N., Mayor, M., et al. 2008, Spectroscopic parameters for 451 stars in the HARPS GTO planet search program, A&A, 487, 373

[140] Still, M. & Barclay, T. 2012, PyKE: Reduction and analysis of Kepler Simple Aperture Photometry data, ascl:1208.004

[141] Stumpe, M., Smith, J., Catanzarite, J., et al. 2014, Multiscale Systematic Error Correction via Wavelet-Based Bandsplitting in Kepler Data, PASP, 126, 110

[142] Szabo, G. & Kiss, L. 2011, A Short-Period Censor of Sub-Jupiter Mass Exoplanets with Low Density, ApJL, 727, L44

[143] Thompson, S., Coughlin, J., Hoffman, K., et al. 2018, Planetary Candidates Observed by Kepler. VIII. A Fully Automated Catalog with Measured Completeness and Reliability Based on Data Release 25, ApJS, 235, 38

[144] Torres, G., Fressin, F., Batalha, N., et al. 2011, Modeling Kepler Transit Light Curves as False Positives: Rejection of Blend Scenarios for Kepler-9, and Validation of Kepler-9 d, a Super-Earth-Size Planet in a Multiple System, ApJ, 727, 24

[145] Torres, G., Kipping, D., Fressin, F., et al. 2015, Validation of 12 Small Kepler Transiting Planets in the Habitable Zone, ApJ, 800, 99

[146] Traub, W. 2012, Terrestrial, Habitable-zone Exoplanet Frequency from Kepler, ApJ, 745, 20

[147] Traub, W. 2016, Kepler exoplanets: a new method of population analysis, arXiv:1605.02255

[148] Tremaine, S. & Dong, S. 2012, The Statistics of Multi-planet Systems, ApJ, 143, 94

[149] Udry, S., Mayor, M., & Santos, N. 2003, Statistical properties of exoplanets. I. The period distribution: Constraints for the migration scenario, A&A, 407, 369

[150] Van der Walt, S. et al. 2011, The NumPy array: a structure for efficient numerical computation, Computing in Science & Engineering, 13, 22

[151] Van Eylen, V., Agentoft, C., Lundkvist, M., et al. 2018, An asteroseismic view of the radius valley: stripped cores, not born rocky, MNRAS, 479, 4786

[152] Vanderburg, A., Latham, D., Buchhave, L., et al. 2016, Planetary Candidates from the First Year of the K2 Mission, ApJS, 222, 14

[153] Wang, J., Fischer, D., Barclay, T., et al. 2013, Planet Hunters. V. A Confirmed Jupiter-size Planet in the Habitable Zone and 42 Planet Candidates from the Kepler Archive Data, ApJ, 776, 10

[154] Wang, J., Fischer, D., Barclay, T., et al. 2015, Planet Hunters. VIII. Characterization of 41 Long-Period Exoplanet Candidates from Kepler Archival Data, ApJ, 815, 127

[155] Wolszczan, A. 1994, Confirmation of Earth-Mass Planets Orbiting the Millisecond Pulsar PSR B1257 + 12, Science, 264, 538

[156] Wolszczan, A. & Frail, D. 1992, A planetary system around the millisecond pulsar PSR1257 + 12, Nature, 355, 145

[157] Wright, J., Upadhyay, S., Marcy, G., et al. 2009, Ten New and Updated Multiplanet Systems and a Survey of Exoplanetary Systems, ApJ, 693, 1084

[158] Wright, J., Fakhouri, O., Marcy, G., et al. 2011, The Exoplanet Orbit Database, PASP, 123, 412

[159] Wu, Y., Murray, N., & Ramsahai, J. 2007, Hot Jupiters in Binary Star Systems, ApJ, 670, 820

[160] Wu, Y. & Lithwick, Y. 2011, Secular Chaos and the Production of Hot Jupiters, ApJ, 735, 109

[161] Xie, J., Dong, S., Zhu, Z., et al. 2016, Exoplanet orbital eccentricities derived from LAMOST–Kepler analysis, PNAS, 113, 11431

[162] Youdin, A. 2011, The Exoplanet Census: A General Method, Applied to Kepler, ApJ, 742, 38

[163] Ziegler, C., Law, N., Morton, T., et al. 2017, Robo-AO Kepler Planetary Candidate Survey. III. Adaptive Optics Imaging of 1629 Kepler Exoplanet Candidate Host Stars, AJ, 153, 66

[164] Ziegler, C., Law, N., Baranec, C., et al. 2018, Robo-AO Kepler Survey. IV. The Effect of Nearby Stars on 3857 Planetary Candidate Systems, AJ, 155, 161

[165] Zink, J., Christiansen, J., & Hansen, B. 2019, Accounting for incompleteness due to transit multiplicity in Kepler planet occurrence rates, MNRAS, 483, 4479

[166] Zink, J. & Hansen, B. 2019, Accounting for Multiplicity in Calculating Eta Earth, MNRAS, 487, 252

Appendix A

Additional AO Observations

As discussed in §3.6.4, I obtained observations for an additional 56 targets across Gemini programs GN-2018B-Q-134 (45 targets with NGS-AO) and GN-2019A-FT-213 (11 targets with LGS-AO) that did not become part of my final candidate list. I present the results of these observations for completeness, showing contrast curve data in Table A.1 and potential companions in Table A.2.

21 of these targets are KOIs that were also observed by Robo-AO and had detected companions. My motivation behind these observations was to compile multiband photometry and confirm the existence of potential companions. Another 12 of the targets are KOIs that have not been observed by Robo-AO. While the corresponding planet candidates I detected have since failed, these observations are still useful for follow-up analysis of known *Kepler* candidates around these stars.

Table A.1: Contrast curve data for all 56 additional targets observed with Gemini NGS-AO and LGS-AO in the K_s band. Only a portion of this table is shown here. textcolorredThe full dataset is available in Kunimoto et al. (2020) [82].

KIC	KOI	Guide Star System	UT Obs. Date	Sep. ($''$)	ΔK_s
7747103	7847	NGS-AO	01 July 2018	0.20	0.38955
				0.35	2.23853
				0.51	3.83136
				0.66	4.6674
				...	...
11350634	8050	NGS-AO	01 July 2018	0.20	0.71592
				0.35	3.93214
				0.51	4.90273
				0.66	5.14522
				...	...

Table A.2: AO results from the additional Gemini North observations, reporting all companions within $4''$.

KIC	KOI	Guide Star System	UT Obs. Date	Comp?	Sep. ($''$)	PA ($^\circ$)	ΔK_s
7747103	7847	NGS-AO	01 July 2018	Y	3.0980±0.0001	2.2708 ± 0.0002	2.276 ± 0.003
11350634	8050	NGS-AO	01 July 2018	Y	0.881 ± 0.001	270.501 ± 0.001	6.64 ± 0.02
7134626	7818	NGS-AO	01 July 2018	N	-	-	-
5894182	7750	NGS-AO	01 July 2018	N	-	-	-

KIC	KOI	Guide Star System	UT Obs. Date	Comp?	Sep. ($''$)	PA ($^\circ$)	ΔK_s
8182107	7870	NGS-AO	11 July 2018	N	-	-	-
11152511	5874	NGS-AO	11 July 2018	N	-	-	-
11360571	2069	NGS-AO	30 July 2018	Y	1.257 ± 0.003	112.737 ± 0.002	2.27 ± 0.02
6938264	4180	NGS-AO	11 Oct 2018	Y	2.4738 ± 0.0001	35.0627 ± 0.0001	1.047 ± 0.001
					3.440 ± 0.0005	37.6856 ± 0.0001	3.763 ± 0.010
10684670	2317	NGS-AO	11 Oct 2018	Y	1.5105 ± 0.0007	113.3984 ± 0.0005	4.28 ± 0.01
7103919	4310	NGS-AO	11 Oct 2018	N	-	-	-
4141593	7685	NGS-AO	11 Oct 2018	Y	1.5573 ± 0.0001	221.4698 ± 0.0001	2.541 ± 0.003
9898447	2803	NGS-AO	17 Oct 2018	Y	3.8182 ± 0.0002	60.4070 ± 0.0001	2.066 ± 0.006
7749773	2848	NGS-AO	03 Nov 2018	Y	2.1854 ± 0.0006	29.8321 ± 0.0003	3.72 ± 0.01
7983117	3214	NGS-AO	03 Nov 2018	Y	0.4855 ± 0.0001	318.3203 ± 0.0001	1.362 ± 0.001
					1.3119 ± 0.0001	199.7957 ± 0.0001	2.222 ± 0.003
6837283	2914	NGS-AO	14 Nov 2018	Y	3.804 ± 0.002	231.2994 ± 0.0004	5.15 ± 0.04
7097965	2083	NGS-AO	14 Nov 2018	Y	0.2517 ± 0.0001	164.7515 ± 0.0003	1.646 ± 0.001
1161345	984	NGS-AO	14 Nov 2018	Y	1.7747 ± 0.0001	41.9867 ± 0.0001	0.1787 ± 0.0008
7449136	1890	NGS-AO	14 Nov 2018	Y	0.4070 ± 0.0001	143.6840 ± 0.0002	2.042 ± 0.002
11869052	120	NGS-AO	14 Nov 2018	Y	1.5793 ± 0.0001	129.5042 ± 0.0001	0.624 ± 0.001
9469494	7938	NGS-AO	06 Dec 2018	Y	0.2917 ± 0.0004	266.04 ± 0.001	3.610 ± 0.002
4252322	396	NGS-AO	08 Dec 2018	Y	1.906 ± 0.003	184.500 ± 0.002	5.76 ± 0.06
10198225	7991	NGS-AO	08 Dec 2018	Y	3.393 ± 0.002	95.5980 ± 0.0006	5.54 ± 0.05
7976520	687	NGS-AO	14 Dec 2018	Y	0.7012 ± 0.0001	12.3221 ± 0.0002	1.360 ± 0.002

KIC	KOI	Guide Star System	UT Obs. Date	Comp?	Sep. ($''$)	PA ($^\circ$)	ΔK_s
10905911	2754	NGS-AO	15 Mar 2019	Y	0.7859 ± 0.0001	260.1817 ± 0.0001	1.564 ± 0.001
5796675	652	NGS-AO	16 Mar 2019	Y	1.239 ± 0001	267.2008 ± 0.0001	0.5812 ± 0.0006
10199984	5776	NGS-AO	21 Mar 2019	N	-	-	-
11401253	4823	NGS-AO	21 Mar 2019	Y	1.3055 ± 0.0001	153.2864 ± 0.0001	0.2808 ± 0.0007
					1.218 ± 0.002	336.697 ± 0.001	4.90 ± 0.01
7287028	7832	NGS-AO	22 Mar 2019	N	-	-	-
10932270	7389	NGS-AO	23 Mar 2019	Y	1.921 ± 0.003	70.205 ± 0.001	5.53 ± 0.06
8332521	4567	NGS-AO	22 May 2019	Y	1.3275 ± 0.0001	141.8448 ± 0.0001	1.619 ± 0.001
8765560	3891	NGS-AO	24 May 2019	Y	1.973 ± 0.001	138.0832 ± 0.0005	4.40 ± 0.02
					0.956 ± 0.003	241.774 ± 0.004	5.79 ± 0.02
4770174	2971	NGS-AO	24 May 2019	Y	0.2350 ± 0.0004	273.012 ± 0.001	3.681 ± 0.002
7190107	-	NGS-AO	31 May 2019	N	-	-	-
3662290	-	NGS-AO	31 May 2019	Y	1.632 ± 0.001	154.4306 ± 0.0006	4.78 ± 0.02
5628770	-	NGS-AO	31 May 2019	Y	1.307 ± 0.003	201.39 ± 0.15	5.35 ± 0.02
6139884	-	NGS-AO	31 May 2019	Y	3.761 ± 0.001	25.79 ± 0.01	2.591 ± 0.003
7020834	-	NGS-AO	31 May 2019	Y	2.4194 ± 0.0006	10.2823 ± 0.0003	4.29 ± 0.01
11565976	-	NGS-AO	31 May 2019	Y	0.8245 ± 0.0002	162.8190 ± 0.0003	2.827 ± 0.004
3345775	-	NGS-AO	01 June 2019	N	-	-	-
7186892	-	NGS-AO	01 June 2019	Y	0.8192 ± 0.0001	179.0674 ± 0.0001	0.282 ± 0.001
9823433	-	NGS-AO	01 June 2019	N	-	-	-
12505309	-	NGS-AO	02 June 2019	N	-	-	

KIC	KOI	Guide Star System	UT Obs. Date	Comp?	Sep. ($''$)	PA ($^\circ$)	ΔK_s
3531436	-	NGS-AO	11 June 2019	N	-	-	-
6380164	-	NGS-AO	12 June 2019	Y	2.752 ± 0.004	257.99 ± 0.08	6.65 ± 0.09
8172679	-	NGS-AO	12 June 2019	N	-	-	-
2985262	-	LGS-AO	30 June 2019	N	-	-	-
4551429	-	LGS-AO	30 June 2019	N	-	-	-
4569091	-	LGS-AO	30 June 2019	Y	3.7120 ± 0.0001	245.5372 ± 0.0001	1.557 ± 0.002
5803540	-	LGS-AO	30 June 2019	Y	2.5206 ± 0.0001	294.8547 ± 0.0001	2.318 ± 0.002
7119412	-	LGS-AO	30 June 2019	N	-	-	-
9274173	-	LGS-AO	30 June 2019	N	-	-	-
11092463	-	LGS-AO	30 June 2019	Y	0.7069±0.0002	247.1112±0.0003	2.936±0.004
5095499	-	LGS-AO	01 July 2019	N	-	-	-
12307455	-	LGS-AO	01 July 2019	N	-	-	-
4346258	-	LGS-AO	03 July 2019	N	-	-	-
6937870	-	LGS-AO	03 July 2019	N	-	-	-

Appendix B

ExoPAG SAG13 Recommended Grids

Table B.1 gives my occurrence rate results for FGK-, F-, G-, and K-type stars over the ExoPAG SAG13 recommended period-radius grid. The median and 68.3% credible interval of each ABC posterior is shown. For bins with zero planet detections, I report only the upper limit (84.1th percentile).

Table B.1: Occurrence rate results for F-, G-, and K-type stars over the ExoPAG SAG13 recommended period-radius grid. Results are given in % (10^{-2}).

Period (days)	Radius ($R_\oplus$)	F (%)	G (%)	K (%)
$10.0 - 20.0$	$0.67 - 1.00$	$0.22^{+0.26}_{-0.16}$	$0.5^{+0.39}_{-0.31}$	$0.76^{+0.85}_{-0.52}$
$10.0 - 20.0$	$1.00 - 1.50$	$1.83^{+0.66}_{-0.58}$	$2.72^{+0.7}_{-0.71}$	$4.8^{+1.53}_{-1.49}$
$10.0 - 20.0$	$1.50 - 2.25$	$1.83^{+0.6}_{-0.52}$	$6.21^{+1.13}_{-0.99}$	$10.75^{+2.47}_{-2.19}$
$10.0 - 20.0$	$2.25 - 3.38$	$2.19^{+0.52}_{-0.51}$	$6.4^{+1.08}_{-0.84}$	$9.32^{+2.02}_{-1.85}$
$10.0 - 20.0$	$3.38 - 5.06$	$0.35^{+0.24}_{-0.19}$	$1.09^{+0.5}_{-0.44}$	$0.89^{+0.76}_{-0.54}$
$10.0 - 20.0$	$5.06 - 7.59$	$0.34^{+0.24}_{-0.18}$	$1.07^{+0.45}_{-0.43}$	$0.86^{+0.72}_{-0.54}$
$10.0 - 20.0$	$7.59 - 11.39$	$0.17^{+0.14}_{-0.1}$	$0.27^{+0.23}_{-0.16}$	$0.52^{+0.48}_{-0.32}$
$10.0 - 20.0$	$11.39 - 17.09$	$0.1^{+0.1}_{-0.06}$	$0.25^{+0.2}_{-0.15}$	$0.55^{+0.46}_{-0.33}$
$20.0 - 40.0$	$0.67 - 1.00$	$0.22^{+0.26}_{-0.16}$	$0.44^{+0.49}_{-0.27}$	$1.4^{+1.45}_{-0.91}$
$20.0 - 40.0$	$1.00 - 1.50$	$0.46^{+0.36}_{-0.25}$	$1.08^{+0.68}_{-0.57}$	$5.05^{+2.35}_{-1.89}$
$20.0 - 40.0$	$1.50 - 2.25$	$1.78^{+0.64}_{-0.6}$	$3.58^{+1.19}_{-1.02}$	$9.13^{+3.08}_{-2.59}$
$20.0 - 40.0$	$2.25 - 3.38$	$4.81^{+0.82}_{-0.77}$	$10.89^{+1.68}_{-1.55}$	$10.18^{+2.52}_{-2.18}$
$20.0 - 40.0$	$3.38 - 5.06$	$0.53^{+0.43}_{-0.31}$	$0.97^{+0.61}_{-0.53}$	$1.24^{+1.19}_{-0.81}$
$20.0 - 40.0$	$5.06 - 7.59$	$0.57^{+0.45}_{-0.33}$	< 1.59	$1.15^{+1.16}_{-0.74}$

Period (days)	Radius ($R_\oplus$)	F (%)	G (%)	K (%)
$20.0 - 40.0$	$7.59 - 11.39$	< 0.69	$0.16^{+0.24}_{-0.12}$	$1.14^{+0.98}_{-0.66}$
$20.0 - 40.0$	$11.39 - 17.09$	$0.13^{+0.23}_{-0.09}$	$0.37^{+0.27}_{-0.22}$	< 1.49
$40.0 - 80.0$	$0.67 - 1.00$	< 0.84	< 1.07	$1.32^{+1.69}_{-0.95}$
$40.0 - 80.0$	$1.00 - 1.50$	$0.29^{+0.31}_{-0.2}$	$0.61^{+0.61}_{-0.41}$	$3.13^{+2.28}_{-1.73}$
$40.0 - 80.0$	$1.50 - 2.25$	$0.56^{+0.49}_{-0.34}$	$4.64^{+1.59}_{-1.37}$	$10.08^{+3.63}_{-3.23}$
$40.0 - 80.0$	$2.25 - 3.38$	$3.78^{+0.96}_{-0.95}$	$12.29^{+2.08}_{-1.9}$	$10.65^{+3.71}_{-3.17}$
$40.0 - 80.0$	$3.38 - 5.06$	$1.22^{+0.73}_{-0.6}$	$1.74^{+0.96}_{-0.85}$	$0.95^{+1.18}_{-0.67}$
$40.0 - 80.0$	$5.06 - 7.59$	$1.23^{+0.78}_{-0.63}$	$1.73^{+1.05}_{-0.83}$	< 2.2
$40.0 - 80.0$	$7.59 - 11.39$	$0.85^{+0.52}_{-0.41}$	$0.48^{+0.46}_{-0.31}$	$0.78^{+1.09}_{-0.58}$
$40.0 - 80.0$	$11.39 - 17.09$	$0.24^{+0.25}_{-0.16}$	$0.47^{+0.47}_{-0.3}$	< 2.45
$80.0 - 160.0$	$0.67 - 1.00$	< 2.03	< 2.77	< 4.34
$80.0 - 160.0$	$1.00 - 1.50$	< 1.21	$0.86^{+1.09}_{-0.56}$	$2.01^{+1.83}_{-1.28}$
$80.0 - 160.0$	$1.50 - 2.25$	$1.13^{+0.94}_{-0.71}$	$3.74^{+1.79}_{-1.65}$	$7.46^{+4.31}_{-3.24}$
$80.0 - 160.0$	$2.25 - 3.38$	$4.25^{+1.64}_{-1.26}$	$10.87^{+2.57}_{-2.31}$	$11.44^{+4.45}_{-3.91}$
$80.0 - 160.0$	$3.38 - 5.06$	$0.67^{+0.68}_{-0.45}$	$2.74^{+1.51}_{-1.2}$	$1.85^{+2.03}_{-1.25}$
$80.0 - 160.0$	$5.06 - 7.59$	$0.6^{+0.56}_{-0.4}$	$2.65^{+1.51}_{-1.16}$	$2.08^{+2.12}_{-1.47}$
$80.0 - 160.0$	$7.59 - 11.39$	$0.96^{+0.68}_{-0.5}$	$0.87^{+0.79}_{-0.57}$	< 2.97
$80.0 - 160.0$	$11.39 - 17.09$	$0.99^{+0.69}_{-0.51}$	$2.29^{+1.2}_{-1.0}$	$1.11^{+1.49}_{-0.79}$
$160.0 - 320.0$	$0.67 - 1.00$	< 6.03	< 7.67	< 12.51
$160.0 - 320.0$	$1.00 - 1.50$	< 2.08	$1.78^{+1.8}_{-1.17}$	$3.62^{+4.17}_{-2.43}$
$160.0 - 320.0$	$1.50 - 2.25$	< 1.73	$2.44^{+2.17}_{-1.43}$	$4.95^{+5.24}_{-3.13}$
$160.0 - 320.0$	$2.25 - 3.38$	$2.54^{+1.58}_{-1.22}$	$9.26^{+3.02}_{-2.79}$	$17.12^{+9.42}_{-6.86}$
$160.0 - 320.0$	$3.38 - 5.06$	$1.15^{+1.07}_{-0.77}$	$2.27^{+1.79}_{-1.37}$	$3.63^{+4.11}_{-2.44}$
$160.0 - 320.0$	$5.06 - 7.59$	$1.1^{+1.1}_{-0.73}$	$2.3^{+1.93}_{-1.36}$	< 7.18
$160.0 - 320.0$	$7.59 - 11.39$	$1.47^{+1.1}_{-0.8}$	$3.46^{+2.11}_{-1.7}$	$2.1^{+3.55}_{-1.52}$
$160.0 - 320.0$	$11.39 - 17.09$	$0.77^{+0.85}_{-0.51}$	$3.65^{+1.95}_{-1.69}$	< 7.77
$320.0 - 640.0$	$0.67 - 1.00$	< 45.0	< 34.14	< 50.7
$320.0 - 640.0$	$1.00 - 1.50$	< 10.9	< 11.68	< 24.36
$320.0 - 640.0$	$1.50 - 2.25$	< 6.61	$4.72^{+5.15}_{-3.13}$	< 20.17
$320.0 - 640.0$	$2.25 - 3.38$	< 5.24	$6.57^{+6.21}_{-4.1}$	$19.06^{+15.63}_{-11.29}$
$320.0 - 640.0$	$3.38 - 5.06$	< 6.82	$6.88^{+5.54}_{-4.3}$	$13.62^{+14.93}_{-8.71}$

Period (days)	Radius ($R_\oplus$)	F (%)	G (%)	K (%)
$320.0 - 640.0$	$5.06 - 7.59$	$2.99^{+4.35}_{-2.19}$	$6.87^{+5.49}_{-4.52}$	< 29.49
$320.0 - 640.0$	$7.59 - 11.39$	< 7.23	$7.1^{+6.29}_{-4.56}$	< 17.87
$320.0 - 640.0$	$11.39 - 17.09$	< 6.48	< 11.45	$7.07^{+9.97}_{-5.21}$