Title
Utilizing Kepler and K2 to Advance Exoplanet Demographics

Author
Zink, Jonathon Kenneth

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Utilizing *Kepler* and *K2* to Advance Exoplanet Demographics

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Astronomy and Astrophysics

by

Jonathon Kenneth Zink

2021

ABSTRACT OF THE DISSERTATION

Utilizing *Kepler* and *K2* to Advance Exoplanet Demographics

by

Jonathon Kenneth Zink
Doctor of Philosophy in Astronomy and Astrophysics
University of California, Los Angeles, 2021
Professor Bradley M. S. Hansen, Chair

To understand the uniqueness of our solar system, we must assess the system architectures and demographic features existing in the known planet population, but such a task requires a homogeneous set of candidates. Fortunately, the *Kepler* spacecraft continuously collected photometry from a single patch of the sky, which in turn produced a well characterized catalog of transiting exoplanets. However, previous studies assumed multi-planet systems were subject to the same selection effects as their single-planet counterparts. I investigate this assumption, finding that a proper completeness accounting significantly increases the underlying occurrence of multi-planet systems to 5.86 ± 0.18 planets per FGK dwarf and only requires a $1.5°$ average system mutual inclination. Using this correction I provide an updated extrapolation of the occurrence of Earth analogs and find that $5.9 \pm 0.5\%$ of GK dwarfs have more than one planet within their habitable zone.

Additionally, the *K2* mission collected photometry from 18 fields along the ecliptic plane, providing a unique opportunity to understand how exoplanet formation may be affected by galactic latitude, stellar metallicity, and stellar age. For my thesis, I developed a fully automated pipeline able to detect and vet transit signals in *K2* photometry, enabling assessment

of sample completeness and reliability. This catalog contains 768 planets, with 235 newly identified candidates. Correspondingly, I present the first uniform analysis of small transiting exoplanet occurrence outside of the *Kepler* field, finding a metallicity dependence in small planet occurrence.

I also discuss how the sun's late stage evolution and the existing solar system architecture will capture Jupiter and Saturn into a mean-motion resonance. Eventually perturbations from passing stars will trigger large-scale instability, culminating in the disassociation of all outer planets. Extrapolating this result to other systems indicates a temporal dependence on bound planet occurrence in the Galaxy.

The dissertation of Jonathon Kenneth Zink is approved.

Jean-Luc Margot

Hilke E. Schlichting

Erik A. Petigura

Bradley M. S. Hansen, Committee Chair

University of California, Los Angeles

2021

For Samantha.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I thank Professor Brad Hansen for guiding me through the research presented in my thesis and for his help navigating the perils of academia; Dr. Jessie Christiansen for inspiring my fervor of exoplanet demographics and for playing a supportive role in my scientific explorations; Professor Erik Petigura for challenging me to be a better science communicator and researcher through refinement of my oral and written presentation skills; and Professor Konstantin Batygin for his insightful discussion of solar system dynamics and his guidance in understanding the broader implications of my research.

I also want to thank my family for their love and support through my pursuit of higher education. My beloved wife, Samantha, for her constant encouragement to pursue a happy and meaningful life; without her I would not have completed this thesis. My parents, Sue and Ken, for enabling me to follow my passions and for fostering my intellectual aspirations. My sister, Kate, for always being smarter than me, despite my 5 year seniority, and for tirelessly supporting me. My nephew, Lukas, who I have not yet met, but very much look forward to sharing the wonders of the cosmos with.

I want to thank all my fellow astronomy grad students at UCLA, who have been an invaluable resource for generations of institutional knowledge and who have always been relatable and supportive through out this process. Despite being a re-entry student, I always felt welcome to engage and participate in necessary-for-mental-health social activities and have made lasting connections that will extend long beyond my residency at UCLA.

I acknowledge that the findings presented in this thesis are based on published manuscripts with additional coauthors. In particular, Chapter 2 is a version of Zink et al. (2019b), Chapter 3 is a form of Zink & Hansen (2019), Chapter 4 is a version of Zink et al. (2020a) with excerpts from a paper currently in preparation with additional coauthors, Chapter 5 is a version of Zink et al. (2020b), and Chapter 6 is taken from Zink et al. (2020c).

VITA

2014	**B.S., Astrophysics**, University of California, Los Angeles
2016	**M.S., Physics**, California State University, Northridge
2016–2020	**Teaching Assistant**, Department of Physics and Astronomy, University of California, Los Angeles
2018	**M.S., Astronomy**, University of California, Los Angeles
2020–present	**Graduate Student Researcher**, Department of Physics and Astronomy, University of California, Los Angeles

PUBLICATIONS

On the Potentially Dramatic History of the Super-Earth Rho 55 Cancri e

Bradley Hansen and Jon Zink 2015, MNRAS, 450, 4505-4520

Accounting for Incompleteness due to Transit Multiplicity in Kepler Occurrence Rates

Jon Zink, Jessie Christiansen, and Bradley Hansen 2019, MNRAS, 483, 4479-4494

Catalog of New K2 Exoplanet Candidates from Citizen Scientists

Jon Zink, Kevin Hardegree-Ullman, Jessie Christiansen, et al. 2019, RNAAS, 3, 2, 43

Accounting for Multiplicity in Calculating Eta Earth

Jon Zink and Bradley Hansen 2019, MNRAS, 487, 246–252

Scaling K2. I. Revised Parameters for 222,088 K2 Stars and a K2 Planet Radius Valley at 1.9 $R_\oplus$

Kevin Hardegree-Ullman, Jon Zink, Jessie Christiansen, et al. 2020, ApJS, 247, 1, 28

Scaling K2. II. Assembly of a Fully Automated C5 Planet Candidate Catalog Using EDI-Vetter

Jon Zink, Kevin Hardegree-Ullman, Jessie Christiansen, et al. 2020, AJ, 159, 4, 154

Scaling K2. III. Comparable Planet Occurrence in the FGK Samples of Campaign 5 and Kepler

Jon Zink, Kevin Hardegree-Ullman, Jessie Christiansen, et al. 2020, AJ, 160, 2, 94

The Great Inequality and the Dynamical Disintegration of the Outer Solar System

Jon Zink, Konstantin Batygin, and Fred Adams 2020, AJ, 160, 5, 232

CHAPTER 1

Introduction

As the *Voyager 1* space probe jettisoned out of the solar system, it captured an image of the Earth from a distance of 6 billion kilometers. Notoriously known as the ”Pale Blue Dot” image, this photo depicts the Earth as a small hardly noticeable speck when viewed from the vantage point of the outer solar system (Sagan, 1994). As we continue our search for life beyond the Earth, it's important that we first understand how unique this small blue dot is in our vast Galaxy. In a broader context, we must grasp the distinctive features that make the solar system capable of harboring such life. Buried beneath this mystery lies two un-resolved questions: (1) what are the underlying physical mechanisms responsible for the formation of planets and (2) how do these orbits evolve over time. It appears our system has a rich dynamic history, likely sculpted by the migration of Jupiter and Saturn (Chiang et al., 2002; Tsiganis et al., 2005; Hansen, 2009), making it difficult to unveil the system's true genesis. Therefore, we must examine external planetary systems, which provide unique snapshots into the formation and evolution processes. Through study of these architectures we can piece together our origin story and understand our place in the Galaxy.

My thesis focuses on using the known population of planetary systems to uncover the features attributed to formation and dynamics. In Chapter 2 I will describe the automated detection software associated with the NASA *Kepler space telescope*, and its importance in creating a homogeneous sample of planets for occurrence rate calculations. In this work I identify a multiplicity completeness correction needed for demographic analysis of the *Kepler* multi-planet sample. This finding invalidates previous completeness assumptions and helps explain an anomaly in the observed multiplicity distribution. In Chapter 3 I use the corrections discussed in Charter 2 to examine the impact on our measurements of Earth-analog occurrence. In Chapter 4 I describe the automated pipeline I assembled to combat the *K2* systematics, culminating in a homogeneous catalog of *K2* planets with corresponding measurements of completeness and reliability. In Chapter 5 I select a subset of this catalog to perform a comparative occurrence estimate with the *Kepler* planet catalog. In Chapter 6, I discuss the eventual fate of the solar system and its impact on planet occurrence. I then provide concluding remarks in Chapter 7.

1.1 Exoplanet Detection

In order to use these external systems, as a tracer of multi-planet formation and dynamics, a sample of extrasolar planets (exoplanets) must be identified. The first detection of a planet orbiting a main-sequence star, like our own, was achieved by measuring variations in the radial velocity (RV) of the host star, 51 Pegasi (Mayor & Queloz, 1995). This Jovian mass planet (51 Pegasi b) was identified with an orbital period of 4.2 days, sharply contrasting the 12 year orbit of the solar system's shortest-period gas-giant, Jupiter.

The RV planet detection technique requires periodic spectral measurements of the host star over a span of time comparable to the planet's orbital period. Doppler shifts in these element absorption and emission lines are indicative of stellar acceleration (Campbell, 1983). Since planet's produce a gravitational pull on their stellar host as they mutually orbit around the system's center of mass (barycenter), strong periodic trends in the host RVs are signature of an orbital companion. Additionally, the overall change in velocity over the signal's period provides information on the planets mass:

$$K \sim \frac{M_{\text{Star}}}{M_{\text{Planet}}} \left(\frac{1}{P}\right)^{\frac{1}{3}}. \tag{1.1}$$

Here, K is the semi-amplitude of the RV signal, P is the planets orbital period, and M

represents the respective mass of the star and planet. Additional information on the planet's orbital eccentricity can also be attained for well-sampled RV signals. This method of detection requires the planet's orbital plane be near the line of sight, maximizing the amplitude of the star's relative velocity changes. Detection capabilities reduce for more orthogonal orientations, and in the extreme case of a completely face-on orbit, no RV signal will be produced. Furthermore, this technique is limited in its ability to detect low-mass ($< 50 M_{\oplus}$) long-period (> 100 day) planets —similar to the planets in the inner solar system— due to their weak RV signals. Despite these limitations, nearly 800 exoplanets had been identified using this methodology (Akeson et al., 2013). To identify and characterize smaller exoplanets, other detection techniques have been developed.

One such method is the transit method, which entails monitoring the brightness of stars for apparent dips. These brightness reductions correspond to a planet passing in front of the star (transiting) and blocking a small fraction of the expected stellar flux (Borucki & Summers, 1984). Upon detection, the depth of this event describes the radius ratio of the planet and star:

$$\Delta F \sim \left(\frac{R_{\mathrm{Planet}}}{R_{\mathrm{Star}}}\right)^2. \tag{1.2}$$

Here, ΔF is the change in observed flux during transit and R is the radius of the host star and planet respectively. Unlike RVs, this method is strictly limited to systems oriented such that the planet passes along our line of sight with respect to the host star (edge-on), significantly reducing the number of detectable systems. Moreover, these transit events are rare (once per orbit) and require continuous monitoring to avoid missing the brief (lasting only a few hours) occultation events. However, this method remains fruitful due to the large number of stars that can be observed simultaneously, overcoming these chance detection limitations through large scale surveys. With very precise photometry the discovery of small planets ($< 2R_{\oplus}$) around distant stars is achievable. Thus, transiting planet represent the largest fraction of known exoplanets, including more than 3,000 known candidates (Akeson et al., 2013).

1.2 Exoplanet Occurrence Rates

Calculating the occurrence of an underlying planet population remains difficult due to the inherit biases in each detection method and the small number of detectable planet signals. In brief, planet occurrence (f) can be calculated as follows:

$$f = \frac{N_{\text{Planets}}}{N_{\text{Stars}}} \frac{1}{\eta}, \tag{1.3}$$

where N_{Planets} is the number of planets detected and N_{Stars} is the number of stars surveyed. η represents the sample completeness fraction and will be 1 if all the existing planets are detected. Therefore, it's essential that selection effects be thoroughly quantified. Thus, there are four main elements required to carry out a meaningful exoplanet occurrence rate calculation: (1) a well parameterized selection of stars that can potentially host detectable planets; (2) a homogeneous sample of planets, with well characterized detection procedures, hosted by stars in the selected stellar sample; (3) corresponding measurements of the planet sample false-negative rate (sample completeness) and (4) the samples false-positive rate (sample reliability). These ingredients can be combined to identify the underlying occurrence of planets around the surveyed stars.

The stellar target list for statistical studies is often limited by proximity. Our ability to collect photons is largely dependent on the stars distance and brightness. More distant cool stars provide reduced photon collection capabilities, making subtle changes difficult to parse. For photometric transit surveys, longer cadences are required to account for this lack of signal. However, for very dim targets the background noise reaches parity with the signal, rendering any planet detection attempts moot, which reaches just beyond ~ 3000 light-years for space-based surveys (Borucki et al., 2010). This problem is greatly amplified for RV surveys, where the signal must be distributed across the star's spectra, limiting the potential stellar sample to targets within ~ 1000 light-years for most ground-based instruments (Howard et al., 2010; Mayor et al., 2011). Thus, our current sample of planets, and all population inferences, are limited to our local galactic neighborhood.

Planet detection biases are technique dependent, but overall smaller (and less massive) planets with long periods are more difficult to identify. Understanding the observational features that make these planets difficult to detect is essential for extrapolation to regions of parameters space with low detection efficiency. The physical limits for detection (Transit - the probability that the planet does occult the star; RV - the probability that the planet is not in a face-on, orthogonal orbital, configuration) have simple analytic equations, which make correcting straight forward. However, the probability of detection is both instrument and sample dependent, making such calculations challenging (Cumming et al., 2008; Fressin et al., 2013a). Furthermore, many signals require some amount of human intervention to ensure the detection is meaningful, complicating bias calculations. The development of automated detection algorithms reduced this complication, by providing a uniform methodology that could be exhaustively tested for sample completeness (Petigura et al., 2013a; Christiansen et al., 2015). Additionally, this automation enables testing for false-positives, which indicates the purity of the planet sample (Coughlin et al., 2016). If a majority of the planet signals are specious, the ability to extract meaningful statistics will be extremely limited. Thus, surveys must optimize their algorithm to maximize completeness and reliability to achieve the most accurate statistical results.

1.3 The *Kepler* spacecraft

Ground-based transit surveys face two major problems. First, photometric collection is only accessible during the night, leaving large gaps in the time series where the transit may or may not have occurred. Thus, a larger number of transits may be needed to accurately pinpoint the planet's orbital period. Second, the transit of a smaller rocky planet around a solar radius star will only produced a flux difference of 0.01%, requiring precise photometric measurements. Even for the brightest stars, with high signal strength, the Earth's atmospheric variations, the instrument instability, and photon counting variations contribute to fluctuations on order 0.5% (Pepper et al., 2003) for large ground-based surveys, making it difficult to examine this meaningful region of parameter space. In order to detect an Earth analog the instrument must be place in orbit, where noise contributions are significantly reduced and continuous observing is achievable.

In March of 2009 the *Kepler* spacecraft was launched with the primary goal of measuring the occurrence of Earth analogs. Focused on one small region, accounting for only 0.25% of the sky, for 4 years, *Kepler* provided continuous photometry for $\sim 200,000$ stars (Borucki et al., 2010; Koch et al., 2010). This uniform sampling enabled the identification of over $3,000$ transit candidates, making the transit method the most prominent source of exoplanet detections (Thompson et al., 2018). Unfortunately, the mission concluded upon a malfunction in two reaction-wheels, which nullified the spacecraft's ability to remain focused on a single patch of the sky for an extended period. Since the identification of an Earth radius and period planet required slightly more than 4 years of baseline photometry, direct measurements of analogous planet occurrence was not fully realized. Despite this limited data spans, several studies have extrapolated to find an occurrence of 0.1-0.4 Earth-like planets per Sun-like star (Catanzarite & Shao, 2011; Petigura et al., 2013a; Silburt et al., 2015; Traub, 2012, 2016; Bryson et al., 2020a).

Since the main goal of this mission was to characterize the occurrence of Earth analogs, the field was selected such that a majority of the targets would be Sun-like stars (Brown et al., 2011). This skewed sample is not representative of the galactic stellar abundance, which largely consists of smaller M-dwarfs. While solar system analog observing preferences can provide meaningful insight, sampling a wider stellar parameters space helps identify features that are important in planet occurrence, and, by proxy, planet formation. Fortunately, such photometry was collected following the conclusion of the prime mission.

1.3.1 The *K2* Mission

The spacecraft was equipped with four reaction wheels, which use angular momentum conservation to control the telescope pointing. It is essential that all three axis of rotation can be manipulated to maintain the telescopes field alignment. On July 2012, the first of these four reaction wheels failed. With three wheels still in working condition the spacecraft remained stable, meeting the minimum three-axis requirement. Unfortunately, a second wheel malfunctioned in May of 2013, enabling the continuous stream of solar radiation to slowly reorient the telescope away from the *Kepler* field. Without stable pointing, the *Kepler* prime mission concluded.

With fuel remaining in the tank and an operational telescope, an alternative plan was established to maximize the spacecraft output. In May 2016 the *K2* mission was approved. Using the remaining two reaction wheels and focusing on targets along the ecliptic plane, the radiation pressure drift could be minimized (Howell et al., 2014). Corrections for the remaining drift could be achieved with a thruster boost every 6 hours (Van Cleve et al., 2016). Through this mission, 18 different campaigns (fields) across the sky where continuously observed for 80 day periods. Each campaign delivered a unique glimpse of a different part of the local galaxy, enabling consideration of how differences in stellar age, mass, and composition may play a role in exoplanet occurrence. After more than two years of observing with only two working reaction wheels (9 years in total), the spacecraft officially ran out of fuel and was retired in October of 2018. Overall, nearly $1,000$ additional transiting candidates where discovered through this secondary mission, providing an additional subset of planets ripe for demographic analysis.

1.3.2 Demographics

One of the most challenging aspects of estimating an occurrence rate is understanding the sample completeness. The automation provided by the *Kepler* pipeline produced a systematic method of detecting transiting exoplanets and thus offers the prospect of a rigorous determination of the survey completeness (Thompson et al., 2018). Here, artificial transits were injected into the *Kepler* light curves, and the recovery fraction was used to understand the *Kepler* detection efficiency (Christiansen et al., 2015; Christiansen, 2017; Christiansen et al., 2020). Using the output of this analysis, inferences about the underlying exoplanet population where achieved (FGK Dwarfs: Youdin 2011; Howard et al. 2012; Petigura et al. 2013b; Mulders et al. 2018; Hsu et al. 2019; He et al. 2019; M Dwarfs: Dressing & Charbonneau 2013; Hardegree-Ullman et al. 2019), finding that a majority of the existing short-period planets are smaller than $2R_{\oplus}$.

Kepler also identified roughly 500 multi-planet systems, most with astonishing uniformity in planet size and orbital spacing, colloquially known as the "Peas-in-a-Pod" finding (Ciardi et al., 2013; Millholland et al., 2017; Weiss et al., 2018a). The formation of these optimized configurations suggests some mass and orbital-regularization process must be at play (Adams, 2019). Additional demographic features have been identified in the *Kepler* photometry, indicating underlying formation mechanisms. The apparent deficit of planets between $1.5-2R_{\oplus}$ (Fulton et al., 2017), separating super-Earths from sub-Neptunes is the apparent outcome of planetary envelope removal from either stellar photoevaporation (Owen & Wu, 2017) or core-powered mass-loss (Gupta & Schlichting, 2019). Furthermore, the high stellar flux "Sub-Neptune Desert", appears to be a symptom of photoevaporation, producing a deficit of short-period large-envelope planets (Lopez & Fortney, 2013). These stellar-dependent processes remain untested due to the focused stellar sample of *Kepler*. Fortunately, the *K2* mission uniformly monitored a wider dynamical range of stars across the ecliptic, enabling assessment of these formation theories by varying stellar parameters.

Before the additional candidates identified in the *K2* campaigns can be incorporated in to a demographic analysis, the sample must undergo some uniform detection procedure. The vetting software developed for the *Kepler* catalog (`RoboVetter`; Thompson et al. 2018) was not optimized for the unique systematic issues experienced by the spacecraft during the *K2* mission, which created artificial dips that mimic transit signals. For my thesis, I examined subtle completeness features in the *Kepler* automated detection pipeline and compiled an analogous automated software for the *K2* photometry, enabling combined demographic analysis for the first time.

CHAPTER 2

Accounting for Incompleteness due to Transit Multiplicity in Kepler Planet Occurrence Rates

2.1 Introduction

The *Kepler* mission has revolutionized our understanding of the frequencies and properties of planets around Sun-like stars. With the final data release DR25, providing all of the data up until the failure of two reaction wheels (Mathur et al., 2017), the primary phase of the project has officially concluded. Within this span, *Kepler* has provided evidence for $\sim 4,500$ transiting exoplanets.[1] Nearly 50% of these candidates have been confirmed or validated (Rowe et al., 2014; Morton et al., 2016), demonstrating that planets are common and widespread in the Milky Way.

There have been many attempts to quantify the frequency of planetary systems and the properties (radius and orbital period) of the planets themselves (Borucki et al., 2011; Catanzarite & Shao, 2011; Youdin, 2011; Howard et al., 2012; Batalha et al., 2013; Fressin et al., 2013b; Petigura et al., 2013a; Dong & Zhu, 2013; Dressing & Charbonneau, 2013; Mullally et al., 2015; Dressing & Charbonneau, 2015; Burke et al., 2015; Mulders et al., 2015; Silburt et al., 2015), with a special attention given to attempting to characterize the frequency of planets with Earth-like properties. One of the most challenging aspects of estimating these

[1] `https://exoplanetarchive.ipac.caltech.edu`

occurrence rates is understanding the completeness of the known exoplanet sample. The automation provided by the *Kepler* pipeline has produced a systematic method of detecting transiting exoplanets and thus offers the prospect of a rigorous determination of the survey completeness. With 3.5 years of nearly continuous light curves of $\sim 200,000$ stars, it is possible to investigate period ranges out to 500 days. Furthermore, the high precision of the *Kepler* light detector has permitted the discovery of planets with radii $r < 1r_{\oplus}$.

Since the completion of the *Kepler* survey, several studies have used this data set to extract population parameters. Petigura et al. (2013a), using their own *TERRA* pipeline, implemented an *Inverse Detection Method*, where the population CDF (Cumulative Distribution Function) is divided by the detection efficiency. This study also introduced the idea of synthetic planet injections into the *Kepler* light curves to map completeness. Here, artificial transits were injected into the *Kepler* light curves, and the recovery fraction in the *TERRA* pipeline was used to understand the *Kepler* detection efficiency. To avoid confusion from multiple planet transits the Petigura et al. (2013a) occurrence rate calculation only included the highest SNR (Signal to Noise Ratio) planet in each system, ignoring any multiplicity. To characterize the official *Kepler* completeness, Christiansen et al. (2015) performed a pixel-level transit injection test to empirically measure how well the pipeline would detect various types of planets. This is discussed in more detail in Section 2.4. The results of this study were then used by Burke et al. (2015) to perform a *Poisson Process Analysis*, where a Bayesian framework is implemented to determine the best population model parameters. The current work employs a similar method.

Planet multiplicity introduces detection biases above and beyond those to which single transit systems are subject. When faced with a system of multiple transiting planets, the *Kepler* pipeline will typically find the largest MES (Multiple Event Statistic; comparable to SNR) signal, fit the transit function, and then discard the corresponding data points. The width of discarded data is 3× the transit duration, with 1.5× removed on each side of the transit center. Very few TTVs (Transit-Timing Variations) are large enough to escape this window. Such deletion is necessary to avoid confusion when looking for additional planets, but introduces data gaps into the light curve as noted by Schmitt et al. (2017). These gaps becomes more invasive in higher multiplicity systems where significant data is being discarded. With each planet removed, the available data set shrinks. This effect creates "swiss cheese"-like holes in the light curves, where the number of holes increases with each detected planet. Beyond possible gaps in the light curve, the *Kepler* pipeline fails to detect some short-period planets because of a harmonic fitting function (Christiansen et al., 2013). Here the pipeline attempts to remove sinusoidal variations in the light curve caused by stellar activity, but in doing so, the procedure can overfit a true planet signal and make low SNR planets difficult to detect. To clarify, the baseline wobble from the dataset is removed using a spline smoothing function. The harmonic fitting function is specifically looking for sinusoidal variations in the light curve. This function may or may not be applied, depending on whether the pipeline is able to detect such periodic variation in the light curve. In multiple-planet systems, the harmonic fitter can also overfit the periodic variations caused by transits and remove true signals. Because the pipeline follows these procedures, the order of planet detection can affect it's detectability.

Our goal in this paper is to assess the effect of planet multiplicity and detection order on the completeness of the Kepler results. In Section 2.2 and 2.3 we describe our methods of stellar and planet selection. In Section 2.4 we show that detection order affects the detection efficiency for a given planet. In Section 2.5 we describe how we account for mutual inclination within this study. In Section 2.6, we lay out our process of accounting for overall detection efficiency. In Section 2.7 we present our expanded likelihood function used to calculate the posterior for the population parameters. In Section 2.8, we discuss the results of our fitting method and the implications of our multiplicity parameters. We provide concluding remarks in Section 2.9.

2.2 Stellar Selection

Using the final release of *Kepler* data (DR25) which includes Q1-Q17, we select a stellar sample for use in creating a detection efficiency map that accounts for *Kepler* completeness. We use the stellar parameters provided by Mathur et al. (2017) with improved radius values derived from *Gaia* DR2 (Berger et al., 2018). The updates from *Gaia* DR2 have yet to provide updated corresponding mass values. Thus we must still utilize the *Kepler* DR25 stellar mass parameters (200,038 stars in total). To focus on the occurrence of planets around solar-like GK dwarfs, we only include stars with $T_{\rm eff} > 4200, K$ and $T_{\rm eff} < 6100K$ (135,494 stars remain). It is also important for completeness mapping that each star has a stellar radius and mass measurement available. “Null” values for either of these fields result in omission (133,056 stars remain). To avoid the inclusion of giants we limit the sample to $log(g) \geq 4$ and $R_{\star} \leq 2R_{\odot}$ (96,167 stars remain). We also place requirements on the duty cycle (f_{duty}) and the time length of the light curve ($data_{span}$). These are $f_{duty} > 0.6$ and $data_{span} > 2$ years are made (86,679 stars remain). The f_{duty} limit requires that 60% of $data_{span}$ has been collected. This ensures that a significant portion of the light curve is filled, while still including stars lost in the Q4 CCD loss (Batalha et al., 2013). Time-varying noise measurements have been provided in the DR25 dataset through a value known as CDPP (Combined Differential Photometric Precision; Christiansen et al. 2012). This parameter has been calculated for every field star over 14 different time periods: 1.5, 2.0, 2.5, 3.0, 3.5, 4.5, 5.0, 6.0, 7.5, 9.0, 10.5, 12.0, 12.5, and 15.0 hours (Mathur et al., 2017). These values correspond to the amount of noise a planet signal will need to exceed, given a transit duration, to generate a 1σ detection. By requiring stars to have a $\mathrm{CDPP}_{7.5h} < 1000$ ppm, we minimize the inclusion of stellar and instrumental fluctuations (74 stars exceed this limit). From this we produce a stellar sample of 86,605 solar-like stars.

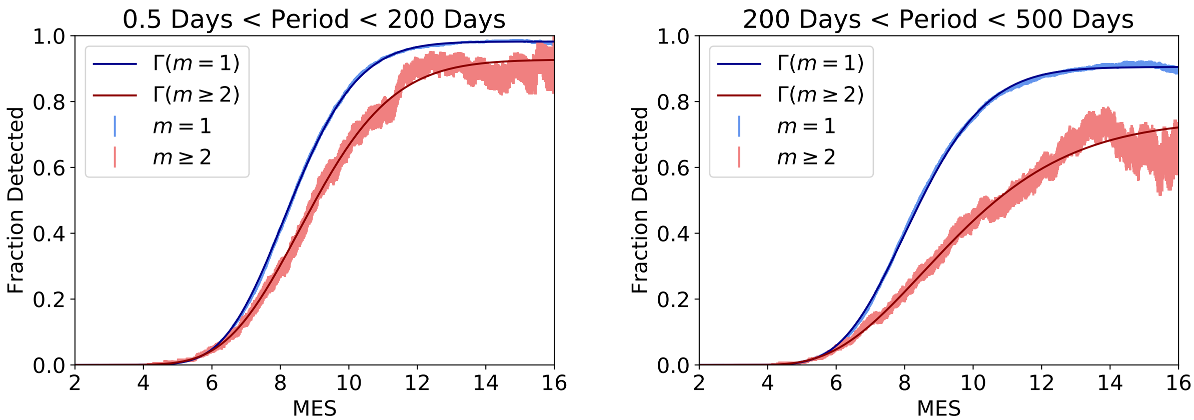

Figure 2.1: The smoothed recovery fraction at each MES bin. The vertical lines (light blue and red) represent the uncertainty in each bin under the assumption of a binary distribution. The bin values are plotted at the center of each bin. The solid lines (dark blue and red) represent the Γ_{CDF} distribution fit. The parameters of this model were fit using a χ^2 minimization.

2.3 Planet Selection

When available we utilize the updated planetary parameters provided by the California *Kepler* Survey (CKS) (Petigura et al., 2017; Johnson et al., 2017) and the asteroseismic updates provided by Van Eylen et al. (2018). One of the main advantages for the inclusion of these updates is the improved planet radius measurements. Since our study, like others, does not account for parameter uncertainty, such improvements are essential for accurate occurrence rates. Where CKS and asteroseismic data are unavailable, the measurements provided by the *Kepler* DR25 catalog (Thompson et al., 2018), in conjunction with the *Gaia* DR2 radius updates (Berger et al., 2018), are implemented. Through private communication, it was indicated that this early release of *Gaia* data may contain some planet radius outliers. To combat this issue, we test the radius values against the *Kepler* DR25 catalog. When the updated *Gaia* measurements differ from the *Kepler* DR25 data by $> 3\sigma$, we utilize the *Kepler* DR25 radius measurements. Overall, 19 planets exceed this outlier limit (statistically, we would expect only 8). All period measurements are drawn from the light curves; thus, improved measurements from *Gaia* and CKS have no effect on the inferred period measurements. We use the periods provided in the *Kepler* DR25 catalogs. Both the CKS and *Kepler* DR25 provide flags for false positives. We include data from both CONFIRMED and CANDIDATE planets in DR25 and $CKS_{fp} = FALSE$ in the CKS update. To further avoid contamination from false positives, we only include planets with periods $.5 < p < 500$ days and radii $.5 < r < 16r_{\oplus}$. Periods beyond 500 days have been noted to be highly contaminated by false positives because they barely meet the three transit limit of the pipeline (Mullally et al., 2015).

Our period and radii range exceeds the conservative cutoffs adopted by many previous studies, but is necessary when exploring the effects of multiplicity. Often planetary systems span the entire range of the *Kepler* parameter space, thus the inclusion of nearly all the planets is needed for an accurate calculation. There exist 3 multi-planet systems (KIC: 3231341, 11122894, and 11709124) where one planet within the system fall beyond the range of this study. We only select the planets from these system that lie within our radius and period cuts. The inclusion of these planets is useful in providing a stronger statistical argument. Although some of the known planets, in these 3 systems, extend beyond the bounds of this study, we expect many other systems within the dataset to contain planets beyond the range of our selection bounds. Furthermore, if we include the planets that lay beyond our radius and period cuts, our analysis we will artificially inflate the number of inferred planets within this range.

The accuracy of the *Kepler* detection order ("TCE Planet Number") can be affected by systems with existing false positives. When removing these data points, we manually ensure that the detection order only reflects the order in which valid KOIs (*Kepler* Objects of Interest) are detected. For example, a system with 5 "real" KOIs and 1 false positive would have detection orders ranging from 1-5 regardless of order at which the false positive was detected. It should be noted that these false positives do create cuts in the data, similar to that of a planet and therefore affect the detection order. However, without reordering these systems we artificially inflate our multiplicity calculation in Section 2.7. Higher multiplicities are especially sensitive to mild increases as their detection probabilities are very low. Further discussion in Section 2.4 shows that we use the same detection efficiency for all planets found after the first detected planet, thus only planets artificially being re-assigned to 1 are of concern. Since most false positives provide relatively weak signals, only 14 systems experience this artificial re-ordering. After making the discussed cuts we find that the highest detection order existing in the parameter space is 7. This means that the highest system multiplicity we consider in this study is a 7 planet system. We find 3062 KOIs meet the indicated period and radius requirements.

It has been suggested that gas giants eject companion planets while migrating inward (Beaugé et al., 2012). Their large Hill radius forces the planets to become unstable as the Hill radius ratio falls below 10. These hot Jupiters create an independent population of single planet systems (Steffen et al., 2012). If it forms via a distinct channel, this population has the ability to skew the inferred distribution of the model for the generic underlying population. To minimize such contamination, we remove all single planet systems with $r > 6.7r_\oplus$ as indicated by Steffen et al. (2012). Further evidence of this independent population was discussed by Johansen et al. (2012), who showed that multi-planet systems with one planet of mass > 0.1 Jupiter mass are dynamically unstable on short timescales. This 0.1 Jupiter mass limit roughly corresponds to the $r = 6.7r_\oplus$ limit used here. We find that 120 of these single hot Jupiters exist in the dataset, leaving us with 2942 KOIs that fit all the parameter requirements described. Our final catalog of planets and their corresponding parameters can be found online.[2]

[2] https://github.com/jonzink/ExoMult

Table 2.1: The Γ function parameters used to fit the recovery CDF displayed in Figure 2.1.

Period Range	Maximum Detection (c)	Shape (a)	Scale (b)	Offset (x_0)
$\Gamma_{CDF}^{m=1}$				
$.5 < p < 200$ days	0.9825	29.3363	0.2856	0.0102
$200 < p < 500$ days	0.9051	18.4119	0.3959	1.0984
$\Gamma_{CDF}^{m\geq 2}$				
$.5 < p < 200$ days	0.9276	21.3265	0.4203	0.0093
$200 < p < 500$ days	0.7456	5.5213	1.2307	2.9774

2.4 Injection Recovery

Here we shall discuss how we can account for the detection efficiency as a function of detection order. Christiansen (2017) injected artificial planet signals into the calibrated pixels of each of the *Kepler* field stars and processed the altered light curves with the standard detection pipeline. This allows the recovery fraction to be assessed, producing a probability function based on transit MES (Multiple Event Statistic; a detailed description of MES can be found in equation 2.14). A Γ_{CDF} (Cumulative Distribution Function) was fit to the empirical probability of recovery, of the form:

$$\Gamma_{CDF}(MES) = \frac{c}{b^a(a-1)!}\int_0^{MES}(x-x_0)^{a-1}e^{\frac{-(x-x_0)}{b}}dx. \tag{2.1}$$

The purpose of this test was to establish an average detection efficiency function for the *Kepler* pipeline as determined by the properties of the target star sample. Therefore planet detection order was not considered. However, many of the target stars are known to host real KOIs, and these signals will remain in the Christiansen (2017) analysis. This provides an opportunity to consider the effects of detection order on recovery. Here we define detection order by the variable m, where m=1 indicates the first planet discovered in the system (i.e. highest MES). Likewise, planet m=2 and m=3 corresponds to the second and third planets found by the *Kepler* pipeline. The highest detection order existing in the parameter space is 7, thus we shall work in the range of $m = 1 : 7$.

We split the data from Christiansen (2017) into injection with a $.5 < p < 200$ days or $200 < p < 500$ days. The break at 200 days was selected by testing different values. Beyond 200 days, we find that the distributions begin to change significantly. To focus on the relevant parameter space of our study, we remove all injections with periods beyond 500 days and only consider stars within $4200K < T_{\text{eff}} < 6100K$ and $log(g) \geq 4$.

Because the goal for the original Christiansen (2017) experiment was to find an overall detection probability, only one artificial signal was injected into each light curve with a radius and period uniformly sampled from $.25 - 7r_{\oplus}$ and $.25 - 500$ days respectively. To understand the effect of multiple planets we therefore need to investigate systems with existing transit signals in the light curve. Over 30,000 unique signals are found within the *Kepler* data pipeline. Although most of these were later deemed false positives by external checks, the pipeline treats them no differently than an actual planet. It is even very likely that some of them are in fact "real" planets. Therefore, injections in these systems will be subject to the same systematic issues as that of an actual multiple-planet system. This offers a far greater number of $m \geq 2$ injections than those provided by the KOI list alone. For the system with injected signals, for $200 < p < 500$ days we find 2,099 $m \geq 2$ systems and 1,579 $m \geq 2$ systems for $.5 < p < 200$ days. We also separated the injections into $m \geq 3$, but this data is extremely limited and cannot produce meaningful results without further injections. Thus, we shall focus only on m= 1 and $m \geq 2$ systems. The data are then binned in MES and the recovery fraction is determined at each binned region of MES space. Because the available data is relatively small compared to the original number of primary injections (31,302 for $200 < p < 500$ days and 29,083 for $.5 < p < 200$ days), a smoothing technique is utilized. The bin width is set to 2 MES, but instead of moving each bin by steps of width 2, the bins were recalculated at steps of 0.01 MES. This produced 800 data points across a parameter space of 0-16 MES. Utilizing this technique avoids artifacts produced when binning smaller data samples. One issue that can arise from such smoothing is an artificial distribution skew. In acknowledging this possibility, we have tested various bin widths while smoothing and find little deviation from the results with the adopted binning. Since each injection within a bin can have two possible outcome, a detection or a failed detection, the distribution within each bin will follow a binomial model. Here the number of trials corresponds to the number of injections within the bin. Thus, the uncertainty for each bin is calculated assuming a binomial distribution. The recovery CDF is then fit with a 4-parameter Γ distribution using a χ^2 minimization. The results of the fit can be seen in Table 2.1 and Figure 2.1.

One of the main motivations for creating these additional detection efficiency curves was a preliminary search of the results of the Christiansen (2017) injection test. This showed that 61 previously detected KOIs were lost when the injection of additional planets was made. Thirty-nine of these KOI planets had a low "Disposition Score", indicating that small perturbation to the light curve could easily disrupt their detectability. One KOI was lost because of transit interference, where a higher MES injection with overlapping transits caused some of the transits of the weaker signal planet to be missed. Twenty-one systems had indicated that the harmonic fitting function was triggered when the injection was made, likely overfitting to the transits themselves. Ten of these light curves had no detections of planets at all. Both the injection and the KOI were missed when the artificial planet was placed into the system. This indicates that multiple-planet systems are constrained by additional detection biases not experienced by single planet systems.

2.5 Effects of mutual inclination

Here, we shall discuss how the effects of mutual inclination are handled within our model. The initial recovery study (Christiansen, 2017) was performed without consideration of higher multiplicity planets. Thus, there was no accounting for mutual inclination. The artificial planets were injected with a random impact parameter (b) from 0 to 1. To understand the effects of mutual inclination on detection efficiency we look at the difference of impact parameters (Δb) for recovered planet systems. Δb is calculated by taking the difference of the artificial planet and the largest MES KOI impact parameter in each system. Since an existing KOI is required for this test, we only look at systems with known planets. We find that the detected planets do not significantly differ in Δb than the difference of two randomly drawn populations of b values. Because the artificial planets were injected with uniformly drawn impact parameters, we conclude that the Δb, and therefore mutual inclination, plays an insignificant role in detection efficiency. However, larger mutual inclinations can cause certain planets to geometrically avoid transit completely.

2.5.1 Transit Probability

Analytic models of transit probability have been found for double transit systems as a function of mutual inclination (Ragozzine & Holman, 2010). However, larger multiplicity systems are more difficult and require semi-analytic models (Brakensiek & Ragozzine, 2016). In order to simplify our calculation, we simulate various semi-major axis to stellar radius ratios ($a_p/R_\star$) and look at 10^6 lines of sight to predict the probability of transit. To determine the period population we need a function for m transit probability at some semi-major axis value (a_p). In order to create a function for probability of transit in addition to $m-1$ other transits, it is essential that we know the distributions of exoplanet periods. Clearly, this argument is circular in nature. We deal with this issue by using a non-uniform method of sampling from the empirical period population. This is performed for detection order $m = 2:7$, since the analytic probability ($R_\star/a_p$) is sufficient for m=1.

To establish the desired detection order, the required number of planets are drawn from the empirical *Kepler* period data. For example, when looking at the case of m=3, ($a_p/R_\star$) is selected and then the two additional planets are drawn from the known *Kepler* period sample. The periods of the additional two planets are redrawn at each line of sight. This is the same as saying we marginalized the additional two planets over the *Kepler* period population. In order to properly account for the transit probability of higher detection orders, we need to know the unbiased underlying populations of periods. To approximate this, we sample the empirical distribution of *Kepler* planet periods, but weighted with a probability $\propto p^{2/3}$. This is done to account for the geometric bias against the detection of longer period planets. To account for the mutual inclination between orbits, we follow the σ_σ distribution provided by Fang & Margot (2012). This mild distribution ($\langle\sigma\rangle = 1.6^o$) was found by looking at the impact parameter ratios within *Kepler* systems. Once all orbits have been selected, the number of lines of sight where all planets transit is divided by 10^6 to establish the transit probability. To determine whether a planet is transiting this equation must be satisfied:

$$\begin{aligned} & cos(i) * cos(\omega) - sin(i) * sin(\omega) \geq R_\star/a_p \\ \text{or} \quad & sin(i) * sin(\omega) - cos(i) * cos(\omega) \leq R_\star/a_p, \end{aligned} \tag{2.2}$$

where i is the inclination of the of the system and ω is the ascending node. Each line of sight is drawn uniformly over $sin(i)$ and the nodes of each orbit are also drawn uniformly over $sin(\omega)$. For nodes between planets within the same system, we sample uniformly over $sin(\Delta\omega)$. We note that Equation 2.2 is only valid for circular orbits. Consideration of eccentric orbits is presented in Section 2.8.5. To avoid the creation of unstable systems, we check the planet separations ($|a_{p2} - a_{p1}|$). If any separation is $< 10\%$ the semi-major axis of the outer planet we resample the entire system. This process is repeated until no separations fall below the 10% threshold. Although mutual Hill radius would provide a better measure of stability, our metric requires no assumptions about the mass of the planets. Furthermore, we find that changing (or removing) this threshold makes little difference to the probabilities calculated, indicating that stability accounting has little effect statistically. The results of this simulation can be seen in Figure 2.2.

It is worth noting that equation 2.2 does not account for grazing transits. To properly account for this, $R_\star$ must becomes $R_\star \pm r$, where r is the radius of the transiting planet. Using a uniform distribution of r values from $0.5r_\oplus$ to $16r_\oplus$, we find that grazing transits provide an increase of 0.2% to the overall transit probability. However, this uniform distribution is weighted far more heavily towards large planets than the underlying planet radius distribution, thus we expect the true correction to be much smaller. To properly account for grazing transit one must have some understanding of the underlying radius population. Any attempt to do so here would add more uncertainty to the calculation and provide a very minimal correction. Thus, we ignore such complications here.

2.6 Detection Efficiency Grid

To represent the *Kepler* survey detection efficiency a grid is created in period and radius space. Both $log_{10}p$ and $log_{10}r$ are divided into 100 bins, creating 10,000 regions of the parameter space. For every region we uniformly sampled in log space for period and radius, all 86,605 stars are assigned m planets based on the detection order of interest. For example, in the detection grid for the first transiting planet (m=1), the probability of detecting at least one planet is calculated at each bin. Similarly for m=2, the probability of detecting at least one planet at each bin in addition to finding another planet in some other arbitrary bin. The average detection probability for each region is calculated using these planetary assignments and the procedures provided in the next Sections (2.6.1; 2.6.2). This process is then repeated for each of the 10,000 regions. We calculated 7 detection efficiency grids: first planet probability (m=1), second planet probability (m=2),..., and the seventh planet probability (m=7). This procedure is similar to that of Burke et al. (2015) and Traub (2016), but now with 7 different detection order grids.

2.6.1 Probability of Detection for m=1

We shall begin with the formula for the detection of the first planet and then discuss the modifications made for the detection of higher order systems. In our base model we assume all planets have perfectly circular orbits and consider the effects of eccentricity in Section 2.8.5. This assumption of little or no eccentricity is reasonable for the typical multiple systems

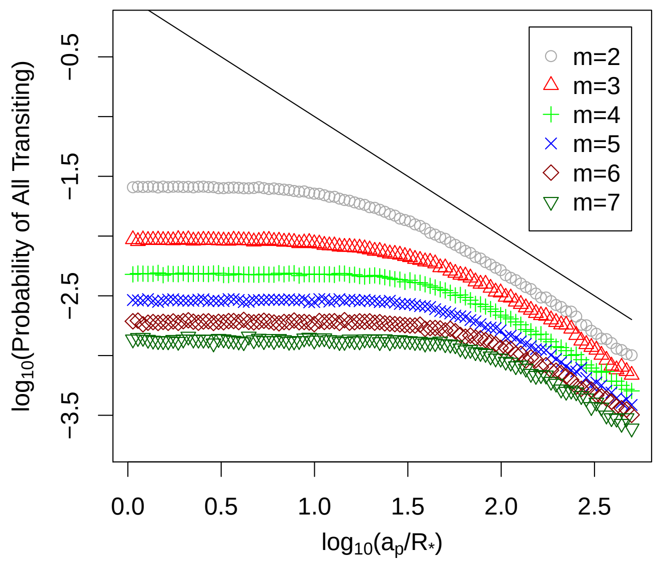

Figure 2.2: The probability for transit of high multiplicity systems using the Fang & Margot (2012) mutual inclination model. The solid black line represents the probability function used for an $m = 1$ planet transit ($R_\star/a_p$). A machine-readable version of this data is available online.

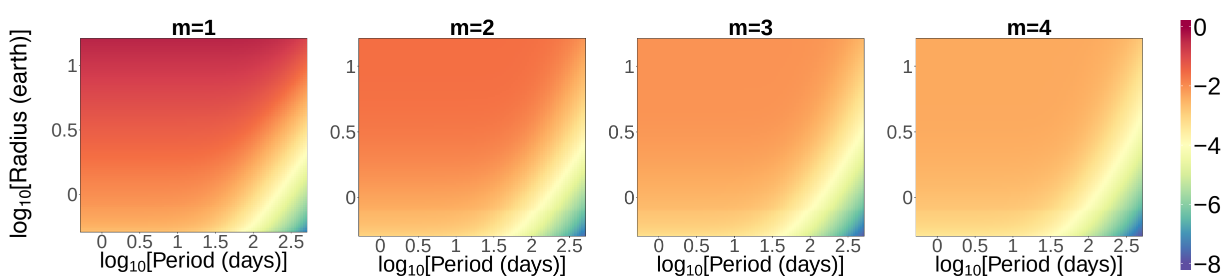

Figure 2.3: The detection efficiency maps for m=1:4 exoplanet discovery orders. The color map is representative of log_{10}(Detection Probability). The fading of color across detection order (m) shows the decreasing detection probability. A machine-readable version of this data is available online.

sampled by *Kepler*, where non-circular orbits would result in unstable system architecture. To account for the geometric probability of transit we use:

$$P_{tr} = \frac{R_\star}{a_p}, \tag{2.3}$$

where $R_\star$ is the radius of the star and a_p is the semi-major axis of the planet orbit. The chord at which the planet transits across the stellar host is given by

$$f_{tr} = \sqrt{1 - b^2}, \tag{2.4}$$

where b is the impact parameter of the planet transit. b is assigned by uniformly sampling between 0 and 1 for each planet. The duration of the transit can be calculated as

$$t_{dur} = \frac{R_\star * f_{tr}}{a_p * \pi}(\frac{p}{1\text{day}}) * 24\text{hr}, \tag{2.5}$$

where p is the orbital period of the planet. The expected number of transits can be found with

$$n_{tr} = \frac{data_{span}}{p}, \tag{2.6}$$

where $data_{span}$ is the span of the data within the *Kepler* survey. Because of various shut downs and data downloads throughout the *Kepler* mission, it is possible that some of the transits may have been missed. To account for the probability of the transit occurring in the window of the *Kepler* mission we adopt the window function provided by Burke et al. (2015).

$$j = \frac{data_{span}}{p} \tag{2.7}$$

$$\begin{aligned} P_{win} = 1 - (1 - duty)^j - j * duty(1 - duty)^{j-1} \\ - \frac{j(j-1)duty^2(1 - duty)^{j-2}}{2}, \end{aligned} \tag{2.8}$$

where $duty$ is the duty fraction of the targeted stellar source. The *Kepler* pipeline requires at minimum 3 transits for candidate consideration; P_{win} is the probability that at least 3 transits will be detected by the available *Kepler* data. Since most targets have a $duty = .95$, short period transits ($j >> 3$) produce a P_{win} nearly 1 and approach 0 as $j < 3$. Almost all of our sample have data throughout the full data set span of 1458.931 days. The mean $data_{span}$ for this study is 1427.445 days.

Other studies have used various way to account for the effects of limb darkening such as that of Claret & Bloemen (2011). We attempt to mimic the pipeline by looking at the empirical limb darkening values chosen for existing KOIs (with the same stellar parameters discussed in Section 2.2). We find that the two limb darkening parameters (u_1, u_2) used to fit planet transits within the pipeline are strongly correlated to stellar temperature ($T_{\rm eff}$). The best fit line to this correlation is as follows:

$$\begin{aligned} u_1 &= -1.93 * 10^{-4} * T_{\rm eff} + 1.5169 \\ u_2 &= 1.25 * 10^{-4} * T_{\rm eff} - 0.4601. \end{aligned} \tag{2.9}$$

We warn that these correlations mimic the choice of the pipeline rather than the true stellar features and should not be used for more evolved stars with $log(g) < 4$. With the given calculated parameters, it is now possible to calculated the expected MES of the *Kepler* pipeline as presented by Burke & Catanzarite (2017).

$$k_{rp} = \frac{r}{R_\star} \tag{2.10}$$

$$c_0 = 1 - (u_1 + u_2) \tag{2.11}$$

$$\omega = \frac{c_0}{4} + \frac{u_1 + 2 * u_2}{6} - \frac{u_2}{8} \tag{2.12}$$

$$\begin{aligned} depth = 1 - (\frac{c_0}{4} + \frac{(u_1 + 2 * u_2) * (1 - k_{rp}^2)^{\frac{3}{2}}}{6} \\ - \frac{u_2(1 - k_{rp}^2)}{8})\omega^{-1} \end{aligned} \tag{2.13}$$

$$MES = \frac{depth}{CDPP * 10^6} * 1.003 * n_{tr}^{\frac{1}{2}}, \tag{2.14}$$

where $CDPP$ is in ppm from the *Kepler* stellar catalog, interpolated by the transit duration. Finally, we account for the systematic detection efficiency using the Gamma distribution CDF described in Section 2.4.

$$P_{tip}^{m=1} = \Gamma_{CDF}^{m=1}(MES), \tag{2.15}$$

where the parameters for Γ_{CDF} are the given in Table 2.1. Combining all of the discussed probabilities provides an estimate of the detection likelihood of the highest MES planet within the system. This probability is given as follows:

$$P_{det}^{m=1} = P_{tr} * P_{win} * P_{tip}^{m=1}, \tag{2.16}$$

This equation provides a metric for understanding the bias of the highest MES planet. This probability is dependent on detection order and we shall now discuss in the next section how higher multiplicity planets ($m \geq 2$) can be accounted for.

2.6.2 Probability of Detection for m=2

For $m \geq 2$ planets we follow much of what is described in the previous Section (2.6.1), with a few mild changes to better model the differences in detection probability.

We change the transit probability to reflect the probability of m planets transiting, accounting for the probability of finding this planet with at minimum $m-1$ other planets. To best capture the probabilities of our simulation in Section 2.5.1, we interpolate between simulated data points for the transit probability.

$$P_{tr}^{m} = \text{Linear Interpolate}(m, \frac{a_p}{R_\star}) \tag{2.17}$$

For example, if we are looking at a planet with m=3 (the third planet detected) with $a_p/R_\star = 32$, we would expect a transit probability of ~ 0.008. This can be clearly seen in the data provided by Figure 2.2. Since no such simulated value exist at this exact point, we interpolate between the the two neighboring estimations to establish this value. Here we use the new detection efficiency for higher m planets.

$$P_{tip}^{m\geq 2} = \Gamma_{CDF}^{m\geq 2}(MES) \tag{2.18}$$

$$P_{det}^{m} = P_{tr}^{m} * P_{win} * P_{tip}^{m\geq 2}, \tag{2.19}$$

where equation 2.8 is again used for P_{win}. In reality, there are differing window functions for each detection order; when tested, we find that $\approx 0.4\%$ of the light curve is lost with the addition of each planet. One can see that varying the *duty* parameter of equation 2.8 by even 3% has negligible effects on the P_{win} value. Because the detection efficiency is the same for $m \geq 2$, the only difference between the $m = 2 : 7$ probability maps is the transit probability. This now produced 7 distinct detection grids ($m = 1 : 7$). The first four grids can be seen in Figure 2.3. The detection order of the exoplanet in question will dictate which grid is most appropriate for application.

To summarize, we have described how the recovery probabilities (CDF) are a function of detection order (m). We use this to create 7 different detection efficiency maps (Figure 2.3). In order to create a map for m=1 planets, we sample across planet period and radius space. Doing so, we calculate the probability of detection and averaged over all stars within the *Kepler* stellar sample. We expand upon this idea, creating a map for m=2 planets. Here the new recovery CDF is implemented to account for the additional loss of planets at higher detection orders. Furthermore, we account for the probability of two planets within the system transiting using a mild mutual inclination model (Figure 2.2). Jumping from m=1 to m=2 we lose an additional 5.5% and 15.9% of the planets for periods $<$ 200 days and periods $>$ 200 days respectively. This is due to properties of the pipeline when fitting multiple transit systems. This procedure is repeated for m=3:7 each accounting for the appropriate number of transiting planets according to the data in Figure 2.2 (3-7 respectively). There is an additional loss of nearly 70% at each respective discovery order due to the unlikely event of multiple orbital alignment with our line of sight. It is clear that these two factors, geometric transit likelihood and pipeline recovery, have a significant effect on the multiplicity extracted from the *Kepler* data set.

2.7 The Likelihood Function

Using the efficiency grids derived in the previous section, we can infer properties of the underlying planetary population. Here we will discuss the likelihood function required to

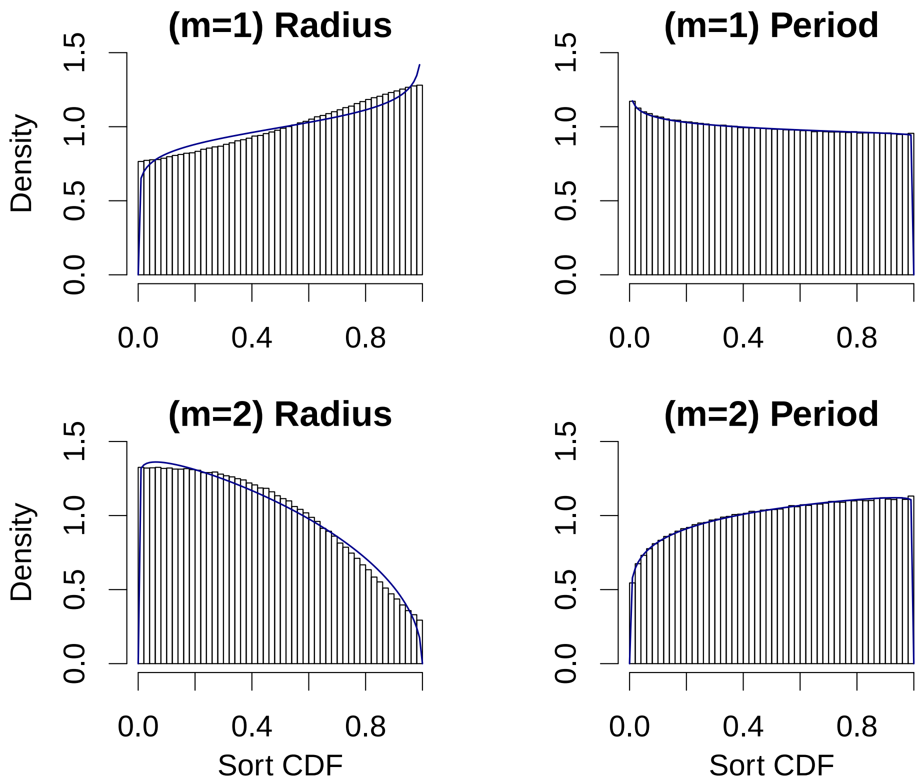

Figure 2.4: The sorting simulation for m=1 and m=2. The solid blue line represents the Beta distribution fit to the respective data set. The boxes are a histogram of the simulated data after being sorted. It is apparent that sorting has a more dramatic effect on radius than period. This is expected as $MES \propto r^2/p^{1/3}$. A mild deviations from the model is noted in the radius skew. This discrepancy dissolves as we move into higher detection orders. Furthermore, the effects of these deviations are insignificant, given the cuts on duty cycle, data span, and stellar type already made.

Table 2.2: The parameters found in testing the sorting effects of MES. These parameters correspond to a Beta distribution skew expected for the CDF of each multiplicity population.

	m=1	m=2	m=3	m=4	m=5	m=6	m=7
a_{rad}	1.095	1.030	1.028	1.013	0.998	1.065	0.951
b_{rad}	0.923	1.470	2.206	3.063	4.013	4.898	6.614
a_{per}	0.957	1.152	1.172	1.184	1.183	1.166	1.234
b_{per}	1.004	1.010	1.000	0.999	0.997	1.006	0.994

implement Bayes theorem and extract these population parameters.

We adopt the approach of previous studies (e.g. Youdin 2011; Petigura et al. 2013a; Burke et al. 2015), modeling the underlying population as characterized by independent power-law distributions in period and radius. We also make explicit the assumption that there is a single planetary population – assuming that systems which show only one transit are drawn from the same underlying distribution as those which show multiple transits. We will examine the validity of this assumption in Section 2.8.1. Our focus on multiple systems also means that we include more of the *Kepler* parameter space than was used in most previous papers.

The population of exoplanets is modeled as follows:

$$\frac{d^2N}{dpdr} = fg(p)q(r) \tag{2.20}$$

$$g(p) = \begin{cases} C_{p1}p^{\beta_1} & \text{if } p < p_{br} \\ C_{p2}p^{\beta_2} & \text{if } p \geq p_{br} \end{cases} \tag{2.21}$$

$$q(r) = \begin{cases} C_{r1}r^{\alpha_1} & \text{if } r < r_{br} \\ C_{r2}r^{\alpha_2} & \text{if } r \geq r_{br}, \end{cases} \tag{2.22}$$

where $f, \alpha_1, \alpha_2, \beta_1, \beta_2, p_{br}$, and r_{br} are all fit parameters. We require continuity at r_{br} and p_{br} through the normalization constants for $q(r)$ and $g(p)$.

Our method expands on the Poisson process likelihood used by Youdin (2011). The main difference is the separation of planets by detection order (m). In doing so, we require different occurrence factors (f) for each m, increasing the required number of parameters. Previous studies such as Burke et al. (2015) have used a single occurrence value, providing an average occurrence factor. By separating the occurrence factor as a function of detection order, we can allow for differences in detection efficiency while simultaneously fitting for the occurrence of planet multiplicity.

$$Likelihood = \prod_{m=1}^{7}[\prod_{i=1}^{n_m} f_m \eta_m(p_i, r_i) g(p_i) q(r_i)] e^{-N_m} \tag{2.23}$$

$$N_m = 86,605 f_m \int_{.5\text{days}}^{500\text{days}} \int_{.5r_\oplus}^{16r_\oplus} \eta_m(p, r) O_m(p_i, r_i) g(p) q(r) dr dp, \tag{2.24}$$

where N_m represents the expected number of planets detected for each discovery order (m) and f_m is an occurrence factor for each m. This value provides information on the occurrence of each m multiplicity. However, to find meaningful information from these values, they must be disentangled from each other as discussed in Section 2.7.0.3. The 86,605 accounts for the number of stars in our test sample and $\eta_m(p, r)$ is the detection probability at the given detection order. The function $O_m(p_i, r_i)$ is the sorting order correction for the (PDF) Probability Distribution Function. This function is necessary to account for the bias in the PDF introduced by our sorting in terms of detection order (discussed further in 2.7.0.2).

It is often more useful to consider the natural log of the likelihood, which can be simplified:

$$ln(Likelihood) \propto \sum_{m=1}^{7}[\sum_{i=1}^{n_m} ln(f_m g(p_i) q(r_i))] - N_m. \tag{2.25}$$

Using the $ln(Likelihood)$ is common practice with fitting algorithms, where the ratio of likelihoods are compared to determine the best fit (maximum likelihood). Since $\eta_m(p, r)$ is not dependent on the fitting parameters it can be considered constant.

Table 2.3: The mixture probabilities for each detection order. For example, this accounts for the possibility that two and three planet systems may only be found with a single planet. These values were found using our transit probability model described in Section 2.5.1.

	m=1	m=2	m=3	m=4	m=5	m=6
$\frac{P(2\lvert\overline{(m,1)})}{P(m)}$	0.67	-	-	-	-	-
$\frac{P(3\lvert\overline{(m,2)})}{P(m)}$	0.68	0.50	-	-	-	-
$\frac{P(4\lvert\overline{(m,3)})}{P(m)}$	0.53	1.05	0.50	-	-	-
$\frac{P(5\lvert\overline{(m,4)})}{P(m)}$	0.53	1.12	1.52	0.46	-	-
$\frac{P(6\lvert\overline{(m,5)})}{P(m)}$	0.37	1.07	1.85	1.69	1.22	-
$\frac{P(7\lvert\overline{(m,6)})}{P(m)}$	0.33	0.71	1.64	1.90	1.25	1.22

2.7.0.1 Calculating N_m

To find $\eta_m(p, r)$ we use the detection maps found in Section 2.6.1 and 2.6.2. Here we are assuming an average probability of detection over the stellar population. To properly treat this integral, one would have to compute the detection probability for each star. Such a procedure would be computationally expensive and provide a minimal increase in precision.

2.7.0.2 Sorting Order

Here we will provide a brief overview of order statistics and why it is an important feature of this model. As mentioned previously, the *Kepler* pipeline finds planets in order of decreasing MES. Such ordering will skew the distribution of planets found in each m. Larger, short period, planets will tend to be found in order $m = 1$ or $m = 2$, because there are more transits and deeper transit depths. Smaller, long-period planets will tend towards orders $m = 6$ or $m = 7$. To account for such a skew, a joint distribution model ($P_m(x)$) can be utilized (David & Nagaraja, 2003).

$$P_m(x) \propto P_0(x) C_0(x)^{a_m - 1} (1 - C_0(x))^{b_m - 1} \tag{2.26}$$

Here, $P_0(x)$ is the true underlying probability distribution function and $C_0(x)$ is the true cumulative distribution function. a_m and b_m can range from (0,inf) and forces the skew of the distribution. Essentially, the PDF of the distribution is skewed by a Beta distribution of the CDF. In the case of $a_m = b_m = 1$ the sorting skew returns the original PDF ($P_0(x)$).

The parameters a_m and b_m can be found analytically for equally sampled orders, but becomes far more complex in the decreasing case at hand (each m has fewer planets than the last). To determine the best values for this case, we choose to simulate this sorting mechanism on a uniform distribution, where the skew can be clearly isolated and extracted. In doing so, we force the ratio of each m sample to mimic that of the empirical population. Each system is then sorted by $r^2/p^{1/3}$, imitating *Kepler's* MES sorting. For example, if a system of (r=1.2,p=25), (3.5,20), (4.1,150) were randomly drawn into $m = 1, 2, 3$ detection orders, they would be re-sorted as (3.5,20), (4.1,150), (1.2,25) corresponding to $m = 1, 2, 3$. As we can see, the highest MES will always rise to $m = 1$. This is then repeated for 10^7 systems. Figure 2.4 shows how the first two detection orders are skewed by this procedure. If sorting were not an issue, these distributions would maintain the uniform flat appearance. Fitting a Beta distribution to this skew, we can determine the best a_m and b_m parameters for our sample. These parameters are provided in Table 2.2.

Since the this joint distribution is separable, we define the skew portion of the distribution as $O_m(p, r)$.

$$\begin{aligned} O_m(p, r) = N * C_r(r)^{a_{m,r}-1}[1 - C_r(r)]^{b_{m,r}-1} \\ * C_p(p)^{a_{m,p}-1}[1 - C_p(p)]^{b_{m,p}-1} \end{aligned} \tag{2.27}$$

where $C_r(r)$ and $C_p(p)$ represents the CDF of the radius and period distributions respectively and N represents a normalization factor that we find numerically within the MCMC.

2.7.0.3 Occurrence Factor

As noted, the value f_m is an integrated occurrence factor. In order to extract meaningful values, we realize that many $m = 6, 7$ planet systems will only provide detectable transits for one or two planets within the system. This will lead to an increased contribution to lower detection orders. Thus we adopt the following method for disentangling the true occurrence factors (F_m):

$$f_m = F_m + \sum_{n=m+1}^{7} F_n \frac{P(n|\overline{(m:n-1)})}{P(m)}. \tag{2.28}$$

Here $P(n|\overline{(m:n-1)})$ represents the probability of finding planet n given that planets $(m : n-1)$ are not found and $P(m)$ is the probability of finding planet m. This ratio accounts for the dependence between occurrence factors. If the mutual inclination is purely isotropic and planets are truly independent this ratio would be one. We use our transit simulation from Section 2.5.1 to extract these marginalized probabilities. Table 2.3 contains the results from this simulation. This model indicates that each multi-planet system will have more than one opportunity to find an f_1 planet. The physical interpretation of the F_m values is the fraction of stars that have at least m planets.

2.7.1 Fitting the Data

We employ EMCEE (Foreman-Mackey et al., 2013), an affine-invariant ensemble sampler (Goodman & Weare, 2010), to explore the parameter space of our study. To better constrain the 13 fit parameters, a Bayesian framework is implemented. Linear space uniform priors are used for all parameters. For $\alpha_1, \alpha_2, \beta_1$, and β_2 the priors range from -30 to +30. For r_{br} and p_{br} the priors range from r_{min} and p_{min} to r_{max} and p_{max} of our planet sample respectively. One unique restriction for our prior is that F_m must be larger than F_{m+1}. It is not possible to have a higher occurrence of $m + 1$ than m planets. To avoid truncation bias and maintain order, all F_m priors range from 0 to F_{m-1}. In the special case of $m = 1$, the prior ranges from 0 to 1. It is important to remember that F_m represents the fraction of the population containing m planets. Therefore, this cascading prior still allows for larger multiplicity systems to be more common than smaller multiplicity systems.

2.8 Discussion

In this section, we now apply the formalism we have developed to infer the revised occurrence rate parameters for planets orbiting GK dwarfs. This sample includes data from the final *Kepler* release DR25 and updated planet radius measurements from the CKS and *Gaia* DR2. Beyond these recent data improvements, we now include a corrected detection efficiency for multiple-planet systems. Given that many multiple-planet systems span much of the *Kepler* Parameter space, we include planets within $.5 < r < 16r_{\oplus}$ and $.5 < p < 500$ days. In implementing two detection efficiencies, this study expands on the Poisson process likelihood function used by other authors, allowing for the treatment of planet multiplicity. This Bayesian framework is fit using an MCMC, where 20,000 steps are used to model the posterior of each parameter. The resulting posteriors are presented in Figure 2.8. From this model we infer best fit power-law values of $\alpha_1 = -1.65\pm^{0.05}_{0.06}$, $\alpha_2 = -4.35 \pm 0.12$, $\beta_1 = 0.76 \pm 0.05$, and $\beta_2 = -0.64 \pm 0.02$. The breaks in our best fit model occur at $p_{br} = 7.08\pm^{0.32}_{0.31}$ days and $r_{br} = 2.66 \pm 0.06r_{\oplus}$.

One novel feature of our fitting method is the ability to extract exoplanet multiplicity. This information is provided through the F_m parameters. These values indicate the probability of a system having at least m planets. We find the following value best fit our model: $F_1 = 0.72\pm^{0.04}_{0.03}$, $F_2 = 0.68 \pm 0.03$, $F_3 = 0.66 \pm 0.03$, $F_4 = 0.63 \pm 0.03$, $F_5 = 0.60 \pm 0.04$, $F_6 = 0.54\pm^{0.04}_{0.05}$, and $F_7 = 0.39\pm^{0.07}_{0.09}$.

2.8.1 Forward Modeling the Results

Thus far, we have accounted for various parameter and population dependencies. To ensure that this process yields meaningful results, we choose to sample the extracted population and

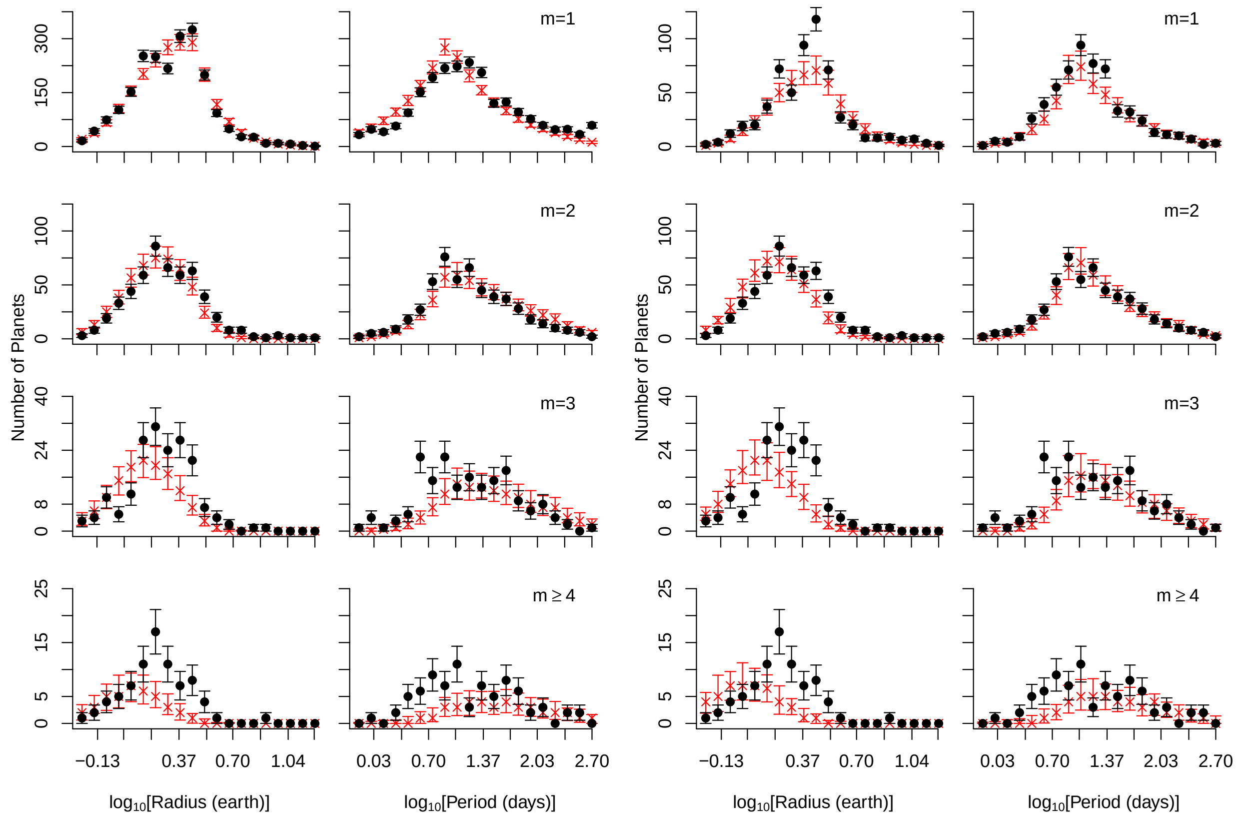

Figure 2.5: A plot of the forward modeled population derived by our Bayesian analysis. The red x marks symbolize the model values with their corresponding 68.3% confidence intervals. To find this interval the model is sampled 50 times using the posterior parameter distributions, the uncertainty reflects the fluctuations we find from these trials. The black points show the *Kepler* data with poisson uncertainty. For m=5:7 many of the bins have 1 or 0 planets, where small number statistics cause significant variations. In order to minimize this variations we present the resulting combination of $m \geq 4$. However, it should be noted that our forward model does differentiate between these detection orders. **Left**: Forward model of multiple and single planet systems. **Right**: Forward model of only multiple-planet systems. This model was produced by only fitting to the data of multiple-planet systems.

subject it to the detection constraints described in Section 2.6. Here we present the **ExoMult** forward modeling software. This code, developed in *R*, simulates these detection effects and produces a population of detected planets. Using this program, we can make far fewer assumptions and directly recover the expected population. For example, the probability of transit for all 7 planets can be directly accounted for by sampling system inclination, mutual inclination and the argument of periapsis directly. Furthermore, the detection probability will not be marginalized over all stars, but rather reviewed for each system independently.

The first step in our forward model is drawing each system of planets according to the population parameters given in Figure 2.8. Each system is randomly oriented with mutual inclinations drawn from a Rayleigh distribution. For planets with detectable impact parameters ($b < 1$), the planets within each system are sorted in decreasing MES. The probability of recovery is assigned to each planet according to the procedure laid out in Sections 2.6.1 and 2.6.2. Based on the calculated probability of detection, the planet is either detected or lost by drawing from a random number generator. Figure 2.5 shows the best fit model to the observed population obtained with this forward model. It is clear the our Bayesian method provides a reasonable model, where nearly all data points are within a 1σ deviation of the observed distribution.

Fulton et al. (2017) and Berger et al. (2018) have provided evidence for a dip in the radius population around $1.5 - 2r_{\oplus}$. This gap is apparent in the m=1 case of Figure 2.5. While the deviation from a broken power-law is mild, we explore the effects here. When we remove the single planet systems from the data set, this gap is no longer apparent. One plausible explanation for this gap is a unique population of single planet systems (Although evidence from Weiss et al. (2018b) shows that a weak gap can be seen in the multi-planet systems when aggregated). To explore this theory, we isolate the multi-planet systems and run our fitting procedure again. We find a mild difference in the extracted α or β power law values ($\alpha_1 = -1.98 \pm 0.08$; $\alpha_2 = -3.90 \pm 0.16$; $\beta_1 = 0.96 \pm 0.08$; $\beta_2 = -0.79 \pm 0.03$). This indicates that if a separate population does exist, the population parameters are weakly affected by their inclusion in our dataset. The resulting forward model of this fit is presented in Figure 2.5. Furthermore, the increase in uncertainty seen in these parameters is due to the reduced samples used for fitting (1305 multiple system candidates vs. 2942 total candidates). It is notable that the empirical *Kepler* data set is sharply peaked, while the model does not provide a similar sharpness for the $m = 1$ radius population (Figure 2.5 Right). This could be due to the existence of the mentioned radius gap. Furthermore, it is possible that a true accounting for planet period and radius covariance could produce such a peak. Millholland et al. (2017) and Weiss et al. (2018a) show that the planets within multiple systems tend to have similar mass and radius components. Although these features are not properly accounted for here, Figure 2.5 (Left) shows that these mild population characteristics remains small and do not deviate greatly from a simple broken power-law model. We hope to include such features in the next iteration of this software.

It is possible that future studies may use this forward modeling technique to directly determine the population parameters. Unfortunately, it remains computationally expensive to properly account for all detection features. Traub (2016) overcame this cost by ignoring multiplicity.

2.8.2 Comparison with Prior Work

We use a Bayesian method to infer population parameters for the *Kepler* exoplanet population, following much of the procedure presented in Youdin (2011). However, we build upon this method to extract information about the population multiplicity. Using a broken power-law distribution we find that population parameters of $\alpha_1 = -1.65\pm^{0.05}_{0.06}$, $\alpha_2 = -4.35 \pm 0.12$, $\beta_1 = 0.76 \pm 0.05$, and $\beta_2 = -0.64 \pm 0.02$ provide the best replication of the empirical population. The best fit breaks in these distributions are as follows: $p_{br} = 7.08\pm^{0.32}_{0.31}$ days and $r_{br} = 2.66 \pm 0.06 r_{\oplus}$.

Many prior studies have examined the occurrence of planets as determined by *Kepler*. Youdin (2011) provided an early estimate of the occurrence rate using a Poisson process likelihood, finding that the PDF exhibited a power law break at periods ~ 7 days, with $\alpha = -2.44$ and $\beta = 3.23$ at short periods, and $\alpha = -2.93$ and $\beta = -0.37$ at longer periods (we have converted his numbers into the definitions of α and β adopted here). These suggest a steep rise towards smaller radius planets at all periods, and a sharp rise with increasing periods to the break, followed by a gradual decline to longer periods. This is consistent with other analysis at the same time (Catanzarite & Shao, 2011; Howard et al., 2012; Dong & Zhu, 2013). With the accumulation of additional data and more detailed treatment of selection effects, subsequent analyses favored a flatter distribution extending to smaller radii (Fressin et al., 2013b; Petigura et al., 2013b; Silburt et al., 2015; Traub, 2016), and a distribution falling off inversely with period ($\beta \sim -1$) at longer periods (Petigura et al., 2013a; Silburt et al., 2015). The plateau at small radii is also found around lower mass hosts (Dressing & Charbonneau, 2013, 2015; Mulders et al., 2015).

Burke et al. (2015) have presented an extensive discussion of planet occurrence using the Q1-Q16 Kepler sample. For their baseline model, they found corresponding values of $\alpha_1 = -1.54 \pm 0.50$ and $\beta_2 = -0.68 \pm 0.17$, with only weak evidence for a break in radius and assuming no break in period (they considered only periods > 50 days and radii $< 2.5 r_\oplus$). This is perhaps the most directly comparable to our analysis, as it uses the completeness estimates from Christiansen et al. (2015); where this study uses the updated Christiansen (2017) completeness data. We find very similar values ($\alpha_1 = -1.65 \pm^{0.05}_{0.06}$; $\beta_2 = -0.64 \pm 0.02$) in a comparable regime. In particular, we note that both of these studies find an increasing occurrence of small radius planets down to the detection threshold, a result also supported by another Bayesian methods estimate in Hsu et al. (2018).

Previous studies have used more limited parameter ranges to avoid issues of parameter covariance and susceptibility to completeness mapping. We approach the problem with

Table 2.4: A representation of the expected empirical multiplicity as a function of selection effects. Each column shows the expected population using the best fit model from this study (see Figure 2.8). Starting from the left, moving right, each effect is adding in addition to all previous effects. The Multiple Detection Efficiency (DE) is broken into two columns. The Data column directly used the multiplicity values shown in Figure 2.6. In contrast, the Model column uses the modified Poisson distribution inferred from the multiplicity data ($\lambda = 8.40 \pm 0.31$ and $\kappa = 0.70 \pm 0.01$).

	Geometric	Mutual Inclination	Single DE	Multiple DE (Data)	Multiple DE (Model)	Real *Kepler* Data
Singles	1870	1910	1558	1649 ± 71	1629 ± 61	1637
Doubles	686	816	397	374 ± 29	375 ± 33	346
Triples	354	483	115	103 ± 15	113 ± 15	119
Quadruples	207	282	30	26 ± 6	25 ± 6	43
Quintuples	127	159	8	5 ± 3	5 ± 3	13
Sextuples	132	77	1	1 ± 1	1 ± 1	2
Septuples	167	28	0	0 ± 1	0 ± 1	1

a rigorous treatment of completeness mapping and a larger parameter space, recovering a similar power-law distribution. This congruity is an encouraging sign as it shows that the inclusion of a larger parameter space does not largely effect the model being inferred. Our inclusion of a broader range of periods and radius allow us to constrain the power-law uncertainty for radius and period to 3.8% and 5.4% respectively.

We find breaks in our period and radius distributions occur at $p_{br} = 7.08 \pm^{0.32}_{0.31}$ days and $r_{br} = 2.66 \pm 0.06 r_{\oplus}$. These results are consistent with those found by prior authors.

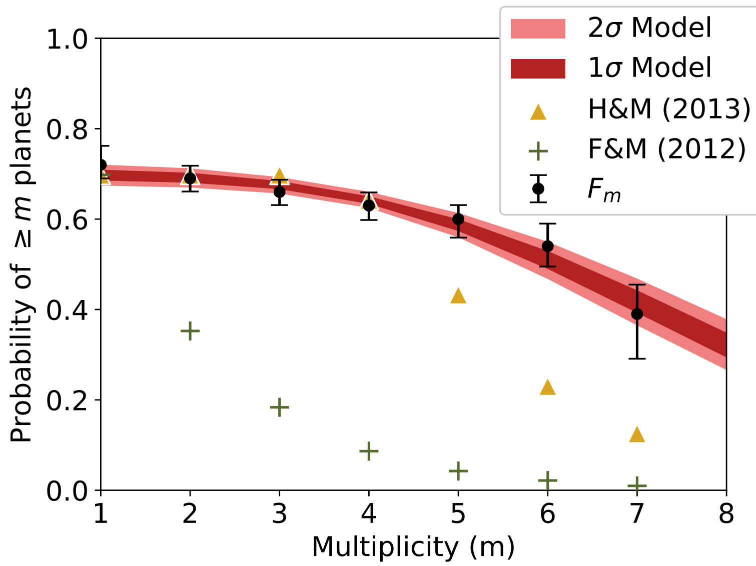

Figure 2.6: A plot of the modified Poisson survival function and the system fraction provided by our Bayesian analysis. This model is fit using a likelihood maximization technique, with the assumption of Gaussian uncertainty (essentially, a χ^2 minimization). The posterior distribution for the model is plotted by sampling 5000 models from the parameter posterior distributions in red. The dark red represents the 1σ range and the light red indicates the extent of the 2σ range. We have included the models provided by Hansen & Murray (2013) (gold $\triangle$ makers) and Fang & Margot (2012) (olive green $+$ markers) for comparison. Both Hansen & Murray (2013) and Fang & Margot (2012) models have been renormalized by our best fit κ value.

2.8.3 Survival Function

Within this study, we only use planets provided by the *Kepler* pipeline. The highest multiplicity seen is m=7 for a GK type star. This is certainly not the actual highest multiplicity within this parameter space. Shallue & Vanderburg (2018) use a deep convolutional neural network to extract an 8th planet from the *Kepler*-90 light curve, proving this assertion to be true. Using a Poisson survival function we can extrapolate the probability of existence for these higher multiplicity systems. The F_m values found by this study represent the fraction of stars with at minimum m planets. This lends itself well to a survival function, where the probability of existing up to a certain value (or multiplicity) is obtained. Survival functions ($S(x)$) can simply be written as:

$$S(x) = 1 - CDF(x), \tag{2.29}$$

where $CDF(x)$ is the cumulative distribution function of the model. In this case we use a modified Poisson distribution to model multiplicity. Poisson distributions are ideal for planet multiplicity as these distributions are used for counting statistics. The modification is that the distribution is not truly normalized, but rather some fraction κ of one. Now that that distribution is no longer normalized the survival function must be modified slightly ($S(x) = \kappa - CDF(x)$). This modification allows for an excess or scarcity of zero planet systems. We are only interested in stars that do harbor planets, thus this modification is necessary. The CDF for this modified function is given as:

$$CDF(m) = \sum_{n=1}^{m} \kappa \frac{\lambda^n e^{-\lambda}}{(n)!}, \tag{2.30}$$

where κ and λ are both fit parameters. Further discussion of this modified Poisson distribution can be found in Section 2.3 of Fang & Margot (2012). The results of this fit are presented in Figure 2.6. We find that $\lambda = 8.40 \pm 0.31$ and $\kappa = 0.70 \pm 0.01$ provide the best match for this distribution. This large λ value incorporates a non-negligible fraction of systems with $m > 10$. Since Equation 2.30 allows for an inflated number of star without planet, the global average for GK dwarfs in the *Kepler* parameter space (denoted as $\langle N_{pl} \rangle$) can be found by multiplying λ, an estimate of the average number of planets a planet harboring system will contain, by κ, the fraction of stars that do harbor planets. We find that $\langle N_{pl} \rangle = 5.86 \pm 0.18$ planets per star. This is likely a lower bound as we have excluded the single Jupiter sized planets that have cleared their systems through migration. Since these stars are currently assumed to have zero planets by this paper, inclusion of these additional planets would increase the κ value. However, we would expect our λ parameter to slightly decrease, with the inclusion of these additional singles, as this value only considers systems that do harbor planets. Overall the increase in κ will dominate, leading to an overall increase in $\langle N_{pl} \rangle$.

Previous studies have averaged over multiplicity and inferred the $\langle N_{pl} \rangle$ value alone. These values are more difficult to compare as they are strongly dependent on the range of planet radius and period include in each study. Looking at short period ($p < 50$ days) planets, Youdin (2011) found $\langle N_{pl} \rangle = 1.36$. Using our population parameters and making similar cuts we find a comparable value ($\langle N_{pl} \rangle = 1.34 \pm .06$). Turning the focus towards small planets ($.75 < r < 2.5 r_\oplus$) and long periods ($50 < p < 300$ days), Burke et al. (2015) found $\langle N_{pl} \rangle = 0.73 \pm^{.19}_{.07}$. When we apply these same bounds to our model we again find a slightly larger value ($\langle N_{pl} \rangle = 1.15 \pm .03$). The most comparable parameter space to our study is that of Traub (2016), who finds $\langle N_{pl} \rangle = 5.04 \pm .23$ using a nearly identical parameter range. While there appears to be a mild tension with this value, we note that Traub (2016) includes a much broader stellar temperature range and pre-*Gaia* radius measurements, likely leading to this deflated $\langle N_{pl} \rangle$ value.

With this function in hand, we can extrapolate to higher multiplicity. For example, our model suggests that $32.3 \pm 2.7\%$ of GK stars will harbor at least 8 planets within the *Kepler* parameter space. In the parameter space of the *Kepler* survey, our solar system has two planets (Venus and Earth). The radius of Mercury is slightly smaller ($.387 r_\oplus$) than our range allows. Since we find $\langle N_{pl} \rangle = 5.86 \pm 0.18$ planets per solar-like star in this range, it appears that our system is more underpopulated than most other systems within $p < 500$ days. We would expect $30 \pm 1\%$ systems harbor zero planets, $4.0 \pm 4.6\%$ harbor just one planet, and $2.0 \pm 4.2\%$ harbor only two planets within the range of this study. This lack of multiplicity in our solar system could be important for habitability, but such claims still lack strong evidence.

2.8.4 *Kepler* Dichotomy

Analysis of the statistics of the *Kepler* multiple planet systems (Lissauer et al., 2011; Fang & Margot, 2012; Hansen & Murray, 2013; Ballard & Johnson, 2016) suggest that the underlying planetary population requires a two component model. One component is composed of systems with high planet multiplicity and a low inclination dispersion, while the other requires either low intrinsic multiplicity or a large inclination dispersion to reduce the frequency of transits by multiple planets. This has been termed the *Kepler* dichotomy. Lissauer et al. (2011) inferred that the two populations had roughly equal frequencies and subsequent analyses confirmed this. There have been several models proposed to explain this on dynamical grounds (Johansen et al., 2012; Moriarty & Ballard, 2016; Hansen, 2017). The simplest solution is to consider a single population of planets in which some fraction have experienced excitation of their mutual inclinations. However, to meet the requirements of the transit statistics, the excitation is sufficiently large that dynamical stability is hard to maintain (Hansen, 2017). Thus, the *Kepler* results seem to imply the existence of a low multiplicity population of planetary systems, whether due to formation or later dynamical instability.

However, this finding rests on the relative frequencies of systems with single transiting planets versus multiple transiting planets. If the completeness is a function of the detection order, this may weaken the claim for a *Kepler* dichotomy. In Figure 2.6 we show that a single Poisson distribution can account for the multiplicity probabilities (F_m) extracted from our analysis. We find a much smaller fraction of intrinsically single systems than Fang &

Margot (2012) and find a distribution broadly similar to the model for a single, dynamically motivated population described in Hansen & Murray (2013). However, we still find $\sim 6\%$ of stars harbor intrinsically single or double planet systems. To test the robustness of this low multiplicity contribution we forward model the inferred population using the Poisson multiplicity model. In Table 2.4 under the label "Multiple Detection Efficiency (Model)" we present the multiplicity results of this model. We can see that almost all of the empirical population fall within 1σ of the multiplicity model. This indicated that that apparent deviations in our infer F_m values can be described by statistical fluctuations in population. Additionally, our F_m are very dependent on the choice of mixture values displayed in Table 2.3. A proper accounting of these values would require distribution dependence. Averaging over these parameters, as done here, can cause mild deviations in the inferred F_m values.

In extracting the population F_m values, we have only employed a mild Rayleigh distribution to account for mutual inclination of each system as directed by Fang & Margot (2012) and have no larger inclination component. It appears that accounting for systematic loss of planets at higher multiplicity substantially reduces the low multiplicity population inferred as per the *Kepler* Dichotomy. We shall now discuss how this works.

Using the forward model presented in Section 2.8.1, we look at how the inclusion of detection efficiency affected the gap seen between systems with one transiting planet and those with two transiting planets. The population provided by the parameters in Figure 2.8 is modeled 20 times and the median from each group is recorded in Table 2.4. Using our population parameters and a mild mutual inclination model show that this anomaly is largely due to *Kepler* detection efficiency. Table 2.4 shows how the frequency of detected systems of different transit multiplicity changes as we include different systematic effects. In the first column, we include only the correction of the probability of transit due to geometric alignment. For a simple numerical comparison, this results in a ratio of double transit to single transit systems of 0.37, to be compared to the observed value of 0.21 (the rightmost column). The inclusion of a small mutual inclination dispersion, comparable to that of Fang & Margot (2012), does not improve the ratio (second column). In the third column, we show the model in which we include the completeness corrections from Christiansen (2017) without the multiplicity treatment discussed here. This results in a partial improvement of the ratio to 0.25. It is also notable that the number of expected high transit multiplicity systems also drops significantly with the inclusion of this effect. Finally, in the fourth and fifth column, we show the expected numbers including the full, multiplicity-dependent completeness correction discussed here (Section 2.6.2). We find that the expected number of different transit multiplicities are now very well matched to the observed numbers, substantially weakening the need for an additional population to explain the observations.

The ultimate reason for this is that high transit multiplicity systems usually contain several planets that lie in the low MES region of parameter space, so that the incompleteness (especially when including the detection order effects) knocks planets down the multiplicity scale, resulting in many single transit systems that, in an ideal world, would show two or three transiting planets. Furthermore, the improved stellar radius measurements from *Gaia* suggests that many stars have larger radii than previously believed (Berger et al., 2018). Increasing the stellar radius of system will decreases the probability of detection for an exoplanet. This correction will overall increase the inferred occurrence measurements.

It is important to remember that our dataset does not include single hot Jupiter planets as discussed in Section 2.3. This observed population of 120 planets does not follow our power-law trend and appears to be uniquely single (Steffen et al., 2012). While these outliers do provide some type of population dichotomy, their presence is not the most prominent cause of the excess of singles.

Our extracted population parameters F_1 and F_2 indicate that $4.0 \pm 4.6\%$ of the underlying population does have only one planet, and that this contribution can be described by the modified Poisson distribution used to fit the higher multiplicity systems. There is dynamical evidence that single transiting systems are more dynamically excited than multiple systems (Morton & Winn, 2014; Xie et al., 2016; Van Eylen et al., 2019) and this is consistent with the notion that some fraction of compact planetary systems are dynamically perturbed by the existence of giant planets on larger scales. Previously, Hansen (2017) found that explaining the original excess of single transits required a frequency of giant planets on large scales that was roughly double that found by radial velocity surveys. The reduction found here substantially alleviates that discrepancy.

Other recent work also supports the notion that single transiting systems are drawn from the same underlying planetary population as multiple harbor systems. Weiss et al. (2018b) find that both populations share essentially the same stellar and planetary properties, while Zhu et al. (2018) use transit timing variations to infer that there is a strong correlation between multiplicity and dynamical excitation. They reject the notion that this is driven by giant planet excitation because they see no correlation with the metallicity of the host star, but such a correlation would be difficult to see at the level of 4% as found here. This is further supported by Munoz Romero & Kempton (2018), who find no metallicity difference between hosts of single and multiple transiting systems, but could easily accommodate mixtures at the 50% level.

2.8.5 Considering Eccentricity

Including eccentricity into our model increases the number of detected planets. We find that the best fit multiplicity parameters are as follows: $F_1 = 0.72 \pm 0.05$, $F_2 = 0.66 \pm 0.03$, $F_3 = 0.63 \pm 0.03$, $F_4 = 0.60 \pm 0.03$, $F_5 = 0.56 \pm 0.03$, $F_6 = 0.51 \pm 0.04$, and $F_7 = 0.43 \pm 0.07$. These parameters are fit using an analog to the Hansen & Murray (2013) eccentricity model. The original modified Gamma distribution (scale=0.055) is unique to Hansen & Murray (2013). We map this model to a Beta distribution ($a = 1.80$ and $b = 14.46$), widely used among recent authors, for consistency. This model was inferred by simulating in situ gravitational assembly of planetary embryos and observing the resulting eccentricity population of the fully formed

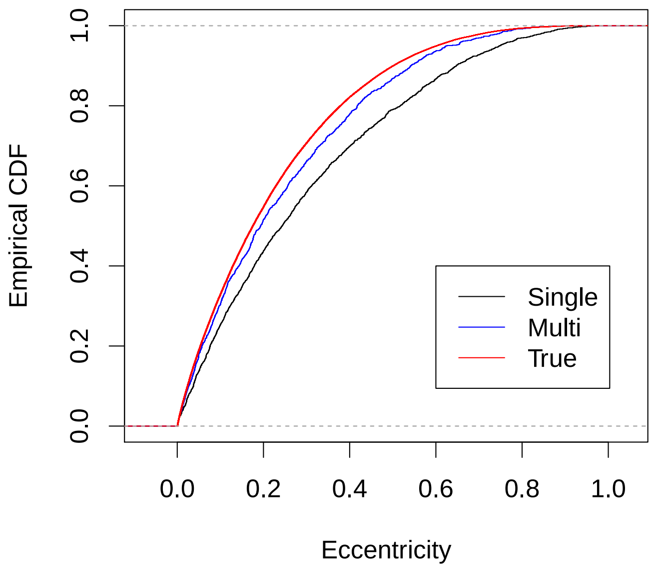

Figure 2.7: A CDF showing the retrieved eccentricities from our forward modeling pipeline. The red line illustrates the eccentricities used to draw the underlying the Beta distribution (Kipping, 2013). The black line represents the empirical CDF of the detected single planet systems and the blue line represents the eccentricities of the detected multi-planet systems.

planets. Although derived within a specific scenario, this distribution matches well with a model in which planets explore the full range of available phase space subject to the constraint of dynamical stability (Tremaine, 2015). As such, it represents a plausible description of the level of eccentricity to be expected in such systems. The average eccentricity of this population is $\langle e \rangle = 0.11$. Comparing these values to those of our base model, we find that eccentricity flattens the CDF of planet multiplicity, slightly decreasing $\langle N_{pl} \rangle$ to 5.69 ± 0.17 planets.

Recently, Van Eylen et al. (2019) provided evidence for two distinct populations of eccentricity (multi-planet systems and single planet systems). Using our forward modeling software (**ExoMult**), we test the strength of this hypothesis. Implementing only one true underlying eccentricity model, we inspect the detected eccentricity populations from both the single and multi-planet populations. When tested with the Hansen & Murray (2013) model ($\langle e \rangle = 0.11$), we find no significant difference between the the observed eccentricities of multi-planet and single planet systems. This indicates that the differences noted by Van Eylen et al. (2019) may be real. However, Van Eylen et al. (2019) suggests a Beta distribution for single planet systems with $\langle e \rangle = 0.26$. This is a significantly larger average eccentricity than expected by the Hansen & Murray (2013) model.

When larger eccentricities are tested we do find observable differences between the single and multi-planet systems. The Kipping (2013) model ($a = 0.867$ and $b = 3.03$) was calculated using radial velocity discoveries and contains a significant fraction of massive planets. This distribution is probably too eccentric ($\langle e \rangle = 0.22$) for the tightly packed model discussed here, but illustrates the effects of detection bias on the eccentricity population. In Figure 2.7 we present the results of our test on the Kipping (2013) model. We find that multi-planet systems tend to produce more low eccentricity detections than single planet detections despite being drawn from the same underlying population. Analyzing the statistical difference with an *Anderson-Darling* test produces a P-value of 10^{-7}, suggesting these differences would appear statistically significant. Furthermore, we can see that neither of the detected populations closely mimic the true Beta distribution, highlighting the importance of detection efficiency consideration when performing eccentricity occurrence measurements. This effect is caused by the increased transit duration for higher eccentricity transits. Increasing the transit duration improves the planet MES, making the signal easier to detect. Since the highest MES planets are the most likely to be detected, this biases the empirical population toward higher eccentricity. The sorting order in combination with the multiplicity detection efficiency of the *Kepler* pipeline will further exaggerate this bias in the single planet systems.

It is clear that low eccentricity distributions are less affected by this bias. Manually tuning the Beta distribution we find that models with $\langle e \rangle \geq 0.18$ will produce statistically significant (P-value$\leq$ 0.001) differences between the empirical eccentricity population of singles and multiple planet systems. Since Van Eylen et al. (2019) suggests a $\langle e \rangle = 0.26$ model for the singles and a $\langle e \rangle = 0.05$ model for the multi-planet systems, it is difficult to determine the effect of detection bias on their eccentricity model. At this point we cannot rule out that two distinct populations of eccentricity exist between the single and multi-planet systems, but propose that such claims require further evidence.

2.8.6 Extrapolation to Longer Periods

As mentioned above, our general populations parameters do not differ greatly from those of previous studies. The quantity $\Gamma_\oplus$ is often quoted to avoid any need for understanding the habitable zone or habitable radius range.

$$\frac{dN}{d\ln p_\oplus \; d\ln r_\oplus} = \Gamma_\oplus \tag{2.31}$$

We find $\Gamma_\oplus = 1.31 \pm 0.07$, consistent with the previous value of Burke et al. (2015) ($\Gamma_\oplus = 0.6$ with a range of 0.04 to 11.5). Youdin (2011) found a much higher value of $\Gamma_\oplus = 2.75 \pm 0.3$, when extrapolating from periods $<$ 50 days. The lack of long period planets provided a weaker power-law, producing the inflated $\Gamma_\oplus$ value. Furthermore, we find tension with

Foreman-Mackey et al. (2014) (with $\Gamma_\oplus = 0.019\pm^{0.019}_{0.010}$). Foreman-Mackey et al. (2014) avoid the assumption of a particular functional form for the extrapolation to longer periods, by using a Gaussian process regression to determine the shape of the distribution. However, they use the results of the *TERRA* pipeline in it's original form, in which it only reported the highest signal to noise candidate around each star. Although they back out an estimate of the detection efficiency from the results of Petigura et al. (2013b), we have shown in Section 2.7 that detection order can bias the results. In particular, we expect Foreman-Mackey et al. (2014) to undercount small planets and long period periods. Both of these biases will lower the $\Gamma_\oplus$ value and we should regard the Foreman-Mackey et al. (2014) result as a lower limit.

For the occurrence of habitable planets we follow the procedure provided by Burke et al. (2015). This $\zeta_\oplus$ value is found by integrating the population distribution by 20% of $r_\oplus$ and $p_\oplus$ in both directions. We find $\zeta_\oplus = 0.217 \pm 0.014$ using our inferred population parameters, similar to the $\zeta_\oplus = 0.10$ (with a range of 0.01 to 2) found in Burke et al. (2015).

2.9 Conclusion

We present a new method for determining the frequency of exoplanet multiplicity within the *Kepler* dataset. In doing so we provide the following new fitting features and conclusions:

1. Previous studies have discussed and provided methods for calculating high multiplicity transit probabilities (Ragozzine & Holman (2010); Brakensiek & Ragozzine (2016); Read et al. (2017)). For occurrence calculations these procedures are often too complex and computationally expensive to carry out. We provide a new method which marginalizes over mutual inclination and the empirical *Kepler* period set to determine the transit probabilities for *Kepler* multi-planet systems. Using this, we provide the transit probabilities for multiple systems containing up to 7 planets. This simplification is important and useful when trying to fit multiplicity parameters via MCMC or some other fitting method that requires 10^4 calculations.

Our method does make some simplification assumptions in the interests of speed. We assume the measurements of planet radius and period are perfect. The uncertainty in period is negligible, however the radius measurements retain significant uncertainty and the present dispersion may yet mask finer features in the distribution. In accounting for mutual inclination, we adopt the model provided by Fang & Margot (2012). This is derived using a different multiplicity model than that found here. All orbits are assumed to be circular in our base model. Because many of the systems are very compact, circular orbits are required for any type of stability. Tidal circularization will also force many of these planets into circular orbits. However, it is possible that some portion of the population, investigated here, contains varying amounts of eccentricity. We show that any amount of eccentricity will increases our the overall multiplicity values, but decreases the fraction of systems with planets. We have assumed the appropriate model for exoplanet occurrence is a broken power-law. Furthermore, we assume period and radius and uncorrelated. It has been shown by Owen & Wu (2013) and Weiss et al. (2018a) that a mild correlation exist between period and radius at short periods where photoevaporation can take effect. Nevertheless, the fact that our forward modeling matches the data inspires confidence that the model provides a coherent description of the data.

2. In systems with more than one detected planet, we find that detection efficiency decreases for higher detection order planets. This conclusion was achieved by re-visiting the Christiansen (2017) injections and looking at systems with pre-existing planets. Multiple planets systems experience an additional loss, for lower MES planets within each system, of at least 5.5% and 15.9% for periods < 200 days and > 200 days respectively. This type of increased selection effects indicates that a larger fraction of the population is being missed. Being able to infer a larger population of multiple exoplanet systems significantly decreases the gap between single and double planet systems. The initial motivation for additional detection efficiencies for multi-planet systems, was the 61 known KOIs lost during the Christiansen (2017) injections. When testing our additional selection effects, for multiples, we expect

41 ± 7 planets should be lost due to a similar type of injection test. Because we find that 61 KOIs are lost (rather than 41) we suspect higher order detection efficiencies may be necessary for an accurate accounting of the true underlying populations.

3. Using Bayesian statistics, we expand the Poisson process likelihood to account for variations in detection order. Furthermore, we are able to infer population multiplicity from this fitting process. The results from this fit match that of Burke et al. (2015), but provide an improved measurement with reduced uncertainty from *Gaia*, CKS, and asteroseismology (Petigura et al., 2017; Johnson et al., 2017; Berger et al., 2018; Van Eylen et al., 2018). Furthermore, by looking at the occurrence of single and double-planet systems, we only find a 0.9σ difference between these two populations ($4.0 \pm 4.6\%$). This disparity can be explained by a modified Poisson distribution with $\lambda = 8.40 \pm 0.31$ and $\kappa = 0.70 \pm 0.01$, indicating that the *Kepler* Dichotomy (discussed by Lissauer et al. (2011); Fang & Margot (2012); Hansen & Murray (2013); Ballard & Johnson (2016)) may largely be an artifact of detection efficiency and statistical fluctuation.

Using a Poisson process likelihood requires that each planet is drawn independently, which is clearly not the case for planets in multiple systems. Much of the work in this study is accounting for these dependencies. Ignoring the independence requirement of Poisson process could be suspect, but is again justified by the success of our forward model, where this assumption is not necessary. The independence of radius between planets within a system has also not been accounted for within this study.

4. Given our inferred multiplicity model we can extrapolate to higher multi-planet systems. We find that $32.3 \pm 2.7\%$ of solar-like stars should contain at least 8 planets within 500 days. The existence of a single 7 planet system and a single 8 planet system (Kepler 90) indicates these systems should be rare but still detectable. We would expect to find < 1 eight planet systems within the constraints of this study.

5. We introduce (**ExoMult**) and demonstrate that forward modeling a broken power-law distribution can still provide a reasonable model for the exoplanet population, despite growing evidence for a gap in $1.5-2r_{\oplus}$ range (Fulton et al., 2017; Berger et al., 2018; Weiss et al., 2018b). We find that our fitting model also produces similar populations of multiplicity to that of the empirical *Kepler* data set, indicating the success of this method.

6.Using the the eccentricity model of Hansen & Murray (2013), we show that eccentricity can affect the multiplicity occurrence by slightly decreasing the expected number of planets around each star. We also find that for eccentricity models with $\langle e \rangle \geq 0.18$ the *Kepler* pipeline will significantly skew the empirical population of eccentricity for single transiting systems, suggesting that differences seen between the single and multiple planet systems may be artificial.

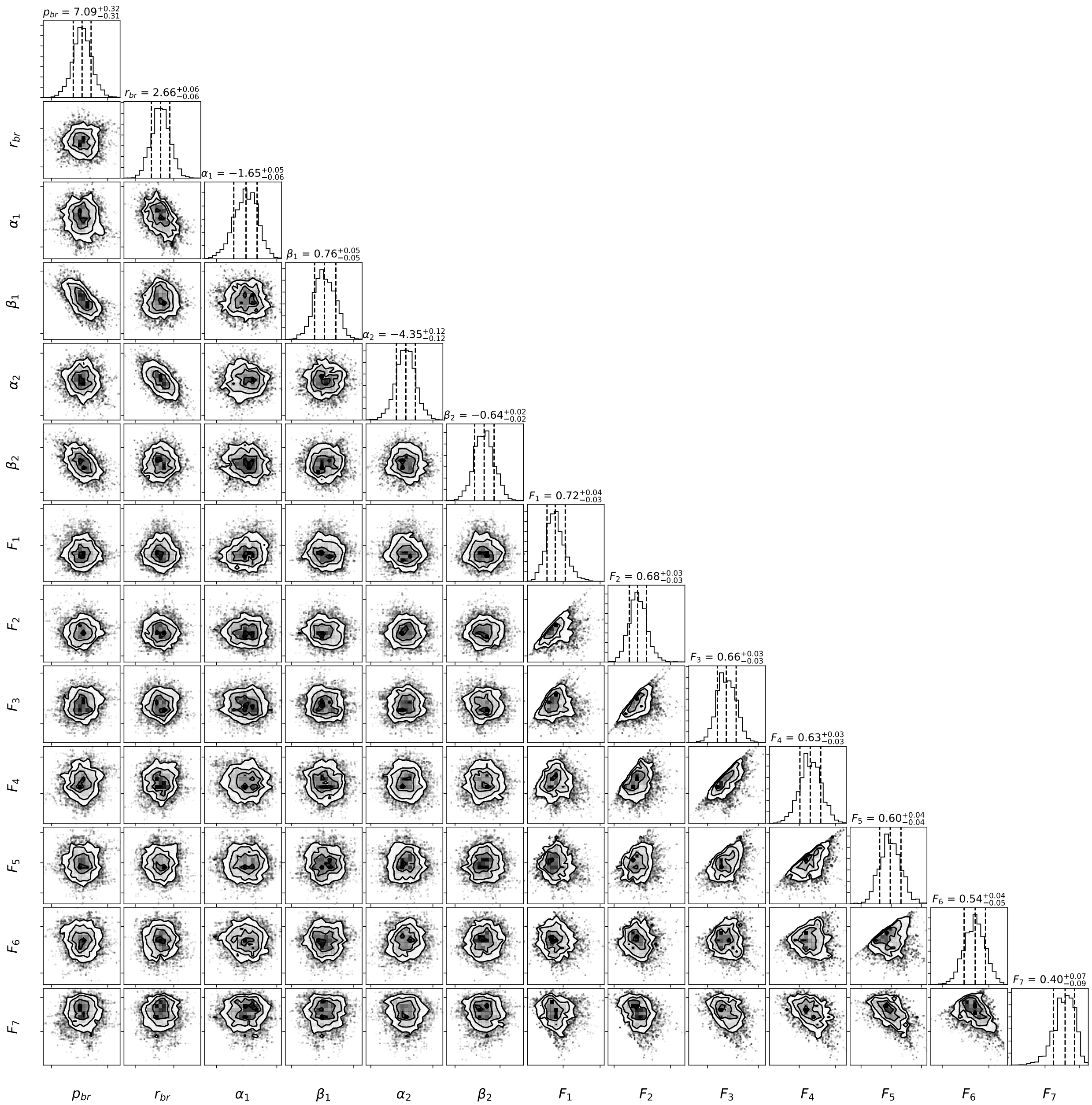

Figure 2.8: The posterior distributions for the 13 parameters varied in this study. This is achieved using a burn-in of 100,000 steps and 20,000 steps to sample the posterior. The results of the fit are presented above the marginalized distribution of each parameter. The uncertainty is presented with a 68.3% confidence interval.

CHAPTER 3

Accounting for Multiplicity in Calculating Eta Earth

3.1 Introduction

As the number of known exoplanets continues to grow, understanding the frequency of habitable planets remains a major area of interest. Assuming extraterrestrial life will share our fundamental characteristics, the search for habitability is focused on finding rocky planets orbiting a star at a distance that allows for the existence of liquid water. It is currently unclear whether the choice of stellar host plays an important role in habitability. Several studies have considered habitability around M dwarf stars (Yang et al., 2013; Kopparapu, 2013; Dressing & Charbonneau, 2013, 2015; Luger & Barnes, 2015), but tidal effects and UV radiation make habitability unclear. Knowing that life is possible around solar-like stars, this study focuses on GK dwarf hosts and the possibility of retaining habitable planets.

The *Kepler* space mission set out to explore the parameter space of Earth analogs. In finding $\approx 4,500$ transiting exoplanets candidates, the continuous viewing of a fixed field makes *Kepler* an ideal data set for occurrence measurements. One of the mission goal was to establish an estimate for $\eta_\oplus$, which is defined as the expected occurrence of "Earth-sized" planets within the habitable zone. Several studies have sought to produce such estimates (Catanzarite & Shao, 2011; Petigura et al., 2013a; Silburt et al., 2015; Traub, 2012, 2016). With the complete *Kepler* DR25 now available (Mathur et al., 2017) and improved planetary and stellar parameters from *Gaia* DR2 (Berger et al., 2018) and asteroseismology (Van Eylen et al., 2018), updated occurrence measurements can be determined.

Our previous study (Zink et al., 2019b) found that detection efficiency is a function of multiplicity. Using the injection recovery results of Christiansen (2017), we determined that the first detected TCE (Transit-Crossing Event) within each systems was recovered with a higher efficiency then any subsequent TCEs within a given light curve. In other words, when using the *Kepler* pipeline the first detected planet will be easier to detect than another other additional planets within the system. From this result it is apparent that the population of single planet systems is over represented in the empirical *Kepler* data set. It is likely that additional planets exits within the light curves, increasing the expected overall planet occurrence. Building on this conclusion, we used the updated stellar radius measurements from *Gaia* to determine current occurrence parameters for the planet radius and period power-laws. With a better understanding of the completeness of multi-planet systems, we calculated a multiplicity distribution for the *Kepler* candidates. With our updated occurrence estimates –which now accounts for the significant loss of multi-planet systems within the *Kepler* data set– it is now possible to consider system architecture when calculating the probability of a true Earth analog.

Additionally, the recent discovery of TRAPPIST-1 (Gillon et al., 2016), with four planets orbiting in the habitable zone, has generated interest in the concept of multiple habitable planets within a single system. This system of rich multiplicity, provides a unique opportunity to study the differences in atmospherical conditions and the corresponding effects on habitability (Lincowski et al., 2018). With the improved multiplicity measurements from Zink et al. (2019b), it is feasible to estimate the likelihood of multiple habitable planet systems.

Our goal in this paper is to provide an updated $\eta_{\oplus}$ value and discuss how this value is effected by multiplicity within the habitable zone. In Section 3.2 we describe our method of stellar selection for this calculation. In Section 3.3 we provide a method of planetary sampling to ensure we minimize orbital instabilities. In Section 3.4 we discuss the criterion we deem necessary for a planet to be considered habitable. In Section 3.5 we present the results of our occurrence calculation and discuss probability of multiple planets within the habitable parameter space. In Section 3.6 we discuss the limitations of our calculation. In Section 3.7 we provide our concluding remarks on this paper.

3.2 Stellar Selection

Because we are implementing the results of Zink et al. (2019b), we use the same stellar selection methods. We shall briefly describe this method here, but suggest that the reader consult Zink et al. (2019b) for a more thorough description. The complete stellar sample is provided by the *Kepler* DR25 stellar parameters (Mathur et al., 2017) with correction from *Gaia* DR2 (Berger et al., 2018). We make various cuts in the sample to ensure the occurrence measurement reflects planets around solar-like GK dwarfs. Our stellar sample includes stars with $4200 < T_{\rm eff} < 6100K$, $log(g) \geq 4$, $R_\star \leq 2R_\odot$, and available mass measurements for each star. To avoid under sampled systems we only include stars with a data span > 2 years with $> 60\%$ coverage during that period. To avoid contamination from especially noisy sources, we exclude stars which exceed 1000 ppm in the CDPP$_{7.5h}$. CDPP (Combined Differential Photometric Precision) is a measure of the amount of noise any planet, transiting for a given period of time, must overcome to provide a 1σ detection (Christiansen et al., 2012). After making the described cuts, we find that 86,605 solar-like stars remain. A machine-readable version of this data is available online.

3.3 Method of Sampling Planets

To determine the probability of habitable planets existing, we draw from the power-law distributions and the multiplicity found in Zink et al. (2019b) for each of the stellar samples.

The multiplicity of the system is drawn from a modified Poisson distribution:

$$p(m) = \kappa\Big(\frac{\lambda^m e^{-\lambda}}{(m)!} - e^{-\lambda}\Big), \tag{3.1}$$

where m in the multiplicity of the system. Using the $\lambda = 8.40 \pm 0.31$ and $\kappa = 0.70 \pm 0.01$ values found in our previous study, we randomly assign the number of planets each system will contain. To avoid extremely unstable planetary archetypes, we allow a maximum of 10 planets per system.

To determine the radius and period of each of these planets we draw from two independent broken power-laws ($g(p)$ and $q(r)$):

$$g(p) = \begin{cases} C_{p1}p^{\beta_1} & \text{if } p < p_{br} \\ C_{p2}p^{\beta_2} & \text{if } p \geq p_{br} \end{cases} \tag{3.2}$$

$$q(r) = \begin{cases} C_{r1}r^{\alpha_1} & \text{if } r < r_{br} \\ C_{r2}r^{\alpha_2} & \text{if } r \geq r_{br}, \end{cases} \tag{3.3}$$

where β_1 and β_2 indicate the slope of the period power-laws and α_1 and α_2 indicate the slope of the radius power-laws. r_{br} and p_{br} correspond to the breaks in the radius and period distributions receptively. Here, C_{r1}, C_{r2}, C_{p1}, and C_{p2} are normalization constants which force continuity at each of the corresponding breaks. From Zink et al. (2019b) we use $\alpha_1 = -1.65\pm^{0.05}_{0.06}$, $\alpha_2 = -4.35 \pm 0.12$, $\beta_1 = 0.76 \pm 0.05$, and $\beta_2 = -0.64 \pm 0.02$ for the power-laws and $p_{br} = 7.08\pm^{0.32}_{0.31}$ days and $r_{br} = 2.66 \pm 0.06 r_\oplus$ for the corresponding breaks. The distributions have been fit to planets with periods $.5 \leq p \leq 500$ days and radii $.5 \leq r \leq 16 r_\oplus$, thus we shall adopt the same sampling range here.

Once each planetary system has been drawn, we determine the stability of the system. With the independent draw of period and radius, most system will be overly crowded and lack the necessary spacing required for stability. To quantitatively understand this, we calculate the mutual Hill radii (ΔH) of each planet separation as follows:

$$\Delta H = \frac{a_2 - a_1}{(\frac{M_1+M_2}{3*M_\star})^{1/3} * \frac{a_1+a_2}{2}}, \tag{3.4}$$

where a_1 and a_2 are the semi major axis of the inner and outer planet orbits respectively, and M_1, M_2, and $M_\star$ correspond to the masses of the inner, outer, and stellar masses respectively. Since our sampling distribution provides a radius value, we must convert the planet radius into mass. This is done by adopting the empirical mass-radius relationship found by Chen & Kipping (2017):

$$M = 0.972 * \left(\frac{r}{r_\oplus}\right)^{3.584} M_\oplus \qquad ; \qquad r < 1.23 r_\oplus \tag{3.5}$$

$$M = 1.436 * \left(\frac{r}{r_\oplus}\right)^{1.698} M_\oplus \qquad ; \qquad 1.23 r_\oplus \leq r < 14.31 r_\oplus \tag{3.6}$$

$$M = 131.581 M_\oplus \qquad ; \qquad r \geq 14.31 r_\oplus . \tag{3.7}$$

The empirical data set contains a significant dispersion in this relationship, making it apparent that planets of similar mass can correspond to different radii. We ignore this issue and assume a perfect correlation between mass and radius. This assumption is justified by the fact that ΔH scales as $M^{-1/3}$, making small deviations in M insignificant. Additionally, Chen & Kipping (2017) finds a weak negative power law for $r > 14.13 r_\oplus$ ($r \propto M^{-0.044}$), while we assume $M = 131.581 M_\oplus$ in this range of r. This simplification is necessary in order for the inverse equation ($M(r)$) to remain a function over the parameters space of interest.

Having provided a method for calculating mass from radius, we can now calculate the ΔH of each planet spacing. Upon our initial draw we find that roughly 50% of all spacings fall above $\Delta H = 10$. Empirically it has been shown that 93% of *Kepler* multi-planet spacings have a $\Delta H \geq 10$ (Fang & Margot, 2012; Pu & Wu, 2016; Weiss et al., 2018a). This threshold indicates a spacing large enough to avoid instabilities cause by two planets in opposition. It is worth noting that orbits with larger eccentricity will require even greater spacing to avoid such interactions (Dawson et al., 2016).However, most models for the formation of these systems favor low eccentricities. In situ assembly models yield $\langle e \rangle = 0.11$ (Hansen & Murray, 2013), while chain migration models require small eccentricities to maintain the stability of the chain (Goldreich & Schlichting, 2014). For this study we ignore such issues and only consider planets with circular orbits. To avoid testing unstable systems, we resample all systems with any spacings of $\Delta H < 10$. In this resample we maintain the initial multiplicity (to avoid the artificial exclusion of high multiplicity systems), but re-draw all period and radius measurement within the system, regardless of the spacing that failed to meet this requirement. This type of complete resampling ensures that we maintain an independent broken power-law distributions of both period and radius. The ΔH values are then re-calculated and if all spacing now exceeds or meets the $\Delta H = 10$ requirement, the system is no longer re-drawn. However, if any spacing in the system remains below this threshold, the planets are again re-drawn. This process will continue until either all spaces meet the ΔH requirement or 100 iterations have occurred. We stop after 100 iterations to mimic the $\Delta H < 10$ spacings that exist in the empirical data set. After this procedure has been completed we find that roughly 98% of the spacings now meet or exceed this ΔH threshold.

3.4 Habitability Requirements

In order to determine which planets could be habitable in our simulation we must impose some habitability criterion. We shall focus on two main factors that determine the habitability of a planet: location and size.

The mass of the planet plays an essential role in habitability. If a planet is too small it will not be able to retain an atmosphere and maintain ongoing plate tectonic activity. Since planet cooling rates scale as r^{-1}, smaller mass/radius planets will cool much faster. This will lead to hardening of the planet interior, which in effect will significantly decrease the planets magnetic field that shields from cosmic radiation and atmospherical stellar wind stripping. Furthermore, the thick crust of the planet will halt any plate tectonic activity. This lack of active geology will not benefit from the long-term climate stabilization produced by the CO_2 cycle (Rushby et al., 2018). Using the radiogenic flux model provided by Williams et al. (1997), Raymond et al. (2007) established that planets with roughly $M < 0.3M_\oplus$ will lack this important geological activity. Therefore, we only consider planets habitable with a $M \geq 0.3M_\oplus$. Using our conversion model in Section 3.3, this mass cut-off corresponds to $r \geq 0.72r_\oplus$.

The maximum mass limit for habitability is set by the limit in which the planet is no longer rocky, but exists as a gas giant. This limit is difficult to determine as the composition of the planet is really the important factor to considers. Observationally, there exists low mass planets like Kepler-11f ($M = 2.3 \pm^{2.2}_{1.2} M_\oplus$) with large radii ($r = 2.61 \pm 0.25r_\oplus$), indicating a gaseous planet (Lissauer et al., 2011). Conversely, the existence of large mass rocky planets like Kepler-10c ($M = 7.4 \pm^{1.3}_{1.2} M_\oplus$; Aigrain et al. 2017) with a radii of $r = 2.35 \pm^{0.09}_{0.04} r_\oplus$ have made mass limits difficult to pin down (Dumusque et al., 2014). Considering the entirety of the known population, the mass-radius relationship for known exoplanets produces a break around $4M_\oplus$ (Weiss & Marcy, 2014; Wolfgang et al., 2016). This break indicates a general transition from rocky to gaseous. However, using a interior planet model and Bayesian analysis, Rogers (2015) found a 95% confidence limit for the rocky planet transition at $r = 1.62 \pm^{0.67}_{0.08} r_\oplus$, with a best fit transition occurring at $r = 1.48 \pm^{0.08}_{0.04} r_\oplus$. Using the empirical mass-radius sample, Chen & Kipping (2017) found a similar transition value of $r = 1.23^{0.44}_{0.22} r_\oplus$. We also know that the Earth is rocky and habitable, therefore providing a pessimistic limit of $1r_\oplus$. Because it remains unclear where this transition takes place, we shall provide the results for an upper radius limit of $1r_\oplus$, $1.23r_\oplus$, $1.48r_\oplus$, and $1.62r_\oplus$ in Section 3.5.

The location of the planet is important for habitability as it receive enough stellar incident flux to support liquid water. Furthermore, the planet must not be so close to the star as to undergo a runaway greenhouse effect. This region is known as the Habitable Zone (HZ hereafter; Huang 1959; Kasting et al. 1993; Kopparapu et al. 2013). Current versions of this region account for the mass (surface gravity) of the planet and how atmospheric H_2O, CO_2, and N_2 will affect the planets ability to retain heat (Kasting et al., 1993; Kopparapu et al., 2014; Ramirez & Kaltenegger, 2014). The effective incident stellar flux (S_{eff}) is give by a fourth order polynomial fit of the effective stellar temperature (T_{eff}):

$$\begin{aligned} S_{\text{eff}}(M,T) = S_\star(M) + a(M) * T + b(M) * T^2 \\ +c(M) * T^3 + d(M) * T^4, \end{aligned} \tag{3.8}$$

where $S_\star(M), a(M), b(M), c(M), d(M)$ are all parameters fit the account for the mass of the planet and T is the normalized stellar temperature ($T = T_{\text{eff}} - 5780K$). We adopt the polynomial values provided by Kopparapu et al. (2014), linearly interpolating with mass between the provided $(.1, 1, 5M_\oplus)$ parameters. Using the "Runaway Greenhouse limit" provides a conservative estimate for the inner edge of the HZ. The physical distance (d) of this limit can be found by using the S_{eff} value derived in Equation 3.8:

$$d = \sqrt{\frac{L/L_\odot}{S_{\text{eff}}}}\text{AU}, \tag{3.9}$$

where L is the luminosity of the host star. The outer HZ limit is defined by the "Maximum Greenhouse limit", where CO_2 partial pressure remains just high enough to produce any amount of greenhouse heat retention. This border is independent of planet mass and fixed only by the host star's T_{eff}. We consider planets, that lay between these two limits, habitable.

In summary, we only consider a planet habitable if the semi-major axis of the planet lays within the bounds provided by the Kopparapu et al. (2014) HZ model and the planet mass lays between $0.72 \leq r \leq 1.62M_\oplus$.

3.5 Results

Using the updated radius, period, and multiplicity distributions from Zink et al. (2019b), we randomly assigned planets to the *Kepler* stellar sample. From this sampling we determined

Table 3.1: The probability that a solar-like star will process the indicated number of habitable planets. The upper radius limit was varied to account for the lack of a clear cut off for rocky planets and therefore habitability. Each row indicates the probability of finding at least the indicated multiplicity of planets within the habitable parameters of this study. The blank cells indicate multiplicities so rare that their occurrence provided no statistical significance ($> 3\sigma$). The $\eta_\oplus$ values correspond to the sum of all multiplicities. All $\eta_\oplus$ values are given in units of $\frac{\text{planets}}{\text{star}}$.

	$r \leq 1.00r_\oplus$	$r \leq 1.23r_\oplus$	$r \leq 1.48r_\oplus$	$r \leq 1.62r_\oplus$
One	0.132 ± 0.005	0.189 ± 0.007	0.227 ± 0.008	0.243 ± 0.009
Two	0.017 ± 0.002	0.036 ± 0.003	0.052 ± 0.005	0.059 ± 0.005
Three	0.0017 ± 0.0003	0.0051 ± 0.0008	0.010 ± 0.001	0.012 ± 0.002
Four	-	0.00065 ± 0.0001	0.0012 ± 0.0003	0.0017 ± 0.0004
Five	-	-	0.00021 ± 0.00006	0.0003 ± 0.0001
$\eta_\oplus$	0.151 ± 0.007	0.23 ± 0.01	0.29 ± 0.01	$0.32 \pm .02$

which planet are habitable using the criterion discussed in Section 3.4. Taking the number of habitable planets and dividing by the total number of stars in our stellar sample (86,605), provided us with a statistical value for habitability. This process was completed 25 times to account for variations in the distribution models and system archetypes. In Table 3.1 we present the results of this simulation.

Using the multiplicity distributions of Zink et al. (2019b) provides an opportunity to understand the occurrence of multiple habitable planets within a given system. In our most optimistic upper radius limit ($0.72r_{\oplus} \leq r \leq 1.62r_{\oplus}$), we find that multiplicity within the HZ should occur for GK dwarfs with a probability of 0.073 ± 0.005. This limit also provides statistical significance for systems with as many as five habitable planets. These extremely rare systems (probability=0.0003 ± 0.0001) consist mostly of tightly packed small mass planets that spread across the entirety of the HZ. Within our stellar sample we would expect 26 systems to harbor this unique architecture. However, even in our most pessimistic mass limit ($0.72r_{\oplus} \leq r \leq 1r_{\oplus}$), we still expect the existence of multiple habitable planet systems (probability=0.017 ± 0.002) with as many as three habitable planets within a single system (probability=0.0017 ± 0.0003). Although more rare, we should still expect to find multiplicity within the HZ for a significant fraction of GK dwarf systems. The intermediate upper radius limits are provided to show how these values change as we consider the inferred transitions from Chen & Kipping (2017) ($r \leq 1.23r_{\oplus}$) and Rogers (2015) ($r \leq 1.48r_{\oplus}$).

The $\eta_{\oplus}$ value corresponds the frequency in which you would expect to find a habitable planet. Since all previously studies have marginalized over multiplicity, we use this value for comparison. Using our updated population parameters we find an optimistic value of $\eta_{\oplus} = 0.32 \pm 0.02 \frac{\text{planets}}{\text{star}}$ for solar-like stars. This is consistent with Traub (2012), who found $\eta_{\oplus} = 0.34 \pm 0.14$ using the first 136 days of *Kepler* data. Using the same 134 day data set, Catanzarite & Shao (2011) found a range of 0.01-0.03 for $\eta_{\oplus}$. This discrepancy was caused by a lack of accounting for completeness and the inclusion of planets beyond 42 day periods.

A more complete *Kepler* sample (Q1-16) study came from Petigura et al. (2013a), who found $\eta_\oplus = 0.22 \pm 0.08$ implementing a 0.25-4 solar flux limit for habitability. Petigura et al. (2013a) only used the largest SNR planet from each system, avoiding any issues with multiplicity. For proper comparison we should consider our optimistic model with only one habitable planet (0.243 ± 0.009), which is within the uncertainty of Petigura et al. (2013a). Again using the same data set as Petigura et al. (2013a) but with a much more narrow HZ range (0.99-1.7AU), Silburt et al. (2015) found a smaller probability ($0.064 \pm^{0.034}_{0.011}$). Using a forward model method, Traub (2016) found an $\eta_\oplus$ value of 0.75 ± 0.11 and 1.03 ± 0.10 for K and G type stars respectively. Here, the wide rage of 0.8-1.8AU was considered habitable, inflating the calculated values. Additionally, the separation of planets by stellar spectral type reduces the number planets used for each extrapolation, producing significant deviations from the previously calculated values. Various assumptions about the habitable radius range and HZ cause much of the dispersion seen among these values. Recently, Barbato et al. (2018) used a simple population extrapolation with *Gaia* updates and found $\eta_\oplus = 0.35 \pm^{0.25}_{0.16}$. This simple method produced a value comparable to our more detailed calculation. Our clear habitability assumptions and updated distributions parameters provide the most current and complete $\eta_\oplus$ value.

3.6 Discussion

The radius distribution inferred from the *Kepler* data set provides sufficient coverage of the important radius range for habitability, extending from $0.5r_\oplus \leq r \leq 16r_\oplus$. This is well beyond the limits of $0.72r_\oplus \leq r \leq 1.62r_\oplus$ that we considered habitable. However, it is important to remember that almost all of the known exoplanets of this size have short periods, due to the low signal produced by such small radii transits. Our understanding of these small planets at longer periods is due to extrapolation from the period population of larger radius planets. It is possible that these small planets follow a unique distribution model, independent of their larger radius counter parts. Without a significant population of detected long period, small radius planets, extrapolations provide our best method of estimation.

Further concern lays in the fact that the period distribution (derived by Zink et al. (2019b)) fails to extend beyond the outer limits of the HZ in most cases (see Figure 3.1). Only 22% of the stellar systems have an outer HZ limit within a 500 day period (the maximum HZ extends to a period of 2782 days). This indicates that we are not entirely covering the parameter space available for habitable planets. It should be noted that the period power-law scales as $p^{-.64}$ in this region and simple extrapolation would indicate that very few planets exist at these long periods. At most, we expect a 21% increase when including this longer period parameter space. Furthermore, the outer HZ limit is set by the furthest location where the greenhouse effect can be experienced. This could be an optimistic limit, thus decreasing the HZ outer limit and reducing the need for the correction. Because of this lack of period coverage, our habitability calculations only provides a lower limit, but given the extrapolated power-law we do not expect significant changes as we extend out beyond 500 day periods.

The power-laws derived by Zink et al. (2019b) were also calculated under the assumption of total validation for planet candidates within the Kepler DR25 sample. As pointed out by Hsu et al. (2019), ignoring the reliability of these planets will lead to the extrapolation of an inflated $\eta_{\oplus}$ values. However, this change should be small as most of the planets within this sample appear to have high validation scores (Thompson et al., 2018). Because we expect this correction to be less significant than that caused by the lack of period space coverage, we still conclude that our $\eta_{\oplus}$ values provide a lower limit.

We also recognize that habitability is not limited to planets. It is possible that moons may provide the crucial ingredients for habitability. Using the fact that the gas giants of our solar system harbor several moons, Hill et al. (2018) argues that the existence of gas giants within the HZ of 70 *Kepler* stars may provide evidence for a significant population of terrestrial moons in the HZ. Since the calculations within the current study only considers exoplanets,

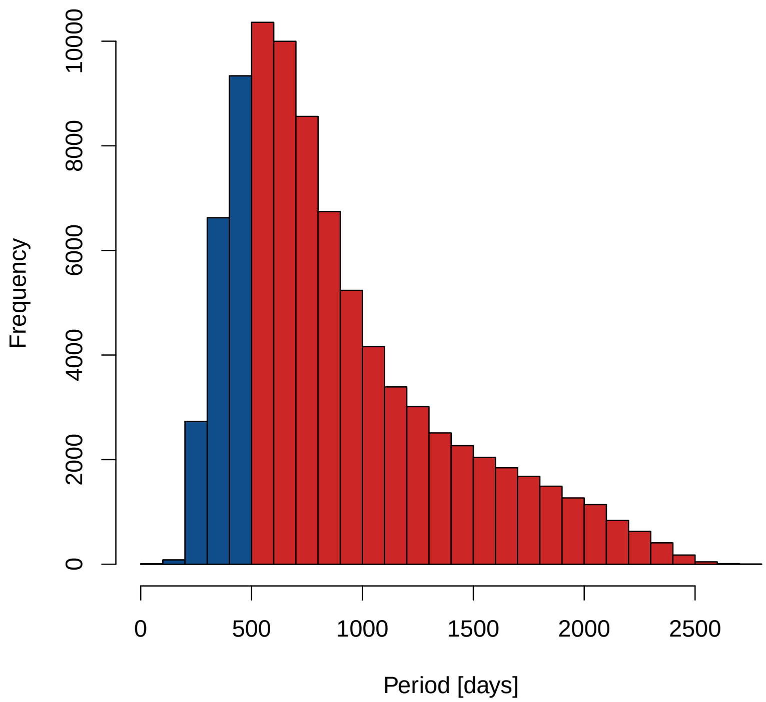

Figure 3.1: A histogram showing the outer HZ limit for the stellar sample of this study. The blue corresponds to the HZ limits within the 500 period range of our calculation and the red corresponds to the stars that have a HZ extending beyond a 500 day period.

the discovery of a sizable number of exomoons could largely inflate the amount of habitability expected within each system. Without some understanding of the exomoon population, it is difficult to estimate how other systems may differ from our own.

The results of this study indicate that optimistically $18 \pm 1\%$ of stars like the sun should only harbor one habitable planet. In this respect, our solar system is somewhat common. It is important to remember that (almost by definition) we have 2 planets (Mars and Venus) just beyond the parameter space of habitability. If these limits are correct, we nearly harbor a three habitable planet system, putting us among the rare $1.0 \pm 0.2\%$ of solar-like stars. The importance of such an architecture remains unclear, but could be crucial for sustaining life.

3.7 Conclusions

Using the updated population parameters of Zink et al. (2019b) and optimistic planetary traits required for habitability, we have provided a frequency estimate for habitable planets in the *Kepler* field. we break this frequency down to account for multiplicity within the HZ (see Table 3.1). Using our most optimistic radius cuts ($r \leq 1.62 r_{\oplus}$) we find $\eta_{\oplus} = 0.32 \pm 0.02$. This value could be larger in reality, as only 22% of the stellar sample provide a HZ that is contained within a 500 day period, but we expect such corrections to be small.

We find that multiplicity within the habitable parameter space should be some what common. Our calculation estimates that $5.9 \pm 0.5\%$ of solar-like stars should have more than one habitable planet with $0.03 \pm 0.01\%$ containing as many as five habitable planets.

This non-negligible fraction of systems expected to contain multiple habitable planets is good news for the *Interplanetary Eavesdropping* program run by SETI (The Search for Extraterrestrial Intelligence).[1] Here, as multiple habitable planet systems (such as Trappist-1)

[1]SETI Eavesdropping

come in conjunction with the Earth, an attempt is made to obtain leaked radio transmissions between the habitable planets. Our results indicate that $5.9 \pm 0.05\%$ (≈ 5000 stars within our sample) of GK dwarfs harbor this type of architecture. It is hopeful that many more systems of multiple habitable planets should be found in the near future, providing more potential targets for this search.

CHAPTER 4

Scaling K2: A Uniform Planet Sample for Campaigns 1-8 and 10-18

4.1 Introduction

The *Kepler* space telescope collected continuous photometric data for nearly 3.5 years from a small 0.25% patch of the celestial sphere (Koch et al., 2010; Borucki, 2016). This uniform monitoring of 150,000 stars enabled the discovery of over 4700 transiting exoplanet candidates through an automated detection pipeline (Jenkins et al., 2010a).[1] Removing the human component from the signal vetting process (`Robovetter`; Thompson et al. 2018), enabled the first homogeneous transiting small planet sample suitable for exoplanet population studies. Through full automation, the sample completeness can be measured via transit injection/recovery procedures (Petigura et al., 2013b; Christiansen et al., 2015; Christiansen, 2017; Christiansen et al., 2020). Here, artificial transit signals are injected into the raw light curves, enabling quantification of the pipeline's detection efficiency as a function of, for instance, the injection signal strength. In addition, the final *Kepler* catalog (DR25) also provided a measure of the sample reliability (the rate of false alarm signals) by testing the software's ability to remove systematics from the final catalog (Coughlin, 2017).

With these catalog measurements available, the underlying planet population can be ex-

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

tracted by correcting for the detection and orbital selection effects. Many studies have used *Kepler* automated planet catalogs to identify a remarkable surplus of sub-Neptunes and super-Earths at short periods (FGK Dwarfs: Youdin 2011; Howard et al. 2012; Petigura et al. 2013b; Mulders et al. 2018; Hsu et al. 2019; Zink et al. 2019b; He et al. 2019; M Dwarfs: Dressing & Charbonneau 2013; Hardegree-Ullman et al. 2019), despite their absence in the solar system. Several other population features have been discovered using the *Kepler* catalog. For example, the apparent deficit of planets near the 1.5–$2R_{\oplus}$ range, colloquially known as the "radius valley" (Fulton et al., 2017). This flux-dependent population feature indicates some underlying formation, or evolution, mechanism must be at play, separating the super-Earth and sub-Neptune populations (Owen & Wu, 2017; Gupta & Schlichting, 2019). Nearly 500 multi-planet systems were identified in the *Kepler* data, enabling examination of intra-system trends. Remarkably, very minor dispersion in planet radii is observed within each system (Ciardi et al., 2013). By examining the intra-system mass values from planetary systems that have measurable transit-timing variations (TTVs), a similar uniformity has been identified for planet mass (Millholland et al., 2017). Moreover, the orbital period spacing of planets appears far more compact than the planets within the solar system. These transiting multi-planet systems also appear to have relatively homogeneous spacing, indicative of a unvaried dynamic history. This combination of intra-system uniformity in orbital spacing and radii is known as the "peas-in-a-pod" finding (Weiss et al., 2018a).

One of the main objectives of the *Kepler* mission was to provide a baseline measurement for the occurrence of Earth analogs ($\eta_{\oplus}$). Despite the scarce completeness in this area of parameter space, several studies have extrapolated from more populated regions, providing measurements ranging from 0.1–0.4 Earth-like planets per main-sequence star (Catanzarite & Shao, 2011; Traub, 2012; Petigura et al., 2013b; Silburt et al., 2015; Burke et al., 2015; Zink & Hansen, 2019; Kunimoto & Matthews, 2020; Bryson et al., 2020a). Currently, it remains unclear what planet features are important for habitability, making precise measurements unattainable. Moreover, these extrapolations are model dependent and require a more empirical sampling of the Earth-like region of parameter space to reduce our uncertainty. Unfortunately, the continuous photometric data collection of the *Kepler* field was terminated due to mechanical issues on-board the spacecraft after 3.5 years, limiting the completeness of these long-period small planets.

Upon the failure of two (out of four) reaction wheels on the spacecraft, the telescope was no longer able to collect data from the *Kepler* field, due to drift from solar radiation pressure. By focusing on fields (Campaigns) along the ecliptic plane, this drift was minimized, giving rise to the *K2* mission (Howell et al., 2014; Van Cleve et al., 2016). Each of the 18 Campaigns were observed for roughly 80 days, enabling the analysis of transiting exoplanet populations from different regions of the local Galaxy. However, the remaining solar pressure experienced by the spacecraft still required a telescope pointing adjustment every ~ 6 hours. This drift and correction led to unique systematic features in the light curves, which the *Kepler* automated detection pipeline was not designed to overcome. As a result, all transiting *K2* planet candidates to date have required some amount of visual assessment. Despite this barrier, nearly 1000 candidates have been identified in the *K2* light curves (e.g., Vanderburg et al. 2016; Barros et al. 2016; Adams et al. 2016; Crossfield et al. 2016; Pope et al. 2016; Dressing et al. 2017a; Petigura et al. 2018; Livingston et al. 2018; Mayo et al. 2018; Yu et al. 2018; Kruse et al. 2019; Zink et al. 2019a). However, this assortment of planet detections lacks the homogeneity necessary for demographics research. Attempts have been made to replicate the automation of the *Kepler* pipeline for *K2* planet detection (e.g., Kostov et al., 2019; Dattilo et al., 2019), but these procedures still required some amount of visual inspection. Zink et al. (2020a) developed the first fully automated *K2* planet detection pipeline, using the `EDI-Vetter` suite of vetting metrics to combat the unique *K2* systematics. This pilot study was carried out on a single *K2* field (Campaign 5; C5 henceforth) and detected 75 planet candidates. The automation enabled injection/recovery analysis of the stellar sample, providing an assessment of the planet catalog completeness. Additionally, the reliability (rate of false alarms) was quantified by passing inverted the light curves—nullifying existing transit signals—through the automated procedure, analogous to what was done for the *Kepler* DR25 catalog. Any false candidates identified through this procedure are indicative of the underlying false alarm contamination rate.

With corresponding measures of sample completeness and reliability for C5, Zink et al. (2020b) carried out the first assessment of small transiting planet occurrence outside of the *Kepler* field, finding a minor reduction in planet occurrence in this metal-poor FGK stellar sample. This provided evidence that stellar metallicity may be linked to the formation of small planets; however, the weak trend detected requires additional data for verification.

This paper is a continuation of the Scaling *K2* series, which aims to leverage *K2* data to expand our exoplanet occurrence rate capabilities and disentangle underlying formation mechanisms. The intent of this study is to derive a uniform catalog of *K2* planets for all 18 Campaigns (with exception to C9, which observed the galactic plane), and provide corresponding measurements of the sample completeness and reliability. The underlying target sample and the corresponding stellar properties are discussed in Section 4.2. In Section 4.3 we outline the automated pipeline implemented in our uniform planet detection routine. We then procure a homogeneous planet sample and consider the interesting candidates and systems in Section 4.6. In Sections 4.7 and 4.8 we provide the corresponding measures of sample completeness and reliability for this catalog of transiting *K2* planet candidates. Finally, we offer suggestions for occurrence analysis and summarize our findings in Sections 4.9 and 4.10.

4.2 Target Sample

We first downloaded the most up-to-date raw target pixel files (TPFs) from MAST[2], which recently underwent a full reprocessing using a uniform procedure (Caldwell & Coughlin, 2020). To ensure consistency among our data set, we used the final data release (V9.3) for all available campaigns considered in this paper.[3] We began our search using the entire EPIC target list (381,923 targets; Huber et al. 2016). Of this list, we found 212 targets have FITS file issues our software could not rectify. For all remaining targets we used the `K2SFF` aperture #15 (Vanderburg et al., 2016), which is derived from the *Kepler* pixel response function and varies in size in accordance with the target's brightness. Apertures exceeding $10''$ radii are prone to significant contamination from nearby sources. Our vetting software, `EDI-Vetter`, is able to account for this additional flux, but the software's accuracy begins to decay beyond a radius of $20''$. Additionally, large apertures have a tendency to contain multiple bright sources, providing further complications. To address this issue, we put an upper limit on the target aperture size (79 pixels; $\sim 20''$ radius). This cut removed 4548 targets that have apertures we deemed too large. Of these rejected targets, 492 are in Campaign 11, which contains the crowded galactic bulge field. After applying these cuts 377,163 targets remained and were then used as our baseline stellar sample.

To parameterize this list of targets we relied on the values derived by Hardegree-Ullman et al. (2020), which used a random forest classifier, trained on LAMOST spectra (Su & Cui, 2004), to derive stellar parameters from photometric data (222,088 unique targets). Additionally, the available Gaia DR2 parallax information was incorporated to significantly improve our understanding of the stellar radii. From this catalog, additional measurements of stellar mass, metallicity, effective temperature ($T_{\rm eff}$), and surface gravity ($\log[g]$) were provided and

[2]https://archive.stsci.edu/k2/

[3]C8, C12, and C14 were released after our analysis was performed. Thus, a previous version (V9.2) was used for this catalog. Overall, we expect the differences to be minor and leave inclusion of these updated TPFs for future iterations of the catalog.

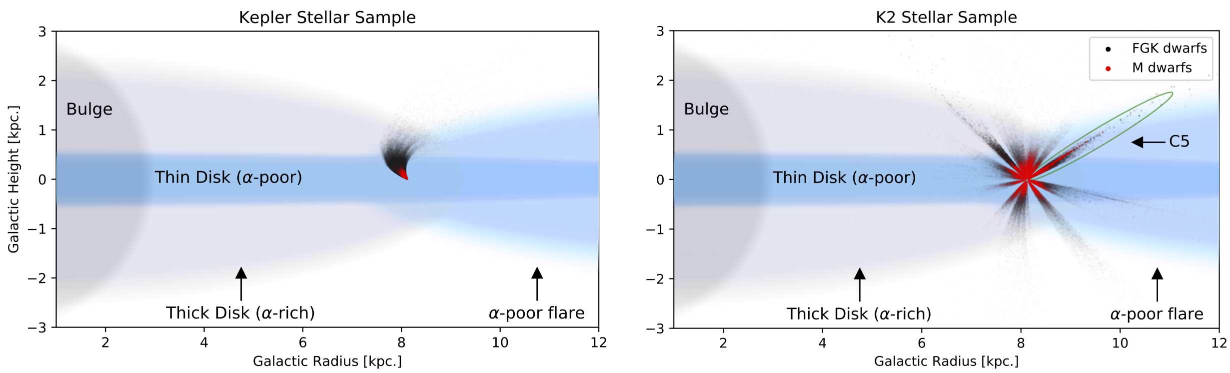

Figure 4.1: The position of the M dwarf and FGK dwarf population as a function of the Galactic radius and height for both the *Kepler* (left) and *K2* (right) stellar sample (spectral classification was established using the Berger et al. (2020b) and Hardegree-Ullman et al. (2020) stellar sample). The Galactic disk structure presented here follows the Hayden et al. (2015) interpretation. With respect to iron, the α-chain elements (O, Ne, Mg, Si, S, Ar, Ca, and Ti) exist in higher abundance in the thick disk (which truncates near the solar neighborhood) and lower abundance in the thin disk. Additionally, the α-poor disk begins to flare up beyond the solar neighborhood. The *K2* Campaign 5 stars, which are the focus of the current study, have been circled in green. The Galactic locations presented here are calculated using the *Gaia* DR2 (Gaia Collaboration et al., 2018).

used in this study. In cases where stellar parameters were not available in Hardegree-Ullman et al. (2020) (typically due to their absence in the Gaia DR2 catalog or not meeting strict photometric selection criteria), we used the stellar parameters derived for the EPIC catalog (Huber et al., 2016) (94,769 targets). Finally, we used solar values for targets that lack parameterization in both catalogs (41,061 targets).[4]

Excluding the 41,061 targets without stellar parameters, we isolated 223,075 stars that appear to be main-sequence dwarfs based on their surface gravity (log[g]$>$ 4). Within this subset we identified 48,702 targets as M dwarfs ($T_{\rm eff} < 4000K$), 164,569 as FGK dwarfs ($4000 < T_{\rm eff} < 6500K$), and 9,731 as A dwarfs ($T_{\rm eff} > 6500K$). This abundance of M dwarfs is nearly 17 times that of the *Kepler* sample (2808 M dwarfs; Berger et al. 2020b). Another noteworthy feature of the *K2* stellar sample is the wide range of galactic latitudes covered by the 18 campaigns. This enabled *K2* to probe different regions of the galactic sub-structure. In contrast, a majority of the *Kepler* stellar sample was bounded within the thin disk (see Figure 1 of Zink et al. 2020b for a comparison). In Figure 4.1 we show this galactic substructure span for *K2*. Making broad cuts in galactic radius (R_g) and height (Z) we can distinguish thick disk ($R_g <$ 8kpc and $|Z| >$ 0.5kpc) from thin disk stars ($|Z| <$ 0.5kpc). Overall, we found 191,002 dwarfs located in the thin disk, while 17,123 dwarfs reside in the thick disk. This sample distinction is interesting and may warrant further consideration in future occurrence studies.

The underlying stellar sample is important as it represents the population in which the planet candidates are drawn from. However, the stellar parameters are subject to change as more comprehensive data and more precise measurements become available (i.e., the upcoming Gaia DR3). To ensure our catalog remains relevant upon improved stellar parameterization, we take an agnostic approach to the available stellar features throughout our pipeline. In

[4]Zink et al. (2020b) showed that systematic differences between stellar catalogs are minor. Additionally, we emphasize our pipeline's agnosticism to stellar paramterization.

other words, each light curve is treated consistently regardless of the underlying target parameters. The only caveat to this claim is our treatment of the transit limb-darkening. Our transit model fitting routine requires quadratic limb-darkening parameters. We derived the appropriate values using the ATLAS model coefficients for the *Kepler* bandpasses (Claret & Bloemen, 2011), in concert with the available stellar parameters. This minor reliance on our measured stellar values should have little effect on the presented catalog. All other detection and vetting metrics in our pipeline are independent of this parameterization.

4.3 The Automated Pipeline

Our automated light curve analysis pipeline consists of four major components: pre-processing, detrending, signal detection, and signal vetting. In Figure 4.2 we provide a visual overview of this procedure, and shall now briefly summarize the execution of each step.

4.4 Data Processing

In the following section we will discuss how our pipeline takes the raw photometric data and converts it into usable transit signals. In brief, the target pixel files are downloaded from MAST and processed with the `EVEREST` Python package (Luger et al., 2016; Luger et al., 2018). The light curves then underwent two Gaussian Process regressions and a harmonic signal removal. Finally, the fully detrended data are examined for transit signals using the `TERRA` software (Petigura et al., 2013b).

4.4.1 `EVEREST` Pre-processing

The *K2* data set provides unique systematics that make the detection of exoplanets more difficult than that of the original *Kepler* mission. The spacecraft thrusters and roll cause periodic fluctuations in the photometric data. In a preliminary study, we compared different *K2* detrending software: `K2SFF` (Vanderburg & Johnson, 2014), `K2PHOT` (Petigura et al.,

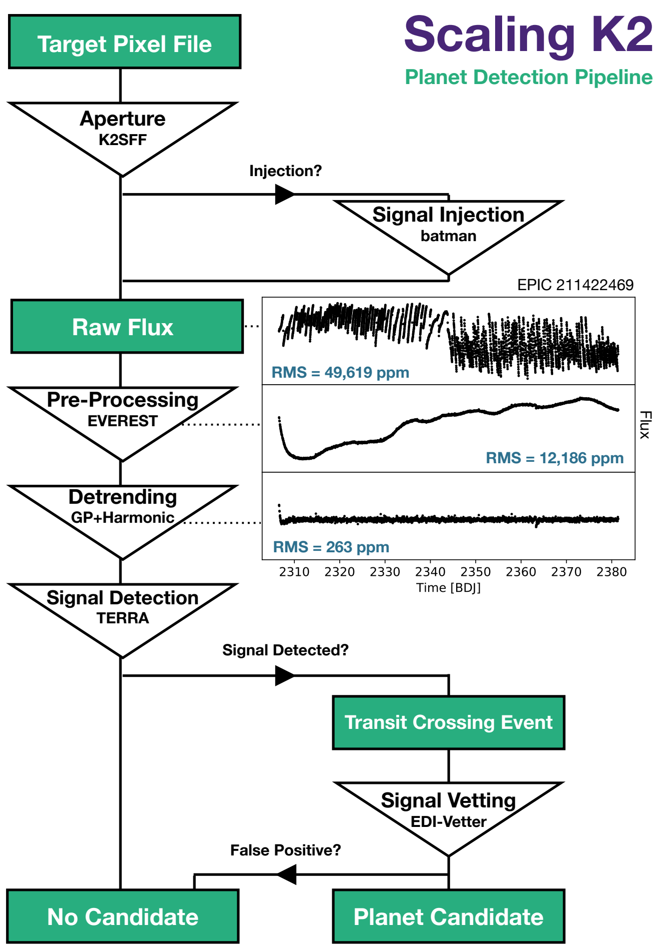

Figure 4.2: An overview schematic of the automated detection pipeline used in the current study to identify *K2* planet candidates. This follows the same procedure described in Zink et al. (2020a). The affects of pre-processing and detrending have also been depicted for EPIC 211422469, illustrating the importance of the corresponding steps.

2013b), and EVEREST[5] (Luger et al., 2016; Luger et al., 2018). We found that all of these algorithms had different performance issues and strengths. The DAVE vetting software (Kostov et al., 2019) considers several different detrended light curves simultaneously for each target, but such a task is computationally expensive when considering the desire to directly execute each detrending algorithm. We found that the EVEREST software provided good minimization of signal RMS and the most user-friendly software for spacecraft fluctuation removal. In overview, EVEREST uses the calibrated pixel-level data to produce time-series photometric light curves and corrects for systematic noise in those light curves by examining pixel-level correlations. All the *K2* Campaigns have been processed with EVEREST (however the calibrated pixel-level data files are currently undergoing reprocessed as part of the *K2* Global Uniform Reprocessing Effort[6], warranting an updated processing with EVEREST). Light curves with EVEREST processing are publicly accessible[7]. However, to ensure a uniform treatment, we undertook the task of reprocessing the Campaign 5 light curves with EVEREST, using the K2SFF aperture # 15, as suggested by Luger et al. (2018). This aperture uses a model of the *Kepler* Pixel Response Function to determine the size and shape of the target aperture (Vanderburg & Johnson, 2014; Bryson et al., 2010). This gave us the ability to later inject signals into the raw flux data (light-curve-level) before EVEREST processing (see Section 2.4). By directly running EVEREST on the calibrated pixel-level data (downloaded from MAST)[8], we could ensure the transit injection recovery is processed using the exact same conditions as that of the candidate sample.

On average, EVEREST masks about 8% of the light curve due to spacecraft anomalies or outliers. Occasionally, EVEREST will remove a large fraction of the data points within a light curve as the excessive noise is seen as numerous outliers. We remove targets where more that

[5]https://rodluger.github.io/EVEREST/

[6]https://keplerscience.arc.nasa.gov/k2-uniform-global-reprocessing-underway.html

[7]https://archive.stsci.edu/hlsp/everest

[8]https://archive.stsci.edu/k2/

50% of the light curve is masked. This severe masking only occurs in 37 of the Campaign 5 light curves, in most cases where the target is near the edge of the CCD. These 37 targets represent a very small fraction of the 25,040 light curves considered for this campaign.

4.4.2 GP and Harmonic Detrending

The `EVEREST` software removes the systematics caused by the thrusters and roll of the spacecraft, but an additional global detrending is needed to flatten the light curves. We achieve this through Gaussian Process (GP) regression and harmonic removal.

The first step in detrending is to establish the white noise level (σ). It is important that the detrending knows what is signal versus noise, and an accurate σ value will help achieve this goal. In well-behaved flat light curves this is an easy task, and σ is the standard deviation of the normalized flux. However, if the time series contains correlated noise, the standard method will artificially increase the measured uncertainty. To complicate this even further, deep transit signals or a large number of outliers can again inflate the RMS. We use the median absolute deviation (MAD), which is a measure of the median absolute value distance from the median value in the time series. This value can be scaled to estimate σ for the time series:

$$\sigma_{\mathrm{MAD}} = \frac{\mathrm{MAD}}{\Phi^{-1}(3/4)} \approx 1.4826\ \mathrm{MAD} \tag{4.1}$$

where $\Phi^{-1}(3/4)$ is the normal inverse cumulative distribution function evaluated at probability of $3/4$. This measure of σ is far more robust to outliers and transit signals. However, light curves that require significant red-noise removal will still inflate the inferred σ value. To limit this issue we break the light curve into bins of width 1 day, as the daily flux trend should be minimal. We measure the σ_{MAD} value for each of these bins and take the median σ_{MAD} value from these bins as our measured value for σ. It is important to note that < 1 days stellar variability can still inflate our measured σ value, but this effect is insignificant when looking for long term trends as done here.

One challenge of GP detrending is that GPs can fit out real transit signals. A thorough discussion of this issue can be found in Hippke et al. (2019), with several alternative detrending methods which we hope to consider in future iterations of our pipeline. To avoid this issue, we attempt to mask out outliers and potential transits. We again bin the time series by single day intervals. Within each bin we look for points that exceed $3\sigma_{\mathrm{MAD}}$ from the median of each bin. These points are only masked for the detrending process and unmasked for the rest of our pipeline. By manually adjusting the threshold, we found that a 3σ clipping was sufficient in masking most significant transits. Additionally, any transit existing below this limit was likely to be seen as noise by the GP, and therefore not removed. An additional restriction we place on our GP is that we do not allow it to choose detrending periods < 5 days. Periods below this range are more prone to over-fitting transit signals as the transit duration is often a significant fraction of a day. Putting this limitation in place decreases our chances of modifying the transit depth. We discuss this issue more thoroughly in Section 4.6.

To implement our GP regression the pipeline uses `PyMC3` (Salvatier et al., 2015) with a wrapper built by the `Exoplanet` software (Foreman-Mackey et al., 2019). Our detrending uses the "Rotation" kernel which is a combination of two simple harmonic oscillators, meant to mimic stellar rotation/noise. Our pipeline uses this GP detrending in two steps: First, it looks for long term trends (periods > 10 days and general flux drifting), then after removing the long term trends, it looks for short term trends (5 days $<$ period < 10 days) and subtract away the best fit GP. However, stellar oscillations can also exist on very short timescales (often < 0.5 days). Any attempt to remove these oscillations with a GP will almost certainly remove transits as well, thus a harmonic fitter is used to address this issue.

Unlike a GP, a harmonic fitter does not have the flexibility to fit and remove long period transits. It can only affect the signals with strong similarities to a sine function and the long periods of flat base line between transits makes this unlikely. This can become problematic when you have a system with multiple transiting planets, decreasing the gaps between transits and mimicking a sine curve (Christiansen et al., 2013; Zink et al., 2019b). Additionally, short period giant planets can also be affected by this harmonic removal process, by either decreasing the transit depth or removing the signal completely. By choosing an upper limit of 0.5 days for our harmonic fitter, we minimize such occurrences. We begin by unmasking the GP sigma-clipped data and search for the strongest periodic signal below 0.5 days using a Lomb-Scargle periodogram. We then use a χ^2 fitting algorithm to measure the amplitude (A) of the signal. It is important that we do not introduce additional noise into the data set, so we require $A > 10\sigma_A$ before removing the harmonic signal. We found in practise that this limit rarely over-fit the data, but we address such possibilities in our vetting metrics (Section 4.5.9). In Figure 4.2 we show how the raw flux data will change at each step of our light curve processing.

Looking at our GP and harmonic fitter, there is an apparent gap in our considered period range for detrending ($0.5 <$ periods < 5 days). Any attempt made to fit out this period range either introduced additional artificial signals into our data, or artificially reduced transit depths. Fortunately, most stars have rotational periods > 5 days (McQuillan et al., 2013) and most variable stars oscillate at periods < 0.5 days, indicating that few non-planetary periodic signals occupy this parameter space. Additionally, our vetting algorithm rejects any signals found with such harmonic features, minimizing the possibility of a False Positive (FP) detection (Section 4.5.9).

4.4.3 `TERRA`

Once the light curve has been fully processed, we begin our search for Threshold Crossing Events (TCEs). These are signals with at least three transit-like events that reach or exceed a specific statistical significance. We discuss our process of selecting this threshold in the next Section (4.4.4). To search for these signals, we use the `TERRA` search algorithm (Petigura et al., 2013b). This software begins by masking outliers and then uses a finely-spaced grid-based search algorithm to find the largest signal-to-noise ratio (SNR) event in phase-folded period space of the light curve. This 3D grid search provides a measure of period (P), the transit duration (t_{dur}), the transit emphemeris (t_0), and SNR. For transiting planets detected by the *Kepler* pipeline, the SNR is measured using the Multiple Event Statistic (MES; Jenkins, 2002), which indicates the strength of the whitened signal assuming a linear emphemeris. We will use MES in reference to the strength of the signal henceforth.[9]

To recover systems with more than one transiting exoplanet, we consider the five signals with the largest MES in the light curve. We do not expect many systems to contain more than five distinct detectable transits within a period of 38 days (*Kepler*-80 and *TRAPPIST*-1 are currently the only known exceptions). However, Kruse et al. (2019) found evidence for one such system (EPIC 210965800) with six planets in Campaign 4 of *K2*. We acknowledge that we may lose some of these higher multiplicity systems by limiting our search to five planets, but such a restriction is necessary to minimize our computational cost. After the largest MES signal has been detected, we then mask 2.5× the transit duration ($1.25t_{dur}$ on either side of predicted transit midpoint), allowing the next largest signal to be detected. This process continues until either five TCEs have been found, or the largest MES value is less than the detection threshold. Zink et al. (2019b) showed that this type of masking will make higher multiplicity systems more difficult to detect, as nearby transits have the potential of being partially or fully hidden by this type of masking. However, removing the signal with a transit model is impractical. A majority of the signals are FPs that do not match well with the transit model. Thus, subtracting the best fit model will not actually remove the signal, but rather morph the data so that the grid search continues to detect the same anomaly repeatedly. In addition, any poorly-fit real transit signal will not be completely removed through model subtraction, making smaller MES signals very difficult

[9] We note that the SES and MES used by `TERRA` and *Kepler* project Transiting Planet Search (TPS) are analogous, but not strictly equivalent. Differences in the detailed constructions of these statistics may be found by comparing Jenkins et al. (2010a) and Petigura et al. (2013b).

to find. Therefore, masking is the only way to ensure the previous signal is not being re-detected at each iteration of multiplicity (m). To minimize the effects of masking we choose to only remove 2.5× the transit duration, whereas the *Kepler* pipeline used a mask of 3× the transit duration, which should decrease the probability of masking neighboring transits.

Searching for transits, we limit the orbital period search range to [0.5, 38] days. The upper limit of 38 days is set by the span of the data (nominally 74.84 days for Campaign 5), which permits the existence of three transits within the given window. The lower limit was set to minimize contamination from the abundance of harmonic signals that exist at periods < 0.5 days (`TERRA` can detect signals with sufficient MES and periods < 0.5 days at a multiple of the true period, but in practise we found these are more often FPs than true planets). Additionally, we wanted to minimize the probability of finding signals artificially introduced by the harmonic fitter (Section 4.4.2). Finally, of the over 4000 confirmed exoplanets, only 15 have orbital periods < 0.5 days (0.37%), including six planets found in the *K2* fields. Thus, limiting our search should have a small effect on the extracted period population. Users concerned with the occurrence of ultra-short period (USP) planets are encouraged to perform their own custom search.

4.4.4 Selecting a TCE Limit

The *Kepler* pipeline used a default TCE limit of 7.1σ (Jenkins et al., 2002). Any signal that produces a MES value of this magnitude or greater was considered for vetting. Assuming Gaussian statistics, this limit would only allow one false alarm signal through in the entirety of the *Kepler* mission. However, the distribution of TCEs has long tails, where false alarm detections are more common at 7.1σ than originally expected. *Kepler* found, on average, that 16% of the targets produce a $\geq 7.1\sigma$ event (198,706 light curves; 32,534 TCEs), and a majority of them were later classified as FPs (Thompson et al., 2018).[10]

[10] In reality *Kepler* detected TCEs in about 8% of the light curves. A large fraction of these light curves contained multiple TCEs, inflating the average.

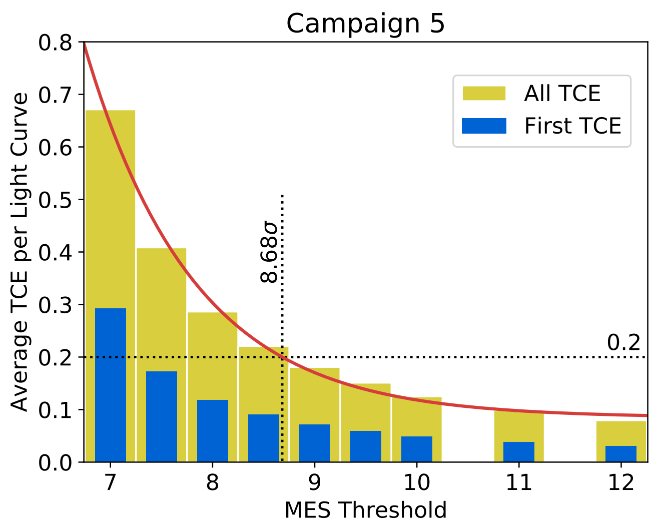

Figure 4.3: MES threshold plotted against the average number of TCEs found for each light curve in Campaign 5. Many of the light curves will produce more than one TCE, thus the yellow bars (All TCE) represent the total number of detected TCEs divided by the number of light curves searched. The blue bars (First TCE) only consider the first TCE found in each light curve. The red line represents the best fit exponential to the All TCE bars. We select the threshold (8.68) where the average light curve will contain 0.2 TCEs.

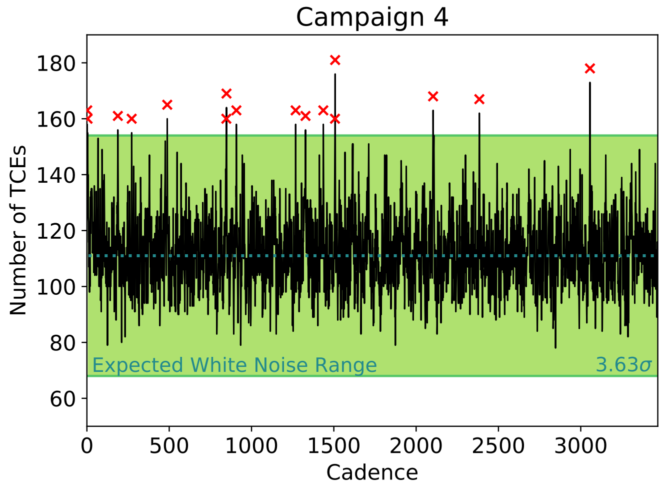

Figure 4.4: An illustration of the Skye excess TCE identification process for Campaign 4. The expected white noise range for this number of cadences is plotted in green (3.63σ). The median number of TCEs is shown by a central blue dotted line. We highlight the rejected cadences with a red x. These 16 cadences exceed the Skye excess limit and are masked in our final signal search.

K2 photometry contains far more outliers than *Kepler* prime photometry, so we tested different MES thresholds to look at the occurrence of TCEs at each limit (Figure 4.3). The number of TCEs appears to match well with an exponential distribution, and the multiplicity of TCE occurrence does not appear to be affected by a difference in MES threshold. The number of "First TCE" detections, where we only consider the first TCE trigger for each light curve, is roughly 50% of the "All TCE" detections at each of the tested thresholds. This indicates that a light curve where a TCE was triggered, was likely to trigger more than one TCE. If this ratio changed significantly at each tested threshold, an argument could be made to select a threshold at some level of similarity between the two values. Since no clear convergence occurs, we select a threshold (from our exponential fit) in which on average 20% of the light curves will contain a TCE (8.68σ; determined by the fit exponential function). Since our threshold is higher than the *Kepler* 7.1σ, we sacrifice some sensitivity in order minimize the number of false alarm signals (Thompson et al. (2018) found that reliability significantly increased for signals with MES>8 for the *Kepler* pipeline). Previous studies have selected even higher thresholds (9σ; Kruse et al., 2019), which indicates that our survey will be able to detect smaller and noisier signals at the cost of a higher false positive rate.

4.4.4.1 Skye Excess TCE Identification

Certain cadences are prone to triggering TCEs within the population of light curves. These can usually be attributed to spacecraft issues and are likely not astrophysical. Inspired by the "Skye" metric usied in Thompson et al. (2018), we minimized this source of contamination by considering the total number of TCEs with transits that fall on each cadence. The median value and expected white noise range was then calculated using:

$$\sqrt{2}\,\mathrm{erfcinv}(1/N_{\mathrm{cad}})\,\sigma, \tag{4.2}$$

where erfcinv is the inverse complementary error function, N_{cad} is the total number of cadences in the campaign, and σ is the median absolute deviation in the number of TCEs detected at each cadence. This calculated value represents the largest deviation expected from this number of cadences under the assumption of perfect Gaussian noise. Any cadences that exceed this limit are likely faulty and warrant masking. In Figure 4.4 we provide an example of this procedure for Campaign 4. To further assist future studies we made the Skye mask for each campaign publicly available.[11].

Once established, the Skye mask was used to reanalyse all light curves, removing problematic cadences. The resulting list of significant signals were then given the official TCE label and allowed to continue through our pipeline. Overall, we found 140,046 TCEs from 52,192 targets. However, a vast majority of these signals are false positives and required thorough vetting.

[11] http://www.jonzink.com/scalingk2.html

4.4.5 Prioritizing Reliability Over Completeness

We selected parameters for our pipeline which maximize the completeness and the reliability of our sample. Adopting the philosophy of the *Kepler* DR25 catalog (Thompson et al., 2018), we willingly allow some known planet candidates to achieve a FP label in order create a uniformly vetted, highly reliable catalog. The parameters described in Section 4.5 have been tuned to minimize such misclassification, without allowing additional FPs through. We hope to maintain these same parameters across all campaigns, so that an aggregate occurrence calculation using all 18 campaigns can be made seamlessly.

4.5 EDI-Vetter

One of the more difficult tasks in any automated transit search is the vetting process. This is the final defense against transit FPs and must be robust against many different types of light curve anomalies. Automatted vetting processes have been created for the *Kepler* pipeline (McCauliff et al., 2015; Thompson et al., 2018), but the *K2* data set is far noisier and additional tests are required. We build upon the criteria used by the `RoboVetter` (Thompson et al., 2018) to create `EDI-Vetter` (Zink, 2019), which is more robust to the issues unique to the *K2* data set. A user-friendly version of this software is made publicly available on GitHub[12]. Here we discuss the metrics `EDI-Vetter` uses to ensure our candidate list is pure with high reliability.

To begin, our software uses the maximum likelihood parameters found by `TERRA` to fit a transit model to the detected TCE signal. Using the affine invariant sampler of Goodman & Weare (2010), as implemented in Python by Foreman-Mackey et al. (2013) (`emcee`), and the `batman` transit model (Kreidberg, 2015; Mandel & Agol, 2002), we confirm that the `TERRA` values found a global maximum likelihood and determine the maximum likelihood of additional transit parameters. We produce our transit model with the `batman` Python

[12] https://github.com/jonzink/EDI-Vetter/

package (Kreidberg, 2015). Using 100 semi-independent walkers, 250 burn-in steps, and 100 parameter test steps (35,000 total step; 10,000 test samples), we estimate the maximum likelihood and uncertainty for the transit emphemeris (t_0), the radius ratio ($R_{pl}/R_\star$), the impact parameter (b), the signal period (P), the semi-major axis to stellar radius ratio (a_{pl}), the transit duration (t_{dur}), and the transit depth. We assume a circular orbit and derive two quadratic limb darkening coefficients using the stellar parameters available on ExoFOP (Huber et al., 2016) along with the ATLAS model table for *Kepler* bandpass limb darkening coefficients from Claret & Bloemen (2011). According to the literature for `emcee` [13], 50 times the integrated auto-correlation time (τ) is the lowest limit in which the MCMC sampler results can be trusted. We find $\tau \sim 50$ samples for our 5 parameter model, indicating we should require at least 2,500 samples before our results can be considered meaningful. Clearly, the 10,000 samples taken by our MCMC is near this lower limit and likely insufficient for a thorough parameter estimation, but we sacrifice some accuracy to significantly improve the computational speed of our software. For vetted candidates we run a more thorough parameter estimation as described in Section 4.6.

In the next 13 sub-sections (4.5.1-4.5.13) we provide a detailed description of the filters used by `EDI-Vetter`.

4.5.1 Previous Planet Check

This test ensures that the detected TCE is not a repeat of a previous signal (Section 3.2.2 of Coughlin et al., 2016). TCE re-occurrence is common for FPs, where masking is inefficient at removing the entirety of the light curve anomaly. Additionally, true signals which have periods misidentified by some integer multiple will produce multiple TCE signals. This test helps remove such redundancies. If the pipeline finds more than one TCE for a given light curve, this test will be enacted in decreasing signal detection order, i.e. the first signal

[13] https://emcee.readthedocs.io/en/latest/tutorials/autocorr/

detected will not be subject to this test, the second signal will be tested against the first, and the third signal will be tested against the first and the second signal.

We define ΔP as the separation in periods, normalized by the smaller of the two periods:

$$\Delta P = \frac{P_B - P_A}{P_A} \tag{4.3}$$

where P_A and P_B represent the shorter and longer periods respectively. To then determine the offset from the nearest integer multiple δP is calculated.

$$\delta P = |\Delta P - \mathrm{round}(\Delta P)| \tag{4.4}$$

where the round() function will round to the nearest integer value. We can then determine how statistically significant (σ_P) a period separation is using the erfcinv() (inverse complementary error function)

$$\sigma_P = \sqrt{2}\ \mathrm{erfcinv}(\delta P) \tag{4.5}$$

where larger δP values will produce small σ_P values. For consideration as a candidate we require high detection order signals (signals found after the first signal within the light curve) to be separated with a $\sigma_P \leq 2$ for candidate consideration. If a signal matches with a previous detection with a $\sigma_P > 2$, it is likely a repeated detection. However, we also must consider the case of resonant systems, where integer period separations are common. To avoid falsely eliminating such systems, we also consider the ephemeris separation for the two signals (Δt_0).

$$\Delta t_0 = \frac{|t_{0A} - t_{0B}|}{t_{dur}} \tag{4.6}$$

where t_{dur} is the transit duration of the signal in question and t_{0A} and t_{0B} are the ephemeris times for the first transit within the data set. If Δt_0 is large (≥ 1) and ΔP is small (≤ 2), it is possible that the signals are from a resonant orbit. The caveat to this is the case where the higher detection order signal is fitting to the secondary eclipse of the transit. To ensure this is not the case ΔSE_1 and ΔSE_2 are calculated:

$$
\begin{aligned}
\Delta SE_1 &= \frac{|t_{0A} - t_{0B} + P_A/2|}{t_{dur}} \\
\Delta SE_2 &= \frac{|t_{0A} - t_{0B} - P_A/2|}{t_{dur}}
\end{aligned}
\tag{4.7}
$$

Again, a small ΔSE_1 or ΔSE_2 (< 1) indicates the higher order detection is likely fitting to the secondary eclipse. This metric will only be sensitive to circular orbit eclipsing binaries, where the secondary eclipse is at a phase of 0.5. Future iterations of this pipeline will consider non-circular orbits in this metric.

All of these metrics are assuming the previous planet was a real signal. If the previous detection was deemed a FP and $\sigma_P > 2$, this signal is likely just a repeated detection from an under masked anomaly. Thus, the signal will be flagged as a FP.

For clarity, we summarize here the full process in which a signal will be classified as a FP. If $\sigma_P > 2$ and the previous signal was deemed a FP, this detection will be flagged as a FP. If $\sigma_P > 2$ and $\Delta t_0 < 1$, this detection will be flagged as a FP. If $\sigma_P > 2$ and $\Delta SE_1 < 1$ or $\Delta SE_2 < 1$, this detection will be flagged as a FP. Otherwise, this signal will continue to be considered for candidacy.

4.5.2 Binary Blending

In this metric we seek to identify cases where the signal is due to an eclipsing binary, either from the target or from a nearby source, but with such a large impact parameter and/or dilution from the third light source that the observed depth is comparable to that of a transiting planet. During the original *Kepler* mission, flux contamination from nearby sources was ruled out by fitting the pixel response function model (Bryson et al., 2010) to the star in and out of transit, looking for offsets in the photo-center of the light (the centroid). If statistically significant differences appeared, the target was considered contaminated (Mullally, 2017). Two scenarios could result in such a shift. First, contamination from a source within a few arcseconds of the primary target would produce this type of shift during transit. Second, a very deep transit signal from a nearby eclipsing binary could contaminate a static light curve, causing a shift during transit. Such methods are difficult to apply to *K2* because of the roll experienced by the telescope, which moves the target image across several pixels. Using cadences with similar roll angles, the `DAVE` software is able to recoup some of this detection ability (Kostov et al., 2019). However, the shorter data span of each *K2* campaign, and the lack of similar roll angle data, makes finding these statistical differences more difficult. Instead, we use the Gaia DR2 data set (Gaia Collaboration et al., 2018) to look for flux contamination from nearby sources.

Gaia can resolve nearby sources down to 1″ for G-band Δmag $\lesssim 3$ (Ziegler et al., 2018). Since the *K2* pixel scale is 3.98″, we can achieve sub-pixel resolution using cross-matching alone. We utilize the gaia-kepler.fun[14] cross-match database with a 20″ search radius to look for these contaminants. To ensure we have correctly matched the target star, we require the *"phot_g_mean_mag"* parameter to be within 1 magnitude of the assigned *"k2_kepmag"*. These values are often almost indistinguishable because the *Gaia* G band and *Kepler* band probe nearly the same wavelength range. If this criteria is not met, we check the next closest source within the 20″ search radius to see if it meets this criteria. This process will continue until a target of the correct magnitude is found, or all of the sources in the 20″ search radius have been tested. If no source is within this 1 magnitude limit, we will select the source that provides the closest magnitude match. This closest match scenario was applied to 301 of the C5 targets (1.2% of the total tested targets).

As used for the `EVEREST` pipeline (Luger et al., 2018), we adopted aperture #15 from the `K2SFF` catalog (Vanderburg & Johnson, 2014). Using the size of the aperture, we can determine the amount of flux contamination from neighboring binaries. Often the apertures select pixels in a circular manner around the target, but in cases where additional point sources are nearby, the aperture may be elongated, capturing additional flux from the nearby source.

[14] https://gaia-kepler.fun

To account for this possible elongation, we consider flux contamination within this radius (Δap):

$$\Delta ap = 3.98'' \sqrt{\frac{N_{pix}}{\pi} + 1} \tag{4.8}$$

where N_{pix} is the number of pixels for the given aperture. This Δap is likely overestimating the flux contamination for many of the well-behaved apertures, where the aperture is large and circular, but is important to ensure high reliability. To determine the amount of flux contamination, we compute the flux ratio for the nearest potential contaminating source (F^{Ratio}_{Gaia}):

$$\frac{F_{neighbor}}{F_{\star}} = F^{Ratio}_{Gaia} = 10^{\frac{\Delta m_{Gaia}}{-2.5}} \tag{4.9}$$

where $F_{neighbor}$ is the expected flux from the nearby source, $F_{\star}$ is the expected flux from the target star, and Δm_{Gaia} is the difference in magnitude between the target source and the possible contaminant in the Gaia G band. To determine the impact of this external source on our target, we calculate the ratio of total flux to the target flux (F^{Tot}_{Gaia}):

$$F^{Tot}_{Gaia} = \frac{F_{\star} + F_{cont}}{F_{\star}} = 1 + \frac{F^{Ratio}_{Gaia}}{2}\left(1 + \mathrm{erf}\left(\frac{\Delta ap - \Delta d}{2.55''\sqrt{2}}\right)\right) \tag{4.10}$$

where F_{cont} is the fraction of flux from $F_{neighbor}$ within the aperture of the target and Δd represents the angular distance between the target and the contaminating source. Using the error function in this method assumes a 1D Gaussian PSF with a standard deviation of 2.55″ (corresponding to the 6″ FWHM of the *Kepler* PSF). The 1D assumption will certainly overestimate the amount of flux contamination, but this cautious approach will improve our reliability against astrophysical FPs. For targets with multiple neighboring stars we consider the flux contamination from all the sources within a 20″ radius of the source.

One limitation to this approach is that our search limit is bound by the 20″ search radius of the gaia-kepler.fun cross-match. In some rare instances, an aperture will extend beyond this radius ($N_{pix} > 78$). Upon visual inspection, most of these targets are in very crowded fields with several stars within a single aperture. Implementing this aperture limit, we remove 54

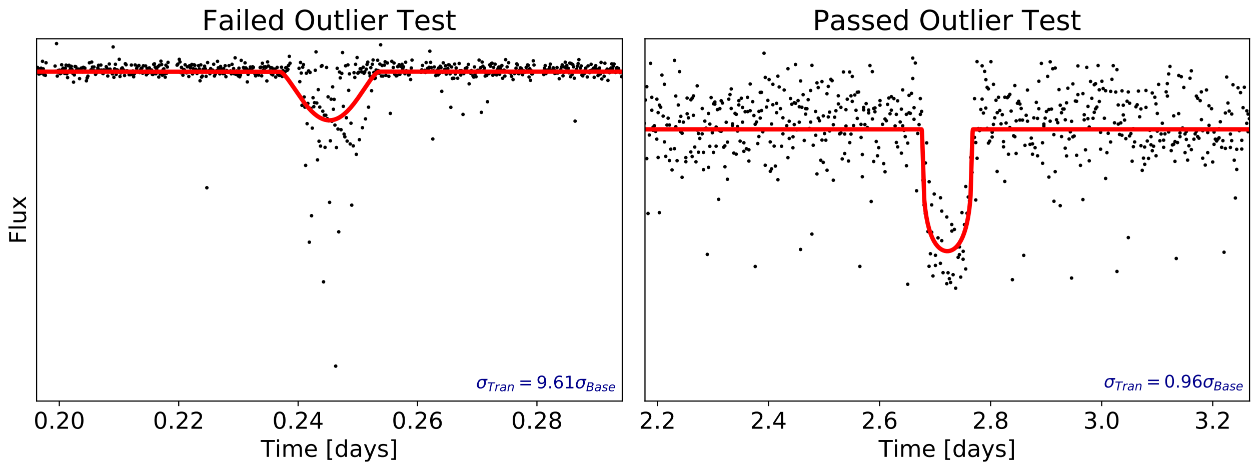

Figure 4.5: **Left** A TCE with a period of 0.5 days, likely caused by the alignment of measurements made near thruster firings. The large in-transit RMS (σ_{Tran}) is an indicator that the singal is a FP. This TCE was rejected using the outlier variance test. **Right** EPIC 211331236.01 with a period of 5.4 days. The in-transit deviations match well with the out of transit fluctuations. Thus, this signal was not removed by our outlier variance test.

targets from our search. Additionally, contaminating sources can also exist beyond the 20″ search radius. To help account for these cases, we also use the 2MASS J-band photometry. Pulling from ExoFOP[15], which provides 2MASS cross-matches beyond 20″, we calculate equation 4.9 for 2MASS and find the potential contamination (F^{Tot}_{2MASS}):

$$F^{Tot}_{2MASS} = 1 + \frac{F^{Ratio}_{2MASS}}{4}\left(1 + \mathrm{erf}\left(\frac{\Delta ap - \Delta d}{2.55''\sqrt{2}}\right)\right) \tag{4.11}$$

Here, it is important to note that we assume the flux contamination will be half as strong as a similar contaminant in the Gaia band. This is to account for the fact the 2MASS J-band photometry cannot be directly compared to the expected *K2* flux. Howard et al. (2012) suggests a flux ratio of $\sim 1/4$ for a J-band to K-band conversion, however we choice 1/2 to err on the side of over-estimating potential contamination. We find that F^{Tot}_{2MASS} is usually very close to unity, only in extreme cases where the J-band contamination is significant, and

[15]https://exofop.ipac.caltech.edu

$F_{Gaia}^{Tot} = 1$, will the 2MASS flux ratio play a role.

To determine the impact of this flux contamination on our source, we consider how F_{Gaia}^{Tot} and F_{2MASS}^{Tot} affect the inferred planet radius. One of the biggest concern with binary contamination is the potential to decrease the transit depth and falsely present an eclipsing binary as a planet candidate. Additionally, in systems with a transiting planet, the planet radius is derived from the depth of the measured transit. When flux from an additional star is present, the depth of the transit will be diluted and result in an underestimation of the planet radius (Ciardi et al., 2017; Fulton et al., 2017; Matson et al., 2018). We adopt the same radius and impact parameter requirements suggested by Batalha et al. (2013), but now include a flux contamination correction.

$$\frac{R_{Pl}}{R_\star}\sqrt{\max(F_{Gaia}^{Tot}, F_{2MASS}^{Tot})} + b \leq 1.04 \tag{4.12}$$

where b is the impact parameter and $R_{Pl}/R_\star$ is the radius ratio fit to the transit in question. The max function chooses the higher of the two flux ratios. This ensures we do not double count the flux contamination from a source. The median F^{Tot} value for Campaign 5 targets is 1.00003, indicating that most targets are well isolated. Any signals that exceed the provided limit (Equation 4.12) will be flagged as a FP. This method of FP flagging allows us to remain agnostic to the stellar parameters. Additionally, we put an upper limit on $R_{Pl}/R_\star$. Any TCE found with $R_{Pl}/R_\star > 0.3$ will automatically be flagged as a FP. This limit is imposed to eliminate low impact parameter (b) eclipsing binaries.

4.5.3 Transit Outliers

One common issue seen in many of the TCEs found in the *K2* light curves is the chance alignment of several outliers. This metric is aimed to identify when such cases occur. The general increase in noise caused by the spacecraft roll motion and thruster firing caused a significant number of outliers to occur in the flux data set. Much of this is removed by the `EVEREST` processing, but occasionally a few will make it through (evidence of this can be seen in the right-hand panel of Figure 4.4). When this occurs, the TCE search is likely to find a period where the outliers line up when phase folded. These chance alignment signals are hard to distinguish with previously developed vetting tools, but easy to spot by eye. Here we present a unique test to eliminate these anomalies.

We use the best fit model to subtract the signal from the folded data set. We first check to see if the in-transit residuals have a larger variance (σ_{Tran}) than the rest of light curve (σ_{Base}). However, this calculation is complicated by larger MES signals. We found that many high MES transit signals will produce larger variances in the residuals because of either poor fitting or poor detrending. Fortunately, many of these chance alignment outliers produce a low MES signal. Thus, we deem any transit residual with $\sigma_{\text{Tran}} > (0.461 * \text{MES} - 2)\sigma_{\text{Base}}$ a FP. This threshold was tuned in correspondence to an acceptable increase of $2\times$ the in-transit RMS at the detection limit and $4\times$ the RMS at MES$= 13$. In Figure 4.5 we show two signals, one which passed this test and one which failed.

We often find that several aligned outliers will trigger a TCE, but retain a low σ_{Tran} due to a larger number of non-outlier baseline flux measurements included in the σ_{Tran} calculation. To account for these cases, we count up the number of in-transit cadences that produce outliers (residual$> 3\sigma_{\text{Base}}$) in the transit residuals. If this number (Out) exceeds the following limits, the signal is deemed a FP:

$$\begin{aligned} \text{if MES} < 40: &\quad Out > N_{\text{Tran}} \\ \text{else}: &\quad Out > 0.3 \times \text{MES} + 4 \end{aligned} \tag{4.13}$$

where N_{Tran} is the number of transits. Equation 4.13 allows, on average, one outlier per transit for signals with MES< 40. For stronger signals, outliers are more likely to be triggered due to poor model fitting. This problem becomes more significant as MES increases, thus we use the linear function of MES as a threshold for these cases. This function was tuned to allow 16 outliers at MES$= 40$ and and 25 outliers at MES$= 70$. This metric performs well against outlier alignments, but has a mild tendency of flagging high MES, but poorly detrended planets, as FP. This is the unfortunate cost of attaining a high reliability catalog.

4.5.4 Individual Transit Check

Much of this part of our pipeline is derived from the "Marshall" test (Mullally et al., 2016). For clarity we will briefly describe how this test works. We look at each transit within the light curve individually. Each the transits detected are then fit to four potential models:

$$\begin{aligned}
f(t) &= y_0 && \text{Flat} \\
f(t) &= y_0 + S(t) && \text{Logistic} \\
f(t) &= y_0 + S(t)(1 - e^{\beta t}) && \text{Logistic-exponential} \\
f(t) &= y_0 + S(t - \tau/2) - S(t + \tau/2) && \text{Double-logistic}
\end{aligned} \tag{4.14}$$

where y_0 is a constant offset and τ and β are tunable parameters (see Figure 1 of Mullally et al. (2016) for reference). We also do not fit an additional GP (as noted in the original Marshall test) because the light curves have already experienced several GP fitters. The function $S(t)$ is a logistic function given as:

$$S(t) = \frac{d}{1 + e^{\gamma(t-t_0)}} \tag{4.15}$$

where t_0, γ, and d are all tunable parameters. These functions are meant to replicate possible FP signals. To gauge the success of each fit, we calculate the Bayesian Information Criterion (BIC) for each function. This value is meant to mimic the Bayesian evidence of a Gaussian likelihood, and penalizes for increasing the number of tunable parameters. We also calculate the BIC for our best fit transit model. If the BIC value for any of the previously noted functions is 10 less than the transit BIC ($BIC_{f(t)} + 10 < BIC_{\text{Transit}}$) and the MES/$\sqrt{N_{\text{Tran}}} > 4$ (this criterion ensures the expected single transit signal strength is large enough to provide a meaningful model fit), we mask that transit signal. However, the MES/$\sqrt{N_{\text{Tran}}} > 4$ requirement is ignored for the constant y_0 model.

We also consider the number of unmasked points in each transit. If more than 50% of cadences within the transit duration are masked, the entire transit is masked. This ensures that a significant fraction of the transit data is available. Here we treat all parts of the transit signal equally, but in reality the egress and ingress are more important for candidacy. We consider this further in Section 4.5.10. The current test has the potential to mask several transits. To see how this affects the signal we reanalyze the light curve to see if the transit still has three observable transits and a MES ≥ 8.68. If the signal now fails to meet either of these requirements, the signal is flagged as a FP.

Looking at the Single Event Statistics (SES) of each transit, we can determine if a single transit is dominating the MES. In an ideal case the SES=MES/$\sqrt{N_{\rm Tran}}$. `EDI-Vetter` will look for SES values that exceed 80% the overall MES value. In the scenario where only 3 transits exist we would not expect any transit to exceed $\sim 60\%$, thus 80% provides an unlikely natural occurrence. As performed in Section 3.2.4 of Coughlin et al. (2016) and Section A.3.5 of Thompson et al. (2018), this test will help eliminate single transit outliers that pushed the MES over the TCE limit. It is not robust to chance alignments of multiple transit outliers as addressed in Section 4.5.3. If this limit (SES$>$ 0.8MES) is exceeded, the signal is deemed a FP by the pipeline.

One potential issue with this test is that a true planet transit which falls on a systematic feature could inflate the SES beyond this threshold, triggering a FP flag. In testing this was not an apparent issue, but something we will continue to monitor as we move forward with the other campaigns.

4.5.5 Even/Odd Transit Test

This classic test will look for cases where the transit primary and secondary eclipse are folded on top of each other. This will occur when the secondary eclipse depth is a significant fraction of the primary eclipse depth and the orbit is circular so that the secondary eclipse occurs at a phase of 0.5. We would only expect such similarity between transit depths for eclipsing binaries. Thus, when we find over-folded light curves of this nature, we flag them as a FP.

To perform this test, we separate the phase-foldings into even and odd transits (i.e. every other transit in one group and the remainder in the other group). The transits are then refit, holding P constant. We compare the inferred R (radius ratio) values from each group (odd vs. even) to look for statistical differences (Z).

$$Z = \frac{R_{even} - R_{odd}}{\sqrt{\sigma^2_{R_{even}} - \sigma^2_{R_{odd}}}} \tag{4.16}$$

where the σ values correspond to the labeled R fit uncertainties. We select $Z > 5$ for FP flagging.

One notable issue arises when the true period of the signal is twice the period detected, folding baseline data on top of the transit signal. This will be flagged as FP despite being a true planet signal. Cases where this occurs will be identified in Section 4.5.12 and re-run at the correct period folding.

One of the main complications of this test are the cases where one of the transits (usually near the end or beginning of the campaign) are poorly detrended. If this occurs in a true signal with very few transits ($N_{Tran} \leq 5$), one of the two groups can be significantly pulled by this faulty transit and cause this flag to be falsely triggered. Many of the known high-MES candidates not found by our pipeline fall victim to this flagging. This problem can be easily spotted by eye, but remains difficult to automate.

4.5.6 Uniqueness Test

This metric identifies cases where the phase folded light curve appears to produce several transit-like dips. This type of noisy light curve is unlikely to reveal a real planet signal, but rather highlight the systematic issues of the spacecraft. We base this metric on the "model-shift uniqueness test" (Rowe et al., 2015; Mullally et al., 2015; Coughlin et al., 2016; Thompson et al., 2018). Here, we provide a brief explanation of this test. For a thorough explanation, we refer interested readers to Coughlin (2017).

The first step in this process is to fold the light curve according to the period of the signal in question. The light curve should remain folded on this period for the entirety of this test. To determine how unique a signal is to the light curve, it is first important to mask the detected signal from the light curve and any possible secondary eclipse. This done by masking the largest signal (MES_1) and the second largest signal (MES_2) in the light curve, given that the second largest signal is within 50% of the transit duration (t_{dur}) of the expected secondary eclipse location. If the signal MES_2 is not within the expected secondary eclipse window, the third largest signal (with the same t_{dur}) is assigned to MES_2 before masking. This allows the second largest signal to remain in the light curve for detection as MES_3. We can then measure the red noise within the period folded data. This is achieved by fitting a transit model to each flux point with a fixed width of the original signal duration (t_{dur}). We simplify this process by using a box shape transit, where the best-fit depth can be analytically shown as the mean of the selected points. This simplification significantly increases the speed of our processing. If the light curve has no red noise, these averaged depth values will have a standard deviation (σ_{Red}) on the order of the light curve noise (σ_{LC}), otherwise it will be much larger. We then calculate the ratio of these two values (F_{Red}):

$$F_{Red} = \sigma_{Red}/\sigma_{LC} \tag{4.17}$$

Additionally, we want to determine the strength of the third largest signal (MES_3) within the light curve, and then the largest flux brightening signal (the largest signal found when the light curve is inverted; MES_4).

For statistical comparison, we consider the threshold in which a signal is statistically significant under the assumption of Gaussian noise:

$$F_1 = \sqrt{2}\ \mathrm{erfcinv}\left(\frac{t_{dur}}{P \times N_{TCE}}\right) \tag{4.18}$$

where P is the period in which the light curve has been folded, and N_{TCE} is the number of TCEs expected in the entire campaign. In an effort to ensure high reliability, we select $N_{TCE} = 25,000$, or roughly one TCE per light curve. Here P/t_{dur} represents the number of events in each light curve and N_{TCE} is the number of light curves considered. For the statistical significance within a single light curve, we consider the threshold as follows:

$$F_2 = \sqrt{2}\ \mathrm{erfcinv}\left(\frac{t_{dur}}{P}\right) \tag{4.19}$$

We consider a signal viable for further vetting if all the following inequalities are satisfied:

$$\begin{aligned} F_1 - (\mathrm{MES}_1/F_{red}) &\leq 1 \\ F_2 - (\mathrm{MES}_1 - \mathrm{MES}_3) &\leq 2 \\ F_2 - (\mathrm{MES}_1 - \mathrm{MES}_4) &\leq 3 \end{aligned} \tag{4.20}$$

If any of the previous equations are false, the signal is flagged as a FP (see Figure 19 of Coughlin (2017) for reference). The inequality threshold values are based on those provided in Thompson et al. (2018), but have been manually tuned to suit the needs of this pipeline.

We also consider the significance of the secondary eclipse detected within the folded light curve. If the phase of the secondary eclipse falls within $0.5t_{dur}$ of $t_0 + P/2$, and $\mathrm{MES}_1 - \mathrm{MES}_2 - F_2 > 0$, the secondary eclipse appears genuine and warrants further testing in Section 4.5.7. This signal will be flagged with a secondary eclipse flag. It is important to note that this test will only be sensitive to companions with near circular orbits, where the secondary eclipse occurs at or near a phase of 0.5. Future iterations of the pipeline will consider eccentric orbits. If the secondary eclipse does not fall at or near a phase of 0.5, the signal is deemed a FP due to lack of uniqueness if all of the following inequalities are satisfied:

$$
\begin{aligned}
\mathrm{MES}_2/F_{red} - F_1 &> 1 \\
\mathrm{MES}_2 - \mathrm{MES}_3 - F_2 &> 0 \\
\mathrm{MES}_2 - \mathrm{MES}_4 - F_2 &> 0
\end{aligned}
\tag{4.21}
$$

If any of the previous equations are false, the signal will be further vetted for candidacy. If the secondary eclipse is not astrophysical, the inequalities in Equation 4.21 make certain that the original MES_1 signal is still sufficiently unique.

4.5.7 Check for Secondary Eclipse

Apparent secondary eclipses are often the hallmark of an eclipsing binary. However, such events can also be seen for certain transiting hot Jupiter-like planets. In the previous section, we discussed the criteria for a secondary eclipse to be considered statistically significant. If the TCE was flagged as a secondary eclipse candidate, we consider several metrics before determining how to classify the signal.

Inspired by the frame work of Section A.3.2 in Thompson et al. (2018), we begin by fitting a transit model to the secondary eclipse in question. If the secondary eclipse model depth is greater than 10% of the depth of the initial transit, the secondary eclipse radius ratio is statistically significant ($R_{SE} > 2\sigma_{R_{SE}}$), and the initial transit impact parameter $b \geq 0.8$, the TCE is flagged as a FP. The large secondary eclipse and high impact parameter are indicators that the transit is likely a grazing eclipsing binary. If the signal fails to meet any of the listed criteria, the TCE remains labeled as a secondary eclipse candidate, and continues for further vetting. Such cases are likely transiting hot Jupiters reflecting a significant amount of light. It is possible that a well aligned, low impact parameter eclipsing binary could potentially slip through this metric, but the occurrence of such an event would likely produce a large enough transit depth to trigger the metrics discussed in Section 4.5.2.

4.5.8 Ephemeris Wandering

The emphemeris value (t_0) found by the grid TCE search (`TERRA`) will be very similar to the value determined by the transit model fitting if the signal is a true planet candidate. In cases where the two values differ significantly, the signal is likely very asymmetric and does not fit well with the transit model. To avoid these contaminants, we flag all signals as FP if the $|\Delta t_0|$ between the grid search and the model fitting is greater than $0.5t_{dur}$.

4.5.9 Harmonic Test

Sinusoidal stellar variability is a common trigger for TCE search algorithms. In Section 4.4.2 we attempted to detrend such periodicity, but our limits of $P < 0.5$ days for the harmonic fitter and $P > 5$ days for the GP fitter create a window in which a sinusoidal signal can sneak in. Furthermore, cases where the sinusoidal signal is not completely removed can cause TCE triggering. To eliminate such contamination, we fit several sine curves to test for such FPs.

When testing the possible harmonic functions, we fit a sine curve to the data while holding fixed the period and allowing the phase and amplitude (A_{sin}) to vary. We attempt to fit six different harmonic periods: P (the TCE period), $P/2$, $2P$, t_{dur} (the transit duration), $2t_{dur}$, and $4t_{dur}$. These potential periods represent the most notorious harmonics found mimicking a TCE.

When considering how well the sine curves fit the data, it is important to only consider the A_{sin} value. Any attempt to use the BIC value, as done in Section 4.5.4, can cause one to misclassify hot Jupiters, which can naturally produce strong sinusoidal oscillations within the folded light curve. If any of the tested periods produce $A_{sin}/\sigma_{A_{sin}} > 50$, the signal is labeled a FP. In cases where the harmonic signal is caused by a hot Jupiter, the primary and secondary eclipse will reduce the sine signal strength below 50σ. Furthermore, a large amplitude in it self may be an indicator that the oscillations are of stellar origin. Our second test checks to see if any of the A_{sin} values are larger than 90% of the fit transit depth and

$A_{sin}/\sigma_{A_{sin}} > 2$, ensuring the signal is real. If this criteria is met, the signal is classified as a FP. Large A_{sin} values correspond to a harmonic that is contributing to a large fraction of the signal depth and is likely the cause of the TCE.

As mentioned in Section 4.4.2, part of our pre-processing attempts to fit out harmonic trends in the light curve. To account for the possibility of artificially introducing a signal which could be detected as a TCE, we calculate the power (Pow_{TCE}) of a Lomb-Scargle periodogram at the transit period (P), and then search for the largest signal (Pow_{max}) with a period less than the transit period but greater than twice the cadence spacing. The goal here is to determine if the TCE is some integer multiplier of the initial harmonic fitter. If $\text{Pow}_{max} > (\text{Pow}_{TCE} + 0.5)$ and the Pow_{max} signal is found with a period less than 0.5 days, the TCE is deemed a FP.

4.5.10 Phase Coverage Test

With the significant processing required to detrend the *K2* light curves, on average $\sim 2\%$ of the data is masked out on any given light curve due to systematic issues causing flux outliers. Although the test performed in Section 4.5.4 will remove transits that have less than 50% phase coverage, it is still possible for phase folded data to contain large gaps in the transit signal. This occurs when all the masked portions of the light curve line up in phase space. This is especially common for planets with periods that are near integer multipliers of the *K2* cadence. It is difficult to properly vet such cases, as the transit appears incomplete and prone to FP contamination. Thus, we have developed a metric for flagging TCEs with such gaps.

Chance outlier alignments lack an important trait, they do not have a pronounced ingress and egress. To capitalize on this feature, we require more phase coverage during these periods than during the transit mid-point. The phase folded gap allowance ($\eta(t_{mid})$) is as follows:

$$\phi(t_{mid}) = \frac{(t_{mid} - t_0) - \text{round}(t_{mid} - t_0)}{t_{dur}} \tag{4.22}$$

$$\eta(t_{mid}) = \begin{cases} t_{cad} \times (16\phi(t_{mid})^4 - 8\phi(t_{mid})^2 + 2); & \text{if } -1 \leq \phi(t_{mid}) \leq 1, \\ \infty; & \text{else} \end{cases} \tag{4.23}$$

where t_{mid} is the midpoint between the nearest neighbor measurements in the phase folded time series, t_0 is the ephemeris time, t_{dur} is the fit transit duration, and t_{cad} is the light curve cadence (nominally 0.0204 days). $\eta(t)$ provides an allowance of t_{cad} during ingress and egress and a value of $2t_{cad}$ at the ephemeris. $\eta(t_{mid})$ quickly becomes very large for t_{mid} values beyond t_{dur} ($\eta(\phi = 1, -1) = 10t_{cad}$), but still requires some baseline coverage immediately before ingress and after egress. If $\Delta t_{mid} > \eta(t_{mid})$ (where Δt_{mid} is the temporal spacing between the two measurements that define t_{mid}) for any value of t_{mid}, the TCE is flagged as a FP. This metric will flag TCEs with large gaps in the phase folded transit signal. As mentioned in Section 4.4.4.1 the `EVEREST` software masks many of the spacecraft operation cadences, but the neighboring cadences can occasionally produce outliers that line up to form TCEs. Such cases will produce gaps in the phase folded transit signal and be flagged as a FP by this metric.

Additionally, to ensure large gaps do not exist in the transit phase coverage, similar to that done in Section A.3.7.1 of Thompson et al. (2018), we implement a minimum 70% in transit phase coverage requirement. Any TCEs that contain phase folded temporal gaps larger than $0.3t_{dur}$ will be flagged as a FP.

4.5.11 Period and Transit Duration Limits

As mentioned in Section 4.4.2, we used a harmonic fitter to remove stellar noise at periods less than 0.5 days. To avoid contamination caused by the introduction of this harmonic, we limit our sample to $P \geq 0.5$ days. This limit has been enforced on `TERRA` in Section 4.4.3, but such grid search algorithms can still provide detections beyond this limit by slightly mis-identifying the true transit period. Furthermore, TCEs found with periods very near

0.5 days can move below this limit when implementing the parameter search method of `EDI-Vetter`. We strictly enforce this limit by again checking that $P \geq 0.5$ days, otherwise the TCE is deemed a FP.

Another common FP is when a stellar oscillation has a short period (< 1 day) falling in the forbidden detrending window (noted in Section 4.4.2). In addition to our harmonic test (Section 4.4.2), a symptom of a sine wave TCE is a very long t_{dur} when compared to the signal period. Thus any TCEs with $t_{dur}/P > 0.1$ are flagged as FPs. This also ensures that the planetary parameters remain physical. Any transit with $t_{dur}/P > 0.1$ will be orbiting very near the surface of the star ($a_{pl} < 3R_\star$).

We acknowledge that ultra-short period planets (Sanchis-Ojeda et al., 2014; Adams et al., 2016) will certainly be rejected by the discussed cuts. However, all attempts made to capture these missed planets introduced a significant number of FPs into our candidate sample, negating our goal of high reliability. Again, readers concerned with USPs should perform their own fine-tuned searches.

4.5.12 Period Alias Check

Occasionally the `TERRA` grid search will find a transit that is some integer multiple of the true signal period. In such cases, it is very common that the signal is deemed a FP because of the significant "secondary eclipse", or the transit is poorly fit causing several other FP flags to be triggered. Alternatively, a signal with a period less than 0.5 days can be found with some larger multiple of the true period and contaminate the candidate population.

To avoid such cases, we measure the Gaussian likelihood ($log\mathrm{L} = -\frac{1}{2}\chi^2$) of the transit model with the fit transit period (P), and compare that with the likelihood of the same model with four other possible periods: $P/2$, $P/3$, $2P$, and $3P$. We also scale the semi-major axis to stellar radius ratio (a_{pl}) parameter accordingly to ensure the transit duration is consistent among all tested likelihoods. To avoid edge cases where noisy data could falsely score one of the likelihoods higher, we give the initially fit period a slight advantage. If any of the alternative periods produce a likelihood greater than 1.05, the likelihood of the initially fit transit period, the signal is flagged as a period alias.

Any signal flagged as a period alias will be refit and revetted with the corresponding highest likelihood period. This cycle will occur at most three times (in most cases it only occured once), allowing for higher order integer multipliers to be tested. Within each cycle, the FP flag will be reset and the vetting metrics will be reprocessed at the new period. We then take the vetting as it stands in the final cycle.

4.5.13 Consistency Score

One cannot help but worry about the edge cases that could potentially slip through the cracks of our vetting metrics. To combat such cases, we implement a consistency score limit. To calculate the consistency score of our vetted candidates, we run each of them through the `EDI-Vetter` fit and vetting pipeline 20 times. Given the statistical nature of the MCMC parameter estimation (and the limited number of burn-in and test steps), each run will process the light curve slightly different, potentially pushing the candidate into the FP bin. We only label a TCE as a candidate if it comes out of the vetting metric as a FP 25% of the time or less (Consistency Score ≥ 0.75). This metric is comparable to the "Disposition Score" made available for *Kepler* DR25 (Thompson et al., 2018).

4.6 The Planet Catalog

Using our fully automated pipeline, we achieve a sample of *K2* transiting planets suitable for demographics. Within this catalog we found:

- 773 Transit signals
- 723 Unique planet candidates

- 41 Multi-planet candidate systems
- 235 Newly detected planet candidates
- 7 Newly identified multi-planet candidate systems.

The majority of new candidates were found in campaigns not exhaustively searched (C10-18), with a few new candidates identified with low MES in early campaigns. Of the 773 transiting signals identified, 50 signals are detected in multiple, overlapping, campaigns. In the succeeding sections we will discuss how the parameters of the planets are derived and the unique systems and candidates found in this catalog.

4.6.1 Planet Parameterization

To maintain the homogeneity of our sample, we used an automated routine to fit and estimate the planet and orbital features. As discussed in Section 4.2 we assume the stellar parameters derived by Hardegree-Ullman et al. (2020), but found 130 of the 723 detected planets did not have stellar characterization available. Fortunately, our pipeline is agnostic to changes in stellar features, enabling subsequent parameter revision. To avoid the assumption of solar values and provide the most current stellar parameters, we used the Hardegree-Ullman et al. (2020) methodology, equipped with APASS (DR9; Henden et al. 2016) and SkyMapper (Onken et al., 2019) photometry, to characterize 59 of the remaining host stars. The final 71 targets were parameterized using the `iscochrones` stellar modeling package (Morton, 2015),[16] ensuring all planets in our catalog have corresponding stellar measurements. Using the `emcee` software package (Goodman & Weare, 2010; Foreman-Mackey et al., 2013), we measured the posterior distribution of the transiting planet parameters: The ephemeris, the radius ratio, the transit impact parameter, the period, the semi-major axis to stellar radius

[16] In order to maintain stellar parameter uniformity we ran the other 652 planet hosts through `isochrones`. We fit a linear offset on these targets between our parameters and the `isochrones` parameters, and applied these offsets to the 71 `isochrones`-only targets.

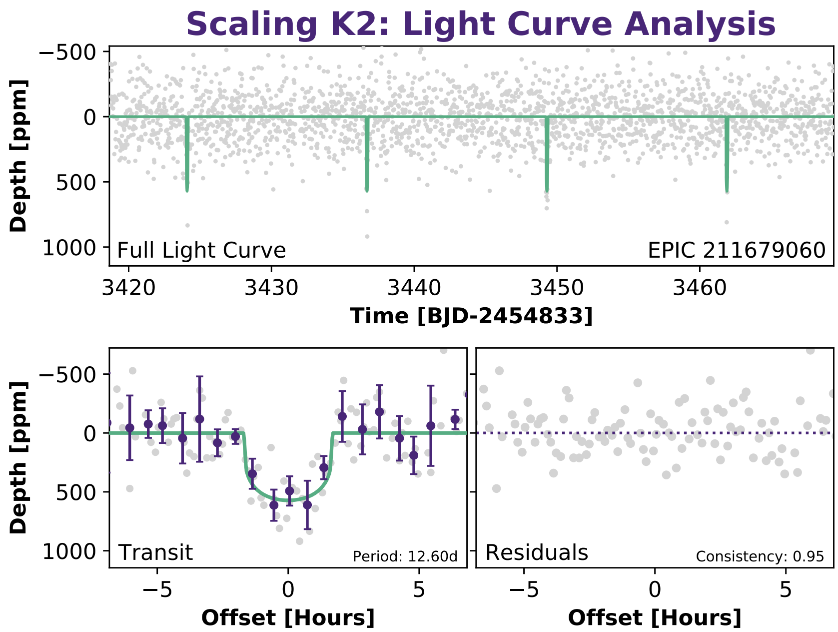

Figure 4.6: A sample diagnostic plot for EPIC 211679060.01; the remaining candidate plots are available online. This planet is a new sub-Neptune found in C18. The grey points represent the light curve data used to extract the signal. The purple points show the binned average (with a bin width equal to 1/6 the transit duration).

ratio (apl), the transit duration (tdur), and the vetting consistency score (score). These values were all derived under the assumption of circular orbits. Furthermore, we provide diagnostic plots for each candidate. In Figure 4.6 we show an example plot for a new sub-Neptune (EPIC 211679060.01), enabling swift visual inspection. However, it is important to clarify that all planets in our sample have been detected through our fully automated pipeline. These plots are only meant to help prioritize follow-up efforts.

With careful consideration of the transit radius ratio (R_{fit}), the planet radius can be extracted. R_{fit} is directly measured by the MCMC routine, but estimates of the true planet radii (R_{pl}) required we take into account the stellar radii measurements ($R_\star$) and potential contamination from nearby sources ($\frac{F_{total}}{F_\star}$). To attain our best estimate we use the following procedure:

$$R_{pl} = R_{fit}\ R_\star \sqrt{\frac{F_{total}}{F_\star}}, \tag{4.24}$$

where $\frac{F_{total}}{F_\star}$ is calculated using the Gaia DR2 stellar catalog to identify contaminants in and near the photometric aperture. In addition to the mentioned complications, Zink et al. (2020a) also showed that the required detrending of *K2* data underestimates the radius ratio by a median value of 2.3%. We do not adjust our estimates of the planet radius to reflect this tendency, as the mode of this distribution indicates a majority of the measured radii ratios are accurate (see Figure 14 of Zink et al. 2020a). However, we do increase the uncertainty in our radii measurements to account for this additional detrending complication. The overall planet radius uncertainty (σ_R) was calculated by adding all of these relevant factors in quadrature,

$$\sigma_R = \sqrt{\sigma_{fit}^2 + \sigma_\star^2 + \sigma_F^2 + \sigma_{\text{Off}}^2}. \tag{4.25}$$

Here, σ_{Off} and σ_F represent the uncertainty due to detrending and flux contamination respectively (see Section 8.2 of Zink et al. 2020a for a thorough explanation of these parameters). Despite these additional contributions, the radius ratio fit (σ_{fit}) and the stellar radius measurement ($\sigma_\star$) account for a majority of the uncertainty, roughly 4% and 6% of the planet radius respectively.

We provide our list of planetary parameters for this catalog with corresponding measurements of uncertainty in Table 4.1. A plot of the resulting planet period and radius distribution is provided in Figure 4.7. The deficit of planets near $2R_\oplus$ is in alignment with the radius gap identified with *Kepler* data (Fulton et al., 2017) and with previously discovered *K2* candidates (Hardegree-Ullman et al., 2020). This gap appears to be indicative of some planetary formation or evolution mechanism, the exact origin of which remains unclear. One theory suggests stellar photoevaporation removes the envelope of weakly bound atmospheres in this region of parameter space (Owen & Wu, 2017), separating the super-Earths from the sub-Neptunes. Alternatively, the hot cores of young planets may retain enough energy to expel the atmosphere for planets within this gap (core-powered mass-loss; Gupta & Schlichting

Table 4.1: Provides the homogeneous catalog of K2 planet candidates and their associated planet and stellar parameters. This table is available in its entirety in machine-readable form.

Column	Units	Explanation
1	—	EPIC Identifier
2	—	Campaign
3	—	Candidate ID
4	—	Found in Multiple Campaigns Flag
5	—	Consistency Score
6	d	Orbital Period
7	d	Lower Uncertainty in Period
8	d	Upper Uncertainty in Period
9	—	Planetary to Stellar Radii Ratio
10	—	Lower Uncertainty in Ratio
11	—	Upper Uncertainty in Ratio
12	$R_\oplus$	Planet radius (R_{pl})
13	$R_\oplus$	Lower Uncertainty in R_{pl}
14	$R_\oplus$	Upper Uncertainty in R_{pl}
15	d	Transit ephemeris (t_0)
16	d	Lower Uncertainty in t_0
17	d	Upper Uncertainty in t_0
18	—	Impact parameter (b)
19	—	Lower Uncertainty in b
20	—	Upper Uncertainty in b
...	...	...

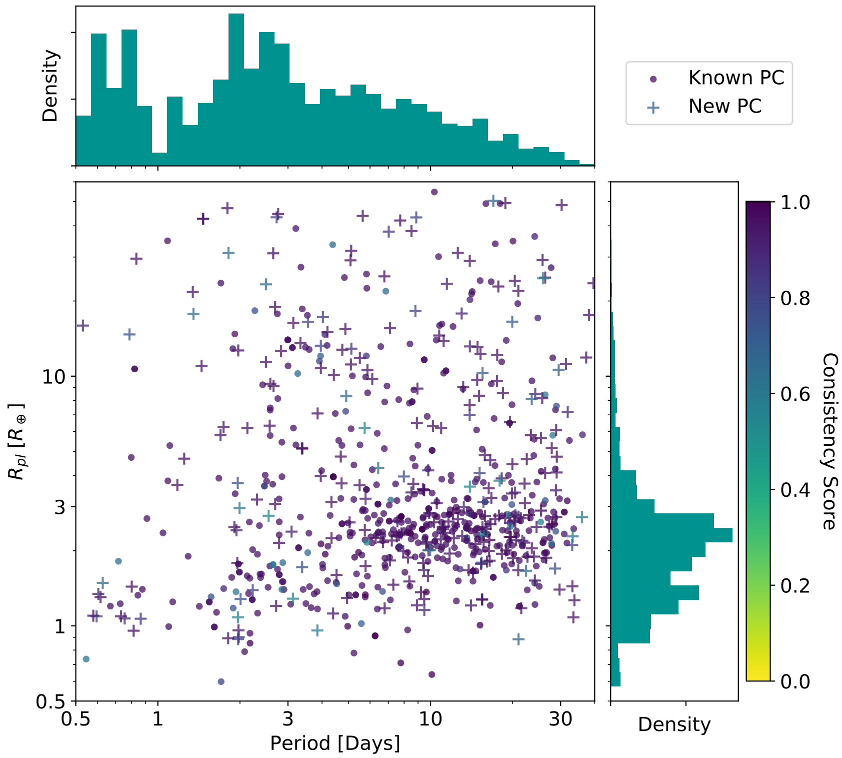

Figure 4.7: Displays the planet sample detected through our fully automated pipeline. The + markers show the new planet candidates (PCs) and round markers show the previously known PCs. The new candidates are uniformly distributed throughout the plot, because many of them come from campaigns not previously examined. The markers have been colored by consistency score to show that most candidates consistently passed the vetting metrics.

2019). However, these two mechanisms require additional demographic data to parse out the main source of this valley. The catalog derived here provides additional planets and the necessary sample homogeneity needed to examine this feature with greater detail, but such a task is beyond the scope of the current work.

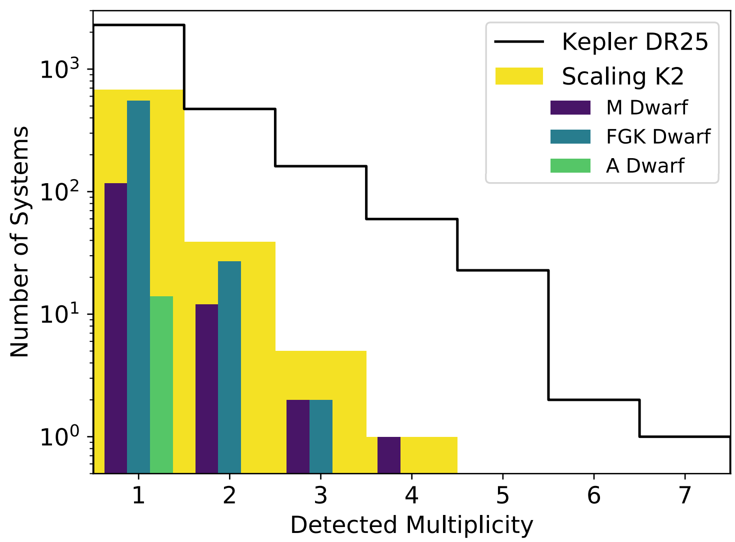

Figure 4.8: A histogram showing the observed system multiplicity distributions of our catalog (Scaling *K2*) and the *Kepler* DR25 (Thompson et al., 2018). The longer data span and reduced noise properties enabled *Kepler* to identify higher multiplicity systems. In addition, we provide the distributions as a function of stellar spectral type (M Dwarf: $T_{\rm eff} < 4000K$; FGK Dwarf: $4000 \leq T_{\rm eff} \leq 6500K$; A Dwarf: $T_{\rm eff} > 6500K$).

4.6.2 Multi-Planet Yield

Multi-planet systems provide a unique opportunity to understand the underlying system architecture and test intra-system formation mechanisms (e.g., Owen & Campos Estrada 2019). Furthermore, these planets are highly reliable, due to the unlikely probability of identifying two false positives in a light curve (Lissauer et al., 2014; Sinukoff et al., 2016). In our sample we detected 41 unique multi-planet systems, with 3 systems independently identified in more than one campaign (EPIC 211428897, 212012119, 212072539). In Figure 4.8 we show the total observed multiplicity distribution for our catalog. Within this sample we do not find any multi-planet systems around A stars, but we identify 13 M dwarfs systems and 28 FGK dwarf systems. In consideration of galactic latitude, we potentially find 12 systems that lay 40° above/below the galactic plane, providing unique coverage of galactic sub-structure.

4.6.2.1 Unique Systems

Our highest multiplicity system was EPIC 211428897, with four Earth-sized planets orbiting an M dwarf (system first identified by Dressing et al. 2017b). Our pipeline found this system independently in two of the overlapping fields (C5 and C18), further strengthening its validity. In addition, Kruse et al. (2019) identified a fifth candidate with a period of 3.28 days. Despite the clear abundance of planet candidates, the *K2* pixels span $3.98''$ and *Gaia* DR2, which `EDI-Vetter` uses to identify flux contamination, can only resolve binaries down to $1''$ for $\Delta\mathrm{mag} \lesssim 3$ (Furlan et al., 2017). Thus, high-resolution imaging is necessary for system validation. Dressing et al. (2017b) observed this system using Keck NIRC2 and Gemini DSSI, and found a companion star within $0.5''$. Since these planets are small, the likely $\sim \sqrt{2}$ radii increase, due to flux contamination, will not invalidate their candidacy (Ciardi et al., 2017; Furlan et al., 2017). However, it remains unclear whether all of the planets are orbiting one of the stars or some mixture of the two. This will require further follow-up, and remains the subject of future work.

Through our analysis we identified seven new multi-planet systems (EPIC 249217834, 249403651, 249559552, 249731291, 211732116, 211765695, and 211502222). EPIC 211502222 had been previously identified as hosting a single sub-Neptune with a period of 22.99 days by Yu et al. (2018). Our pipeline discovered an additional super-Earth at 9.40 days, promoting this G dwarf to a multi-planet host. The remaining six systems are entirely new planet candidate discoveries, existing in campaigns not exhaustively searched (C15 and C16). Notably, the K dwarf (EPIC 249559552) hosts two sub-Neptunes which appear in a 5:2 mean-motion resonance. These resonant systems are of great importance because of the potential detection of additional planets through transit-timing variations (e.g. Holczer et al. 2016). EPIC 249731291 is also interesting as it is an early F dwarf (or sub-giant) system, with two short period gas giants.

4.6.3 Low-Metallicity Planets

The core-accretion model indicates a link between stellar metallicity and planet formation (Pollack et al., 1996). Observational evidence of this connection was first identified with gas giants (Santos et al., 2004; Fischer & Valenti, 2005). However, the recent findings of Teske et al. (2019) complicates this narrative, by showing a lack of correlation between planetary residual metallicity and stellar metallicity. Thus, identification of low-metallicity planet hosts enables us to test our understanding of formation theories and to identify subtle features. In our sample of planets we found five candidates orbiting host stars with abnormally low photometrically-derived stellar metallicity ([Fe/H]< -0.75 dex.). Upon further examination, three of these candidates (EPIC 220489535, 228751724, and 201310077) have radii greater than $30R_{\oplus}$, indicative of an astrophysical false positive. The remaining two planets have radii less than $10R_{\oplus}$, making them some of the lowest metallicity planet-harboring hosts known. EPIC 212417884 is a low metallicity ([Fe/H]$= -0.960 \pm 0.235$ dex.) G dwarf, hosting a Neptune sized planet candidate. Additionally, we found EPIC 246865183, a low metallicity ([Fe/H]$= -0.761 \pm 0.235$ dex.) K dwarf with a Jupiter sized planet candidate. These two outliers appear in tension with the expectations of core-accretion in-situ formation, providing constraints for planet formation models. Further spectroscopic and radial velocity follow-up are needed to confirm the photometrically derived stellar metallicities and the planetary parameters.

4.7 Measuring Completeness

Any catalog of planets will inherit some selection effects due to the methodology of detection, limitations of the instrument, and stellar noise. These biases will effect the sample completeness and must be accounted for when conducting a demographic analysis. The selection of transiting planets can be addressed using analytic arguments, but the instrument and stellar noise contributions to the sample completeness are dependent on the stellar sample and the specifics of the instrument. With an automated detection pipeline this detection efficiency mapping can be achieved through the implementation of an injection/recovery test. Here, artificial signals are injected into the raw data and run through the automated software to test the pipelines recovery capabilities, directly measuring the impact of instrument and stellar noise on the catalog. Many previous studies have used this technique on *Kepler* data, yielding meaningful completeness measurements (Christiansen et al., 2013, 2015, 2020; Burke & Catanzarite, 2017).

4.7.1 Measuring CDPP

The first step in computing the detection efficiency is to establish a noise profile for each target light curve. With these values available, the signal strength can be estimated given the transit parameters. The ability to make these calculations enables one to understand the likelihood of detection for a given signal. Christiansen et al. (2012) described the combined differential photometric precision (CDPP) metric to quantify the expected the target stellar variability and systematic noise. For all targets we provide CDPP measurements for following transit durations: 1, 1.5, 2, 2.5, 3, 4, 5, 6, 7, 8, 9, and 10 hours, spanning the range of occultation timescales expected for *K2* planet candidates.

To compute these values we randomly injected a weak ($\sim 3\sigma$) transit signal, with the appropriate transit duration, into each detrended light curve. The strength of the recovered signal, normalized by the injected signal depth, provided a measure of the target CDPP. This process was carried out 450 times to ensure a thorough and robust examination of the light curve, sampling the full light curve for 4 hour transits and 25% of the light curve for 1 hour transits. To measure the impact of differing transit durations, this process was executed for each respective CDPP timescale. For a detailed account of this procedure see Section 4 of Zink et al. (2020a).

In Figure 4.9 we show the measured 8h CDPP for the targets within this sample. Overall, there is a clear correlation with photometric magnitude, demonstrating our ability to quantify the light curve noise properties. Moreover, despite the introduction of additional systematic noise from the telescope, the detrended *K2* data is still near the theoretical noise limit at the brighter magnitudes. The complete set of measurements are available in Table 4.2.

4.7.2 Injection/Recovery

There are several points along the pipeline at which the signal can be injected. Ideally, injections would be made on the rawest form of data (at the pixel-level), but doing so is computationally expensive and provides a marginal gain in completeness accuracy (see Christiansen 2017 for the effects on the *Kepler* data set). Moving just one step down stream, the injections can be more easily made at the light curve-level. Here, the artificial signal is introduced into the aperture-integrated flux measurements, followed by pre-processing, detrending and signal detection. Finally, the most accessible, but least accurate, method is by injecting signals in the pre-detrending flux (e.g., Cloutier & Menou 2020). Since the pre-processed `Everest` light curves are readily available, this method requires minimal computational overhead. However, it fails to capture the impact of pre-processing on the

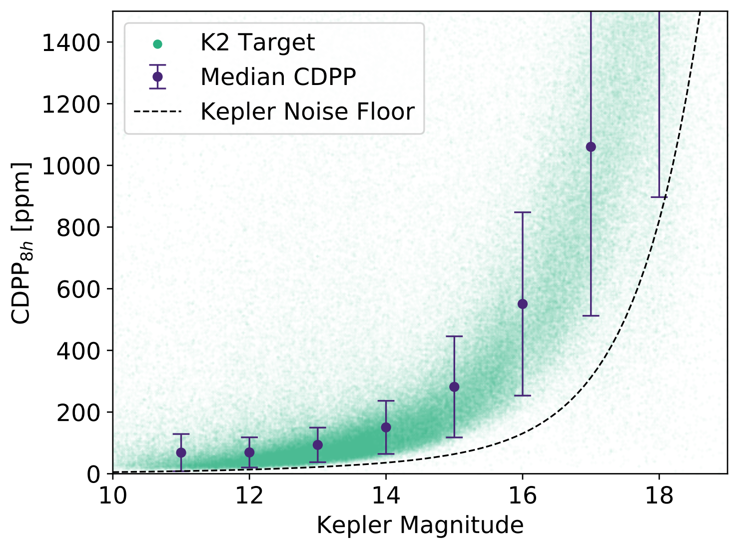

Figure 4.9: Characterizes the light curve noise (the 8-hour CDPP measurements) for our target sample as a function of the *Kepler* broadband magnitude measurement. The median CDPP markers represent the median within a one magnitude bin and the corresponding bin standard deviation. The *Kepler* noise floor represents the shot and read noise expected from the detector alone (Jenkins et al., 2010b).

Table 4.2: Description of the CDPP measurements of each stellar target. This table is available in its entirety in machine-readable form.

Column	Units	Explanation
1	...	EPIC identifier
2	...	Campaign
3	ppm	CDPP RMS Value for Transit of 1.0 hr
4	ppm	CDPP RMS Value for Transit of 1.5 hr
5	ppm	CDPP RMS Value for Transit of 2.0 hr
6	ppm	CDPP RMS Value for Transit of 2.5 hr
7	ppm	CDPP RMS Value for Transit of 3.0 hr
8	ppm	CDPP RMS Value for Transit of 4.0 hr
9	ppm	CDPP RMS Value for Transit of 5.0 hr
10	ppm	CDPP RMS Value for Transit of 6.0 hr
11	ppm	CDPP RMS Value for Transit of 7.0 hr
12	ppm	CDPP RMS Value for Transit of 8.0 hr
13	ppm	CDPP RMS Value for Transit of 9.0 hr
14	ppm	CDPP RMS Value for Transit of 10.0 hr

sample completeness. These effects are especially important for *K2* data, which undergoes significant modification before being searched. Following the procedures of our previous study (Zink et al., 2020a), we injected our artificial signals at the light-curve-level (see Figure 4.2). In the next few paragraphs we shall briefly outline our methodology, but suggest interested readers reference Section 5 of our previous work for a more detailed account.

Using the `batman` Python package (Kreidberg, 2015), we created and injected artificial transits in the raw flux data. For each target we uniformly drew a period from [0.5, 40] days and a $R_{pl}/R_\star$ from a log-uniform distribution with a range [0.01,0.1]. The ephemeris was uniformly selected, with the requirement that at least three transits reside within the span of the light curve, and the impact parameter was uniformly drawn from [0,1].[17] All injections were assumed to have zero eccentricity. This assumption is motivated by the short period range of detectable *K2* planets, which many have likely undergone tidal circularization. In addition, eccentricity only effects the transit duration, making it's impact on completeness minor. The limb-darkening parameters for the artificial transits were dictated by the stellar parameters discussed in Section 4.2. Using the ATLAS model coefficients for the *Kepler* bandpasses (Claret & Bloemen, 2011), we derived the corresponding quadratic limb-darkening parameters based on their stellar attributes. In cases where stellar parameters did not exist, we assumed solar values.

We expect the pipeline's recovery capabilities scale as a function of the signal strength. In order to quantify this effect, one must have a measure of the expected injection signal strength (MES). Equipped with our CDPP measurements, this value is directly related to the transit depth (depth) and can be analytically found using:

$$\mathrm{MES} = C\,\frac{\mathrm{depth}}{\mathrm{CDPP}_{\mathrm{t_{dur}}}}\,\sqrt{N_{\mathrm{tr}}}, \tag{4.26}$$

[17] In the current iteration of this pipeline we have removed the eclipsing binary impact parameter limit mention in Section 5 of Zink et al. (2020a). Doing so, we provide a more accurate accounting of the impact of grazing transits.

where $\mathrm{CDPP}_{\mathrm{t_{dur}}}$ represents the targets CDPP measure for a given transit duration (achieved through interpolation of the measured CDPP values discussed in Section 4.7.1) and N_{tr} is the number of available transits within the data span. The C value is a global correction factor that re-normalizes the analytic equation to match the detected signal values. For this data set we found C was equal to 0.9488. Using this equation, we calculated the expected MES values for all injections and measured the sample completeness.

Once our injections were performed, we passed this altered data through the software pipeline, testing our recovery capabilities. We considered a planet successfully recovered if it met the following criteria: the detected signal period and ephemeris were within 3σ of the injected values and the signal passes all of the vetting metrics. The results of this test can be seen in the completeness map of Figure 4.10. To quantify our detection efficiency as a function of injected MES, we used a uniform kernel density estimator (KDE; width of 0.25 MES) to measure our software's recovery fraction. These values were then fit with a logistic function of the form:

$$f(x) = \frac{a}{1 + e^{-k(x-l)}}, \tag{4.27}$$

where a, k, and l are all tunable parameters. The resulting optimized models are available in Table 4.3.

While MES is the strongest component of completeness, additional signal parameters can play a role. Christiansen et al. (2020) tested the effects of stellar effective temperature (T_{eff}), period, N_{tr}, and photometric magnitude on the completeness of the analogous *Kepler* injections, finding the strongest effect linked to N_{tr}. In Figure 4.10 we show two of these completeness features for the *K2* injections: spectral class and N_{tr}. Christiansen et al. (2015) noted a significant drop ($\sim 4\%$) in detection efficiency for cooler M dwarfs, which exhibit higher stellar variability. Remarkably, we did not find a significant difference between the AFGK dwarfs ($T_{\mathrm{eff}} > 4000K$) and the M dwarfs ($T_{\mathrm{eff}} < 4000K$). It's likely that the increased systematic noise of *K2* blurs this completeness feature, making the two populations indistinguishable. Like the *Kepler* TPS, we found $N_{\rm tr}$ has the strongest effect on our *K2* planet sample completeness. Here, a significant drop in detection efficiency was expected for signals near the minimum transit threshold. In light of the pipeline's three transit requirement, these marginal signals were more susceptible to systematics and vetting mis-classification. In order words, if any of the three transits were discarded by the vetting, it would immediately received a false positive label. We also parsed the data into larger $N_{\rm tr}$ value bins, but found little difference in completeness between $N_{\rm tr}$ equal to four, five, and six; the vetting was unlikely to discard more than one meaningful transit. Furthermore, we considered the effects of the signal period. While this parameter is strongly correlated with $N_{\rm tr}$, it has the potential to describe period dependent detrending and data processing issues. Separating the injected data at a period of 26 days (see Table 4.3), we find a loss of completeness ($\Delta a \sim 0.20$) that is comparable to the $N_{\rm tr}$ partition ($\Delta a \sim 0.21$). This similarity indicates a strong correlation between parameters, suggesting either of the two features would appropriately account for the reduced completeness in this region of parameter space. Since $N_{\rm tr}$ provides a slightly larger deficit, we suggest using this function for future demographic analysis of our catalog.

Completeness measurements provide a natural test of `EDI-Vetter`'s classification capabilities. The introduction of systematics from the telescope makes signal vetting more difficult, when compared to the *Kepler* TPS, requiring more drastic vetting metrics. This could lead to significant mis-classification, discarding an abundance of meaningful planet candidates. Fortunately, in Figure 4.10 we identify a $\sim 13\%$ loss of completeness due to the vetting metrics, which is comparable to the *Kepler* `Robovetter` ($\sim 10\%$ loss; Coughlin 2017). Ideally, this difference would be zero, but such a minimal loss is acceptable and, more importantly, quantifiable.

We have provided the completeness parameters necessary for a demographic analysis of this catalog. However, the injections carried out here are computationally expensive and hold significant value for research beyond the scope of this catalog. Therefore, we provide, in

Table 4.3: Provides the logistic parameters for the corresponding completeness functions shown in Figure 4.10. Additionally, we include the completeness parameters for a separation of 26 day period signals. It's important to highlight that a represents the maximum completeness for high MES signals.

Model	a	k	l
Unvetted	0.7407	0.6859	9.74077
Vetted	0.6093	0.6369	10.8531
>**3 Transits**	0.6868	0.6347	10.9473
=**3 Transits**	0.4788	0.6497	10.5598
<**26d Periods**	0.6619	0.6231	10.9072
>**26d Periods**	0.4635	0.6607	10.5441
AFGK Dwarf	0.6095	0.6088	10.8986
M Dwarf	0.6039	0.8455	10.5636

addition, the injected `Everest`-processed light curves and a summary table of the injection/recovery test.[18] These data products enable users to set custom completeness limits and to test their own vetting software with significantly reduced overhead.

4.8 Measuring Reliability

Despite efforts to remove problematic cadences, instrument systematics pollute the light curve data, creating artificial dips that can be erroneously characterized as a transit signal. To measure the reliability in a homogeneous catalog of planets, the rate of these false alarms

[18] http://www.jonzink.com/scalingk2.html

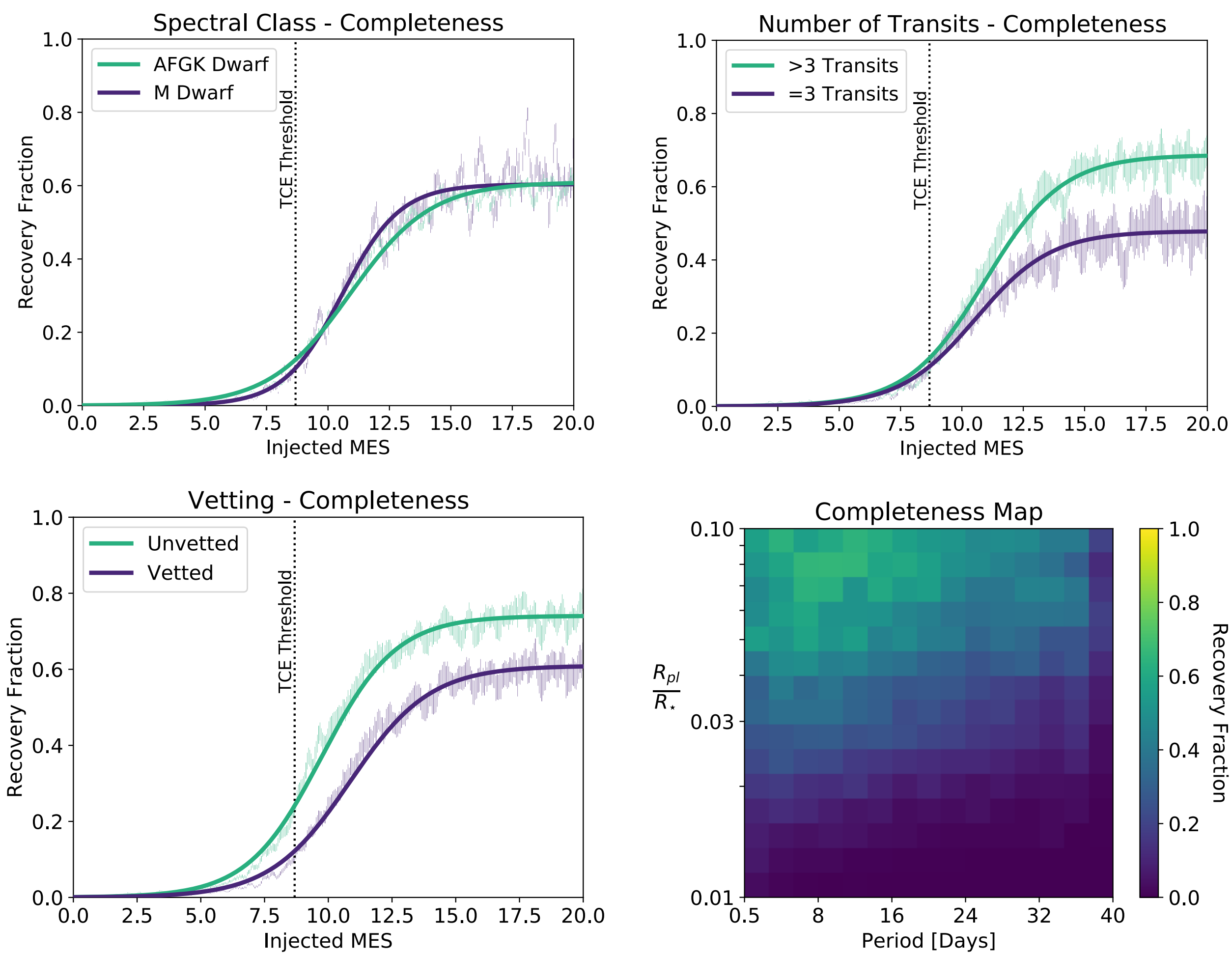

Figure 4.10: Shows the measured completeness, resulting from our injection/recovery test, for the presented planet sample. To show the effects of characteristic stellar noise, the number of transits, and the loss of planets due to our vetting software, we provide several slices of the data. The corresponding logistic function parameters are available in Table 4.3. The heat map shows the overall vetted completeness as a function of planet radius ratio and the transit period.

(FAs) must be quantified. Bryson et al. (2020b) showed that proper accounting of the sample FA rate is essential in extracting meaningful and consistent planet occurrence measurements.

The main goal of `EDI-Vetter` (and its predecessor `Robovetter`) is to parse through all TCEs and remove FAs without eliminating true planet candidates. However, this process is difficult to automate and requires a method of testing the software's capability to achieve this goal. Accomplishing such a task necessitates an equivalent data set that captures all the unique noise properties, that contribute to FAs, without the existence of any true astrophysical signals. With such data available, the light curves can be processed through the detection pipeline. If the vetting algorithm worked perfectly, nothing would be identified as a planet candidate. Therefore, any signals capable of achieving planet candidacy would be an authentic FA and provide insight into the software's capabilities.

Coughlin (2017) explored two methods for simulating this necessary data using the existing light curves: data scrambling and light curve inversion. The first method takes large portions of the real light curve data and randomizes the order, retaining all of the noise properties while scrambling out true periodic planet signatures. Using this method (Thompson et al., 2018) was able to replicate the real long period TCE distribution of the *Kepler* DR25 catalog. The second method inverts the light curve flux measurements. Upon this manipulation, the existing real transit signals will now be seen as flux brightening events, rendering them unidentifiable by the transit search algorithm. Under the assumption that many systematic issues are symmetric upon a flux inversion, the data will retain its noise properties without containing any real transit signals. This method was used by Thompson et al. (2018) to replicate the real short period TCE distribution of the *Kepler* DR25 catalog.

Since the *K2* data has unique instrument systematics that are periodic in nature, we chose the light curve inversion method. In Figure 4.11 we assess our ability to capture the noise features using this method. By comparing the distribution of TCEs from the real light curves with those of the inverted light curves, we can identify regions of parameter space

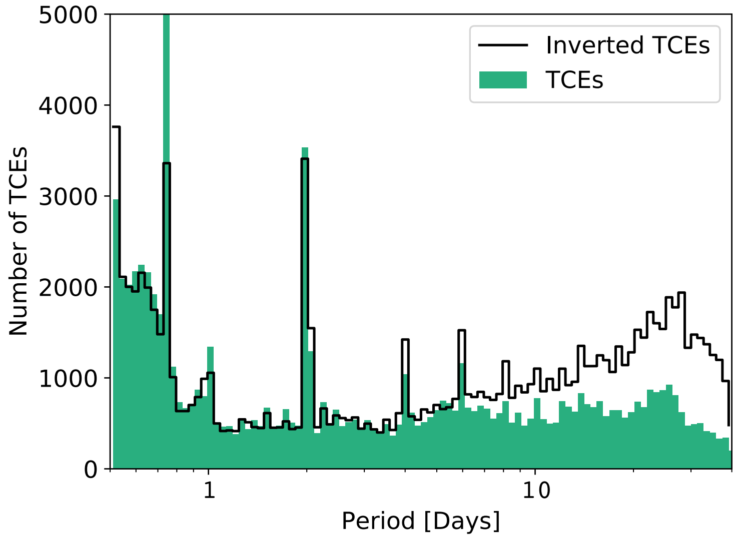

Figure 4.11: Represents the distribution of TCEs from the real light curves and the distribution of inverted TCEs from the FA simulation. This shows consistency among the two distributions with a minor surplus of inverted TCEs at longer periods and a deficiency at the 3rd harmonic of the thruster firing (18 hours).

where the inversions over- or under-represent systematic noise properties. Overall, the two distributions (108,379 TCEs and 110,548 inverted TCEs) are well aligned, providing an adequate simulation of the data set's noise characteristics. However, the inverted TCEs are slightly over-represented at long periods and under-represented at the 3rd harmonic of the 6 hour thruster firing. While we acknowledge these minor discrepancies, their impact will be small on the overall measure of the catalog reliability.

After running the inverted light curves through our pipeline we identify 67 FA signals as planet candidates (773 candidates were identified in the real light curves). We used this information, alongside the formalism described in Section 4.1 of Thompson et al. (2018), to quantify our sample reliability. The first step in achieving this measurement requires an understanding of the vetting routines FA removal efficiency (E). From the inverted light

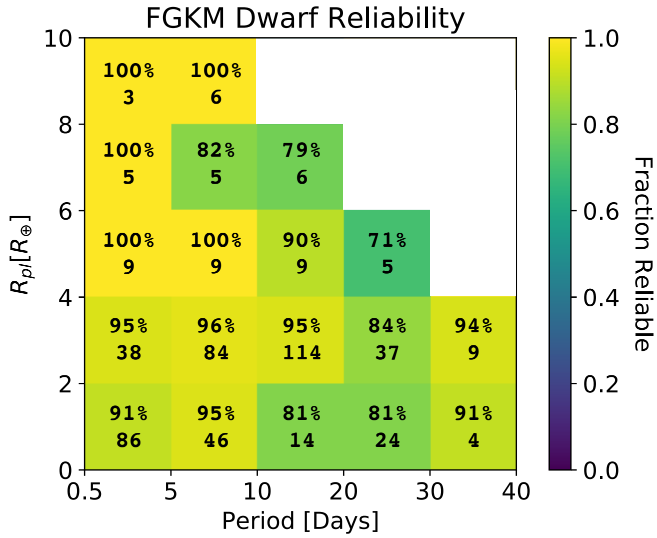

Figure 4.12: Shows the calculated reliability for our planet sample as a function of orbital period and planet radius. The reliability percent and the number of candidates have been listed in each corresponding box. The white regions represent areas of parameters space where the number of candidates and FAs are sparse, making accurate measurements of reliability unachievable.

curve test we can estimate E as

$$E \approx \frac{N_{FP_{\rm inv}}}{T_{TCE_{\rm inv}}}, \tag{4.28}$$

where $T_{TCE_{\rm inv}}$ is the total number of TCEs found in the inversion test and $N_{FP_{\rm inv}}$ is the number of TCEs that were accurately flagged as false positives. For the total data set we found `EDI-Vetter` has a 99.9% efficiency in removing FA signals. However, this extreme competence must be balanced by the abundance of TCEs found by our pipeline.

Using the number of planet candidates (N_{PC}) and the number of TCEs flagged as false positive (N_{FP}) in the real light curves, we can determine the reliability fraction (R) of our data set:

$$R = 1 - \frac{N_{FP}}{N_{PC}}\left(\frac{1-E}{E}\right). \tag{4.29}$$

Overall, we found the planet catalog provided here is 91% reliable. This is slightly lower than the 97% reliability of the *Kepler* DR25 catalog (Thompson et al., 2018). However, this *Kepler* value is greatly improved by the extensive data baseline. A more appropriate comparison would consider a period range with a comparable number of transits (Kepler candidates with periods greater than 10 days have a similar number of transits). In this region of parameter space the `Robovetter` has a reliability of 95%, which is closer to the value reported for `EDI-Vetter`. Moreover, these broad summary statistics fail to capture the complexity of this metric.

We find 22 of the FAs detected in our inversion simulation are hosted by sub-giant and giant stars, which have notably more active stellar surfaces. By excluding these targets, we focus on the dwarf star population (in Figure 4.12) and consider how reliability changes as a function of period and planet radius. At longer periods the reliability drops. The reduced number of transits available for a given candidate more easily enables systematic features to line up and create a FA signal. Smaller planet radii regions are also more susceptible to FAs due to their weak signal strength, which can be replicated by noise within the data set. When accounting for this period and radius dependence we expect 96% of our dwarf host candidates to be real astrophysical signals.

4.8.1 Astrophysical False Positives

It is important to highlight that reliability is a measure of the systematic FA contamination rate. There exist non-planetary astrophysical sources capable of producing a transit signal. For example, eclipsing binaries (EBs) have the ability to dilute their transit depth, by contributing additional unaccounted for stellar flux, leading to signal mis-classification (Furlan et al., 2017; Matson et al., 2018). Furthermore, Ciardi et al. (2017) showed that non-transiting stellar multiplicity can also artificially reduce transiting planet depths, leading to an overestimation in the occurrence of Earth-sized planets by 15-20%. In addition to modifying planet radii measurements, this can also lead to the mis-classification of EBs. `EDI-Vetter` has several metrics that attempted to address the possibility of astrophysical false positive contamination. The most robust EB diagnostic used a threshold on the transit impact parameter and radius ratio (see Section 4.5.2). Using this cut, it is unlikely that a large fraction of these EBs exist in our catalog, but we do not make any attempts here to quantify this potential source of contamination.

4.9 Occurrence Rate Recommendations

All exoplanet demographic analyses require defining an underlying stellar sample of interest. Our catalog used nearly the entire K2 target sample, which may be too broad for future analysis. The completeness measurements given here may not accurately reflect that of a reduced target list. Therefore, we recommend users select their own stellar sample and employ the injection/recovery summary table to assess the completeness of their selected targets for the highest degree of accuracy.[19]. However, the completeness parameters provided in Table 4.3 should be fairly robust to sample selection modifications. Evidence of this claim is provided by the fact that the M dwarf and AFGK dwarf samples provide consistent results, see Figure 4.10, despite known differences.

In addition, 41,061 targets in our sample lacked stellar parameters, requiring the assumption of solar values. This may modify the results once these parameters become available. The remaining stellar parameters were provided by Huber et al. (2016) for 94,769 targets and Hardegree-Ullman et al. (2020) for 222,088 targets. While Zink et al. (2020b) showed the offset between these two methods is minimal, making this mixture of catalogs reasonable, a uniform stellar parameterization would provide more homogeneous results. Fortunately, the pipeline's stellar agnosticism makes parameter updating uncomplicated. As additional data from the forthcoming *Gaia* DR3 becomes available, it will likely modify many of these stellar values and alter the underlying sample of dwarf stars. We suggest users implement the most up-to-date stellar parameters along with the injection/recovery summary table to

[19] www.jonzink.com/scalingk2.html

evaluate sample completeness.

Ideally, the reliability would also be updated as additional information on stellar and planetary parameters are made available. This is possible using the reliability summary table, but given the small number FAs found we expect very minor changes to occur.

It is important to note that our completeness measurements do not address the window function, which requires three transits occur within the available data. All injections where required to have at least three transits occurring within this window, removing this detection probability from the calculated completeness. In testing we found most light curves follow the expected analytic probability formula (*prob*):

$$
\begin{aligned}
prob &= 1; & P &< t_{\rm span}/3 \\
prob &= \frac{t_{\rm span}}{P} - 2; & \frac{t_{\rm span}}{3} \leq P &\leq \frac{t_{\rm span}}{2} \\
prob &= 0; & P &> t_{\rm span}/2,
\end{aligned}
\tag{4.30}
$$

where P is the signal period and $t_{\rm span}$ is the total span of the data (see Figure 11 of Zink et al. 2020a). However, intra-campaign data gaps do exist and should be carefully considered in any occurrence rate calculations.

4.10 Summary

We provide a catalog of transiting exoplanet candidates using *K2* data from Campaign 1-8 and 10-18, derived using a fully automated detection pipeline. This demographic sample includes 723 unique planets, with 235 previously unidentified candidates. Within this homogeneous sample we detect two planets, EPIC 212417884.01 and 246865183.01, which reside around likely low metallicity stellar hosts ([Fe/H]< -0.75). Once verified, these planets may help constrain metallicity-driven planet formation mechanisms (e.g., Pollack et al. 1996). Additionally, we find 41 multi-planet candidate systems, with 7 newly identified. These discovered systems include a K dwarf (EPIC 249559552) hosting two sub-Neptune candidates in a 5:2 mean-motion resonance and an early F dwarf (EPIC 249731291) with two short period gas giant candidates, providing an interesting constraint on formation and migration mechanisms.

We employed an automated detection routine to achieve this catalog, enabling measurements of sample completeness and reliability. By injecting artificial transit signals before `EVEREST` pre-processing, we provide the most accurate measurements of *K2* sample completeness. Additionally, we used the inverted light curves to measure our vetting software's ability to remove systematic false alarms from our catalog of planets, providing a quantitative assessment of sample reliability. Using this planet sample and the corresponding completeness and reliability measurements, exoplanet occurrence rate calculations can now be realized using *K2* planet candidates. With careful consideration of each data set's unique window functions, the *Kepler* and *K2* planet samples can now be combined, maximizing our ability to measure transiting planet occurrence rates throughout the local galaxy.

CHAPTER 5

Comparable Planet Occurrence in the FGK Samples of Campaign 5 and Kepler

5.1 Introduction

The *Kepler* mission continuously collected photometric data of over 150,000 stars for ∼3.5 years (Koch et al., 2010; Borucki, 2016). This data set has provided evidence for nearly 4,500 transiting exoplanet candidates.[1] Most of the candidates were detected by the fully automated *Kepler* pipeline (Jenkins et al., 2010a) and the `Robovetter` (Thompson et al., 2018), an automated vetting software built to distinguish real signals from false positives. Removing the human component of planet detection, the rate of false negatives (completeness) could be calculated with artificial transit injections (Petigura et al., 2013b; Christiansen et al., 2015; Christiansen, 2017). The DR25 team also estimated the rate of false positives (reliability) by inverting the light curves, eliminating any real transit signals, and testing the ability of the pipeline to remove false detections (Coughlin, 2017). With this set of planet candidates detected autonomously and an associated measure of the sample completeness and reliability, this data set has become the gold standard for deriving planet occurrence rates.

Numerous studies have used completeness and reliability to calculate planet occurrence rates

[1] https://exoplanetarchive.ipac.caltech.edu

for FGK dwarfs (i.e. Youdin 2011; Howard et al. 2012; Petigura et al. 2013b; Mulders et al. 2018; Zink et al. 2019b; He et al. 2019) and M dwarfs (i.e. Dressing & Charbonneau 2013, 2015; Hardegree-Ullman et al. 2019). However, one critique of these studies is that they only sample a single 116 square degree patch of the sky –these occurrence rates may be specific to this region of the galaxy. Planet occurrence may change at Galactic latitudes with differing stellar metallicities, masses, radii, multiplicities, and stellar age, potentially highlighting the effects of stellar environment on planet formation. Additionally, these differences will modify the inferred Galactic exoplanet population. This point becomes more relevant when considering the calculated values of $\eta_\oplus$ (Catanzarite & Shao, 2011; Traub, 2012; Petigura et al., 2013b; Silburt et al., 2015; Zink & Hansen, 2019), the probability of a potentially habitable planet being found around a given Sun-like star. Since all current estimates have used the *Kepler* sample to extrapolate this probability, it remains unclear how well the value represents the census of habitable planets throughout the Galaxy.

Fortunately, the *K2* mission (which came into existence after the *Kepler* telescope lost functionality of two reaction wheels to stabilize the spacecraft) collected data from 18 different regions (Campaigns) across the Ecliptic plane (Howell et al., 2014; Vanderburg et al., 2016). In Figure 4.1 we show how the *K2* sample probes different parts of the Galactic disk structure. A majority of the *Kepler* sample is contained in the thin disk, which is α-poor and consists of comparatively younger stars. The *K2* sample probes deeper into the thick disk (older stars that are α-rich; Sharma et al. 2019), providing potential insight into the effects of age and α element abundance on planet occurrence. Additionally, the *K2* sample spans 7-10 kpc in Galactic radius, while the *Kepler* sample is limited to a range of 7.5–8.5 kpc in Galactic radius. This expansion allows us to investigate planet occurrence around stars with varying Galactic radii.

K2 photometry is affected by considerably more systematic pointing issues than the *Kepler* prime mission. Therefore, the automated pipeline built for *Kepler* data was not able to be successfully applied to the *K2* data set. Various studies, which involved manual vetting, have found nearly 800 planet candidates in the *K2* data (Vanderburg et al., 2016; Barros et al., 2016; Adams et al., 2016; Crossfield et al., 2016; Pope et al., 2016; Dressing et al., 2017b; Petigura et al., 2018; Livingston et al., 2018; Mayo et al., 2018; Yu et al., 2018; Kruse et al., 2019; Zink et al., 2019a), but none have provided estimates of completeness and reliability due to the subjective nature of the searches.

With the introduction of the `EDI-Vetter` vetting software, Zink et al. (2020a) were able to fully automate a planet detection pipeline for *K2*. By removing the manual element from planet detection, Zink et al. (2020a) were able to provide a uniform set of planet candidates with corresponding measures of completeness and reliability for the Campaign 5 field (henceforth, C5). In Figure 5.1 we show the separation between these two fields, illustrating the independence of these two samples. As uniform processing continues for the remaining 17 Campaigns, this early sample provides the first opportunity to consider small transiting planet occurrence outside of the *Kepler* field.

In this paper we model the underlying exoplanet population for the *K2* C5 FGK stellar sample using the forward modeling software `ExoMult` (Zink et al., 2019b). We provide discussion of our stellar and planet parameters and the cuts we make to isolate the samples of interest in Section 5.2. The forward modeling methodology is presented in Section 5.3 along with modifications to `ExoMult`. In Section 5.4 we provide the optimized population models and discuss the occurrence parameters derived for each model. In Section 5.5 we consider the implication of the radius gap and other empirical features on our ability to model the population. We comment on the implications of our findings with regard to the effect of stellar metallicity in Section 5.6, and provide concluding remarks of this study in Section 5.7.

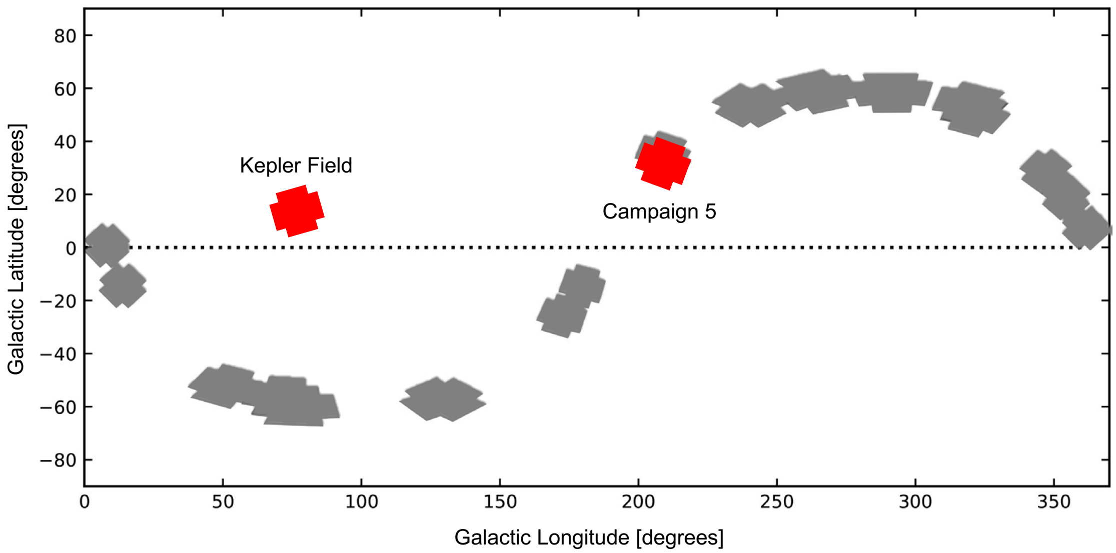

Figure 5.1: The location of the 18 *K2* fields are plotted in grey as a function of Galactic longitude and latitude. The *Kepler* and Campaign 5 field, which are the subjects of this study, have been highlighted in red. These two fields are separated by about 120 angular degrees.

5.2 Sample Selection

In this section we discuss the stellar and planetary parameters used in this study. We also present the cuts made in both samples to ensure purity.

5.2.1 *K2* Stellar Selection

A key part of any planet occurrence rate is understanding the underlying sample of stars from which the detected planets are drawn. This can be accomplished by ensuring the stellar attribute measurements are as accurate as possible. If a significant portion of the sample parameters are inaccurate, we will under- or over-estimate the difficulty of detecting planets, and therefore bias the occurrence measurements. To minimize this effect we use the stellar parameters provided by Hardegree-Ullman et al. (2020), as this provides the most uniform and up-to-date stellar parameters for *K2*. Using *Gaia* DR2 and *LAMOST* spectroscopic data, this sample significantly reduces the uncertainty in stellar radius measurements compared to measurements derived primarily by photometry (Huber et al., 2016). However, due to the requirement of a *Gaia* DR2 parallax, this catalog of random-forest derived parameters is only available for 19,220 stars out of the 25,030 potential targets studied by the *K2* pipeline (Zink et al., 2020a). In Figure 5.2 we show the overall change in stellar radius and effective temperature between the Huber et al. (2016) and Hardegree-Ullman et al. (2020) catalogs. We find a median decrease of 37K in stellar temperature and a median increase of $0.11R_{\odot}$ in stellar radius. While the overall offset between catalogs is minor, we acknowledge that some structure can been seen between parameters. Most noticeably, the temperature spread of M dwarfs. Accurate measurements of this parameter are difficult to achieve for M dwarfs, leading to this spread. Fortunately, these stars are not included in our sample (as this work focuses on FGK stars). Overall, the apparent systematic structure between catalogs is minor and within the uncertainty of the parameter measurements. Since this effect is small, and most targets experience a minor change in parameter values between catalogs, we rely on the Huber et al. (2016) stellar parameters with photometrically derived metallicities for the

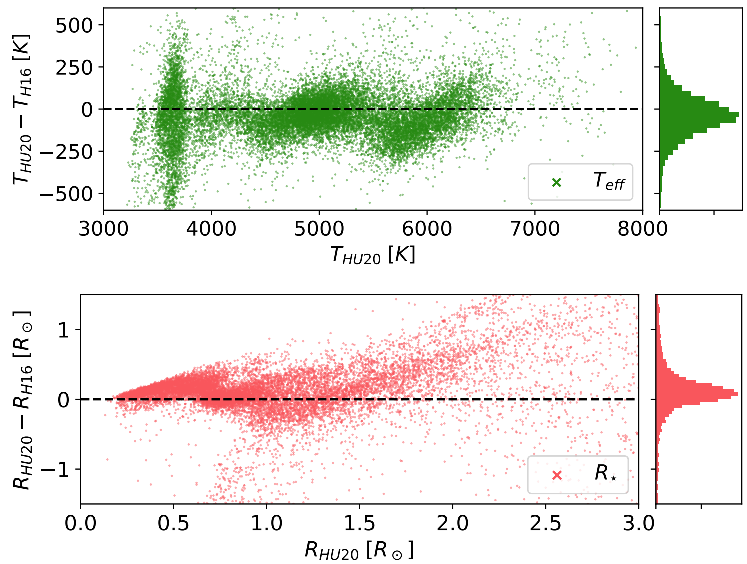

Figure 5.2: Shows the overall change from H16 (Huber et al., 2016) to HU20 (Hardegree-Ullman et al., 2020). The 19,004 C5 targets that overlap both catalogs are compared here. **Top:** We present the overall change in effective temperature ($T_{\rm eff}$). We do find minor variations at different temperatures due to changes in spectral classification, but the overall systematic offset is not significant. **Bottom:** We present the overall change in stellar radius ($R_\star$) and find a very minor trend for small radius stars, but these M dwarfs are not included in our stellar sample. Overall the systematic offset is negligible.

remaining 5,810 stars. By including these additional stars in our sample, we are able to maximize the planet sample. However, we recognize the increased uncertainty introduced by including portions of the Huber et al. (2016) sample.

It is important that all the stars in our sample have measured values of stellar temperature ($T_{\rm eff}$), radius, mass, surface gravity (log(g)), and metallicity ([Fe/H]). The requirement of metallicity helps constrain the limb darkening parameters. Thus, we remove 3,020 targets that do not have these measurements available, reducing the sample to 22,010 stars. In all cases we find the target either has a measure of all five of these attributes or none of them.

Our second cut to the sample is meant to eliminate evolved giant stars. We implement the log(g) threshold derived by Huber et al. (2016):

$$\log(\mathrm{g}) \geq \frac{1}{4.671}\arctan\left(\frac{T_{\mathrm{eff}} - 6300\mathrm{K}}{-67.172\mathrm{K}}\right) + 3.876 \tag{5.1}$$

which approximates the limit for dwarf classification (according to the Parsec models; Bressan et al. 2012) for stars with solar-metallicity. Undoubtedly some of the stars in our sample deviate from solar-metallicity, but this threshold remains sufficient in eliminating evolved stars (see Figure 6 of Huber et al. 2016). For reference, this equation permits log(g)$\geq 4.20, 4.19$, and 3.61 for T_{eff} = 4500K, 5500K, and 6500K respectively. We remove 6,895 stars that do not meet this requirement, leaving 15,115 stars in our stellar sample.

The Hardegree-Ullman et al. (2020) catalog is unique in that it provides an estimate of the spectral type of the star. We utilize this feature to select for F, G, or K spectral types. Unfortunately, the Huber et al. (2016) does not provide a similar spectral measure, thus we rely on the inferred stellar temperature (T_{eff}). We consider a star in the FGK regime if T_{eff} is within the range of 4000-6500K. In Figure 5.3 we compare targets that overlap both catalogs and find this range best represents FGK classification in the Hardegree-Ullman et al. (2020) catalog. After removing 5,060 targets (4,808 M dwarfs and 252 A dwarfs), this cut allows 10,055 stars to remain in our sample.

Our last filters remove stars that were deemed problematic by the *K2* pipeline. These light curves were either unable to be properly smoothed, or the stellar surface is extremely active, making transit detection nearly impossible. We use the Combined Differential Photometric Precision (CDPP) to determine the threshold of transit detection. CDPP is a measure of the average noise found within the light curve, given a window of time (Christiansen et al., 2012). As suggested by Zink et al. (2020a), we remove targets with CDPP$_{8hr}$ > 1200ppm, eliminating 782 targets. Additionally, some targets had very large photometric apertures which likely contained more than one star. We exclude these targets by enforcing

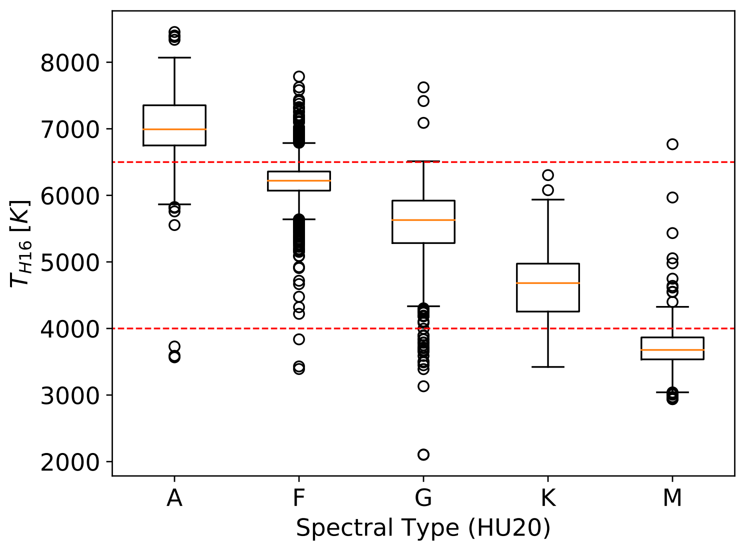

Figure 5.3: Shows the inferred stellar temperature (T_{eff}) from H16 (Huber et al., 2016) vs. the spectral classifications of HU20 (Hardegree-Ullman et al., 2020). The 13,029 C5 targets that overlap both catalogs and meet the log(g) threshold in Equation 5.1 are compared here. The orange line represents the 50th percentile (median) of each set of values. The black box shows the 25th and 75th percentile of the data and the whiskers represented the lower and upper bounds of the data (excluding outliers which are displayed as flier points). The red dotted lines represent the limits selected for FGK classification from stars with only H16 parameters available.

a maximum aperture threshold of 80 pixels, removing 16 stars.

After refining our sample to meet the discussed requirements, we are left with 9,257 stars. Of these stars, only 865 rely on parameters from Huber et al. (2016). The remaining 8,392 stars use parameters derived by Hardegree-Ullman et al. (2020). Finally, we calculate the two quadratic limb darkening values for each target, using the ATLAS model coefficients for the *Kepler* bandpasses tabulated by Claret & Bloemen (2011).

With our stellar sample in hand, we caution that *K2* field selection was guest observer driven. This could potentially bias our sample to focus on regions with bulk stellar properties (i.e. mass, radius, and metallicity) that favor planet detection. Upon examining a 10° radius of the C5 field using the TIC (TESS Input Catalog)[2], we find the stellar parameter distributions of the C5 sample do not deviate from that of the broader field. We can therefore conclude that the guest observer selection effect will be negligible and disregard such issues in this study.

5.2.2 *Kepler* Stellar Selection

As the main objective of this paper is to compare *Kepler* occurrence to that of the *K2* C5, we also use similar cuts to select our *Kepler* sample. We begin with the stellar parameters provided by the Berger et al. (2020b) catalog (186,301 stars). By limiting the sample to stars that meet the log(g) requirements of Equation 5.1, a $T_{\rm eff}$ within the range of 4000–6500K, a measured $\mathrm{CDPP}_{7.5hr} < 1000$ppm, and available measurements of radius and mass, we are left with 104,498 stars. Here, we have relaxed the requirement of metallicity for this sample and instead use Equation 9 of Zink et al. (2019b), which provides a method of determining the limb darkening parameters used by the *Kepler* DR25 detection pipeline using only the star's $T_{\rm eff}$. To ensure our sample only includes stars with significant data available, we remove targets with less than two years between the first and last photometric data points (span >

[2]https://tess.mit.edu/science/tess-input-catalogue/

730 days) and only include targets with more than 60% of the cadences available between these end points (duty > 0.6). This final cut leaves 94,222 stars in our stellar sample.

As noted for the *K2* stellar sample, using catalogs with uniquely derived parameters can introduce potential systematic offsets. The stellar radius measurements are the most concerning, as these have the largest effect on the inferred planet occurrence. To address this issue, we compare the radius measurements of both our *Kepler* and *K2* samples to the radius values uniformly derived by the *Gaia* team (Gaia Collaboration et al., 2018). In Figure 5.4 we find a very minimal systematic offset between the *K2*-*Gaia* sample ($0.02R_\odot$) and the *Kepler*-*Gaia* sample ($-0.04R_\odot$), indicating that differences caused by unique parameter derivation will be relatively small. This finding is not surprising as both catalogs used *Gaia* DR2 to infer their radius parameters. However, the additional photometry data used by both Berger et al. (2020b) and Hardegree-Ullman et al. (2020), to infer stellar parameters, provides increased accuracy when compared to that *Gaia* team measurements. Therefore, we use our original *K2* and *Kepler* catalog values while acknowledging their minor systematic offsets.

5.2.3 *K2* vs. *Kepler* Stellar Sample

The *Kepler* field and the *K2* C5 field represent independent samples in which we can measure planet occurrence in our local Galaxy. Furthermore, the stars in these samples are unique and provide insight into the stellar features that inhibit or encourage planet formation. In Figure 5.5 we compare the distributions of stellar parameters across both samples. Overall, the *K2* C5 sample appears contains stars that have smaller radii, are slightly cooler, and are metal-poor by comparison.

Early evidence from radial velocity surveys found an increased hot Jupiter occurrence around metal-rich stars, suggesting metal-rich protoplanetary disks are able to form planets more efficiently (Fischer & Valenti, 2005). However, the *Kepler* data provided evidence for a

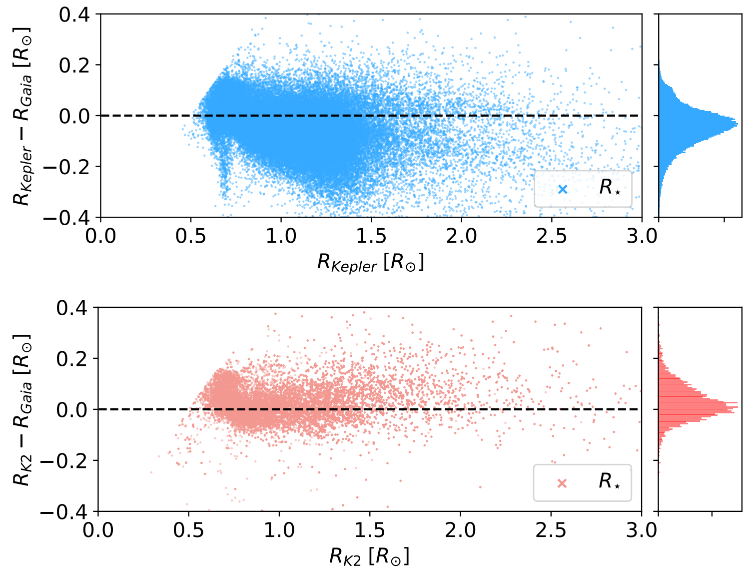

Figure 5.4: A comparison of the radius values used by our *Kepler* (**Top**) and *K2* (**Bottom**) stellar catalogs with the radius values derived by the *Gaia* team (Gaia Collaboration et al., 2018). Overall, both catalogs find similar values with those uniformly derived by *Gaia*.

lower occurrence of hot Jupiters ($0.5 \pm 0.1\%$ of stars; Howard et al. 2012) compared to the local solar neighborhood population ($1.20 \pm 0.38\%$ of stars; Wright et al. 2012). With data from LAMOST, Dong et al. (2014) was able to show that the *Kepler* field has a near-solar mean metallicity ([Fe/H]$= -0.04$ dex), which is comparatively higher than the local solar neighborhood ([Fe/H]$= -0.14 \pm 0.19$ dex; Nordström et al. 2004). Additional evidence of this positive stellar metallicity offset was provided by Guo et al. (2017), indicating that a metal deficiency cannot explain the reduced occurrence of hot Jupiters in the *Kepler* sample. This discrepancy may lead one to minimize the role of metallicity in planet formation. However, an increased sub-Neptune population was found around metal-rich stars within the *Kepler* sample (Petigura et al., 2018), indicating the role of metallicity is more nuanced than previously believed. Clearly, more data are needed to parse out the details of this effect.

Remarkably, the *K2* C5 sample provides a metallicity distribution that is very similar to solar neighborhood distribution ([Fe/H]$= -0.14 \pm 0.18$ dex). If an overall decrease in planet

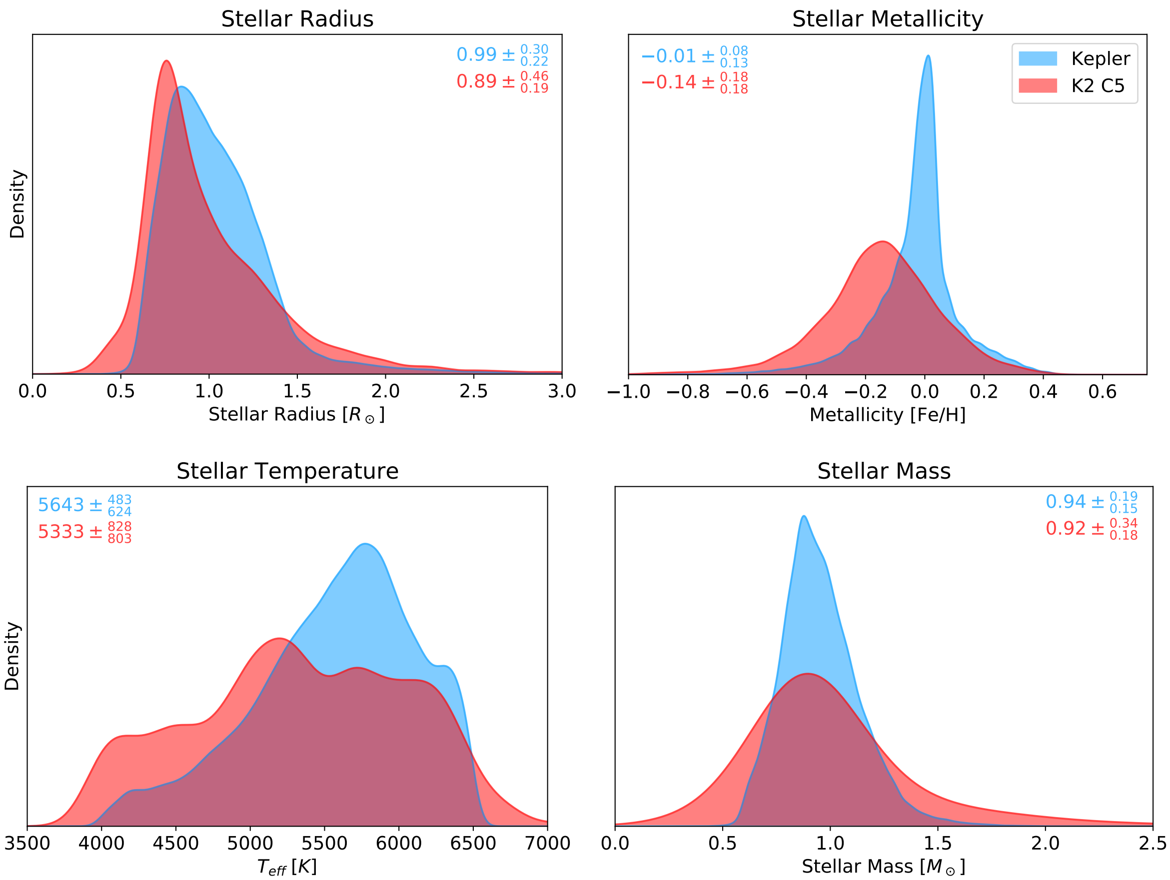

Figure 5.5: Shows the distribution of stellar parameters for our *Kepler* and *K2* C5 FGK dwarf sample. Measurements of radius, mass, metallicity, and $T_{\rm eff}$ use the parameter values provided by Berger et al. (2020b) for *Kepler* and Hardegree-Ullman et al. (2020) & Huber et al. (2016) for *K2* C5. The 50th, 16th and 84th percentiles have been listed in the upper corner of each plot.

occurrence was found in this sample, it would provide additional evidence for a metallicity-dependent formation mechanism. In the absence of an occurrence deficiency, the role of metallicity remains nuanced and beyond the detection of our broad summary statistics. Once additional *K2* Campaigns are available, detailed studies considering the effects of α-chain element abundances can be accomplished with *K2* data (see Figure 4.1) as these elements appear to be correlated with the detection of planets (Adibekyan et al., 2012).

5.2.4 Planet Selection

5.2.4.1 *K2* C5 Planets

In this study we use the *K2* Campaign 5 planet sample obtained by Zink et al. (2020a). This catalog is a uniformly vetted sample with a corresponding measure of completeness and reliability. This sample includes 75 planet candidates that are at least 94.2% reliable (small number statistics only allow for a lower limit on the measure of reliability). For our sample we adopt the planet parameters of radius and period derived in Zink et al. (2020a).

The planet sample is drawn from the subset of stars selected in Section 5.2.1. This cut removes 26 planets from our sample: 18 M dwarf candidates, one high CDPP light curve candidate, six low log(g) stellar host candidates, and one candidate without measured stellar parameters.

In addition, we remove gas giant planets from our sample. It has been shown these giants have a tendency to eject planets as they migrate inward (Beaugé et al., 2012). This inward orbital migration creates an independent population of giant planets that do not share the same population features as the planets formed in-situ (further evidence of this unique population was noted by Johansen et al. (2012), who showed that multi-planet systems with a short period planets greater than 0.1 Jupiter mass were dynamically unstable on short timescales). Additionally, it has been shown empirically that planets with $R > 6.7R_{\oplus}$ form a unique population that deviates from a simple power-law (Steffen et al., 2012). This cut of $R >$

$6.7R_\oplus$ roughly corresponds to planets with a mass greater than 0.1 Jupiter mass, derived dynamically by Johansen et al. (2012). However, the existence of short period gas giants in multi-planet systems has been seen in the *Kepler* data set (i.e Kepler-56; Huber et al. 2013) and the *K2* data set (i.e. WASP-47; Becker et al. 2015). While rare, these multi-planet systems indicate some small fraction of these massive planets can co-exist with other planets. As done in Zink et al. (2019b), we only remove planets with $R > 6.7R_\oplus$ if no other planets were detected in the system. This cut removes six hot Jupiters from our sample. No multi-planet systems with an $R > 6.7R_\oplus$ planet exist in the *K2* C5 sample used for this study, however to maintain consistency with our *Kepler* sample, we allow such systems in our forward model. Cutting the sample in this manner introduces a mild bias, as some of the planets removed may be part of systems with undetected planets. Additionally, the measured planet radius value may differ from the true value, modifying the parameter location of the planet relative to this cut. We account for this bias in Section 5.3.1.

Despite the improved radius measurements provided by Hardegree-Ullman et al. (2020), planet radii are still uncertain to about 16% for most planets in this catalog. We assume the values provided are accurate, but address the biases these uncertainties produce in our forward model (Section 5.3.1).

Overall, our sample consists of 43 planets with radii ranging from 1.2–$6.3R_\oplus$ and orbital periods ranging from 1.54–35.40 days. In Figure 5.6 we present our planet sample and the expected detection probability. Empirically, this sample contains 34 single planet systems, three double planet systems, and one triple planet system.

5.2.4.2 *Kepler* Planets

We use the *Kepler* planet candidate parameters provided by Thompson et al. (2018, 4,612 planets) with radius updates from Berger et al. (2020a). Drawing planets from the stars selected in Section 5.2.2 and applying the same gas giant removal procedure performed on

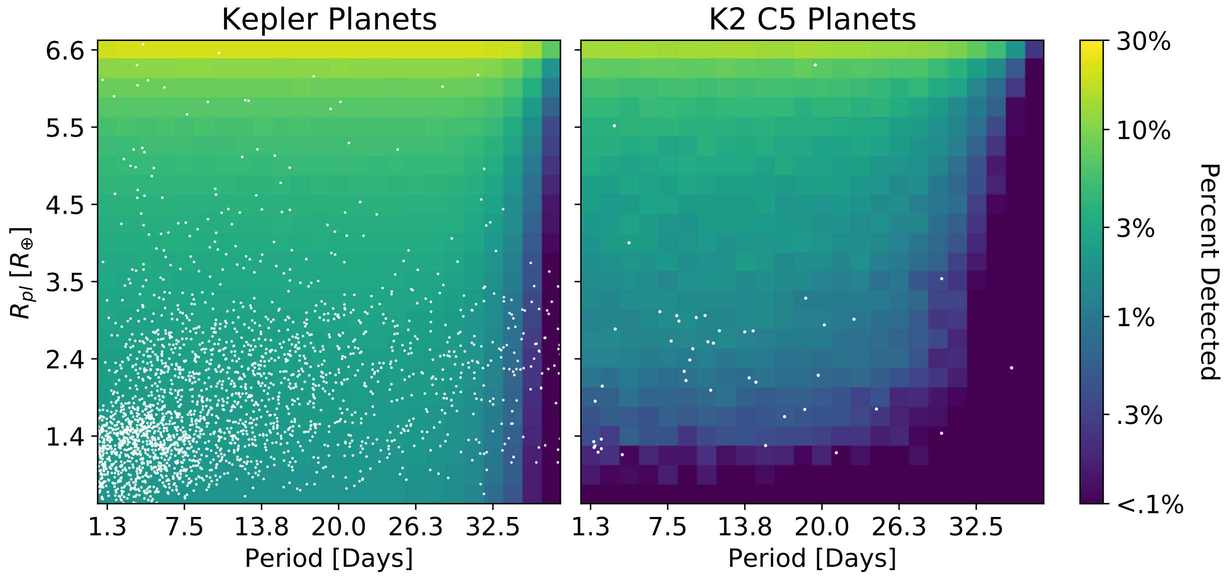

Figure 5.6: Shows the distribution of our planets samples in white for both *Kepler* (**Left**) and *K2* C5 (**Right**). Beneath this plot is a detection probability map. Using the stellar sample from each field a single planet, drawn uniformly over each bin of radius and period, is tested for detection through `ExoMul`. The fraction of planets recovered is then portrayed in this color map.

the *K2* C5 sample, we are left with 3,023 planets. This ensures the samples are comparable. The only difference is that we remove *Kepler* planets with periods greater than 38 days (2,318 planets remain), as this is the longest detectable period for the *K2* C5 data. In Figure 5.6 we present our planet sample and the underlying sample completeness.

Overall, our sample contains 2,318 planets with periods ranging from 0.51–38.00 days and radii ranging from 0.50–11.45$R_\oplus$. This sample has an average reliability of 98.3%, using the values provided by Thompson et al. (2018). Only 8 planets exceed the 6.7$R_\oplus$ single planet radius cut. Again these large multi-planet systems are rare and have little effect on the overall occurrence measurements (as the radius power law decays quickly in this region of parameter space, $\sim R^{-4}$). However their inclusion allows us to account for fluctuations in and out of sample near the 6.7$R_\oplus$ boundary due to measurement uncertainty.

5.3 Forward Model

We use the `ExoMult` software to forward model our population (Zink et al., 2019b). This program takes a population of planets and subjects them to the selection effects of the detection pipeline, providing an expectation for the observed population. `ExoMult` was originally designed to address the issues of the *Kepler* pipeline (Thompson et al., 2018), but the *K2* data set has unique issues which require modification to the base code. We discuss these differences below.

5.3.1 ExoMult

The original `ExoMult` code assumes the planet radius and period distributions are independent, and modeled by broken power laws. Here, we adopt these same assumptions. One of the goals of this study is to make a comparative statement about the *K2* sample versus the *Kepler* sample. Trimming the *Kepler* sample to match our 38 day period limit accomplishes this comparative goal, but we must carefully consider how doing so affects the optimization. A simple cut and re-process could yield inaccurate values, as non-detections provide constraints when optimizing the model. To avoid such issues we remove the window function from our *Kepler* forward model. This function determines the probability of at least three transits appearing in the data. Since the *Kepler* data set spans roughly 3.5 years, all of the planets within a 38 day period range would have experienced more than three transits, making this function unnecessary. In contrast, the *K2* C5 data set spans roughly 75 days. This means that the three transit window function will be important for optimization. Zink et al. (2020a) provides window function data for each light curve, but implementing these data into our forward model is computationally expensive. Instead we use the theoretical window function provided by Zink et al. (2020a), which closely matches the expected window function of most targets.

Several previous studies have empirically found an average mutual inclination around 1–2°

(Lissauer et al., 2011; Lissauer et al., 2012; Fang & Margot, 2012; Mulders et al., 2019). Implementing these non-independent details into our forward model is straight forward and a real advantage to this method of occurrence calculation. However, doing so requires a few additional parameters. This is not problematic for the *Kepler* sample with greater than 2000 data points, but begins to verge on over-fitting when attempted on the 43 data point sample of *K2* C5. For simplicity, we aim to minimize the number of variable parameters in this study. Therefore, we assume a mutual inclination of 0° and that the planets all orbit in a flat disk. Additionally, we assume all the planets exist on a circular orbit. This assumption is reasonable because nearly all of the planets with orbital periods < 38 days will have experienced tidal circularization, resulting in a population with eccentricities near zero.

`ExoMult` was built to mimic the selection effects of the *Kepler* pipeline. However, the Zink et al. (2020a) planet catalog is far less complete than the *Kepler* sample (See Figure 7 of Zink et al. 2020a). Thus, we adopt the vetted completeness function of Zink et al. (2020a) for our *K2* C5 processing. For *Kepler*, Zink et al. (2019b) showed that additional signals in the same light curve had a lower detection efficiency. This effect has not yet been measured for the *K2* sample. In an effort to make this a fair comparison between the two samples, we turn off this additional multiplicity completeness accounting in the *Kepler* forward model optimization.

One new feature introduced to `ExoMult` is the ability to deal with radius uncertainties. Previous studies have addressed this issue using hierarchical Bayesian analysis (Foreman-Mackey et al., 2014; Hsu et al., 2019), but forward modeling provided a straight forward method of accounting for these fluctuations. Each planet population is sampled and subject to all the selection effects of the *Kepler* or *K2* pipeline accordingly. The true stellar radius ($R_\star$) and planet radius (R_p) values are used to calculate the expected depth of the transit ($TD = R_p^2/R_\star^2$). To mimic radius variations caused by inaccurate depth measurements, we draw the measured transit depth ($\overline{TD}$) from a Gaussian distribution centered around the expected depth and a width of 4% the expected depth (the median depth uncertainty determined by Zink et al. 2020a). Independently, we draw the measured stellar radius ($\overline{R_\star}$) from a split normal distribution centered around the true radius value with a spread reflecting the upper and lower radius uncertainty. Using our drawn depth and drawn stellar radius values, we calculate the measured planet radius ($\overline{R_p}$):

$$\overline{R_p} = \sqrt{\overline{TD}} * \overline{R_\star}. \tag{5.2}$$

The $\overline{TD}$ value is drawn independently for each planet in a given system, while $\overline{R_\star}$ is only drawn once per system. This method of drawing measured values allows us to account for fluctuations introduced by poor radius measurements. To address the bias introduced by our giant planet removal, we remove single planet detection systems with $R > 6.7R_\oplus$ after the uncertainty modification has been applied.

5.3.2 MCMC

Using the expected populations provided by `ExoMult`, we can optimize the model to produce an observed population similar to that of our sample. Under the assumptions listed in Section 5.3.1, we have seven model parameters. The six broken power-law parameters (three for radius and three for period), and one overall occurrence factor (f). This f factor tells the forward model what fraction of systems have a planet. In previous versions of `ExoMult`, this factor was broken up to account for the number of stars with a planet and then the multiplicity of these systems. However, we want to minimize the number of parameters in this optimization to avoid over-fitting the *K2* data. Thus, the f parameter represents an overall measure of planet occurrence in these samples. If f is greater than one, each star will be assigned a planet, and the excess fraction will be assigned a second planet. For a detailed discussion of the model parameters we refer the reader to Section 7 of Zink et al. (2019b). We measure the Bayesian posterior for the seven model parameters using the `emcee` affine invariant sampler (Goodman & Weare, 2010; Foreman-Mackey et al., 2013) with 50 semi-independent walkers, 5000 burn in steps, and 10,000 sample steps (500,000 total samples of each posterior).

5.3.3 Priors

One of the features of Bayes theorem is the ability input prior information about your model parameters. For our *Kepler* model and our *K2* model, we assume uniform priors for all parameters. In order to avoid nonphysical cases, we allow f to range from 0–7. All power law parameters are allowed to range from -20 to +20, and the radius break and period break are allowed to range from 0–16$R_{\oplus}$ and 0–35 days, respectively. While these uniform priors have little effect on our current fitting, in Section 5.4.4 we discuss how priors can be used to combine *K2* and *Kepler* samples in future studies.

5.3.4 Likelihood Function

To compute the likelihood function of our model we utilize two test statistics. First, we utilize the K-sample Anderson-Darling test statistic (AD; Anderson & Darling, 1952; Pettitt, 1976) to capture the shape of the distribution. Second, we use a marginalized Poisson distribution to ensure our inferred distribution is properly normalized.

Our shape metric uses the AD test, a non-parametric method of measuring the probability that two samples come from the same distribution. In our case the two samples are the observed *Kepler* or *K2* C5 planet sample and the sample produced by our forward model. In overview, the Cumulative Distribution Function (CDF) of these two samples are compared and the differences are summed at each step, weighted by the location within the distribution. Theoretically, the largest differences should exist near the median of the distributions, therefore these separations are weighted less than those near the minimum and maximum values of the distribution. For a more thorough explanation of this test we refer to Babu & Feigelson (2006). The test statistic $AD \propto ln(\text{probability})$, measuring the probability that the two samples come from the same parent population. A similar metric is used in `SysSim` (He et al., 2019) and a version of this metric using the Kolmogorov–Smirnov (KS; Kolmogorov 1933; Smirnov 1948) test is used in `EPOS` (Mulders et al., 2018).

We compute independent measures of AD for the radius and period population (AD_R and AD_P) of our forward model distribution compared to the empirical *Kepler* and *K2* samples. This ensures the shape of the distributions are optimized at each step of the MCMC. However, we must also optimize the normalization of these distributions.

Since the number of detected planets are discrete values, we rely on Poisson statistics to optimize our normalization factors. The number of planets detected by the forward model population (N_S) are compared to the empirical *Kepler* or *K2* C5 samples (N_E). One issue unique to the *K2* data is that $N_E = 43$ is a small discrete value, which means we do not have a good measure of the expected number of planets. In other words, the number of planets detected in the empirical sample is drawn from some Poisson distribution, but we do not know the true scale parameter (λ) of this distribution. When N_E is large (as is the case for the *Kepler* sample; $N_E = 2,318$) it is reasonable to assume $\lambda = N_E$, but this assumption is less valid when N_E is small. To account for these small number statistics we assume N_S and N_E come from the same Poisson distribution with an unknown λ. By multiplying these two probabilities together we get the probability of drawing N_E and N_S given some λ value. Since we do not care what λ is, just that the two drawn values came from the same distribution, we can then marginalize over the nuisance parameter λ, removing it from the equation:

$$\begin{aligned} P(N_S \cap N_E|\lambda) =& \frac{e^{-\lambda}\lambda^{N_S}}{N_S!} * \frac{e^{-\lambda}\lambda^{N_E}}{N_E!} \\ P(N_S \cap N_E) =& \int_0^\infty \frac{e^{-\lambda}\lambda^{N_S}}{N_S!} * \frac{e^{-\lambda}\lambda^{N_E}}{N_E!} d\lambda \\ P(N_S \cap N_E) =& \frac{2^{-N_S-N_E-1} * \Gamma(N_S+N_E+1)}{N_S! * N_E!} \end{aligned} \tag{5.3}$$

where Γ is the gamma function. This equation ($P(N_S \cap N_E)$) gives us the probability that these two discrete values come from the same Poisson distribution, without needing to know the true underlying λ. The overall difference between making the assumption that $N_E = \lambda$ and using the above equation is that $P(N_S \cap N_E)$ allows for a slightly larger variance between values, as expected from small number statistics.

Putting together our shape metrics (AD_R and AD_P) and our normalization probability ($P(N_S \cap N_E)$) we get the log likelihood function of our posterior:

$$ln(\text{likelihood}) \propto ln(P(N_S \cap N_E)) + AD_R + AD_P. \tag{5.4}$$

This function multiplied by the model priors provides our measure of the posterior distribution for our model parameters.

5.3.5 Reliability

As noted in Bryson et al. (2020b), it is important that occurrence rates consider the sample reliability when optimizing their model. Without such consideration, the inferred model can be affected by our choice of planet candidacy thresholds. To account for such an effect, we calculate the reliability of each planet using the values provided by Zink et al. (2020a) (for *K2* C5) and Thompson et al. (2018) (for *Kepler*). At each step of the MCMC we draw (without replacement) each planet in our sample based on a probability corresponding to the planet's reliability. This means that the empirical *K2* C5 sample will often have less than 43 planets. However, our sample has an overall reliability of 95%, indicating that on average 41 planets will be included in the sample that is compared against the forward model sample. Comparatively, the *Kepler* sample, with 2,318 planets and 98.4% reliability, will on average be drawn with 2,281 planets.

It is important to note that the measurements of reliability provided by Zink et al. (2020a) and Thompson et al. (2018) are measures of systematic reliability which ignore potential astrophysical false positives. Projected double stars with small separations can dilute the transit depth, resulting in an underestimation of the planet radius (Ciardi et al., 2017; Fulton et al., 2017). This can directly lead to an overestimation in the number of small radius planets and potentially contaminate the planet sample with eclipsing binaries. To minimize this potential contamination, Thompson et al. (2018) looked for shifts in the centroid of the target star while the candidate was in transit. Finding such a shift provides evidence that significant flux is being contributed by a secondary source and the candidate warrants rejection from the planet sample. This metric is able to detect contaminants down to $1''$ separations (Bryson et al., 2013). Unfortunately, such measurements are more difficult to establish for *K2*, where spacecraft systematics are constantly shifting the centroid. Consequently, Zink et al. (2020a) relied on the *Gaia* DR2 data to minimize these false positives, which provides contaminant detections down to $1''$ separations (Ziegler et al., 2018). While unique in methodology, both catalogs are robust to contaminants wider than $1''$ separations. However, contaminants within a $1''$ separation (largely gravitationally bound binaries; Horch et al. 2014) will remain undetected and reduce the overall reliability of these catalogs. While such corrections are essential for accurate occurrence measurements, our sample is limited to planets with $R \leq 6.7R_{\oplus}$, minimizing contamination from eclipsing binaries (Fressin et al., 2013a). Work by Matson et al. (2018) found that *K2* planet hosts have a binarity rate of $23 \pm 5\%$ and Furlan et al. (2017) found a similar value of 30% for *Kepler* planet hosts, but implementing such information into an occurrence rate is beyond the scope of this paper and therefore ignored.

5.4 Results

In this section we compare three different models against the empirical *K2* C5 planet sample: the *K2* model, which only uses the 43 planets in our sample to fit the seven population parameters; the *Kepler* model, which uses the 2,318 *Kepler* planet candidates to fit the seven population parameters; and the *K2* w/ *Kepler* model, which uses the six shape parameter posterior distributions derived by the *Kepler* model and fits the normalization factor (f) using the 43 *K2* C5 planet candidates.

Table 5.1: The resulting best fit parameters of our forward model optimization. The *K2/Kep.* model uses the same shape posteriors as that of the *Kepler* model. In Figure 5.9 we plot the posterior distributions for the f values provided here.

Model	α_1	R_{br}	α_2	β_1	P_{br}	β_1	f
		($R_\oplus$)			(days)		($\frac{\text{planets}}{\text{star}}$)
Kepler	$-1.61^{+0.17}_{-0.14}$	$3.03^{+0.38}_{-0.37}$	$-6.56^{+1.77}_{-5.58}$	$0.91^{+0.19}_{-0.17}$	$6.83^{+1.59}_{-1.26}$	$-0.59^{+0.15}_{-0.18}$	$1.10^{+0.05}_{-0.05}$
***K2* C5**	$-0.38^{+1.78}_{-1.29}$	$2.98^{+1.22}_{-1.29}$	$-6.99^{+3.28}_{-10.01}$	$1.66^{+6.49}_{-1.26}$	$6.91^{+2.07}_{-4.03}$	$0.15^{+0.71}_{-0.90}$	$1.00^{+1.07}_{-0.51}$
K2/Kep.	—	—	—	—	—	—	$0.77^{+0.21}_{-0.20}$

5.4.1 *K2* C5 Model

Using the 43 planets in our C5 FGK sample, we optimize the seven population parameters. Since the number of planets is only six times greater than the number of parameters, the uncertainties in our estimates for this model are very large. Nevertheless, we can still provide some measure of the population parameters.

In Table 5.1 we provide the results of our MCMC for the population parameters using only the *K2* C5 data. To ensure the model reflects the data, we look at the model distributions in Figure 5.7. It is clear that the model subjected to the selection effects of our pipeline follows the shape (CDF) of the period and radius distributions. Overall, these model parameters find the same trend that previous occurrence rates have noted: a constantly decreasing radius distribution (α_1 and $\alpha_2 < 0$; Burke et al. 2015; Mulders et al. 2018) and a peaked period distribution around 10 days ($P_{\rm br} \sim 10$ days; Youdin 2011; Howard et al. 2012; Mulders et al. 2019). We can also see in Figure 5.7 that this model produces a proper normalization of the data, producing an expectation of 34^{+11}_{-11} detected planets. This normalization is controlled by the overall occurrence parameters (f), which indicates our stellar sample should host on average $1.00^{+1.07}_{-0.51}$ planets per star. It is not surprising that we find a value near one, as the 38 day period limit of this study removes a significant fraction of system multiplicity. However, it is likely that this number is a lower estimate as we have assumed no mutual inclination between planets for simplicity. The real strength of this value is that it allows us to make a comparative statement to that of the *Kepler* model.

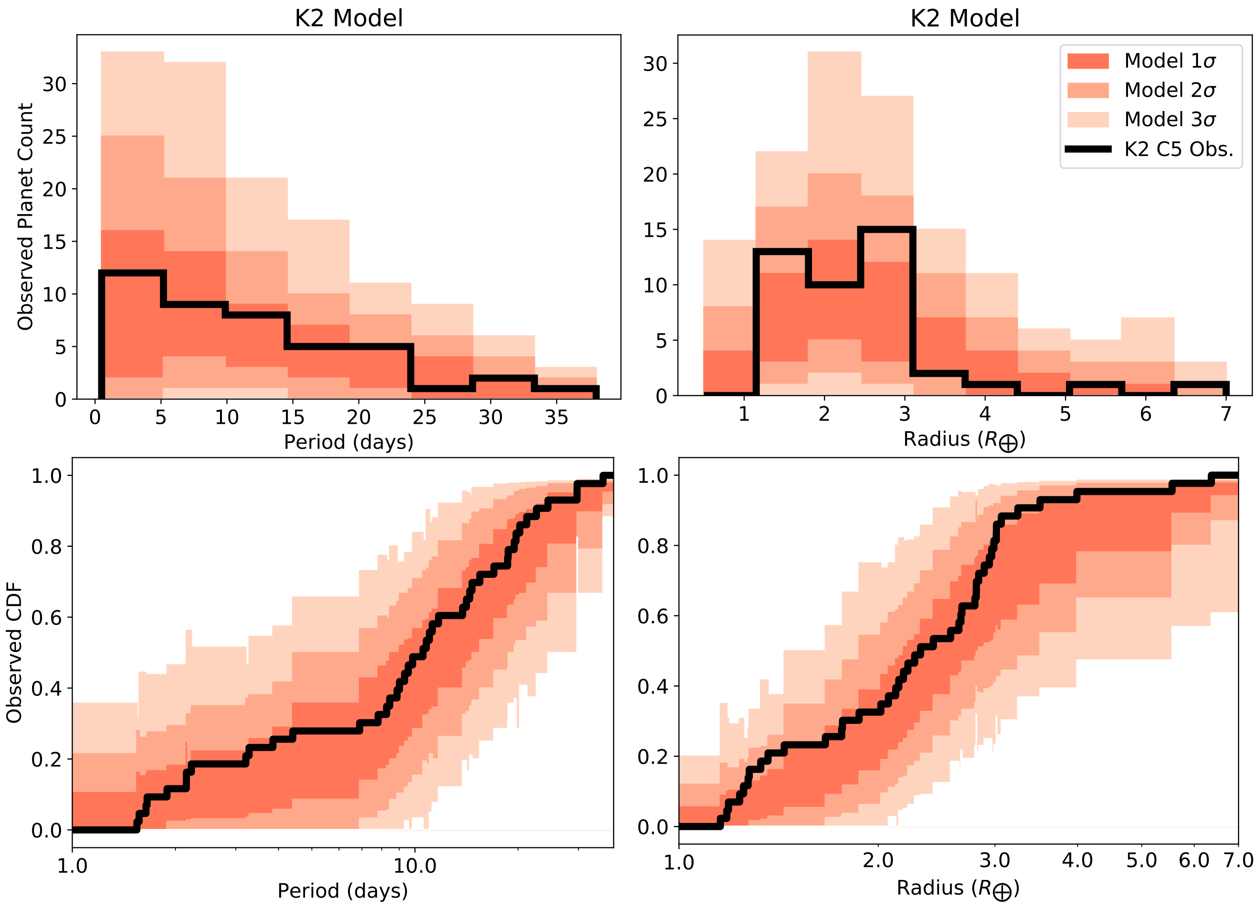

Figure 5.7: **Top** the observed radius/period distributions for the *K2* C5 data set versus the expected detections from the best fit broken power-law model, binned to show that the normalization factor correctly matched the observed number of planets. **Bottom** the observed CDF for the *K2* C5 data plotted against the expected detections from the best fit model, indicating a good match of the distribution shapes. The colored regions reflect the 68, 95, and 99.7 percentiles found after sampling with the model parameters 10,000 times.

5.4.2 *Kepler* Model

Using our *Kepler* planet and stellar sample we report the results of our *Kepler* model optimization in Table 5.1. Clearly the larger planet and stellar samples help reduce the overall uncertainty in our population parameters. Unsurprisingly, the *Kepler* model shape parameters are in agreement with previous *Kepler*-centric studies (i.e. Zink et al. 2019b; $\alpha_1 = -1.65$, $R_{\rm br} = 2.66$, $\alpha_2 = -4.35$, $\beta_1 = 0.76$, $P_{\rm br} = 7.09$, and $\beta_2 = -0.64$), which uses a larger planet sample that spans the full 500 day period range of the *Kepler* data set. By cutting the data and simplifying the processing to match that of *K2* C5, we can now look for potential differences and similarities between these two independent samples.

In Figure 5.8 we present the detected sample for the *Kepler* model population parameters given the selection effect of the *K2* C5 pipeline. This allows us to see how the *Kepler* planet population would have been detected by *K2*. It is apparent that the *K2* C5 sample is within 3σ of nearly all aspects of the *Kepler* population model. Furthermore, in Table 5.1 all of the *K2* C5 inferred population parameters are well within 1σ of their corresponding *Kepler* parameters. This indicates that the *K2* C5 sample is not statistically different from the *Kepler* population. We caution that a lack of statistical significance does not ensure true similarity; a much larger planet sample will be needed to make such a claim. However, it appears that any differences that do exist will be minor. If we consider the overall expected number of detected planets, this model predicts the detection of 58 ± 8 planets (compared to the 43 in our *K2* C5 sample). This similarity further indicates a lack of uniqueness among these samples.

Although differences in the populations are statistically insignificant, the uncertainties in the *K2* C5 data remains high. Should subtle differences exist, a larger *K2* sample would be needed for detection. One method of improving the fit for the overall occurrence factor (f) would be to reduce the number of model parameters. It is likely that the same formation mechanisms are at play in both the *Kepler* and *K2* samples, thus the overall shape of these distributions should be similar and information from *Kepler* can be used to help fit the *K2* data.

5.4.3 *K2* w/ *Kepler* Model

To minimize the number of parameters in the *K2* C5 model we use the posterior values from the *Kepler* shape distribution for optimization. At each step of the MCMC a set of shape values are drawn from the posterior samples inferred by the *Kepler* model. Drawing values in a set ensures we maintain any dependence between parameters. The only parameter that is allowed to roam is the occurrence factor (f), which normalizes the distribution. By reducing the parameter space search we can produce a sharper estimate of the overall planet occurrence in *K2* C5.

We find that reducing the parameter search produces a 22% lower occurrence factor ($f = 0.78^{+0.23}_{-0.21}$) and a 70% reduction of the parameter uncertainty. Although reduced even further, we still do not find a statistically significant difference from that of the *Kepler* model values ($f = 1.21 \pm 0.05$). In Figure 5.9 we show the posterior distributions for our three model occurrence factors and their significant overlap, providing further evidence that differences in these two independent populations will be subtle, if existent.

Since both *K2* models produce statistically indistinguishable occurrence factors (when compared to the *Kepler* model), we can use the posteriors to bound the differences that could exist and remain undetected by our study. By considering the 3σ posterior values for f, we can bound $\Delta f = f_{\mathrm{Kepler}} - f_{\mathrm{K2}}$ to a range of $[-4.47, 0.99]$ planets per star for the *K2* C5 model and $[-0.86, 0.98]$ planets per star for the *K2* w/ *Kepler* model. Clearly, these bounds remain rather weak due to the large uncertainty in *K2* occurrence, but both models are able to rule out $\Delta f = 1$. Should the true f_{Kepler} value be greater than the f_{K2} value–as expected by the median posterior values–the difference will be less than one planet per star in the parameter space of this study.

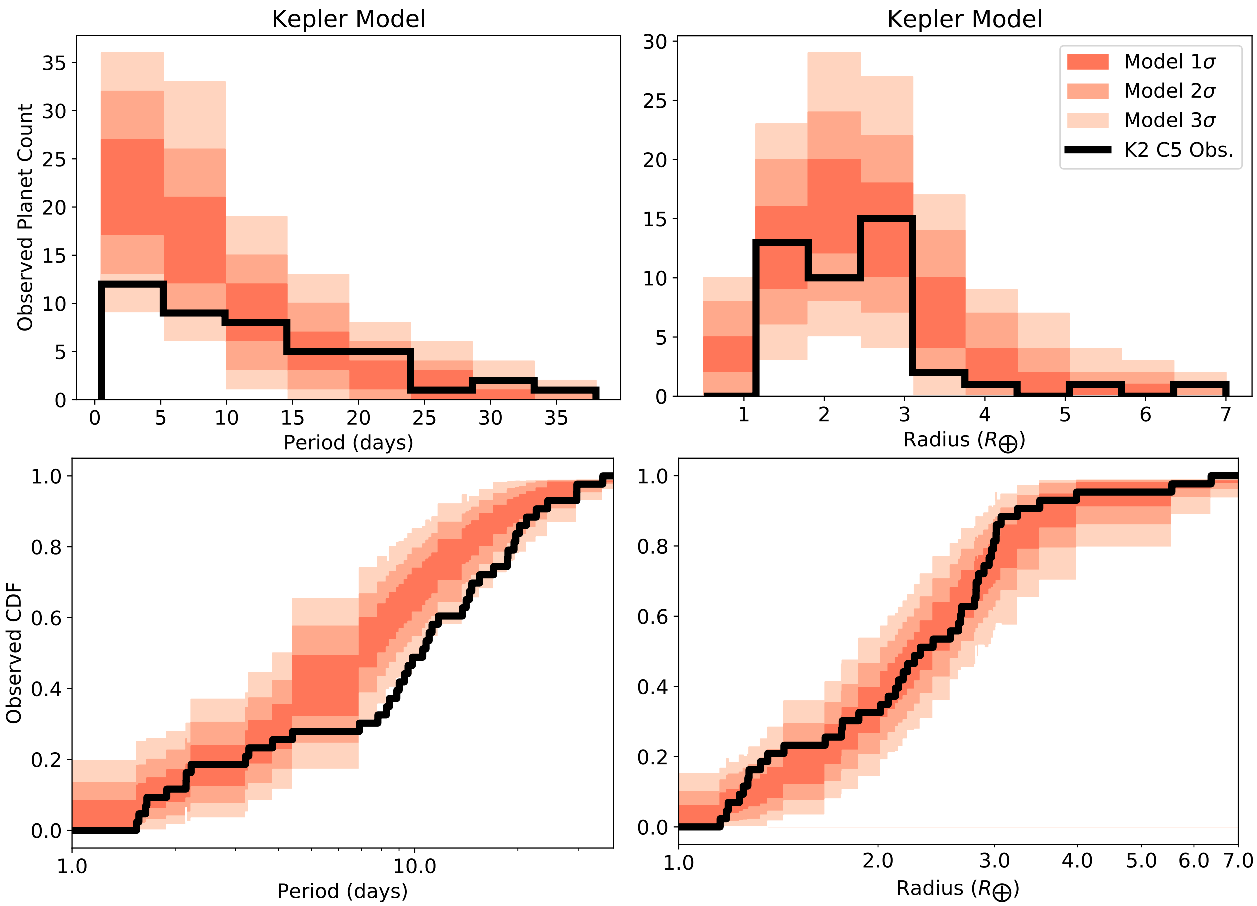

Figure 5.8: **Top** the observed radius/period distributions for the *K2* C5 data set versus the expected detections from the *Kepler* broken power-law model. **Bottom** the observed CDF for the *K2* C5 data plotted against the expected detections from the best fit *Kepler* model, indicating a reasonable match of the distribution shapes. The colored regions reflect the 68, 95, and 99.7 percentiles found after sampling with the model parameters 10,000 times.

5.4.4 *K2* and *Kepler*

Since we do not find an overall statistical difference between these two populations, it seems reasonable to combine these two samples to improve the overall model fit. However, appropriately combining the *Kepler* DR25 sample with the *K2* sample is difficult because both empirical samples have unique selection effects that need to be taken into account. Fortunately, Bayesian analysis provides a direct way of implementing this type of knowledge.

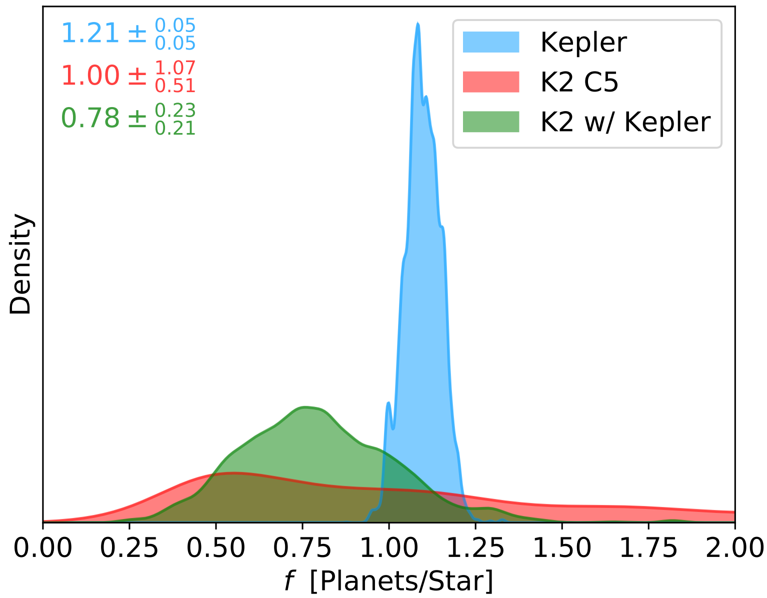

Figure 5.9: The posterior distributions for the occurrence factors (f) derived from our planet samples. The 50th, 16th and 84th percentiles have been listed in the upper corner of the plot for each value. Clearly, there is significant overlap between all of these values.

Since the *Kepler* sample is an independent measurement of the population parameters, we can use the inferred *Kepler* posterior values as priors for our model. This will essentially update the posteriors given the new *K2* data, pulling the posteriors closer to the true population parameters. In a preliminary study, this Bayesian analysis was carried out for the *K2* C5 sample, but the C5 sample only added an additional 43 planets to the *Kepler* sample of 2,318 planets (a 1.9% increase), providing little influence on the overall population parameters. Once a larger portion of the *K2* Campaigns have been processed we will use this methodology to combine data across missions.

5.5 Deviations from the Model

In Figures 5.7 and 5.8 it is clear that the broken power-law model for the planet radii fails to replicate the gap seen in the radius sample near $R = 2R_{\oplus}$. Currently, the cause of this gap remains unclear. Lopez & Fortney (2013) and Owen & Wu (2017) suggest that the gap is caused photoevaporation, while Gupta & Schlichting (2019) indicates such a gap could be caused by core-powered mass loss. Regardless, this gap has been noted in both the *Kepler* (Fulton et al., 2017) and *K2* (Hardegree-Ullman et al., 2020) planet samples, indicating an underlying formation mechanism is at play. As we continue to increase the known planet sample, we will be able to better constrain the underlying mechanism (Loyd et al., 2020).

Without a well defined model available for this gap, it remains difficult to recreate in our population analysis. We acknowledge that our model does not address this issue. We also attribute the increased planet occurrence near $2.75R_\oplus$ and the subsequent decrease in planet occurrence seen near $3.1R_\oplus$, which appears to deviate from a power-law model in the empirical *K2* C5, to this lack of a well defined radius gap model.

Both photoevaporation and core-powered mass loss predict a period dependency to this gap. In the *Kepler* data set, which included planets with periods of 0.5–500 days, the effect of the radius gap is almost completely washed out when considering the period distribution on its own. However, our *K2* C5 sample is limited to periods of 0.5–38 days, where most of the detected planets have periods less than 10 days. Additionally, almost all radius gap models intersect at the junction of $R = 1.7R_\oplus$ and $P = 7$ days, where the gap is most prominent (MacDonald, 2019). Combining these two facts, our smaller under-sampled slice of the observable exoplanet period population is prone to gaps in the period distribution. We see such a gap in our empirical sample between 4–7 days, where the period distribution appears to deviate from the power-law (see Figure 5.8 Bottom). We therefore conclude that this apparent gap is not meaningful.

An additional complexity was introduced by Millholland et al. (2017) and Weiss et al. (2018a), who showed that planets are not independent within a given system. The "Peas in a Pod" result shows that planets within a given system tend to have similar sizes. Modeling this type of radius correlation remains difficult for population analysis. He et al. (2019) proposed a clustered point process model for dealing with system radius similarity. However, a proper accounting for such similarities would require a thorough understanding of the multiplicity completeness, which is currently not available for the *K2* data set. Furthermore, accounting for these empirical features requires additional parameters, which we have tried to minimize due to the small size of the *K2* C5 sample. Fortunately, only seven planets in our *K2* sample exist in multi-planet systems, therefore the overall effect of this correlation on the inferred population is small. Of the *Kepler* sample, 1,020 planets are part of multi-planet systems, and thus accounting for such correlations becomes important. We have ignored such issues here in order to minimize the number of parameters and to make the analysis between *Kepler* and *K2* data equivalent.

5.6 Effects of Metallicity

In Section 5.2 we discussed the differences between the *Kepler* and *K2* C5 stellar samples. The most notable difference is the metallicity of these two samples. The *Kepler* sample represents a more metal-rich sample than the *K2* C5 stellar sample. When considering the trends observed in Petigura et al. (2018), we should expect to find more sub-Neptunes (1.7–4$R_\oplus$) in the metal-rich *Kepler* sample.

While we do not directly consider this metallicity effect, we can discuss the expected consequences in our results. Since a larger number of sub-Neptunes should be found in the *Kepler* sample, we expect this to increase the overall occurrence of planets in the *Kepler* model. Although statistically insignificant, we do find a slight increase in the occurrence of planets produced by our *Kepler* model ($\Delta f = 0.21 \pm 1.07$ planets per star) compared to the *K2* C5 model. This difference is amplified even further when considering the *Kepler* and *K2* w/ *Kepler* model ($\Delta f = 0.33 \pm 0.22$ planets per star), producing a 1.5σ difference.

Additionally, we would also expect the α_2 population parameter to be slightly inflated for the *Kepler* sample as there would be a greater number of sub-Neptunes in the radius range this parameter spans. Again, we find a statistically insignificant increase between our models

($\Delta\alpha_2 = 0.4 \pm 6.47$). While the differences observed are not able to confirm the findings of Petigura et al. (2018), they are in agreement with the expectations of such a metallicity effect. We leave a more thorough consideration of this effect for future studies when a larger *K2* sample is available.

5.7 Conclusions

We used the *K2* C5 fully automated detection pipeline data set to determine the underlying population of planets around FGK dwarfs in the C5 field. In doing so we were able to infer an overall occurrence of $1.00^{+1.07}_{+0.51}$ planets per star in the parameter space of this study. When we compared the population parameters to those of the best-fit *Kepler* model, we found that all values are well within a $\sim 1\sigma$ difference, including the overall occurrence factor (1.10 ± 0.05). Even when using the *Kepler* model shape parameters to improve the optimization of the *K2* C5 occurrence factor, we found only a 1.5σ decrease in planet occurrence in the *K2* C5 field. Despite the C5 field probing a different region of the Galaxy, we infer a population that appears consistent with the *Kepler* sample. This indicates that our knowledge of the *Kepler* field could potentially be extrapolated to a larger part of the Galaxy.

Using Bayesian priors, we also discussed a methodology for combined *K2* & *Kepler* mission data to carry out a Galactic transiting exoplanet occurrence rate. With the currently available data, this analysis would be heavily biased toward the *Kepler* field data (2,318 *Kepler* planets versus 43 planets from C5). A more rigorous Galactic survey would sample various regions of the local Galaxy, with a similar data span at each of them. Fortunately, the *K2* mission did just that, and as more campaigns are fully processed with the automated pipeline we will use this methodology to calculate a representative Galactic occurrence rate for planets.

Finally, we showed that the *K2* stellar sample is metal-poor compared to the *Kepler* stellar sample, but we were unable to find statistical differences in our models. Findings of model similarity may reduce the role metallicity plays in planet formation, however our results found a weak increase in the occurrence of planets in the *Kepler* field. This trend seems to indicate that a larger planet sample–or a more substantial sample metallicity difference–is needed to confirm/refute the importance of metallicity in planet formation. Once the entire *K2* planet sample is made available, a more thorough consideration of metallicity effects can be achieved.

CHAPTER 6

The Great Inequality and the Dynamical Disintegration of the Outer Solar System

6.1 Introduction

Understanding the long-term dynamical stability of the solar system constitutes one of the oldest pursuits of astrophysics, tracing back to Newton himself, who speculated that mutual interactions between planets would eventually drive the system unstable (Laskar, 1996, 2012). Laplace (1799-1825) and Lagrange (1776) successfully challenged this perception by approximating the mutual interactions as perturbations, showing that, to leading order, the long-term evolution of all known solar system planets could be described via cyclic secular variations, thus analytically demonstrating the indefinite stability of the solar system. However, subsequent analyses by Gauss (1809) and Le Verrier (1856) showed that these approximations break down over sufficiently long time intervals, so that more complicated solutions are required. Poincaré (1892) formalized this insight by proving that the full "three-body problem" could not be solved in closed form — a barrier that has only recently been overcome with the advent of modern computing and N-body integration methods.

Unfortunately, even the most precise N-body simulations are only able to produce time-limited prognosis for the evolution of the solar system. Due to the chaotic nature of the planetary orbits, deterministic forecasting is impossible over sufficiently long timescales. In particular, the Lyapunov time for the inner terrestrial planets (Mercury, Venus, Earth, and

Mars) is of order ~ 5 Myr (Laskar, 1989; Sussman & Wisdom, 1992), while the outer Jovian planets (Jupiter, Saturn, Uranus, and Neptune) appear chaotic with a Lyapunov time of order 10 Myr[1] (Laskar, 1989; Murray & Holman, 1999; Guzzo, 2005, 2006). Predictions on timescales significantly longer than these benchmark values are only meaningful in a statistical sense, recasting the question of the solar system's long-term fate as a probabilistic one.

Over the course of the last three decades, the question of whether or not the orbits of the solar system's eight planets can remain immutable has come into sharp focus, with state of the art simulations demonstrating that Mercury has a $\sim 1\%$ chance of becoming unstable within the remaining main sequence (MS) lifetime of the Sun (Laskar, 1994; Laskar, 2008; Batygin & Laughlin, 2008; Laskar & Gastineau, 2009; Zeebe, 2015). The mechanism for the onset of Mercury's instability is well understood: by virtue of locking into a linear secular resonance with the g_5 mode of the solar system's secular solution, Mercury's eccentricity can attain near-unity values, resulting in a collision with the Sun, or even Venus. Intriguingly, General Relativistic effects factor into this estimate, with ancillary apsidal precession providing a stabilizing influence on Mercury's orbit (Lithwick & Wu, 2011; Boué et al., 2012; Batygin et al., 2015). Within the context of this narrative, however, the remaining planets appear unaffected and are currently expected to remain stable for a lower limit of 10^{18} years, when diffusion arising from the overlapping mean motion resonance of Jupiter and Saturn are expected to decouple Uranus (Murray & Holman, 1999).

Although the estimate of Murray & Holman (1999) addresses the intrinsic stability of the solar system, on sufficiently long timescales, extrinsic effects come into play. For example, stellar evolution, which is generally not considered in orbital stability studies, represents an important additional aspect of the problem. In particular, the Sun will undergo significant

[1]The exact value for the Lyapunov exponent is strongly dependent on the initial conditions of the system and has be found to range from 5-20 Myr, using orbital parameters within the observational uncertainty.

mass-loss over the next 7 Gyr, reducing the mass by roughly half, down to $0.54 M_{\odot}$ (Sackmann et al., 1993). Over this time span, it is probable that Mercury, Venus, and Earth will be engulfed by the Sun, as its radius expands during the red-giant branch (RGB) phase of evolution (Rybicki & Denis, 2001; Schröder & Smith, 2008). This epoch thus marks the end of the three innermost planets. Although Mars, Jupiter, Saturn, Neptune, and Uranus will survive this phase of stellar evolution (Veras & Wyatt, 2012), they will experience a ~ 1.85 increase in their semi-major axes (Jeans, 1924; Veras et al., 2011; Adams et al., 2013; Veras, 2016). With their newly expanded orbits, the remaining planets are expected to remain stable for a minimum of 10 Gyr (Duncan & Lissauer, 1998). However, the details of orbital evolution that unfolds on much longer timescales are less well characterized.

A distinct form of external forcing upon the solar system stems from stellar encounters. As the solar system traces through its Galactic orbit it will experience perturbations from passing stars, which will act to excite the orbits. If the solar system maintained its current orbital configuration, passing stars would liberate the planets over timescales of order 10^{14} yr (Dyson, 1979; Adams & Laughlin, 1997). However, the interaction cross-section for these gravitational encounters scale with the semi-major axis of the planet (Li & Adams, 2015). As a result, by accounting for stellar mass-loss and the inflation of the outer planet orbits, these encounters will become more influential. Given enough time, some of these flybys will come close enough to disassociate — or destabilize — the remaining planets. This paper examines the timescales and mechanisms that bring about the demise of the solar system.

This paper is organized as follows. In Section 6.2, we discuss the input parameters of our numerical study, including methods for simulating the evolution of the Sun itself and dynamical perturbations from stellar flybys. We then provide the results of our study and discuss the mechanisms for planetary disassociation in Section 6.3. The paper concludes in Section 6.4 with a summary of our results, a discussion of the significance of these findings, and the implications for free floating planets.

6.2 Numerical Methods and System Parameters

6.2.1 The N-Body Simulation

The timescales of interest for this study greatly exceed the current age of the solar system. To carry out this simulation, we used the IAS15 high-order integrator (Rein & Spiegel, 2015) as implemented in the `Rebound` (Rein & Liu, 2012) software package. This 15th order integrator uses a Gauß-Radau algorithm to numerically solve the equations of motion. Despite the high-order of this calculation, any use of finite time-steps introduces error, which can be characterized by deviations from the initial system energy ($\Delta E/E$). In an idealized system, IAS15 is capable of achieving 10^{-28} precision, but testing of the outer solar system produces a more realistic $\Delta E/E$ of $\sim 10^{-19}$. However, this metric assumes that the system energy is conserved throughout the simulation and the current study considers the effects of extrinsic factors that stochastically modify the energy of the solar system itself.[2] Over sufficient time, the accumulated energy contributions from stellar flybys will dominate the $\Delta E/E$ metric. Without a meaningful measure of the systematic integration error, we rely on adaptive time-steps (see Quinn et al. 1997), which automatically reduces the stepsize when the expected error exceeds machine precision ($\sim 10^{-16}$). Doing so, we acknowledge our inability to provide an accurate measure of this error, but expect $\Delta E/E$ to be of the order $\sim 10^{-16}$, the adaptive time-step limit.

6.2.2 The Aging Solar System

Since this study is focused on the long term effects of the outer solar system, we only include Jupiter, Saturn, Uranus and Neptune in our simulations. Although Mars is likely to survive the red giant phase of the Sun's evolution (Veras, 2016), its dynamical contribution is negligible and can be ignored in order to reduce computational costs. By focusing on the

[2]We find a slow increase in energy ($\Delta E/E$) $\sim 10^{-18}$ after a typical flyby interaction. This value is consistent with the finding of Li & Adams (2015), who did a similar flyby injection test. It is important to note that rare close encounter flybys have the ability to produce larger changes in the system energy.

outer giant planets, we are only limited by the orbital resolution of Jupiter. Furthermore, the orbital period of Jupiter changes as the Sun loses mass, increasing the orbital period by a factor of ~ 3.4 over the lifetime of the Sun. In order to take full advantage of this increased period, we break our simulation into two parts. The first part (Phase I) includes all of the stellar evolution, starting from the present epoch when the Sun is on the main-sequence, continuing through the red giant and mass-loss phases, and ending as the star becomes a white dwarf. The second part (Phase II) includes all of the subsequent temporal evolution, when the Sun has a fixed stellar mass and remains as a quiescent white dwarf.

The Phase I epoch extends from today to the epoch of final mass-loss experienced at the end of the Sun's planetary nebula phase. During this phase, 46% of the current solar mass will be ejected via stellar winds. To ensure that we appropriately account for the orbital effects due to this mass-loss, we implement the MESA (Paxton et al., 2019) solar mass evolution model with zero rotational velocity into our simulation. At each time-step ($\Delta t = 216.63$ days, corresponding to 1/20 the orbital period of Jupiter) we update the mass of the Sun to reflect the value suggested by this model. To ensure our mass-loss is smooth, we linearly interpolate the stellar mass between the output profiles provided by the MESA simulations. Throughout most of the Phase I interval, the Sun retains a majority of its current mass and the orbits of the giant planets remain static. In the final ~ 1 Myr of this period, however, the Sun loses $0.41 M_\odot$, and the semi-major axes (a) of the giant planets expand accordingly. As long as solar mass-loss occurs over many planetary orbits — which is expected to be the case — the orbits will experience adiabatic expansion (for a more in depth discussion of the adiabatic limit, see Veras et al. 2011).

The orbital expansion associated with solar mass-loss can be understood qualitatively as follows. A well-known result of classical perturbation theory (see e.g., Lichtenberg & Lieberman 1983) is that an oscillator, subjected to slow parametric changes, will preserve the ratio of its energy (E) to its frequency (n). By analogy, the adiabatic invariant associated with a

Keplerian orbit (J) has the form:

$$J = \mid \frac{E}{n} \mid \approx \frac{GM_{\star}m/2a}{\sqrt{GM_{\star}/a^3}} \propto \sqrt{GM_{\star}a} \approx \text{constant}, \tag{6.1}$$

which also corresponds to the first Poincaré action (see Morbidelli 2002). Therefore, reduction in stellar mass ($M_{\star}$) will inflate the semi-major axis according to $a \sim 1/M_{\star}$. To ensure this orbital expansion remains within the adiabatic limit, we have repeated this simulation, artificially slowing down the mass-loss rate (and hence orbital expansion) by a factor of 100, and find that the orbital parameters remain consistent with results obtained from the expected real-time expansion.

After the epoch of solar mass-loss concludes, and the Sun remains as a $0.54M_{\odot}$ white dwarf, we enter the second phase of our simulations. We now adjust the time-step to become 1/20th of the increased orbital period of Jupiter (P_{orb} = 40.6 yr, Δt = 740.56 days) at its new expanded orbit (a = 9.62 AU) and lower stellar mass. This time-step increase allows for consistent orbital resolution and speeds up the simulations by a factor of 3.4.

6.2.3 Stellar Flyby Encounters

Throughout the simulations, we introduce stellar flybys which perturb the system, at a rate consistent with the current galactic environment of the solar system. To simulate these flybys, we draw a 10,000 AU sphere around the solar system and introduce incoming stars as described below. Only flybys within this sphere are resolved in our calculation. The expected rate (Γ) for the solar system to encounter passing stars can be written in the form:

$$\Gamma = \langle n_{\star} \rangle \pi B^2 \langle v \rangle \,, \tag{6.2}$$

where $\langle n_{\star} \rangle$ is the local stellar number density (0.14 pc^{-3}; McKee et al. 2015; Bovy et al. 2012), B is the boundary of the interaction region (10,000 AU), and $\langle v \rangle$ is the expected local stellar velocity dispersion ($\sim$ 40 km/s; Binney & Tremaine 2008). With these parameter values, the expected encounter rate Γ is about 4.2×10^{-8} stars/year. In other words, we

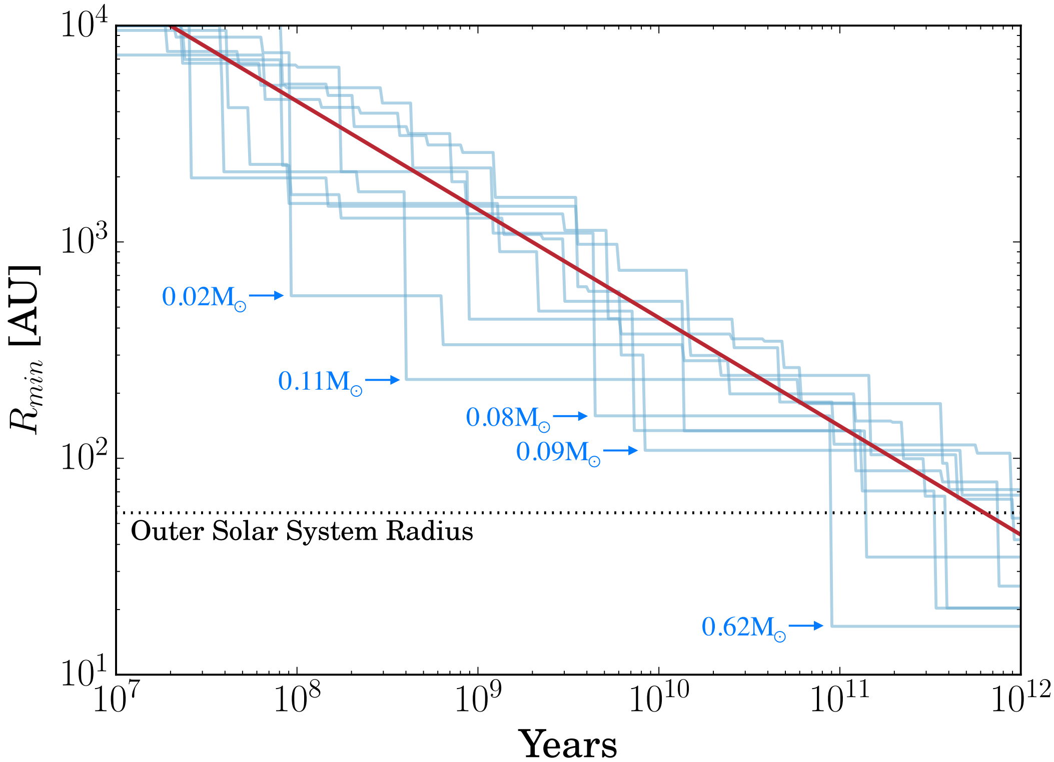

Figure 6.1: The closest approach of flybys as a function of time. The blue lines show the minimum distances as a function of time realized in the 10 simulations and the red line indicates the expected value from Equation (6.4). The outer solar system radius (60 AU) is represented by the semi-major axis of Neptune at the start of Phase II, when the Sun has reached its final (reduced) mass. The blue labels indicate the mass of the corresponding stellar encounter, highlighting that close encounters typically occur with low-mass stars.

expect a star to enter our sphere every 23 Myr. It is important to note that we use a static local galactic environment, which may change over the periods considered in the present study. Further discussion of this issue is provided in Section 6.3.5.2.

To simulate these flyby encounters we follow a procedure similar to that of Heisler & Tremaine (1986). We first randomly select stellar masses from the initial mass function (using the form advocated by Kroupa 2001) within a mass range of $0.08 - 1 M_{\odot}$[3] and assign a relative velocity ($v_{\rm inf}$) drawn from a Maxwell-Boltzmann distribution with a scale parameter $\langle v \rangle$. At each time-step, a new star is selected and the expected probability of encounter is calculated ($\Gamma \Delta t$; exchanging $v_{\rm inf}$ for $\langle v \rangle$) and an independently drawn random value ($R[0,1]$) indicates whether the star will enter the sphere or not. If the star is permitted to dynamically engage with the solar system, the inclination for the flyby is drawn from a distribution of $\arcsin(R[0,1])$. The angular orbital elements Ω and ω are uniformly drawn from a distribution of $R[-\pi/2, \pi/2]$. The impact parameter (b) of the flyby dictates the distance of closest approach (r_p) and is drawn from a distribution of $10000\sqrt{R[0,1]}$ AU, as needed for a uniform sampling of the cross sectional area. For a hyperbolic orbit, the perihelion r_p is related to the impact parameter according to $r_p = \sqrt{a^2 + b^2} - |a|$, where a is the semi-major axis of the encounter. In other words, the relationship between r_p and impact parameter b takes the form

$$b^2 = r_p^2 \left(1 + \frac{2G(M + \langle M' \rangle)}{r_p \langle v \rangle^2} \right) \approx r_p^2 \, , \tag{6.3}$$

where G is the gravitational constant, M is the solar mass, and $\langle M' \rangle$ is the expected mass of the flyby star. The gravitational focusing term represents a small correction (only $\sim 1\%$ for $r_p = 100$ AU) which is negligible for nearly all the encounters considered in this study and can be ignored.

To show that our method of randomly selecting stellar encounters produces the correct

[3]We select an upper bound of $1 M_{\odot}$ as a conservative estimate of the effects of possible stellar encounters. In practice, we found invoking this bound had little effect on the overall outcome of our simulations. Our results thus provide an upper limit, in that the inclusion of more massive stars would reduce the time needed for planetary disassociation.

distribution, we compare our simulations to the expected time τ at which a given minimum distance of closest approach (R_{min}) is achieved. The quantity R_{min} is thus the minimum value of the perihelion r_p experience by the solar system as a function of time. Using the simplification of equation (6.3), along with the relation $\tau = 1/\Gamma$, we can calculate R_{min} as a function of τ:

$$R_{min}(\tau) = [\langle n_\star \rangle \pi \langle v \rangle \tau]^{-1/2} \, . \tag{6.4}$$

Figure 6.1 presents the results of this comparison and shows that our sampling procedure replicates the expected time required to reach a given minimum distance of closest approach R_{min}.

Although we insert stellar flybys throughout the full integration, we find that they have little effect on the solar system during Phase I. The tightly packed planets are effectively immune to these distant perturbations during the Sun's (relatively short) remaining lifetime as a main-sequence star. For example, the cross section for changing the orbit of Neptune enough to create significant disruption (specifically, so that it crosses the orbit of Uranus) is of order 1000 AU^2 (Laughlin & Adams, 2000), which corresponds to a distance of closest approach $R_{min} \sim 20$ AU. Figure 6.1 shows that this value of R_{min} is not achieved until a time $t \geq 10^{12}$ yr, well beyond the time span of Phase I. Although more distant encounters are expected over the remainder of the Sun's main-sequence life, these perturbations will produce only small modifications to Neptune's eccentricity ($\Delta e \sim 0.01$). Moreover, these minor perturbations will not have a significant impact on the inner gas giants. As result, the remainder of the paper focuses on the second part (Phase II) of our simulations.

6.3 Results

We have carried out 10 simulations of the outer solar system's long-term evolution, with each run spanning 10^{12} years. While this ensemble of simulations does not constitute a large statistical sample, we find similar results in each case, indicating that the dynamical picture attained here is representative. In this section we discuss the findings of this study. In addition to determining the timescales for planetary ejection, we also elucidate the mechanisms that cause the solar system to dissolve.

6.3.1 Ejected planets

After the Sun has completed its stellar evolution, including mass loss, the solar system will remain stable with the remaining planets orbiting with semi-major axes 1.85 times larger than their current values. As shown in Figure 6.2, the eccentricities of all the planets remain low ($e \leq 0.2$) during the first 10 Gyr of the Phase II era. This finding is consistent with the results of Duncan & Lissauer (1998), who found no significant eccentricity growth during the initial onset of the Sun's white dwarf phase. However, our results begin to deviate beyond this 10 Gyr timescale.[4] As depicted in Figure 6.1, the occurrence of a stellar flyby with a perihelion less than 500 AU is likely within a 10 Gyr period, and such an encounter provides a significant perturbation to the system. Figure 6.2 indicates that these stellar encounters can significantly increase the eccentricity of the planets, and even lead to the complete disassociation for many of the planets before the 45 Gyr benchmark for stability reported by Duncan & Lissauer (1998).

In all 10 of our simulations, the four gas giants are ejected from the solar system within 10^{12} years, following the end of solar mass-loss. Figure 6.3 presents the times at which each planet was removed from each of the simulations. The overall average ejection time is roughly 65 Gyr after mass-loss (72 Gyr from today). This timescale is far shorter than the 10^{18} yr lower bound predicted by Murray & Holman (1999) for internal instability, and shorter than the timescale of 10^{14} yr predicted for external perturbations with a compact configuration (Dyson, 1979; Adams & Laughlin, 1997). Moreover, if we only consider the first planet ejected, we find an average ejection time of about 30 Gyr. In contrast, the last

[4] Duncan & Lissauer (1998) acknowledged that calculations beyond 10 Gyr would require a more careful accounting of external stellar encounters, as carried out in this present study.

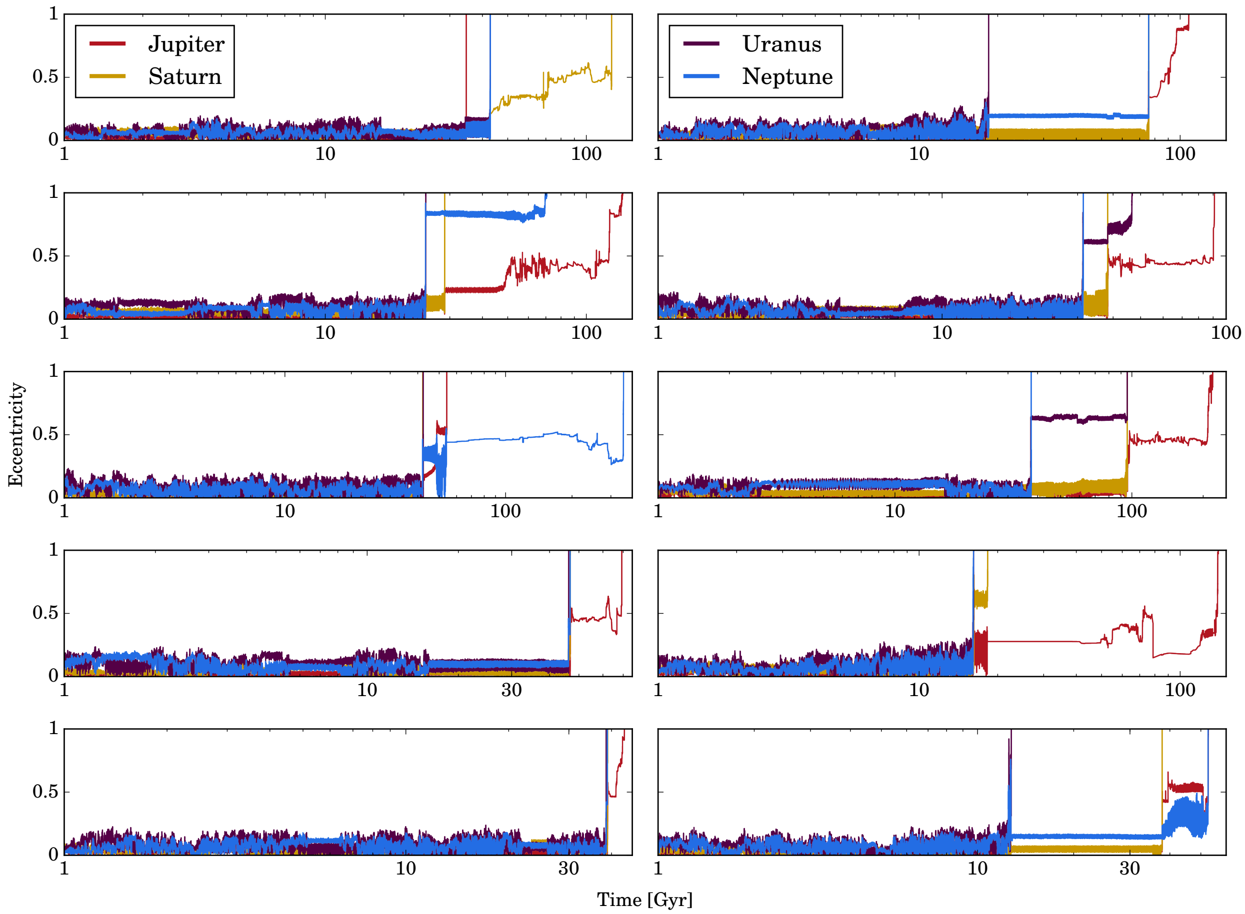

Figure 6.2: Evolution of the outer solar system after solar mass loss. Eccentricity of the gas giants is shown as a function of time. Given sufficient time, perturbations from a stellar encounter will drive the outer solar system planets unstable, ejecting all but one planet. A additional close stellar encounter is needed to remove the final planet. The time at which each planet is dissociated from the system is plotted in Figure 6.3.

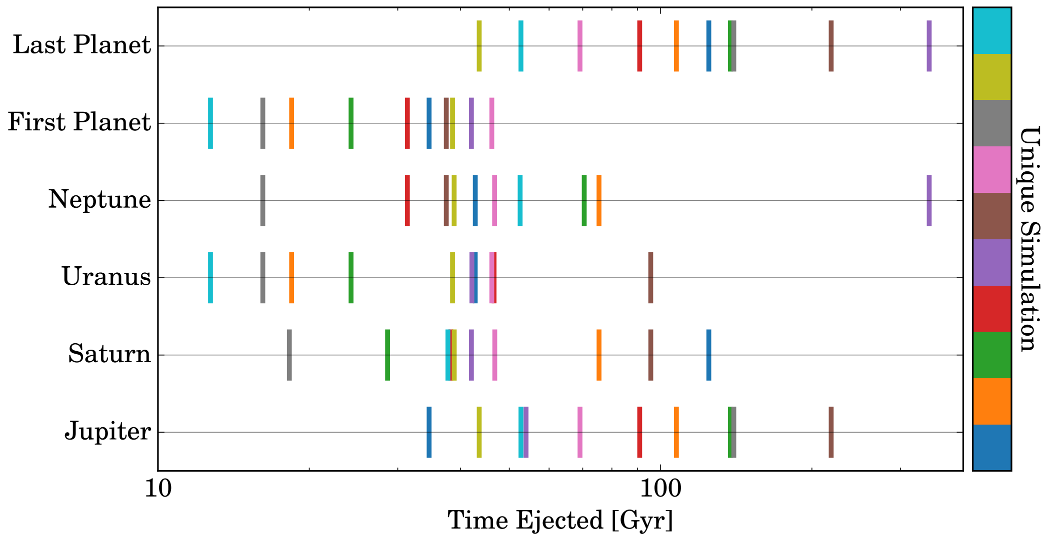

Figure 6.3: The ejection time for each of the gas giants, after the Sun has become a white dwarf. We represent each of our 10 simulations with a different color. The First and Last Planet columns show the time at which the first and last planet was ejected from the system in a given simulation. In all cases the four gas giants were removed before the 10^{12} year limit of this study.

planet is ejected (on average) at a time of ~ 100 Gyr. There is no definitive order in which the planets get removed, but typically the ice giants are removed first, with Uranus's ejection followed by Neptune, Saturn, and Jupiter, respectively. Usually the first three planets are all expelled within 5 Gyr of the first ejection. The remaining planet will then linger for an additional 50 Gyr before being removed from the system (see Section 6.3.4 for further discussion of this final planet). By accounting for the expanded Phase II planetary orbits and external perturbations from stellar flybys, we thus find a significantly reduced expected lifetime for planets to remain bound to the Sun. In the following sections, we discuss the mechanisms that drive this dissolution of the solar system.

6.3.2 Baseline Stability Considerations

Arguably the simplest way to characterize the stability of a planetary system is to require sufficient separation in units of mutual Hill Radii (Chambers et al., 1996). This condition is often written in the form

$$a_2 > a_1 + AR_H \,, \tag{6.5}$$

where a_k are the semi-major axes of adjacent planets and A is a dimensionless constant that depends on the specific architecture of the system. Typically, one requires $A \geq 10$ for stability of multiplanet systems (Pu & Wu, 2016), whereas smaller values are applicable for two-planet systems (Petit et al., 2018). The mutual Hill Radius R_H is given by

$$R_H = \left(\frac{a_1 + a_2}{2}\right)\left(\frac{m_1 + m_2}{3M_\star}\right)^{1/3} \,, \tag{6.6}$$

where the m_k denote the masses of adjacent planets and $M_\star$ is the mass of the Sun. As already discussed above, $M_\star$ varies over the history of the solar system, from $M_\star = 1M_\odot$ at the present epoch down to $M_\star = 0.54M_\odot$ after mass-loss. For a given pair of adjacent planets, we can thus define a dimensionless parameter Δ that provides a measure of system stability, i.e.,

$$\Delta \equiv \frac{2(a_2 - a_1)}{a_2 + a_1}\left(\frac{3M_\star}{m_2 + m_1}\right)^{1/3} \,. \tag{6.7}$$

During the epoch of mass-loss, we expect the system to remain in the adiabatic regime so that $aM_\star \approx$ constant as the Sun loses mass. As a result, the leading coefficient is invariant and the stability parameter scales according to $\Delta \sim M_\star^{1/3}$. As the stellar mass decreases, the stability parameter also decreases, and the system becomes more unstable.

Applying the expected changes in stellar mass over the course of the Sun's lifetime, one finds the stability factor (Δ) is reduced from today's value of 8 to 6 for the Jupiter/Saturn orbital spacing. Likewise, Δ is reduced from 14 to 11 for both the Saturn/Uranus and the Uranus/Neptune stability pairing.

6.3.3 Jupiter and Saturn Resonance

Although the above discussion indicates that planet-planet interactions are expected to grow stronger due to solar mass-loss, the actual source of large-scale instability remains to be identified. Remarkably, our simulations suggest that Jupiter and Saturn's 5:2 near-commensurability may provide the mechanism that triggers this large-scale instability.

To make this argument, we consider the resonance angle $\phi = 5\lambda_{\mathrm{Saturn}} - 2\lambda_{\mathrm{Jupiter}} - 3\varpi_{\mathrm{Saturn}}$, where λ is the mean longitude and ϖ is the longitude of pericentre for the respective orbits. Slow circulation of this angle indicates planets are near (but not in) mean-motion resonance (MMR). In other words, if the planets are not in MMR, the conjunction position will continuously move along the orbit. This behavior is characteristic of Jupiter and Saturn's present-day configuration. In the 'Today' panel of Figure 6.4, we show the circulation of the Jupiter/Saturn resonant angle as seen today. The period of this circulation is about 900 years and is directly associated with the modulation in semi-major axes known as the "Great Inequality" (see Section 6.4.1 for further discussion).

As the Sun loses mass, the semi-major axis ratio is conserved, but the width of the 5:2 MMR expands as $\sim M_\star^{-1/2}$ (Henrard et al., 1986). This growth leads to the adiabatic capture of Jupiter and Saturn into the 5:2 MMR. In such a capture, the resonant angle will transition from circulation to libration. Our simulations indicate this transition occurs during the final $\sim$ 1 Myr of mass-loss (see the 'During Mass-Loss' panel of Figure 6.4). Once the planets have been successfully captured, the resonant angle will execute bounded oscillations (see the 'After Mass-Loss' panel of Figure 6.4). In isolation, this orbital configuration is stable on long timescales, as discussed in Duncan & Lissauer (1998). However, the inflated orbits and weakened gravitational pull from the reduced solar mass render this new configuration more vulnerable to perturbations from stellar flybys.

The large amplitude ($\sim 100°$) of the libration seen in the 'After Mass-Loss' panel of Figure

6.4 is indicative of a MMR system near the separatrix (i.e., the MMR boundary, where the resonant angle changes from libration to circulation). For roughly 30 Gyr after solar mass-loss, the system remains stable. Given sufficient time, however, a close encounter ($R_{\rm min} < 500$ AU) from a stellar flyby will create a perturbation large enough to perturb Jupiter and Saturn into the chaotic region of the 5:2 MMR (see Morbidelli 2002). This event triggers chaotic diffusion as shown in the 'After Perturbation' panel of Figure 6.4, leading to large-scale instabilities in the outer solar system. In most cases the Jupiter/Saturn resonance angle will repeatedly switch from circulation to libration and back, pumping up the eccentricity of Uranus, Neptune, and Saturn until they are ejected from the solar system. This process takes place over $\sim$ 10 Gyr after the onset of accelerated chaos. Jupiter is usually the last planet remaining, but this ordering is not the only possible outcome. In one case Saturn was the final remaining planet and in another Neptune survived the tumultuous instability of the inner gas giants. However this large-scale instability played out, one planet remains orbiting the Sun for an extended period of time in most of the simulations.

6.3.4 The Last Planet Standing

In all but one simulation, we found that a single planet remains orbiting the Sun for about 50 Gyr. (In the exceptional case, orbital diffusion left Jupiter with an eccentricity greater than $e \sim 0.9$, leading to a swift ejection, 200 Myr later.) With an absence of additional planets, the surviving planet lacks a direct mechanism to attain positive energy. The only remaining source of energy exchange is through interactions with passing stars. Significantly, the large-scale instability that led to the ejection of the other three gas giants leaves the final planet with a heightened eccentricity (typically in the range $e \sim 0.2 - 0.5$). As shown by Li & Adams (2015), the dynamical cross-section required for ejection is an exponential function of eccentricity. As a result, the expected timescale for ejection of the post-instability gas giant is decreased by roughly a factor of two (compared to the planetary orbit before the onset of instability).

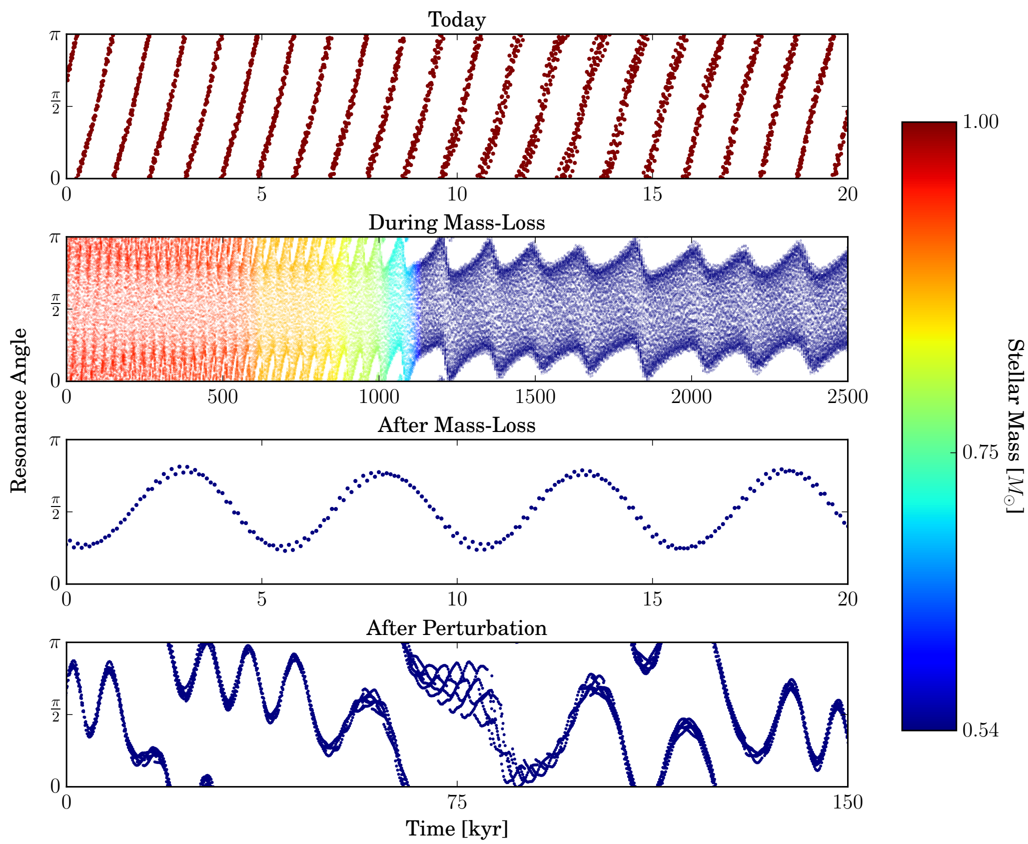

Figure 6.4: The 5:2 mean-motion resonance (MMR) angle for Saturn and Jupiter ($5\lambda_{\mathrm{Saturn}} - 2\lambda_{\mathrm{Jupiter}} - 3\varpi_{\mathrm{Saturn}}$) as a function of time. The data is colored based on the mass of the Sun at that given time-step. In the **Today** panel we show how the current angle circulates with a roughly 900 year period. In the **During Mass-Loss** panel we show the adiabatic capture of Jupiter and Saturn into a 5:2 MMR. The **After Mass-Loss** panel shows the libration that occurs once the planets find their new expanded orbits after the solar mass-loss has been completed. In the **After Perturbation** panel we provide a sample of the chaotic circulation that transpires after being perturbed by a stellar flyby. This period of chaotic motion was followed by the ejection of Uranus 8 Myr later.

Since flyby encounters are rare (entering the 10,000 AU sphere once every 23 Myr), and most interactions will have small dynamical effects on the remaining planet, the process of ejection can in principle occur steadily (e.g., through incremental increases in orbital eccentricity and semi-major axes). On the other hand, given sufficient time, it is also possible that an extremely close encounter will independently liberate the final planet. The underlying mechanism for the removal of the final planet thus represents a competition between these two processes. In other words, will the final planet be ejected by a single major event or many small energy exchanges?

To intuitively understand the timescales over which these possible outcomes take place, we can crudely describe the process as a random walk. The follow derivation is merely an order of magnitude calculation, similar to the dynamical relaxation time derived by Binney & Tremaine (2008). The cumulative change in velocity of ΔU is given by

$$\Delta U \sim \Delta u \sqrt{N}\,, \tag{6.8}$$

where Δu is the change in the velocity of the orbiting planet from a single interaction and N represents the number of interactions. The planet must attain a positive energy in order to be disassociated from the system. To leading order, this requirement can be written in the form

$$\frac{(\Delta U)^2}{2} \approx \frac{GM_\star}{2a}. \tag{6.9}$$

Assuming all interactions follow the impulse approximation, we can express Δu as:

$$\Delta u = \frac{2GM_\star}{vb}\,, \tag{6.10}$$

where v is the velocity of the flyby star and b is the impact parameter for the stellar encounter. In reality, most interactions are much weaker and produce smaller changes in the planets orbital velocity. As a result, this calculation will provide a limiting constraint. Recalling Equation (6.2), we note that the number of interactions can be calculated as $N = \Gamma t$, where t is the expected time needed to achieve N interactions. Finally, we can use Equation (6.8) to solve for the time required to eject the planet, i.e.,

$$t = \frac{v^2}{4\pi G M_\star a \langle n_\star \rangle \langle v \rangle} \,. \tag{6.11}$$

Note that this equation is independent of the flyby impact parameter.[5] Under the assumption that the impulse approximation is valid, we find that a single planet is equally likely to be ejected by a series of distant encounters or by a single close encounter flyby. However, this approximation breaks down for adiabatic interactions, where the the gravitational interaction timescale ($T_{\rm enc}$) is much greater than the planet's orbital period (P),

$$T_{\rm enc} \sim \frac{2b}{v} \gg P. \tag{6.12}$$

In typical cases, the star will enter the 10,000 AU sphere of influence with a velocity of order 40 km/s and an impact parameter of order 7000 AU. These values indicate an interaction time $T_{\rm enc} \sim 1800$ yr, which is nearly five times greater than Neptune's Phase II period (350 years). Most interactions that occur after the large-scale instability has isolated a single planet will thus be adiabatic. In these cases, where perturbations are effectively secular, semi-major axis growth will be suppressed (Batygin et al., 2020) relative to predictions from the impulse approximation. In other words, the distant encounters will be weaker than the limiting case of Equation (6.10). As a result, a single extreme encounter is more likely to liberate the final planet than the accumulated effects of many distant perturbers. Nonetheless, both the cumulative effects of many distant encounters and the impact of rare close encounters are likely to play a role.

In Figure 6.5 we show how the eccentricity and semi-major axis of our simulated final planet changes before eventual disassociation. From the onset of isolation, nearly all cases show slow orbital diffusion due to weak gravitational interactions. Providing consistency with the above discussion, four of the simulations show the eventual disassociation of the final planet by a

[5] A similar calculation can be achieved through a random walk process of the orbital energy, under the assumption of the impulse approximation, culminating in a solution that deviates by only a factor of 4 from Equation 6.11.

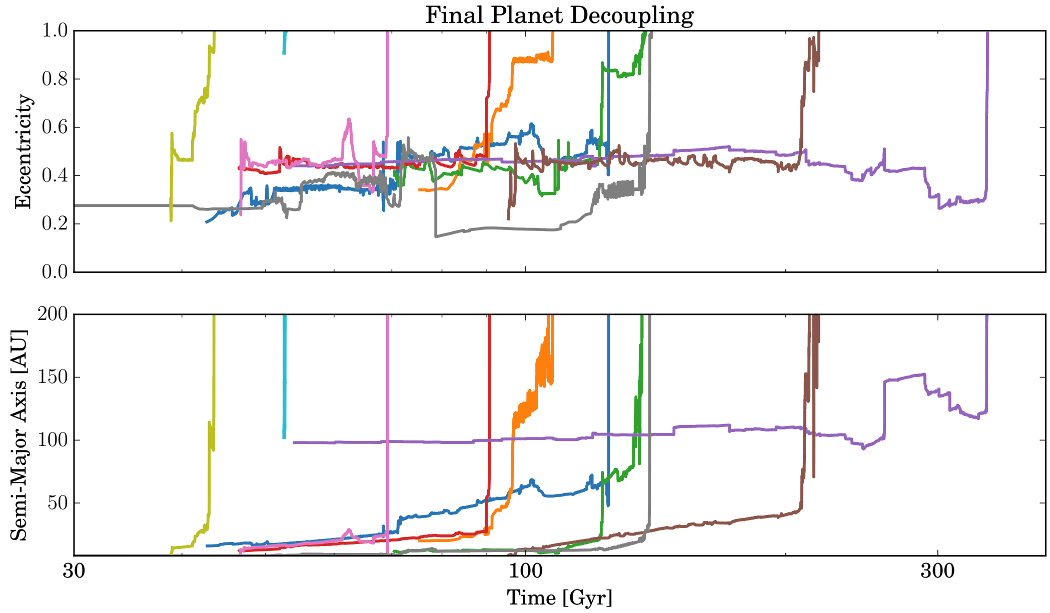

Figure 6.5: The eccentricities and semi-major axes of the final planet, which remains bound after all of the other planets have been ejected from the system. The slow rise in eccentricity and semi-major axis are due to energy exchanges with multiple stellar flybys, but an important mechanism for disassociation of the final planet is an extremely close flyby encounter (with $R_{\rm min} \leq 200$ AU).

single extremely close flyby encounter (with $R_{\rm min} < 200$ AU). In the remaining simulations, a close encounter significantly modifies the orbit, making the planet far more susceptible to perturbations from subsequent stellar flybys. After surviving this initial interaction, the final planet is ejected within a few Gyr. In both scenarios, we find that both major and minor energy exchanges play a role in the removal of the final planet. However, the majority of the liberating energy comes from a single close encounter.

6.3.5 Caveats

6.3.5.1 Binary Encounters

Within the current study we assume all flybys are single star encounters, in reality, however, observations indicate that roughly half of all nearby stars are part of a binary system (e.g., Raghavan et al. 2010; Duchêne & Kraus 2013). Previous dynamical studies, considering these binary interactions, have found that the two stars will appear as independent perturbers when the relative flyby velocity (v) of the binary system is sufficiently large, increasing the interaction cross-section by roughly a factor of two (Laughlin & Adams, 2000; Li & Adams, 2015). However, when the orbital velocity of the interacting planet and the velocity between binary members are near the relative flyby velocity, the gravitational influence is enhanced. In this scenario, the extended interaction time allows motion from the binary system orbit to increase the gravitational cross-section (by a factor greater than 2), thereby increasing the likelihood of a significant perturbation.

The 40 km/s expected flyby velocity used for this study is quite large compared to the orbital speeds of the gas giants (Jupiter's orbital speed today is about 13 km/s and decreases to 9 km/s after solar mass-loss). Therefore, this effect would only be applicable for slow close encounter flybys, which alone would be completely disruptive to the orbiting planets. However, the shear number of binary systems in existence would increase the dynamical cross-section for the interactions. By choosing to exclude these binary encounters, we are making a conservative estimate for the lifetime of the future solar system. In other words, the effect of including binary flybys would further reduce this expected lifetime.[6]

[6] Note that if the binary fraction is 1/2, then the factor by which the cross section is increased would be 3/2.

6.3.5.2 Galactic Evolution

Over the timescales considered in this study, the solar system may undergo radial migration through the Galaxy, encountering regions of differing stellar density and velocity dispersion (Halle et al., 2015). However, the magnitude and direction of this migration remains an active area of discussion (e.g., Roškar et al. 2008 and Martínez-Barbosa et al. 2017). As one example, the latter authors find that the encounter rate could vary by a factor of ~ 3 if the solar system migrates outwards versus inwards. Continued star formation can also increase the stellar density. Acting in the opposite direction, the galactic disk tends to increase its velocity dispersion and hence its scale height over comparable timescales. In addition, the Milky Way is likely to collide with the Andromeda Galaxy over the next Hubble time, or two, again modifying the local galactic environment (Binney & Tremaine, 2008). These changes will impact the rate and velocity of stellar encounters, but accurately estimating these changes remains difficult and is beyond the scope of this present work. In this study, the current local stellar density and velocity dispersion are considered fixed throughout our simulations. One should keep in mind, however, that a more precise accounting of these future changes could modify our results.

6.4 Conclusion

Using a suite of long-term simulations of the solar system, that account for solar mass-loss and extrinsic forcing from passing stars, we have demonstrated that the expected dynamic lifetime of the outer planets is of order 100 Gyr, — significantly shorter than previous estimates. Moreover, we have identified the specific dynamical pathway responsible for the onset of the solar system's final large-scale instability. The narrative emerging from our calculations can be summarized as follows. As solar mass-loss unfolds, the planetary orbits expand adiabatically and maintain their period ratios. At the same time, mutual planet-planet interactions grow stronger for two reasons: The separation of planetary orbits in units of mutual Hill radii decreases, and the width of mean-motion resonances (MMR) expand in concert. The process culminates in the adiabatic capture of Jupiter and Saturn into the 5:2 MMR, giving way to a period of stable resonant motion, one that is characterized by large-amplitude librations of the associated critical angle. In time, however, stellar flybys perturb the giant planets onto the chaotic sub-domain of the 5:2 MMR. Correspondingly, orbital diffusion ensues, leading to eventual orbit crossings, and ejection of the outer planets. This process is responsible for the ejection of all but a single remaining planet, which continues to orbit with an eccentricity that has been excited by the aforementioned large-scale instability. The planet's increased eccentricity and expanded semi-major axis enhances the subsequent dynamical interactions due to stellar encounters, which lead to the expeditious disassociation of the final gas giant.

6.4.1 Historical Context

It is worth noting that speculation regarding large-scale instabilities within the solar system that are driven by the 5:2 resonant interaction between Jupiter and Saturn dates back to the work of Newton (1687, 1713, 1726), as well as the pioneering development of perturbation theory by Laplace (1799-1825). Accordingly, let us contextualize our results against the back-drop of this remarkable saga.

The earliest data on the orbits of Jupiter and Saturn date back to Ptolemy in 228BC (Saturn) and 240BC (Jupiter) and were not recorded again until 1590 (nearly 1800 years later) when Tycho Brahe measured their orbital positions (see the account of Laskar 1996). When Kepler set about mapping the elliptical orbits of our solar system, in 1625, he was unable to reconcile the data collected by Ptolemy, using the model derived from Tycho Brahe's observations. He noted that Ptolemy's observation required a slower mean motion for Jupiter and a faster mean motion for Saturn (Wilson, 1985). In an effort to understand this discrepancy, Halley (1687) extrapolated the semi-major axes of planets over this period of time and determined Jupiter was slowly moving inward while Saturn's orbit was moving outward (Laskar, 1996).

At face value, the data implied that given sufficient time Jupiter would collide with the Sun and Saturn would expand out into deep space.

When Newton (1687) announced the universal law of gravity, he indicated that such deviations from the invariant elliptical orbits were due to mutual gravitational interactions. His belief was that such perturbations, lacking divine intervention, would eventually lead to the demise of the solar system (Laskar, 1996). However, Newton was not able to directly explain these observations since perturbation theory had not yet been developed.

Laplace (1799-1825) finally explained this phenomenon, which became known as the "Great Inequality", by showing this was not a secular effect, but rather a periodic modulation of Jupiter and Saturn's semi-major axes by the 5:2 near resonance. As seen in the 'Today' panel of Figure 6.4, this nearly 900 year circulation leads to a cyclical expansion and contraction of the Jovian and Saturnian orbits, and appears to be stable for the remainder of the Sun's life as a main sequence star.

Our simulations show that the mechanism responsible for the eventual disintegration of the outer solar system is keenly related to the "Great Inequality". That is, the expanded orbits during Phase II (after solar mass-loss) allow the planets to be captured in a 5:2 MMR, producing large-scale instability when perturbed by stellar flybys. As a result, Jupiter and Saturn appear to be responsible for the ultimate demise of the solar system, only at a much later time than originally prophesied by Newton.

6.4.2 Free Floating Planets

Once liberated, the outer solar system planets will independently roam through the Galaxy, becoming Free Floating Planets (FFP). Current estimates, using gravitational microlensing, suggest there are less than about 0.25 FFPs for every main-sequence star in the Galaxy (Mróz et al., 2017). The exact origin of these abundant planets remains unclear, but it is probable that a large fraction of these FFPs were once bound to a host star and disassociated via dynamical instability (Ford et al., 2001).

Here, we have shown that the outer gas giants of the solar system will eventually be ejected and contribute to this reservoir of FFPs. Given that all stars will experience some amount of mass-loss over their lifetimes, it is likely that other planetary systems — that survive stellar evolution — will also eventually experience large-scale instability (e.g. Veras et al. 2011). We can therefore conclude that as the Galaxy, and the stars that reside in it, continue to age the number of FFPs will increase as a function of time.

CHAPTER 7

Conclusions

The *Kepler* mission identified nearly 3,000 exoplanet candidates. The remarkable abundance of these short period small planets, starkly contrasts the structure of the solar system. To understand why these planets are absent in our own system, we must investigate how these known exoplanet systems dynamically evolve and what underlying processes are responsible for their formation. We can use demographics to determine these processes by examining observed trends and using that data to constrain astrophysical formation models. This is the underlying goal that has driven my thesis.

I began by investigating the role that planet detection order plays in the *Kepler* planet detection pipeline to better understand the expected system architectures. The *Kepler* pipeline typically detects planets in order of descending signal strength. I found that the detectability of transits experiences an additional 5.5% and 15.9% efficiency loss, for periods < 200 days and > 200 days respectively, when detected after the strongest signal transit in a multiple-planet system. We provided a method for determining the transit probability for multiple-planet systems by marginalizing over the empirical *Kepler* data set. Furthermore, because detection efficiency appears to be a function of detection order, I discussed the sorting statistics that affect the radius and period distributions of each detection order. My population model is consistent with the results of Burke et al. (2015), but now includes an improved estimate of the multiplicity distribution. From my obtained model parameters, I found that only $4.0 \pm 4.6\%$ of Sun-like GK dwarfs harbor one planet. This excess is smaller than prior studies and can be well modeled with a modified Poisson distribution, suggesting that the *Kepler* Dichotomy can be accounted for by including the effects of multiplicity on detection efficiency. Using my modified Poisson model we expect the average number of planets to be 5.86 ± 0.18 planets per GK dwarf within the radius and period parameter space of *Kepler*.

Using these updated exoplanet population parameters, which includes the planetary radius updates from *Gaia* DR2 and an inferred multiplicity distribution, I provided a revised $\eta_{\oplus}$ calculation. This was achieved by drawing from the inferred population and determining which planets meet our criterion for habitability. My most optimistic criterion for habitability provided an $\eta_{\oplus}$ value of $0.32 \pm 0.02 \frac{\text{planets}}{\text{star}}$. I also considered the effects of multiplicity and the number of habitable planets each system may contain. My calculation indicates that $5.9 \pm 0.5\%$ of GK dwarfs have more than one planet within their habitable zone. This optimistic habitability criterion also suggests that $0.03 \pm 0.01\%$ of Sun-like stars will harbor 5 or more habitable planets. These tightly packed highly habitable system should be extremely rare, but still possible. Even with our most pessimistic criterion we still expect that $1.7 \pm 0.2\%$ of Sun-like stars harbor more than one habitable planet.

In an effort to expand the sample of useable planets for planet occurrence calculations beyond the *Kepler* sample, an additional homogeneous candidate catalog must be established using data from a uniform survey. I provided the first full *K2* transiting exoplanet sample, using data from Campaigns 1-8 and 10-18, derived through a fully automated procedure. I identified 723 unique planet candidates and 41 multi-planet systems. Of these candidates, 235 have not been previously identified, including two low stellar metallicity candidate hosts, one resonant multi-planet system, and one system with two short-period gas giants. By automating the construction of this list, measurements of sample biases (completeness and reliability) can be quantified. I carried out a light curve-level injection/recovery test of artificial transit signals and found a maximum completeness of 61%, a consequence of the significant detrending required for *K2* data analysis. Through this operation we attained measurements of the detection efficiency as a function of signal strength, enabling future population analysis using this sample. I assessed the reliability of my planet sample by testing my vetting software, `EDI-Vetter`, against inverted transit-free light curves. I estimated that 91% of my planet candidates are real astrophysical signals, increasing up to 96% when limited to the FGKM dwarf stellar population. The presented catalog, along with the completeness and reliability measurements, enable robust exoplanet demographic studies to be carried out across the fields observed by the *K2* mission for the first time.

Using a subset of this catalog, Campaign 5 (43 planets), which has corresponding measures of completeness and reliability, I infered an underlying planet population model for the FGK dwarfs sample (9,257 stars). Implementing a broken power-law for both the period and radius distribution, I found an overall planet occurrence of $1.00^{+1.07}_{-0.51}$ planets per star within a period range of 0.5–38 days. Making similar cuts and running a comparable analysis on the *Kepler* sample (2,318 planets; 94,222 stars), I found an overall occurrence of 1.10 ± 0.05 planets per star. Since the Campaign 5 field is nearly 120 angular degrees away from the *Kepler* field, this occurrence similarity offers evidence that the *Kepler* sample may provide a good baseline for Galactic inferences. Furthermore, the *Kepler* stellar sample is metal-rich compared to the *K2* Campaign 5 sample, thus a finding of occurrence parity may reduce the role of metallicity in planet formation. However, a weak (1.5σ) difference, in agreement with metal-driven formation, is found when assuming the *Kepler* model power-laws for the *K2* Campaign 5 sample and optimizing only the planet occurrence factor. This weak trend indicates further investigation of metallicity dependent occurrence is warranted.

Finally, the solar system in its current state can only provide constraints on formation, however, forecasting it's long-term stability, provides insight into the expected lifetime of a typical planetary system in our galaxy. I discussed the eventual fate the outer gas giants (Jupiter, Saturn, Uranus, and Neptune) after the Sun leaves the main sequence and completes its stellar evolution. Due to solar mass-loss – which is expected to remove roughly half of the star's mass – the orbits of the giant planets expand. This adiabatic process maintains the orbital period ratios, but the mutual interactions between planets and the width of mean-motion resonances (MMR) increase, leading to the capture of Jupiter and Saturn into a stable 5:2 resonant configuration. The expanded orbits, coupled with the large-amplitude librations of the critical MMR angle, make the system more susceptible to perturbations from stellar flyby interactions. Accordingly, within about 30 Gyr, stellar encounters perturb the planets onto the chaotic sub-domain of the 5:2 resonance, triggering a large-scale instability, which culminates in the ejections of all but one planet over the subsequent ~ 10 Gyr. After an additional ~ 50 Gyr, a close stellar encounter (with a perihelion distance less than ~ 200 AU) liberates the final planet. Through this sequence of events, the characteristic timescale over which the solar system will be completely dissolved is roughly 100 Gyr. My analysis thus indicates that the expected dynamical lifetime of the solar system is much longer than the current age of the universe, but is significantly shorter than previous estimates.

Bibliography

Adams F. C., 2019, MNRAS, 488, 1446

Adams F. C., Laughlin G., 1997, Reviews of Modern Physics, 69, 337

Adams F. C., Anderson K. R., Bloch A. M., 2013, MNRAS, 432, 438

Adams E. R., Jackson B., Endl M., 2016, AJ, 152, 47

Adibekyan V. Z., et al., 2012, A&A, 543, A89

Aigrain S., Rajpaul V., Buchhave L. A., 2017, MNRAS, 471, L125

Akeson R. L., et al., 2013, PASP, 125, 989

Anderson T. W., Darling D. A., 1952, The Annals of Mathematical Statistics, 23, 193

Babu G. J., Feigelson E. D., 2006, Astrostatistics: Goodness-of-Fit and All That!. p. 127

Ballard S., Johnson J. A., 2016, ApJ, 816, 66

Barbato D., Bonomo A. S., Sozzetti A., Morbidelli R., 2018, arXiv e-prints, p. arXiv:1811.08249

Barros S. C. C., Demangeon O., Deleuil M., 2016, A&A, 594, A100

Batalha N. M., et al., 2013, ApJS, 204, 24

Batygin K., Laughlin G., 2008, ApJ, 683, 1207

Batygin K., Morbidelli A., Holman M. J., 2015, ApJ, 799, 120

Batygin K., Adams F. C., Batygin Y. K., Petigura E. A., 2020, AJ, 159, 101

Beaugé C., Ferraz-Mello S., Michtchenko T. A., 2012, A&A, 12, 1044

Becker J. C., Vanderburg A., Adams F. C., Rappaport S. A., Schwengeler H. M., 2015, ApJ, 812, L18

Berger T. A., Huber D., Gaidos E., van Saders J. L., 2018, ApJ, 866, 99

Berger T. A., Huber D., Gaidos E., van Saders J. L., Weiss L. M., 2020a, arXiv e-prints, p. arXiv:2005.14671

Berger T. A., Huber D., van Saders J. L., Gaidos E., Tayar J., Kraus A. L., 2020b, AJ, 159, 280

Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition

Borucki W. J., 2016, Reports on Progress in Physics, 79, 036901

Borucki W. J., Summers A. L., 1984, Icarus, 58, 121

Borucki W. J., et al., 2010, Science, 327, 977

Borucki W. J., et al., 2011, ApJ, 736, 19

Boué G., Laskar J., Farago F., 2012, A&A, 548, A43

Bovy J., Rix H.-W., Liu C., Hogg D. W., Beers T. C., Lee Y. S., 2012, ApJ, 753, 148

Brakensiek J., Ragozzine D., 2016, ApJ, 821, 47

Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127

Brown T. M., Latham D. W., Everett M. E., Esquerdo G. A., 2011, AJ, 142, 112

Bryson S. T., et al., 2010, ApJ, 713, L97

Bryson S. T., et al., 2013, PASP, 125, 889

Bryson S., et al., 2020a, arXiv e-prints, p. arXiv:2010.14812

Bryson S., Coughlin J., Batalha N. M., Berger T., Huber D., Burke C., Dotson J., Mullally S. E., 2020b, AJ, 159, 279

Burke C. J., Catanzarite J., 2017, Technical report, Planet Detection Metrics: Per-Target Flux-Level Transit Injection Tests of TPS for Data Release 25

Burke C. J., et al., 2015, ApJ, 809, 8

Caldwell D. A., Coughlin J. L., 2020, Technical report, K2 Data Release Notes

Campbell B., 1983, PASP, 95, 577

Catanzarite J., Shao M., 2011, ApJ, 738, 151

Chambers J., Wetherill G., Boss A., 1996, Icarus, 119, 261

Chen J., Kipping D., 2017, ApJ, 834, 17

Chiang E. I., Fischer D., Thommes E., 2002, ApJ, 564, L105

Christiansen J. L., 2017, Technical report, Planet Detection Metrics: Pixel-Level Transit Injection Tests of Pipeline Detection Efficiency for Data Release 25

Christiansen J. L., et al., 2012, PASP, 124, 1279

Christiansen J. L., et al., 2013, ApJS, 207, 35

Christiansen J. L., et al., 2015, ApJ, 810, 95

Christiansen J. L., et al., 2020, arXiv e-prints, p. arXiv:2010.04796

Ciardi D. R., Fabrycky D. C., Ford E. B., Gautier T. N. I., Howell S. B., Lissauer J. J., Ragozzine D., Rowe J. F., 2013, ApJ, 763, 41

Ciardi D. R., Beichman C. A., Horch E. P., Howell S. B., 2017, ApJ, 834, 96

Claret A., Bloemen S., 2011, A&A, 529, A75

Cloutier R., Menou K., 2020, AJ, 159, 211

Coughlin J. L., 2017, Technical report, Planet Detection Metrics: Robovetter Completeness and Effectiveness for Data Release 25

Coughlin J. L., et al., 2016, ApJS, 224, 12

Crossfield I. J. M., et al., 2016, ApJS, 226, 7

Cumming A., Butler R. P., Marcy G. W., Vogt S. S., Wright J. T., Fischer D. A., 2008, PASP, 120, 531

Dattilo A., et al., 2019, AJ, 157, 169

David H. A., Nagaraja H. N., 2003, Order Statistics

Dawson R. I., Lee E. J., Chiang E., 2016, ApJ, 822, 54

Dong S., Zhu Z., 2013, ApJ, 778, 53

Dong S., et al., 2014, ApJ, 789, L3

Dressing C. D., Charbonneau D., 2013, ApJ, 767, 95

Dressing C. D., Charbonneau D., 2015, ApJ, 807, 45

Dressing C. D., et al., 2017a, AJ, 154, 207

Dressing C. D., et al., 2017b, AJ, 154, 207

Duchêne G., Kraus A., 2013, ARAA, 51, 269

Dumusque X., et al., 2014, ApJ, 789, 154

Duncan M. J., Lissauer J. J., 1998, Icarus, 134, 303

Dyson F. J., 1979, Reviews of Modern Physics, 51, 447

Fang J., Margot J.-L., 2012, ApJ, 761, 92

Fischer D. A., Valenti J., 2005, ApJ, 622, 1102

Ford E. B., Havlickova M., Rasio F. A., 2001, Icarus, 150, 303

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Foreman-Mackey D., Hogg D. W., Morton T. D., 2014, ApJ, 795, 64

Foreman-Mackey D., Barentsen G., Barclay T., 2019, dfm/exoplanet: exoplanet v0.1.5, doi:10.5281/zenodo.2587222

Fressin F., et al., 2013b, ApJ, 766, 81

Fressin F., et al., 2013a, ApJ, 766, 81

Fulton B. J., et al., 2017, AJ, 154, 109

Furlan E., et al., 2017, AJ, 153, 201

Gaia Collaboration et al., 2018, A&A, 616, A1

Gauss C. F., 1809

Gillon M., et al., 2016, Nature, 533, 221

Goldreich P., Schlichting H. E., 2014, AJ, 147, 32

Goodman J., Weare J., 2010, Communications in Applied Mathematics and Computational Science, Vol.~5, No.~1, p.~65-80, 2010, 5, 65

Guo X., Johnson J. A., Mann A. W., Kraus A. L., Curtis J. L., Latham D. W., 2017, ApJ, 838, 25

Gupta A., Schlichting H. E., 2019, MNRAS, 487, 24

Guzzo M., 2005, Icarus, 174, 273

Guzzo M., 2006, Icarus, 181, 475

Halle A., Di Matteo P., Haywood M., Combes F., 2015, A&A, 578, A58

Hansen B. M. S., 2009, ApJ, 703, 1131

Hansen B. M. S., 2017, MNRAS, 467, 1531

Hansen B. M. S., Murray N., 2013, ApJ, 775, 53

Hardegree-Ullman K. K., Cushing M. C., Muirhead P. S., Christiansen J. L., 2019, AJ, 158, 75

Hardegree-Ullman K. K., Zink J. K., Christiansen J. L., Dressing C. D., Ciardi D. R., Schlieder J. E., 2020, ApJS, 247, 28

Hayden M. R., et al., 2015, ApJ, 808, 132

He M. Y., Ford E. B., Ragozzine D., 2019, MNRAS, 490, 4575

Heisler J., Tremaine S., 1986, Icarus, 65, 13

Henden A. A., Templeton M., Terrell D., Smith T. C., Levine S., Welch D., 2016, VizieR Online Data Catalog, p. II/336

Henrard J., Lemaitre A., Milani A., Murray C. D., 1986, Celestial Mechanics, 38, 335

Hill M. L., Kane S. R., Duarte E. S., Kopparapu R. K., Gelino D. M., Wittenmyer R. A., 2018, ApJ, 860, 67

Hippke M., David T. J., Mulders G. D., Heller R., 2019, arXiv e-prints, p. arXiv:1906.00966

Holczer T., et al., 2016, ApJS, 225, 9

Horch E. P., Howell S. B., Everett M. E., Ciardi D. R., 2014, ApJ, 795, 60

Howard A. W., et al., 2010, Science, 330, 653

Howard A. W., et al., 2012, ApJS, 201, 15

Howell S. B., et al., 2014, PASP, 126, 398

Hsu D. C., Ford E. B., Ragozzine D., Morehead R. C., 2018, AJ, 155, 205

Hsu D. C., Ford E. B., Ragozzine D., Ashby K., 2019, AJ, 158, 109

Huang S.-S., 1959, American Scientist, 47, 397

Huber D., et al., 2013, Science, 342, 331–334

Huber D., et al., 2016, ApJS, 224, 2

Jeans J. H., 1924, MNRAS, 85, 2

Jenkins J. M., 2002, ApJ, 575, 493

Jenkins J. M., Caldwell D. A., Borucki W. J., 2002, ApJ, 564, 495

Jenkins J. M., et al., 2010a, ApJ, 713, L87

Jenkins J. M., et al., 2010b, ApJ, 713, L120

Johansen A., Davies M. B., Church R. P., Holmelin V., 2012, ApJ, 758, 39

Johnson J. A., et al., 2017, VizieR Online Data Catalog, p. J/AJ/154/108

Kasting J. F., Whitmire D. P., Reynolds R. T., 1993, Icarus, 101, 108

Kipping D. M., 2013, MNRAS, 434, L51

Koch D. G., et al., 2010, ApJ, 713, L131

Kolmogorov A. N., 1933, Giornale dell'Istituto Italiano degli Attuari, 4, 83

Kopparapu R. K., 2013, ApJ, 767, L8

Kopparapu R. K., et al., 2013, ApJ, 765, 131

Kopparapu R. K., Ramirez R. M., SchottelKotte J., Kasting J. F., Domagal-Goldman S., Eymet V., 2014, ApJ, 787, L29

Kostov V. B., et al., 2019, AJ, 157, 124

Kreidberg L., 2015, PASP, 127, 1161

Kroupa P., 2001, MNRAS, 322, 231

Kruse E., Agol E., Luger R., Foreman-Mackey D., 2019, ApJS, 244, 11

Kunimoto M., Matthews J. M., 2020, AJ, 159, 248

Lagrange J.-L., 1776, Sur l'altération des moyens mouvements des planètes, Mémoires de l' Académie de Berlin, 199 Oeuvres complétes VI 255. Paris, Gauthier-Villars (1869)

Laplace P. S., 1799-1825, Trait é de M écanique C éleste (Paris: Gauthier-Villa)

Laskar J., 1989, Nature, 338, 237

Laskar J., 1994, A&A, 287, L9

Laskar J., 1996, Celestial Mechanics and Dynamical Astronomy, 64, 115

Laskar J., 2008, Icarus, 196, 1

Laskar J., 2012, arXiv e-prints, p. arXiv:1209.5996

Laskar J., Gastineau M., 2009, Nature, 459, 817

Laughlin G., Adams F. C., 2000, Icarus, 145, 614

Le Verrier U., 1856, Annales de l'Observatoire de Paris, Vol. 2 (Paris: Mallet-Bachelet)

Li G., Adams F. C., 2015, MNRAS, 448, 344

Lichtenberg A. J., Lieberman M. A., 1983, Regular and stochastic motion

Lincowski A. P., Meadows V. S., Crisp D., Robinson T. D., Luger R., Lustig-Yaeger J., Arney G. N., 2018, ApJ, 867, 76

Lissauer J. J., et al., 2011, ApJS, 197, 8

Lissauer J. J., et al., 2012, ApJ, 750, 112

Lissauer J. J., et al., 2014, ApJ, 784, 44

Lithwick Y., Wu Y., 2011, ApJ, 739, 31

Livingston J. H., et al., 2018, AJ, 156, 78

Lopez E. D., Fortney J. J., 2013, ApJ, 776, 2

Loyd R. O. P., Shkolnik E. L., Schneider A. C., Richey-Yowell T., Barman T. S., Peacock S., Pagano I., 2020, ApJ, 890, 23

Luger R., Barnes R., 2015, Astrobiology, 15, 119

Luger R., Agol E., Kruse E., Barnes R., Becker A., Foreman-Mackey D., Deming D., 2016, AJ, 152, 100

Luger R., Kruse E., Foreman-Mackey D., Agol E., Saunders N., 2018, AJ, 156, 99

MacDonald M. G., 2019, MNRAS, 487, 5062

Mandel K., Agol E., 2002, ApJ, 580, L171

Martínez-Barbosa C. A., Jílková L., Portegies Zwart S., Brown A. G. A., 2017, MNRAS, 464, 2290

Mathur S., et al., 2017, ApJS, 229, 30

Matson R. A., Howell S. B., Horch E. P., Everett M. E., 2018, AJ, 156, 31

Mayo A. W., et al., 2018, AJ, 155, 136

Mayor M., Queloz D., 1995, Nature, 378, 355

Mayor M., et al., 2011, arXiv e-prints, p. arXiv:1109.2497

McCauliff S. D., et al., 2015, ApJ, 806, 6

McKee C. F., Parravano A., Hollenbach D. J., 2015, ApJ, 814, 13

McQuillan A., Mazeh T., Aigrain S., 2013, ApJ, 775, L11

Millholland S., Wang S., Laughlin G., 2017, ApJ, 849, L33

Morbidelli A., 2002, Modern celestial mechanics : aspects of solar system dynamics

Moriarty J., Ballard S., 2016, ApJ, 832, 34

Morton T. D., 2015, isochrones: Stellar model grid package (ascl:1503.010)

Morton T. D., Winn J. N., 2014, ApJ, 796, 47

Morton T. D., Bryson S. T., Coughlin J. L., Rowe J. F., Ravichandran G., Petigura E. A., Haas M. R., Batalha N. M., 2016, ApJ, 822, 86

Mróz P., et al., 2017, Nature, 548, 183

Mulders G. D., Pascucci I., Apai D., 2015, ApJ, 814, 130

Mulders G. D., Pascucci I., Apai D., Ciesla F. J., 2018, AJ, 156, 24

Mulders G. D., Mordasini C., Pascucci I., Ciesla F. J., Emsenhuber A., Apai D., 2019, ApJ, 887, 157

Mullally F., 2017, Technical report, Planet Detection Metrics: Automatic Detection of Background Objects Using the Centroid Robovetter

Mullally F., et al., 2015, ApJS, 217, 31

Mullally F., Coughlin J. L., Thompson S. E., Christiansen J., Burke C., Clarke B. D., Haas M. R., 2016, PASP, 128, 074502

Munoz Romero C. E., Kempton E. M. R., 2018, AJ, 155, 134

Murray N., Holman M., 1999, Science, 283, 1877

Newton I., 1687, Philosophiae Naturalis Principia Mathematica. Auctore Js. Newton, 1 edn, doi:10.3931/e-rara-440.

Newton I., 1713, Philosophiae Naturalis Principia Mathematica. Auctore Js. Newton, 2 edn

Newton I., 1726, Philosophiae Naturalis Principia Mathematica. Auctore Js. Newton, 3 edn

Nordström B., et al., 2004, A&A, 418, 989

Onken C. A., et al., 2019, PASA, 36, e033

Owen J. E., Campos Estrada B., 2019, MNRAS, 491, 5287–5297

Owen J. E., Wu Y., 2013, ApJ, 775, 105

Owen J. E., Wu Y., 2017, ApJ, 847, 29

Paxton B., et al., 2019, ApJS, 243, 10

Pepper J., Gould A., Depoy D. L., 2003, Acta Astron., 53, 213

Petigura E. A., Howard A. W., Marcy G. W., 2013a, PNAS, 110, 19273

Petigura E. A., Howard A. W., Marcy G. W., 2013b, PNAS, 110, 19273

Petigura E. A., et al., 2017, AJ, 154, 107

Petigura E. A., et al., 2018, AJ, 155, 21

Petit A. C., Laskar J., Boué G., 2018, A&A, 617, A93

Pettitt A. N., 1976, Biometrika, 63, 161

Poincaré H., 1892, New Methods of Celestial Mechanics, Vol.1 (Paris: Gauthier-Villarslla)

Pollack J. B., Hubickyj O., Bodenheimer P., Lissauer J. J., Podolak M., Greenzweig Y., 1996, Icarus, 124, 62

Pope B. J. S., Parviainen H., Aigrain S., 2016, MNRAS, 461, 3399

Pu B., Wu Y., 2016, ApJ, 819, 170

Quinn T., Katz N., Stadel J., Lake G., 1997, arXiv e-prints, pp astro–ph/9710043

Raghavan D., et al., 2010, ApJS, 190, 1

Ragozzine D., Holman M. J., 2010, arXiv e-prints, p. arXiv:1006.3727

Ramirez R. M., Kaltenegger L., 2014, ApJ, 797, L25

Raymond S. N., Scalo J., Meadows V. S., 2007, ApJ, 669, 606

Read M. J., Wyatt M. C., Triaud A. H. M. J., 2017, MNRAS, 469, 171

Rein H., Liu S. F., 2012, A&A, 537, A128

Rein H., Spiegel D. S., 2015, MNRAS, 446, 1424

Rogers L. A., 2015, ApJ, 801, 41

Roškar R., Debattista V. P., Quinn T. R., Stinson G. S., Wadsley J., 2008, ApJ, 684, L79

Rowe J. F., et al., 2014, ApJ, 784, 45

Rowe J. F., et al., 2015, ApJS, 217, 16

Rushby A. J., Johnson M., Mills B. J., Watson A. J., Claire M. W., 2018, Astrobiology, 18, 469

Rybicki K. R., Denis C., 2001, Icarus, 151, 130

Sackmann I. J., Boothroyd A. I., Kraemer K. E., 1993, ApJ, 418, 457

Sagan C., 1994, Pale blue dot : a vision of the human future in space Sagan.

Salvatier J., Wiecki T., Fonnesbeck C., 2015, arXiv e-prints, p. arXiv:1507.08050

Sanchis-Ojeda R., Rappaport S., Winn J. N., Levine A., Kotson M. C., Latham D. W., Buchhave L. A., 2014, ApJ, 783, 64

Santos N. C., Israelian G., Mayor M., 2004, A&A, 415, 1153

Schmitt J. R., Jenkins J. M., Fischer D. A., 2017, AJ, 153, 180

Schröder K. P., Smith R. C., 2008, MNRAS, 386, 155

Shallue C. J., Vanderburg A., 2018, AJ, 155, 94

Sharma S., et al., 2019, MNRAS, 490, 5335–5352

Silburt A., Gaidos E., Wu Y., 2015, ApJ, 799, 180

Sinukoff E., et al., 2016, ApJ, 827, 78

Smirnov N., 1948, The Annals of Mathematical Statistics, 19, 279

Steffen J. H., et al., 2012, Proceedings of the National Academy of Sciences, 109, 7982–7987

Su D.-Q., Cui X.-Q., 2004, Chinese Journal of Astronomy and Astrophysics, 4, 1

Sussman G. J., Wisdom J., 1992, Science, 257, 56

Teske J. K., Thorngren D., Fortney J. J., Hinkel N., Brewer J. M., 2019, AJ, 158, 239

Thompson S. E., et al., 2018, ApJS, 235, 38

Traub W. A., 2012, ApJ, 745, 20

Traub W. A., 2016, arXiv e-prints, p. arXiv:1605.02255

Tremaine S., 2015, ApJ, 807, 157

Tsiganis K., Gomes R., Morbidelli A., Levison H. F., 2005, Nature, 435, 459

Van Cleve J. E., et al., 2016, PASP, 128, 075002

Van Eylen V., Agentoft C., Lundkvist M. S., Kjeldsen H., Owen J. E., Fulton B. J., Petigura E., Snellen I., 2018, MNRAS, 479, 4786

Van Eylen V., et al., 2019, AJ, 157, 61

Vanderburg A., Johnson J. A., 2014, PASP, 126, 948

Vanderburg A., et al., 2016, ApJS, 222, 14

Veras D., 2016, Royal Society Open Science, 3, 150571

Veras D., Wyatt M. C., 2012, MNRAS, 421, 2969

Veras D., Wyatt M. C., Mustill A. J., Bonsor A., Eldridge J. J., 2011, MNRAS, 417, 2104

Weiss L. M., Marcy G. W., 2014, ApJ, 783, L6

Weiss L. M., et al., 2018a, AJ, 155, 48

Weiss L. M., et al., 2018b, AJ, 156, 254

Williams D., Kasting J., Wade R., 1997, Nature, 385, 234

Wilson C., 1985, Archive for History of Exact Sciences, 33, 15

Wolfgang A., Rogers L. A., Ford E. B., 2016, ApJ, 825, 19

Wright J. T., Marcy G. W., Howard A. W., Johnson J. A., Morton T. D., Fischer D. A., 2012, ApJ, 753, 160

Xie J.-W., et al., 2016, Proceedings of the National Academy of Science, 113, 11431

Yang J., Cowan N. B., Abbot D. S., 2013, ApJ, 771, L45

Youdin A. N., 2011, ApJ, 742, 38

Yu L., et al., 2018, AJ, 156, 22

Zeebe R. E., 2015, ApJ, 798, 8

Zhu W., Petrovich C., Wu Y., Dong S., Xie J., 2018, ApJ, 860, 101

Ziegler C., et al., 2018, AJ, 156, 259

Zink J. K., 2019, EDI-Vetter, doi:10.5281/zenodo.3585940

Zink J. K., Hansen B. M. S., 2019, MNRAS, 487, 246

Zink J. K., et al., 2019a, Research Notes of the American Astronomical Society, 3, 43

Zink J. K., Christiansen J. L., Hansen B. M. S., 2019b, MNRAS, 483, 4479

Zink J. K., Hardegree-Ullman K. K., Christiansen J. L., Dressing C. D., Crossfield I. J. M., Petigura E. A., Schlieder J. E., Ciardi D. R., 2020a, AJ, 159, 154

Zink J. K., Hardegree-Ullman K. K., Christiansen J. L., Petigura E. A., Dressing C. D., Schlieder J. E., Ciardi D. R., Crossfield I. J. M., 2020b, AJ, 160, 94

Zink J. K., Batygin K., Adams F. C., 2020c, AJ, 160, 232