# The Population of Exoplanets as Sculpted by their Host Stars

**Robert F. Wilson**
San Antonio, TX

B.S. Physics, University of Texas at Austin, 2016
B.S. Astronomy, University of Texas at Austin, 2016
M.S. Astronomy, University of Virginia, 2018

A Dissertation Presented to the Graduate
Faculty of the University of Virginia
in Candidacy for the Degree of
Doctor of Philosophy

Department of Astronomy
University of Virginia
August 2021

**Committee Members:**

Steven R. Majewski
L. Ilsedore Cleeves
James Davidson
Robert R. Jones
Michael F. Skrutskie
Johanna Teske

# Acknowledgements


### "Well... This is what I have to work with... "

### –David Bordenave

Whoops. I meant to be an engineer. Unfortunately, my first year in college I joined a research group that did asteroseismology. And it was cool. Now, eight years later, I have my undergrad advisors, Mike and Don, to thank for my career and a couple of extra degrees. Thanks for everything.

To my cohort: Andy, David, Hannah, Luca, and Yiqing, I never would have gotten through my classes if I hadn't waited for you all to finish the homework first so you could help me. Thanks for all the help. The also applies to the cohorts before and after me. Special thanks to those of you who worked with me on group projects. You all know you dragged me to a good grade, even if you won't admit it to me.

To Chris, thanks for the three years as a roommate and the countless conversations we had about, well anything really. Sometimes even science. Thanks for the days in the office where all I did was keep you from getting your work done. To Allison, thanks for all the Atlas invitations and for being a real human I could talk to when the World was ending. I don't even mind that you were really inviting Leo. To Paul, Trey, Matt, Lauren, and Sean, you all made it impossible to decline my UVA offer after visiting, and

I can't remember a time I clicked so quickly with a group of people. Thanks for the good impression, and for more than living up to it afterwards. To the Nelly's crews, Bruce, David, Yiqing, Caroline, Chris, Ian, Lauren, Trey, Matt, Sean, thanks for some unforgettable late nights and some great times I needed to unwind. To my Brohana. You the real ones. Between wizard staffs, poker, and blowing the bug, I've had countless unforgettable moments with you all, and even more support.

To Joel, Evan, Melody, and Victoria, thanks for everything. It's hard to single out just a few things to be grateful for, but y'all have been a great source of friendship, support, and pizza.

To the women in the front office, Barbara, Whitney and Gabby, thanks for solving all of my problems. Especially Whitney, when my problems were dumb and I was really just looking for an excuse to take a break. Thanks for giving me things to do so my brain could relax guilt-free and I could feel like I helped.

To all the other people I haven't even mentioned yet. I had way too many good times with you all, from being a distraction while I got coffee (not you, Molly), to tail-gating, Foxfield, DnD sessions, cheese balls, late night observing, Chinese lessons, mimosas, watching the crabs, the beer tub, and all the rest. Thanks for the good times and the many memories that fill my brain.

To the others: Ez, Sadie (Jake?), Calvin, Winnie, Ellie, Bowie, Mercury, Emrys, Samson, Sirius, Dougie, Spirit, Max, Ivy, Loken, Lyra, Sorin, Bo, Storm, Snickers, Tom, Fox, Maya, Hannah, Gutter Cat, and especially to Leo. You may not be aware of how happy you make me. Thanks for changing the rough times into the ruff times.

Now professionally: to my advisor, Steve. I'm so grateful that you weren't phased my first year when I joined your group meeting uninvited and presented an idea for a research project. Instead, you encouraged me and gave me the freedom to pursue new ideas as you have throughout my time at UVA. I also want to thank all of my collaborators, especially Johanna, Katia, and Verne, who let me know about a week later that the idea

I had presented to Steve my first year was not a good one. Thank you for working with me and mentoring me anyway over the last five years, and for the many years yet to come. I'm sorry in advance for all of the bad ideas that I will continue to have, and I hope I don't waste too much of your time.

To my committee, thanks for all the support over the past year and especially for the support this Summer. I actually think everything's working out, thanks to your advice.

Lastly and most importantly, to Mom, Dad, and Alix: thank you so much for all of your support. You've been easily the most reliable part of my life since forever, and I can't ever write enough to thank you for just being around and providing unconditional love. And Erin, I wish you were here to make fun of me for being a nerd and buy me too many drinks to celebrate *finally* finishing school. I still miss you every day and relish the time we spend together in my dreams.

# Abstract

Over the last decade, exoplanet studies have experienced a revolution. This boom is in part due to large-scale photometric surveys such as the *Kepler* mission, which to date has discovered nearly 5000 credible exoplanet candidates. This vast dataset has stimulated exoplanet science to advance beyond the stage of discovery and characterization of single systems, to population studies of exoplanets within our Solar neighborhood. Evolving in a similar timeframe over the last decade are large-scale spectroscopic surveys, which provide means for efficient characterization of stars found by *Kepler* to host planets. With the combination of planet-finding missions and host star spectroscopic surveys, we are able to characterize more carefully and precisely the properties of planet-hosting stars, which in turn gives us a more detailed understanding of the planets themselves.

In this dissertation, we combine information about planet hosting stars as gleaned from the SDSS/APOGEE surveys and *Kepler* surveys to uncover the detailed and intertwined links that planets exhibit with the stars that they orbit. We first report on a relationship that planets show between their host star's metallicity and the planets orbital period/semi-major axis known as the period-metallicity relation. We provide a detailed characterization of this relationship, showing that there is a lack of hot planets in stars with low metallicities, and provide evidence that the population of planets found at distances $\lesssim$0.07 au from their host stars are distinctly different from those beyond that limit. We interpret this transition region as evidence for an inner protoplanetary

disk boundary that is controlled, in part, by stellar metallicity.

We then extend the impact of stellar chemistry beyond that of just planet period and stellar metallicity, by examining the relationships between the planet occurrence rate (the average number of planets per star) and the enhancement of ten individual chemical elements (C, Mg, Al, Si, S, K, Ca, Mn, Fe, and Ni). We measure the planet occurrence rate density with orbital period and planet radius, and for the first time ever incorporate a detailed chemical understanding of planet host stars to track changes in the planet population due to chemistry. We find that the enhancement of all elements contributes equally to the enhancement of the planet occurrence rate, with stronger correlations given for planets with shorter orbital distances. This strong correlation presents a challenge in some sense, as it is the source of a strong confounding variable when considering changes in the planetary distribution function with stellar properties such as age and mass. After demonstrating how such a degeneracy can arise, we motivate the need for systematic, homogeneous, high-resolution spectroscopic surveys to properly characterize planet host stars.

In addition to large-scale spectroscopic surveys, the exoplanet field has also seen a large boom with the addition of *Gaia*, an astrometric survey which precisely measures distances to all the stars within the Solar neighborhood, allowing us to determine fundamental stellar properties for all of the stars observed by *Kepler*. To this end, we derive fundamental stellar properties for ∼163,000 stars in the *Kepler* field using a homogeneous spectroscopic and photometric metallicity scale, enabling us to break degeneracies in mass and age. Motivated by the higher than expected fraction of evolved stars in the *Kepler* field and the difficulty in estimating occurrence rates for evolved stars due to an inadequate grid of transit templates searched, we develop a new transit detection algorithm, `TraSH-DUMP` (TRAnsiting planets with Subgiant Hosts – Detection with an Unbiased Matched filter Pipeline). `TraSH-DUMP` utilizes precisely known stellar parameters to optimize a transit search for evolved stars, a host star target class not well

searched by the original *Kepler* software pipeline. We describe the methodology behind `TraSH-DUMP` in detail and compare it to the current state of the art detection pipelines, demonstrating its competitive sensitivity. The combination of `TraSH-DUMP` and the catalog of precise stellar properties derived set the foundations for a detailed study of planet demographics and the evolution of planetary systems with age.

# Table of contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Background

### 1.1.1 Exoplanets in a Historical Context

The first extra-Solar planet (a planet orbiting a star other than the Sun, commonly referred to as an *exoplanet*) around a Sun-like star, was discovered orbiting a bright star in the Pegasus constellation (51 Pegasi; Mayor & Queloz 1995). Though the existence of exoplanets had been considered for decades, and even though planetary-mass objects had already been discovered outside our Solar system (albeit orbiting a pulsar and thus representing a planetary system substantially different from our own; Wolszczan & Frail 1992), this discovery gave the first concrete evidence of what astronomers would come to appreciate two decades later; exoplanets are *ubiquitous*. Nearly every star on in the Galaxy is thought to host planets of some kind.

In the 10-20 years immediately following the discovery of the first exoplanet orbiting a Sun-like star, the discovery of new exoplanets began to fill the literature. The majority of these exoplanets were discovered via radial velocity surveys, which detect planets by measuring a doppler shift in a stellar spectrum consistent with a star orbiting the

common center of mass of a planet-star system. At this time, typical uncertainties in radial velocity measurements ($\sim$15 $\mathrm{m\,s^{-1}}$) allowed for the detection of exoplanets with typical masses ranging from $\sim$1-10$\times$ the mass of Jupiter ($M_{\mathrm{Jup}}$), with detection limits creeping to as low as $\sim$0.1 $M_{\mathrm{Jup}}$ by the year 2010. These planets had orbital separations spanning anywhere from $\sim$0.01-10 au. Even early on, it was clear that the properties of exoplanets displayed a wide range of diversity, and with the discovery of dozens of Jupiter-sized and Jupiter-mass planets, exoplanetary systems could vary dramatically from our own Solar system.

### 1.1.2 Transiting Exoplanets and the Kepler Revolution

An alternative method proposed for detecting exoplanets was via transits. A transit occurs when a planet passes between the observer and star along the observer's line-of-sight (see Figure 1.1). During a transit, the integrated flux of the star appears to dip significantly, as the shadow from the transiting body passes through the observer's line of sight. In this way, the diameter of a planet can be estimated, because the amount of light blocked from the host star during a transit scales as $\sim R_p^2/R_\star^2$ (See Figure 1.1).

The first detection of a transiting exoplanet occurred via follow-up photometric observations of the radial-velocity detected exoplanet, HD 209458b (Charbonneau et al. 2000), allowing for the first time, an estimate of not only the mass of an exoplanet, but also the radius, and by extension, the density and surface gravity. Immediately, puzzling results were uncovered. For instance, the mean density of HD 209458b was measured to be $\rho_p \approx 0.38$ $\mathrm{g\,cm^{-3}}$, about half the mean density of Saturn, the least dense planet in the Solar system. With this discovery, the authors remarked that future work would lead to, " ... an exciting time for those with an interest in the field". The next two decades would prove them correct.

The advantages of transit-search surveys became clear in the coming years. Searching for transits has several advantages compared to searching for planets via radial velocities

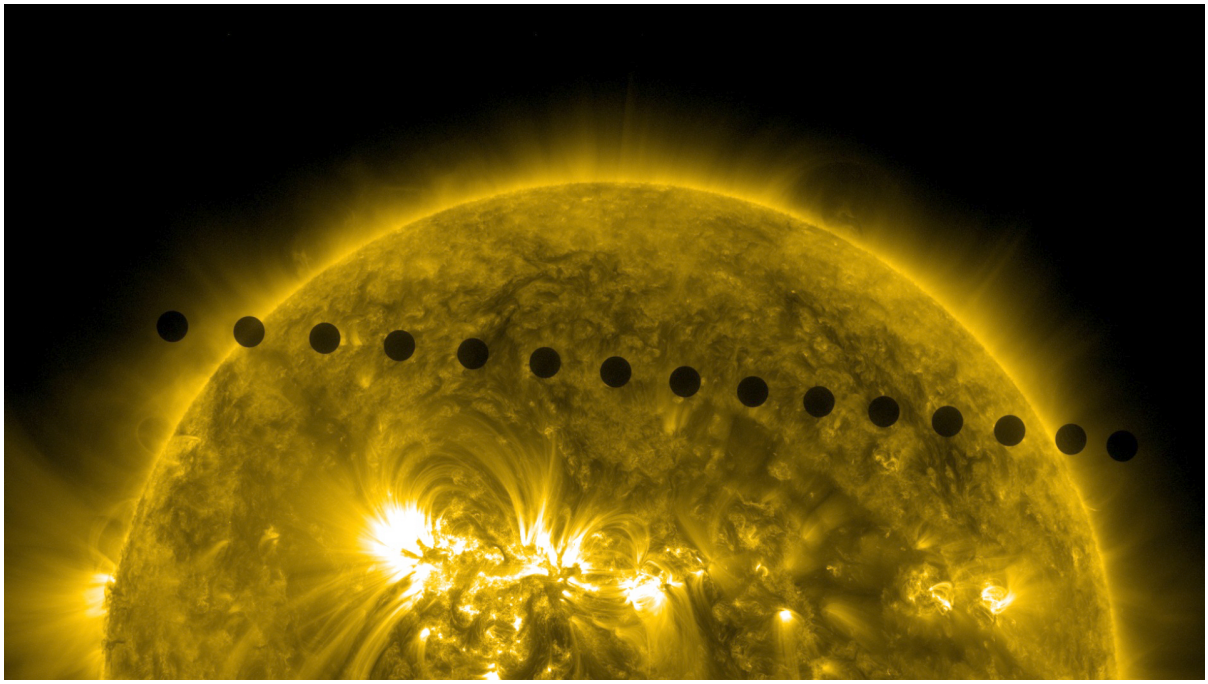

Fig. 1.1.— A composite image of the transit of Venus, captured by NASA's Solar Dynamics Observatory. The dark spots are Venus at different times during its transit and demonstrate the (small) decrease in light caused by the shadow of Venus passing through our line of sight. Unlike with transiting exoplanets, during the transit of Venus we are able to resolve the shadow itself, which appears as a dark disk that blocks a fraction of light, $\delta F$, equal to the projected area of the planet, $\delta F \approx (\pi R_p^2)/(\pi R_\star^2) = R_p^2/R_\star^2$. Image Credit: NASA/SDO.

(RVs), the second most efficient method to date. First, the survey efficiency can be drastically improved, simply because one can observe many more stars at the same time. The ability to observe multiple stars via imaging supercedes even multiplexed spectrographs due to the relative size of the detectors needed. In principle, a star only needs one pixel of a detector to measure a relative brightness, while for a high-resolution spectrum, a single star would need (at least) hundreds to thousands of pixels. In the time period where exoplanet discoveries had just begun, the adoption of CCDs and CMOS detectors by the astronomical community especially opened up the possibility of wide-field imaging, and at a high cadence. This new technology, in contrast with previous technology used for precision photometry such as photomultiplier tubes, made it possible to observe many stars at once, and uniformly. Another advantage of transit-search surveys was that the

precision in photometry needed to detect a transit (at least for the Jupiter-sized exoplanets first discovered) was easy to achieve, meaning that a survey could be designed with relatively cheap instrumentation and small telescopes. Subsequently, many such efforts have taken place with relative success, discovering dozens to hundreds of transiting planets since their inception (e.g., Hartman et al. 2004; Pollacco et al. 2006).

Despite the ease in instrumenting these ground-based photometric efforts, there were still severe limitations in the search for exoplanets via transits. The primary disadvantage of the transit method is the low likelihood that a planet will transit. For a star with a given radius, $R_\star$, the probability that a planet transits is $\sim R_\star/a$, where $a$ is the orbital distance of the planet. Because $R_\star \ll a$ for a typical planetary system, the probability of any one star having a transiting planet is minimal. For instance, the probability that a planet with an orbital distance of 1 au has only a $\sim$0.5% chance of transiting. Therefore, for a survey aimed at detecting planets via transit to be efficient, a very large number of stars must be searched.

Another disadvantage is the duty cycle of a transit. Transits occur relatively quickly, typically on the timescale of a few hours for the types of exoplanets first discovered, which is much smaller than the orbital period of the planets. Without a priori knowledge of the time of transit, only a small fraction ($\sim t_{\rm dur}/P$; where $t_{\rm dur}$ is the duration of the transit, and $P$ is the orbital period of the planet) of the light curve would contain information useful for detecting a planet. Thus, the probability of actually observing a transiting planet, assuming it does in fact transit, is relatively small. This is further complicated by the rotation of the Earth, as the average duty cycle of observations over the course of a year could only be as high as 50%, and will certainly be lower due to inclement weather, meaning that the actual observing time needed to detect and confirm a transit, and then a second (and often third or fourth) transit, needed to unambiguously determine the orbital period, was actually on the order of several times the period. One strategy undertaken by some groups to improve the effective duty cycle of observations was to observe from the

Earth's South pole, where observing can commence nearly uninterrupted in the winter months (Crouzet et al. 2010).

However, from the culmination of many ground-based transit searches, it became clear that the photometric precision of any ground-based observations would be limited by the Earth's atmosphere. Effects such as image wander, scintillation, and other issues caused by turbulence in the Earth's atmosphere all degrade the observational quality, and while strategies would be developed to mitigate some of these effects (Southworth et al. 2009; Stefansson et al. 2017), it was not likely that photometric precision from a large ground-based telescope could ever truly be photon limited, neutralizing the advantage of large aperture telescopes and high signal-to-noise observations.

With the growing scientific impetus for detecting small planets, the motivation for a space-based telescope to search for transiting planets picked up steam. The Hubble Space Telescope made some strides in this direction, providing a proof of concept by continuously observing a region of the sky centered on the Galactic bulge for 7 days and discovering 16 transiting planet candidates (Sahu et al. 2006). However, the field of view of the Hubble Space Telescope was relatively small and as a result, for such a study to observe enough stars to discover a large number of planets, it would be limited to densely populated regions of the sky where follow up observations would be difficult. To make a larger impact, a dedicated planet-hunting space telescope would be needed to monitor a large number of stars with a high cadence, high precision, wide field of view, and a long temporal baseline.

## The Kepler Mission

The need for a high-precision, space-based photometer to discover transiting exoplanets was realized in the form of the *Kepler* space telescope (Borucki et al. 2010). The *Kepler* mission has been, to date, the most important contribution in the field of planet demographics, as it uncovered nearly 5,000 planet candidates over the course of its mission, of

which ∼2400 have been confirmed, accounting for more than half of the confirmed exoplanets currently known.[1] The *Kepler* mission was able to achieve this feat by observing ∼200,000 stars in a $\sim 15^\circ \times 15^\circ$ patch of sky in the Cygnus constellation, nearly continuously for ∼4 years. Without limitations induced by the day/night cycle of the Earth or the degradation in photometric quality caused by the Earth's atmosphere, the *Kepler* space telescope achieved a photometric precision sufficient to detect small, Earth-sized planets for the first time. The first five confirmed exoplanets discovered by the *Kepler* mission are displayed in Figure 1.2.

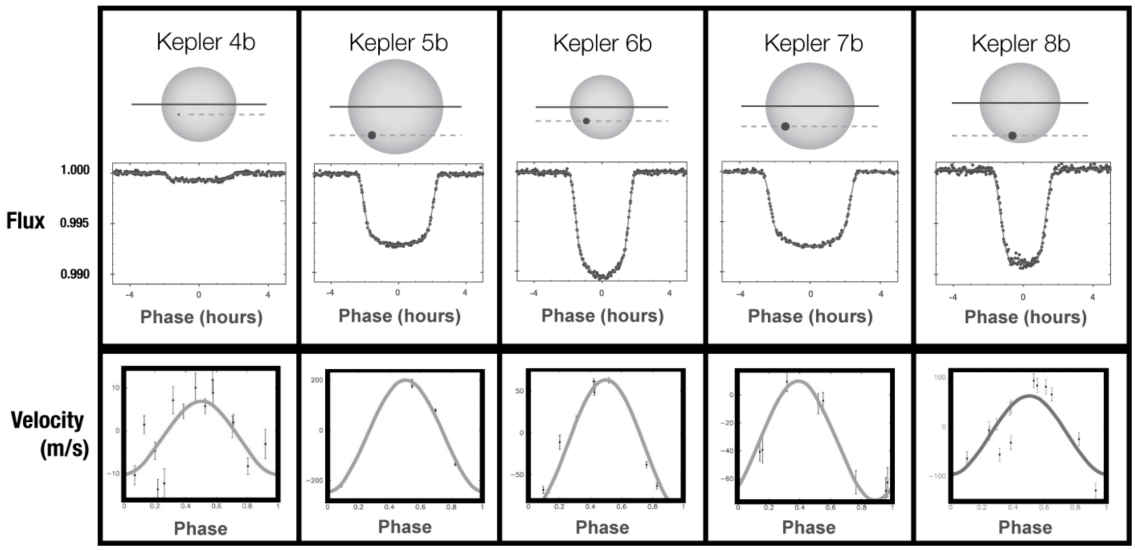


Fig. 1.2.— The first five transiting exoplanets discovered by the *Kepler* mission (Kepler-1, -2, and -3 were already discovered by ground-based transit surveys prior to the launch of the *Kepler* space telescope). The top panel shows the light curve at the time of transit, along with a schematic representing the relative sizes of the planet, host star, and the inclination of the transiting planet. The bottom row shows RV observations folded on the orbital period measured from the light curve, confirming the planetary nature of the transit signal. The comparisons between these systems highlight how the depth of transits scale as $\sim R_p^2/R_\star^2$. Figure credit: Borucki et al. (2010).

Many of the major results from the *Kepler* mission were made possible by the sheer number of planets discovered (see Figure 1.3). Important results for the field of exoplanet demographics included the ubiquity of planets, as mentioned previously, but also that the

[1] as of July 1, 2021: `https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html`

ubiquity of small planets held up in all parts of the Galaxy probed by *Kepler*. No region of the *Kepler* field had any more planets than any other region. One particular discovery, unexpected given the knowledge of exoplanets at the time, was that the most common exoplanets (with the possible exception of planets smaller than the Earth, as such a regime has yet to be adequately probed by transit surveys) were those with diameters between that of the Earth and Neptune, of which there is no analogue in our own Solar system. Such planets, referred to as Super-Earths and Sub-Neptunes (or mini-Neptunes), would become the legacy of the *Kepler* mission.

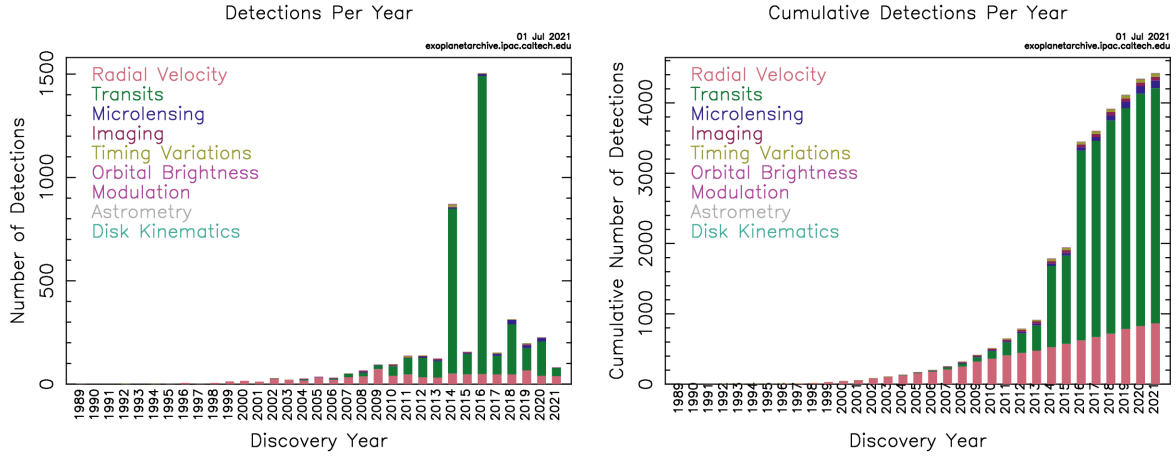


Fig. 1.3.— The number of detected exoplanets by year, where the color denotes the method by which the exoplanets were detected. *Left:* The number of detected exoplanets each year. In particular, note the large spikes in years 2014 and 2016 that correspond to the release of the first *Kepler* planet candidate catalog (2014) and the statistical validation of $\gtrsim$1000 planet candidates by Morton et al. (2016). *Right:* The cumulative number of detections by discovery year, reaching todays total of nearly 4500 known, confirmed exoplanets.

## Biases Inherent in the Detection of Transiting Planets

However, transit search surveys, although responsible for more confirmed exoplanets than any other planet-search survey, suffer from extreme biases. For one, unlike with radial velocities, transit signals have a low probability of being observable. An observer has no control over whether a planet transits, and no amount of technological improvement could overcome the geometry of a planetary system. Because the transit probability

decreases inversely with orbital distance, the distribution of discovered transiting planets are very heavily biased toward short orbital periods and semi-major axes, even more so than RV surveys which are also biased toward short orbital periods (see Figure 1.4).

In addition to being biased toward short orbital periods due to a geometric bias, the detection metric for a transiting planet is the transit depth. The depth of a transit is proportional to $R_p^2/R_\star^2$, and is thus skewed toward large planet radii, for a given stellar radius. Thus, the distribution of detected planets from transit surveys should differ from the distribution of real planets in two important ways. First, they will be extremely heavily skewed toward short orbital periods, and therefore orbital distance, and second, they will be skewed toward larger radii. There is no inherent bias to transiting planets with respect to mass, except in the sense that mass correlates strongly with planet radius.

## 1.2 Exoplanet Demographics and Occurrence Rates

For the purposes of this thesis, we are interested in measuring the demographics of planets as a way to better constrain theories of planet formation and evolution. Because such processes can take significantly longer ($\gtrsim 10^{5-11}$ years) than a human lifetime ($\lesssim 10^2$ years) to occur, understanding the aggregate properties of exoplanets offers a more efficient method of testing such theories, rather than observing and waiting for a planetary system to evolve, although such a strategy may be possible in some very extreme cases (e.g., Patra et al. 2020; Turner et al. 2021). However, even then such a strategy is unlikely to produce interesting results on the timescale needed to finish a dissertation.

In studying the demographics of planets, our goal is to recognize patterns and elements common across all (or a large fraction of) planetary systems. In discovering such properties, we are able to distinguish between events that are natural outcomes of universal truths, and those that are the results of unlikely coincidences. These universal

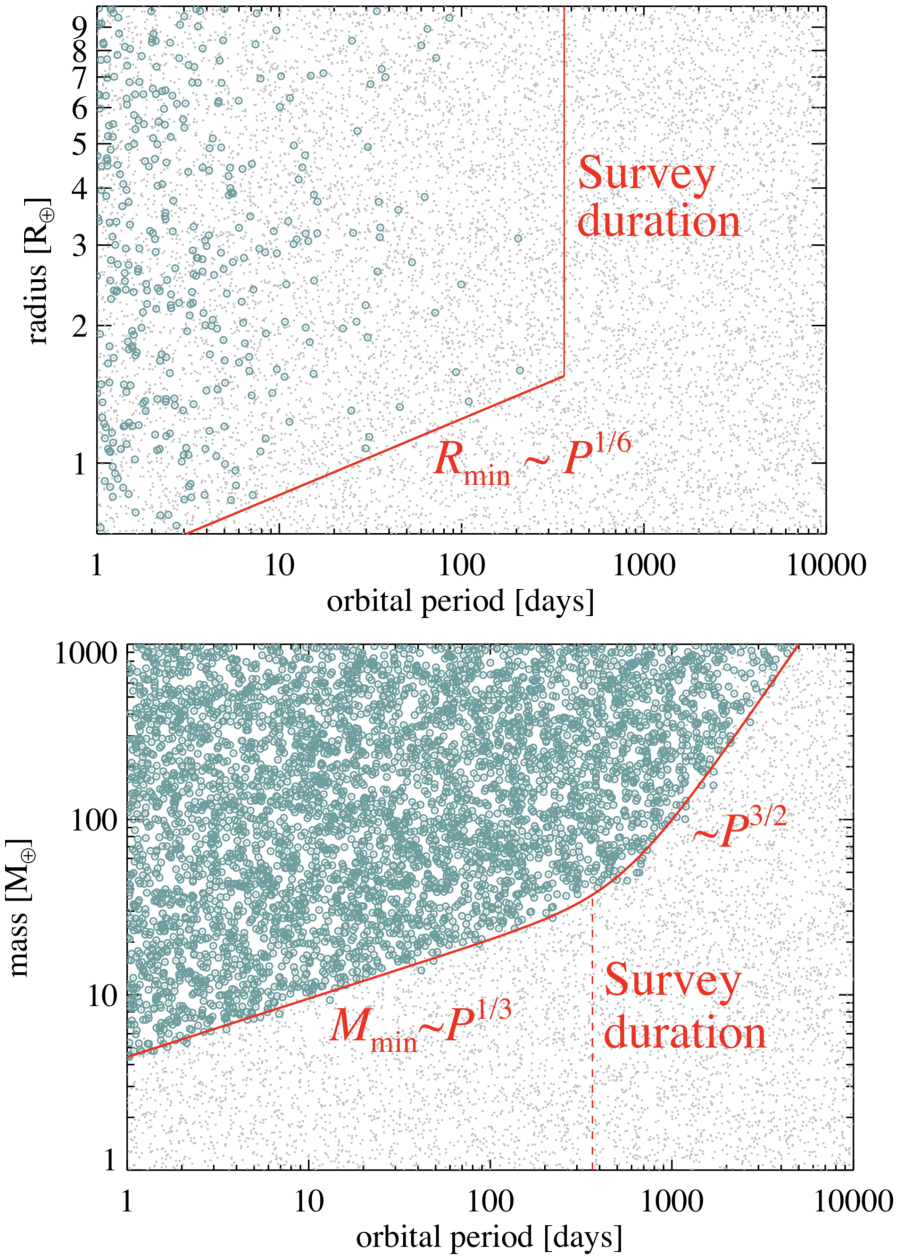


Fig. 1.4.— A simulated, idealized transit (top) and RV (bottom) survey. Each grey dot represents one planet in the simulated survey, and each green circle denotes that the planet was detected. The red lines display the detection limits of the survey. For the transit survey, the detection rate is low due to the geometric bias of a planet transiting its host star from our point of view. The lower envelope in this case is set by the ratio of the transit depth ($\sim R_p^2/R_\star^2$) to the noise in the light curve, and the square root of the number of transits in the survey. The detection rate for an idealized RV survey is significantly higher than in an idealized transit survey because the geometric bias against inclination is significantly less severe. Figure Credit: Winn (2018).

truths provide constraints or clues to the formation or evolution of planetary systems, but, due to the intrinsic diversity in exoplanetary systems, a large number of planets are needed to fully recognize such a pattern.

One fundamental measurement that is particularly important for theory to reproduce is the planet occurrence rate. For the purposes of this work, we define the planet occurrence rate as the average number of planets per star (NPPS). Another definition for a planet occurrence rate that is often adopted may be the fraction of stars with planets (FSWP) which, although related, is not the same measurement. The NPPS takes into account stars that may host multiple planets, while the FSWP does not.

While the planet occurrence rate is a conceptually simple measurement, the ratio of the number of planets to the number of stars, in practice it can be quite complicated to infer. The difficulty comes not from knowing the number of stars in your sample, or even from the number of planets that are detected, but in estimating the number of planets that were *not* detected by your survey. This quantity is typically recast by estimating the number of detected planets divided by the *effective* number of stars in a survey. For example, if a survey is expected to detect one of every ten planets, then each star searched would be the equivalent of 0.1 effective stars searched. This estimate requires a comprehensive understanding of the biases inherent to a planet-search survey, ranging from biases inherent to the detection method itself to the sensitivity of the software used to find planets in the data, and a precise understanding of the quality of data.

To further complicate matters, the detection of planets depends severely on the properties of the planet as well as the properties of the stars being searched. For example, an important consequence of the relationship between transit depth and stellar radius is that planet radii measurements are often limited by the precision with which one can measure the radius of the host star. Because the observations needed to precisely infer stellar radii are resource intensive, the lack of a homogeneous, precisely characterized planet-search sample has historically been a barrier in understanding the properties of

exoplanets. Fortunately, over the past few years, the *Gaia* mission (Gaia Collaboration et al. 2018b) has nearly solved this problem by measuring geometric parallaxes to nearly every star within a few kpc. Using these precise geometric parallaxes with apparent magnitudes and effective temperatures (which are both observationally "cheap" to obtain), one can apply the Stefan-Boltzmann law to infer the angular diameter of a star, and by extension the radius of a star (e.g., Berger et al. 2018).

In the sections below, we give a brief overview of known trends with planet occurrence rates, but first motivate this understanding by providing a brief overview of planet formation. Then, for each of the planet occurrence trends, we discuss common interpretations of the trends in the context of the physical processes governing planet formation.

### 1.2.1 The Classical Picture of Planet Formation

The subject of planet formation is a highly complex topic in its own right, and to describe it in such a short section of an introduction to a larger work would be an insult to those who spend entire careers and lifetimes trying to understand these processes. Unfortunately, we will attempt to do just that.

For the purpose of this work, we will distinguish between two kinds of planets when discussing formation: giant planets ($R_p \gtrsim 4R_\oplus$, $M_p \gtrsim 30M_\oplus$) and small planets ($R_p \lesssim 4R_\oplus$, $M_p \lesssim 30M_\oplus$). Small planets can be further described by those with a H/He atmosphere ($R_p \gtrsim 2R_\oplus$) and those without a H/He atmosphere ($R_p \lesssim 2R_\oplus$). Because these planets share relatively similar formation pathways at the beginning of their lifetimes, we will first discuss the general environment in which planets form, before discussing the differences between the processes that shape these two kinds of planets. Note, that while binary stars are a natural process of star formation, we refrain from considering binary stars in this brief discussion, instead focusing on the formation of planets around single stars. For the purpose of providing a brief review of planet formation that will motivate the physical interpretations of exoplanet demographics, we will

loosely follow the discussion from Armitage (2018), and the references therein.

## Protoplanetary Disks and the Planet-Forming Environment

As a natural consequence of a molecular cloud collapsing via gravitational instability to form a star, a disk is generated around the star with which to conserve angular momentum. It is from this disk of material that planets form and coalesce. In this way, planet formation is thought to be natural consequence of star formation, which explains the ubiquity of planets in our Galaxy, as stars themselves are ubiquitous. The disk of material orbiting the protostar is referred to as the protoplanetary disk, and it is within this disk that the planet formation processes discussed below occur.

The protoplanetary disk typically extends out to radii of $\sim 10^{2-4}$ au and is substantially less massive than the star itself ($M_{\rm disk}/M_\star \lesssim 10^{-2}$). The disk is made up of a combination of gas and dust from the molecular cloud from which it collapsed. Because of this, the protoplanetary disk is typically assumed to have, to first approximation, the same concentration of metals (i.e., elements heavier than H/He) as the star and molecular cloud. However, due to changes in the temperature, density, and a myriad of other physical processes, whether metals are in the form of ice, dust, or gas depends on the location in the disk. Generally, the temperature of the disk due to irradiation decreases with orbital radius, $r$, by $T \sim r^{-1/2}$. The disk typically survives for $\sim$1-10 Myr, before high energy radiation from the protostar evaporates and disperses the gas disk.

For the purposes of this work, we are primarily interested in the inner $\lesssim 1$ au of the disk, as that is the region of parameter space effectively probed by *Kepler*. At close-in distances ($\sim$0.01-0.1 au), the boundary of the disk is set by the corotation radius, within which material is accreted onto the central protostar. Another important location to consider is the dust-sublimation radius, within which the surrounding temperature is too high to support dust, which instead sublimates into a gas. At the other end of our orbital separation regime of interest is the water ice line. The ice line is the orbital separation

in the disk where water may exist in solid form. Because the location of the ice line is dominated not by stellar irradiation, but by accretion heating, the location of the ice line changes with the evolution of the disk, and can be as close as ∼1 au, depending on the accretion rate. Planets beyond the ice line are generally thought to form more efficiently because there is more solid material available (ice and dust, instead of just dust), and because ice tends to coagulate with other ice particles more efficiently than dust.

## Planet Building Blocks: From Dust to Planetesimals

To form planets, there must be an efficient mechanism for dust to coagulate together into bodies of planetary mass. The growth of such bodies is still a bit of a challenge in planet formation, but can typically be explained on two different scales: dust and pebbles. We typically take dust to be particles between $\mu$m-mm sizes. Pebbles are not always well-defined in the literature, and are typically taken to mean particles larger than dust, but smaller than planetesimals. For the sake of this discussion, we take pebbles to be ∼cm to km sized particles, and planetesimals to be particles with sizes $\gtrsim$km, similar to asteroids in the solar system.

In the protoplanetary disk, dust particles with $\mu$m sizes grow to ∼mm sizes by aerodynamically-controlled collisional growth. Because the gas of the protoplanetary disk orbits at sub-Keplerian velocities, the dust particles experience drag forces proportional to the difference in velocity between the gas and dust. Locally, the relative distribution of particle velocities is determined by dust coupling to turbulence in the gas, allowing particles to collide. At small sizes ($\lesssim$ 1 cm), these collisions are dominated by van der Waals forces which cause the particles to stick together. However, at larger sizes (∼0.1-1 cm), collisions from dust particles are equally likely to cause fragmentation, so collisional growth stalls.

In addition to a fragmentation-coagulation equilibrium, ∼mm to cm sized particles experience radial drift due to pressure gradients in the disk, the direction of which is typ-

ically assumed to be inward (as the global pressure profile of the disk decreases with radius), which implies that ∼cm sized particles are likely to drift into the star on timescales of $\sim 10^3$ years. Thus, the growth from ∼cm sized pebbles to ∼km sized planetesimals must happen very fast and cannot occur via collisional growth.

One potential solution to this problem is the streaming instability (Youdin & Goodman 2005). In this scenario, pebble-sized objects tend to cluster together, which effectively negates the drag force felt by the cluster (as pebbles are shielded from the headwind) and halts radial drift. Larger numbers of pebbles join these clusters, which then begins a runaway process. Eventually enough pebbles are clustered together to self-gravitate and collapse to form planetesimals. Simulations seem to suggest that this process can form ∼km sized planetesimals on dynamical timescales. The result of this process is a planetesimal mass distribution that is extremely top-heavy, i.e., most of the mass ends up in heavier planetesimals. The radial drift timescale for these larger mass planetesimals is sufficiently large that gas drag is no longer a significant barrier to planet formation.

## Core Accretion

Once a population of planetesimals is available, the outcome of collisions depends on the energy of the collisions, relative momenta between the two objects, and the material properties of the two objects. For planetesimals with higher velocity dispersion, collisions will fragment planetesimals leading to a cascading reaction where interactions increase the velocity dispersion, and grind bodies down to smaller particles.

However, if the planetesimal disk is kinematically cold, then a putative body, which we'll refer to as the planetary core, can accrete nearby planetesimals without exciting the velocity dispersion of other planetesimals in the disk. In this regime, and if the escape velocity of the planetary core is greater than the velocity dispersion of the disk, $v_{\rm esc} \gg \sigma$,

then there is a runaway accretion which can be described analytically by the form,

$$\dot{M}_{\rm core} \propto \Sigma \left(1 - \frac{v_{\rm esc}^2}{\sigma^2}\right) \quad , \tag{1.1}$$

where $\Sigma$ is the surface density of solids in the disk and $\dot{M}_{\rm core}$ is the accretion rate of the planetary core. This phase leads to runaway accretion as $v_{\rm esc}$ increases when $\dot{M}_{\rm core} > 1$, and $\sigma$ remains roughly constant. Eventually, as the accreting body's mass grows, it begins to excite planetesimals in its vicinity, dynamically heating the disk locally, and soon approaches the regime where $v_{\rm esc}/\sigma \sim 1$. In this phase, the accretion rate of the planetary core slows down and allows other planetary cores in the system to grow. In this way, this mechanism is thought to form a system of small planets on nearly circular orbits, and the formation of larger planets is halted by the local excitation of planetesimals in the disk.

This model typically explains the growth of small, terrestrial planets without H/He atmospheres. In addition to the processes described above, if the core is able to grow to $M_{\rm core} \gtrsim M_\oplus$, then the core is able to form a bounded hydrostatic H/He envelope, resembling a planet similar to Neptune or Uranus. Planetary cores with this bound envelope are thought to be the progenitors for planets with $R_p \sim$ 2-4 $R_\oplus$. This envelope is thought to help trap accreting pebbles, which further accelerates the planetary core to grow via pebble accretion.

This model can be further extended to giant planets. The primary difference is that giant planets are thought to form at greater distances ($\sim$3-10 au) where there is more solid material available in the form of ice. Once the core reaches a critical mass of $\sim$10-20 $M_\oplus$, the mass of the core is then sufficient to accrete gas directly from the protoplanetary disk, leading to rapid accretion as the planetary core is massive enough to tap into an additional supply of material.

However, it is worth noting that the picture described above makes several assump-

tions that have been challenged in the literature. First, this picture of planetesimal-driven formation assumes that most of the solid material is locked up in planets, and that there is not significant radial motion. The assumption that dust is mostly locked up in planets is in conflict with observations of planetary systems that have uncovered large amounts of dust, and the radial drift of ∼cm sized pebbles appears more significant in simulations than given in the above picture. In fact, in many regimes, planetary cores may grow more efficiently via pebble accretion than via the planetesimal-driven growth model described here. Second, the assumption that a planetary core of a few $M_\oplus$ contracts a static, bound envelope is likely not valid. 3D simulations seem to indicate that the gaseous envelope formed at this stage is actually recycled with its surroundings, rather than being bound as predicted by 1D models (e.g., Lambrechts & Lega 2017). These effects change the thermodynamic properties of the H/He envelope and the exact consequences of these results are still being explored and debated in the literature.

### In-Situ Formation versus Migration

Another important aspect of planet formation neglected in the discussion above is that of migration. For instance, the above discussion implies that giant plants have to form at long periods. Therefore, the presence of giant planets at short periods must invoke migration for the classical planet formation paradigm to be consistent with observation. Type I migration is typically important to consider for small planets that grow to $\gtrsim$0.1-1 $M_\oplus$ while the gas is still present in the disk, which is thought to be the case for many of the *Kepler* planets. This mechanism generally results from torques in the disk caused by the Linblad resonances and corotation resonances. Migration caused by Linblad resonances generally results in inward migration, while migration caused by corotation resonances is a complicated function that may result in either inward or outward migration, depending on the pecularities of the disk.

Migration is still not a fully understood problem, and the possibility of migration

tends to complicate the interpretation of planetary demographics. One method for understanding the role of migration is via the internal structure of planets. Planets that have formed outside the water ice line are thought to have different interior structures, and therefore different densities, than those formed close-in. This is because the solid material accreted onto the planet itself is a different composition. However, mass measurement needed to make such inferences are difficult due to the limited brightness of typical *Kepler* planet hosting stars and the lower mass of the *Kepler* planets.

### 1.2.2 Trends in Exoplanet Demographics

In this work we tend to classify planets based on their orbital period and their radius, as these are the observable properties from transit surveys. The orbital period, $P$, is simply a proxy for orbital distance, as $P^2 \propto a^3$, where $a$ is the semi-major axis of the planet's orbit. In addition, the orbital period of a transiting planet is measured, at least in the context of this work, with negligible uncertainty (i.e., $\sigma_P \ll P$), providing a useful metric with which to characterize these planets. In the following sections, we tend to describe and classify exoplanets using these properties. In addition, due to the biases inherent in the *Kepler* survey, the majority of the trends discussed in this section are for planets with $R_p \gtrsim 1R_\oplus$, and $P \lesssim 350$ days, as planets with smaller radii and/or longer orbital periods are more difficult to detect, leading to substantial biases in their statistical properties and small number statistics with large uncertainties.

#### Small Planets are more Common than Large Planets

Exoplanets have been found to come in a wide diversity of sizes, ranging from inflated Jupiter-analogs, which can be as large as ~20-25$R_\oplus$, to rocky planets even smaller than Mercury. It's been shown from numerous surveys (e.g., Cumming et al. 2008; Mayor et al. 2011; Fressin et al. 2013) that planets of different sizes have drastically different occurrence rates. Giant planets ($R_p \gtrsim 4R_\oplus$) are significantly less common than small

planets ($R_p \lesssim 4R_\oplus$). For instance, within periods of $P < 100$ days ( $a \lesssim 0.5$ au), giant planets are typically found with an average occurrence of $\sim$1 planet for every 50-100 stars while small planets are typically found with an occurrence of $\sim$1 planet for every 4 stars (Petigura et al. 2013; Winn & Fabrycky 2015). Similar statistics are found with RV surveys, where lower mass planets ($M_p \lesssim 30M_\oplus$) are found with significantly higher occurrence than higher mass planets (See Figure 1.5).

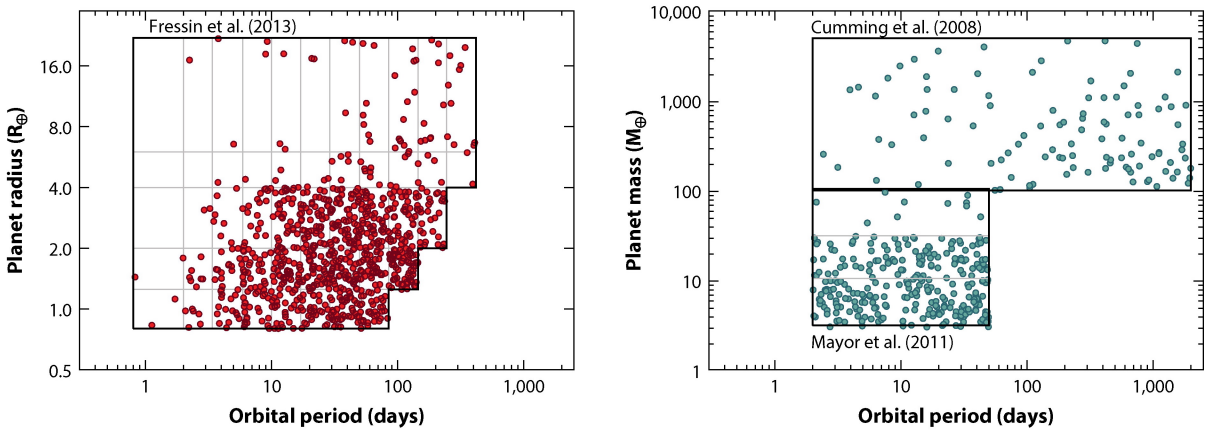


Fig. 1.5.— Figure 2 from Winn & Fabrycky (2015). These figures each show a simulated, representative sample of planets based on the occurrence rates of planets from transit surveys (left) and RV surveys (right). The bolded boxes in each figure denote the parameter space over which the occurrence rate measurements were made. The number of planets is significantly larger for planets with small radii and lower masses, which is in contrast to the distribution of detected planets due to biases inherent in detection methods.

This pattern is relatively easy to understand in the planet formation paradigm described above. Because smaller planets don't need as much material, and because the disk lifetime is limited, planetary cores with slower accretion rates may be able to form small cores and planets while not having enough time to form larger planets. This is particularly true at short orbital distances probed by the *Kepler* mission where there is less solid material, and there may not be enough solid material to form larger planets.

### Planets are More Common at Larger Orbital Distances

The occurrence of small planets is relatively constant from $P \sim$ 10-300 days ($a \sim$ 0.1-1 au), but then drops substantially below $P \sim 10$ days. This drop in occurrence for small planets is thought to coincide with the inner edge of the protoplanetary disk. However, it's less clear exactly what sets that inner radius. This discussion is further complicated by whether one considers in situ formation versus migration. In the former, a promising candidate for the inner radius is the dust sublimation radius, or the maximum distance at which the temperature of the disk is sufficiently hot that dust is sublimated into a gas. Thus, the dust sublimation radius is the innermost region in which dust can exist. The corotation radius is the radius interior to which gas in the protoplanetary disk is accreted onto the central protostar. The corotation radius is commonly invoked to explain a "stopping" distance, interior to which planets can't migrate. This is because many mechanisms for planet migration invoke torques from the gas in the protoplanetary disk.

The occurrence of large planets on the other hand is a relatively constant increase from $P \sim$ 1-300 days. Under the classical paradigm of planet formation, this is likely explained by the need for high mass planets to form at larger distances and migrate inward, so the distribution of giant planets at short periods is controlled solely by how far the planet can migrate inward. In addition to the slowly increasing occurrence at increasing periods, there is a sharp increase of large planets at $P \gtrsim 100$ days (Cumming et al. 2008; Petigura et al. 2018). This increase in large planet occurrence is thought to be the result of the water ice line providing more solid material for such planets to form.

### Low Mass Stars Have More Small Planets, Fewer Giant Planets

The implication of low mass stars hosting more and smaller planets compared to higher mass stars has two interesting implications. In general, the mass of the protoplanetary disk scales with the mass of the host star. Therefore, high mass stars have high mass

disks, and low mass stars have low mask disks. Thus, for high mass stars, there is more planet forming material, which is thought to facilitate the growth of higher mass planets. Within this paradigm, the idea that lower mass stars have fewer giant planets would make sense via the same logic.

However, low mass stars having a higher multiplicity of small planets doesn't quite fit this interpretation. This result has two interesting implications. First, small mass disks can efficiently form small planets. This fits into the picture of small planets being ubiquitous, but why would there not be more small planets around more massive stars? One proposed explanation is that because low mass stars do not have enough dust mass in their disks to form large planets, smaller planets are less likely to be dynamically interrupted. Thus, low mass stars are more likely to host dynamically cold planetary systems, that tend to be co-planar, and have less mass overall.

### The Occurrence of Giant Planets Increases with Stellar Metallicity

The stellar metallicity is typically measured, for historical reasons, as the ratio in the number of iron nuclei on the surface of the star to the number of hydrogen nuclei, normalized to the value of the Sun, and reported as a logarithm, i.e., $[\mathrm{Fe/H}] \equiv \log(N_{\mathrm{Fe}}/N_{\mathrm{H}})_{\star} - \log(N_{\mathrm{Fe}}/N_{\mathrm{H}})_{\odot}$. Giant planets discovered from early radial velocity surveys were generally found to have metal-rich host stars on average compared to stars without detected giant planets (Heiter & Luck 2003a; Santos et al. 2004).

By comparing the fraction of stars that did not have a detected giant planet in RV surveys against the fraction of stars that did have a detected planet within a range of metallicities, Fischer & Valenti (2005) found that the occurrence of giant plants could be characterized by a strong power law in metallicity. This result has held up against numerous studies, from both RV and transit surveys (e.g., Santos et al. 2004; Ghezzi et al. 2010; Buchhave et al. 2014; Petigura et al. 2018).

The most common interpretation for this connection is that stellar metallicities reflect

the compositional imprint of the molecular cloud from which the star formed, and therefore stars with higher metallicities have a higher dust to gas ratio in the protoplanetary disk. Therefore, because there is more dust in the protoplanetary disk, the growth of planetary cores is accelerated, and it is increasingly more likely that a planetary core can grow to a mass in excess of $\sim 10 M_{\oplus}$, allowing it to accrete gas from the protoplanetary disk before dissipation. In this way, stellar metallicity, $Z$, is typically interpreted as a proxy for the surface density of solids in the protoplanetary disk, $Z \propto \Sigma$. Therefore, based on Equation 1.1, $\dot{M} \propto \Sigma$, thus the accretion rate is more rapid in disks with higher solid surface density.

## Other Connections Between Stellar Metallicity and the Planet Population

With the above interpretation in mind, other relationships between stellar metallicity and exoplanet properties have begun to come to light. For instance, planets with higher metallicity host stars tend to have higher eccentricities on average (Dawson & Murray-Clay 2013). This is typically interpreted as evidence for increased planet-planet scatter due to the presence of on-average more massive planets (Winn & Fabrycky 2015).

The connection between planet orbital period and stellar metallicity is quite a bit more complicated. Before this work, such a relationship seemed tentative, and it was unclear that results could be separated from their observational biases. Some results seemed to suggest that low metallicity stars have longer period planets (Adibekyan et al. 2013), while other results seemed to suggest that a correlation only existed in the case of giant planets (Dong et al. 2017), some studies claimed it was only in the case of small planets (Mulders et al. 2016), and finally others questioned any correlation whatsoever with metallicity and orbital period (Winn et al. 2017). These results proved to be quite intricate, and the relationship between stellar metallicity and planet period nuanced. In this work, particularly Chapter 2, we show that indeed metallicity is correlated with

planet period, but mostly at very short orbital periods.

## 1.3 Overview

In this work, we aim to build a better understanding of the intricate relationship between stars and the planets that orbit them. In particular, we utilize large-scale surveys to provide the tools needed to characterize stars with and without detected transiting planets. Some common themes throughout each of these chapters are that both uniformity and precision are necessary to detect subtle trends in the exoplanet population, and that a complete understanding of the planet host star population is necessary for a complete picture of planet demographics.

### 1.3.1 The Planet Period-Stellar Metallicity Correlation

More specifically, in Chapter 2 we characterize the relationship between stellar metallicity, clearing up the debate as to how metallicity influences the distance at which planets orbit their host stars. To accomplish this characterization, we first validate spectroscopic parameters from the Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2017), and apply the metallicities measured from APOGEE to definitively show that the planet orbital period is in fact correlated with stellar metallicity, clearing up the debate in the literature. We take this correlation one step further to show that the population of exoplanets can be divided into two populations, defined by their metallicity, with orbital periods above and below $P = 8.3$ days. We interpret this critical period as a transition region where exoplanets are significantly less likely to exist if orbiting a host star with a depleted metallicity.

### 1.3.2 The Role of Detailed Chemistry on Planet Occurrence Rates

In Chapter 3, we measure the correlation strength between planet occurrence and the stellar abundances of ten different elements, the first ever such measurement. We subdivide our sample into varying planet sizes and orbital distances to measure how the correlation strength changes for differing planet types, and for differing elements. We find that the occurrence of small planets is strongly correlated with the enhancement of any element at short periods, but not strongly correlated with the enhancement of any element at longer periods. In trying to better understand the relationship between planet demographics and stellar chemistry, we consider the possibility that correlations like this could be a proxy for age, or vice versa. In discussing this possibility, we demonstrate how the strength with which metallicities are correlated with planet occurrence can be a strong confounding variable, and advocate for homogeneous, high-precision spectroscopic surveys to better characterize planet host stars and caution interpretations of changes in planet demographics with age due to such confounding variables.

### 1.3.3 Foundations for an Age-Demographics Study in the Kepler Field

Finally, in Chapter 4 we set the foundation for a systematic study of the demographics of planets with stellar age. Such studies have not been widely adopted in the literature because the difficulty in inferring ages for field stars. However, precise geometric parallaxes as measured by *Gaia* (Gaia Collaboration et al. 2018b, 2021), have made the inference of stellar ages via isochrone-modeling tangible. We utilize this information to derive the most precise fundamental stellar parameters (e.g., age, $R_\star$, $M_\star$) in the literature to date.

We discuss these improved parameters in the context of a larger program aimed at an independent transiting planet search in the *Kepler* field to better understand long-

term planetary evolution. We introduce a novel planet-detection algorithm inspired by the difficulty in estimating occurrence rates for evolved stars using the *Kepler* detection pipeline combined with the large fraction of subgiants (∼20%) revealed by *Gaia* DR2 to be in the *Kepler* field. We present this new algorithm, and demonstrate its effectiveness by comparing it to current open-source state of the art transit detection algorithms.

Through the combination of improved stellar ages, a novel detection pipeline that is sensitive and computationally feasible with the long temporal baselines of the *Kepler* mission, we set the foundation for a targeted demographics study of planets in the *Kepler* field, and the changes in the population with age.

### 1.3.4 Contributions

The work was made possible thanks in no small part to large collaborative surveys. The teams behind these surveys, particularly the *Kepler* and APOGEE teams, deserve gratitude and a wealth of credit, not only for making these surveys possible, but also for providing extensive documentation with which to educate the author. In addition to monumental efforts taken by the *Kepler* and APOGEE teams to produce the impressive results upon which this work is built, this work was also made possible by individuals who deserve individual recognition for their contributions to this work. In the interest of giving proper credit, we clarify those contributions below.

For each of these chapters, Steven R. Majewski has given invaluable scientific advice, feedback, and support as an advisor, and as such played a vital role in shaping the message and goals of this work.

Special gratitude is warranted for the team who designed and implemented the APOGEE-KOI survey, upon which the results of Chapters 2 and 3 are based.

The work in Chapter 2 was made possible in part by the efforts of Johanna Teske. Teske led the interpretation of the statistical analysis in the context of current planet formation models, provided the analysis to estimate the uncertainties on the [Fe/H]

measurements in APOGEE, and contributed text in the introduction and discussion sections.

Chapter 2 of this work is published in the Astronomical Journal under Wilson et al. (2018), and as such is publicly available. Chapter 3 is under peer review at the time of writing this work, with the expectation that it will soon be published and available to the public as well. These works were made possible by a group of collaborators and co-authors who contributed as architects of the APOGEE survey and/or provided feedback on the manuscripts of these two chapters. For these efforts, we recognize individual contributions from Carlos Allende Prieto, Andrés Almeida, Rachael Beaton, Chad F. Bender, Jonathan Brinkmann, Caleb I. Cañas, Katia Cunha, Scott W. Fleming, D. A. García-Hernández, Luan Ghezzi, Sten Hasselquist, Christian R. Hayes, Henrik Jönsson, Suvrath Mahadevan, Steven R. Majewski, Christian Nitschelm, Verne V. Smith, Keivan Stassun, Jamie Tayar, Johanna Teske, Nicholas Troup, Michael F. Skrutskie, Diogo Souto, and Olga Zamora.

# Chapter 2

# Kepler Objects of Interest in APOGEE: Two Distinct Orbital Period Regimes Inferred from Host Star Iron Abundances

## Summary

The Apache Point Observatory Galactic Evolution Experiment (APOGEE) has, at the time of this work, observed ∼600 transiting exoplanets and exoplanet candidates from *Kepler* (Kepler Objects of Interest, KOIs), most with ≥18 epochs. The combined multi-epoch spectra are of high signal-to-noise (typically ≥100) and yield precise stellar parameters and chemical abundances. We first confirm the ability of the APOGEE abundance pipeline, ASPCAP, to derive reliable [Fe/H] and effective temperatures for FGK dwarf stars — the primary *Kepler* host stellar type — by comparing the ASPCAP-derived stellar parameters to those from independent high-resolution spectroscopic characterizations for 221 dwarf stars in the literature. With a sample of 282 close-in ($P < 100$

days) KOIs observed in the APOGEE KOI goal program, we find a correlation between orbital period and host star [Fe/H] characterized by a critical period, $P_{\rm crit}$= $8.3^{+0.1}_{-4.1}$ days, below which small exoplanets orbit statistically more metal-enriched host stars. This effect may trace a metallicity dependence of the protoplanetary disk inner-radius at the time of planet formation or may be a result of rocky planet ingestion driven by inward planetary migration. We also consider that this may trace a metallicity dependence of the dust sublimation radius, but find no statistically significant correlation with host $T_{\rm eff}$ and orbital period to support such a claim.

## 2.1 Introduction

With the advent of the *Kepler* mission (Koch et al. 2010; Borucki et al. 2010; Borucki 2016), statistical studies of exoplanets, particularly small planets ($R_p \lesssim 4R_\oplus$), have become possible. While a key finding of such studies is that small planets are common in the Galaxy in general (e.g., Howard et al. 2012; Dressing & Charbonneau 2013; Petigura et al. 2013; Batalha 2014; Burke et al. 2015; Silburt et al. 2015), distinguishing the characteristics of these planets and how they may relate to the properties of their host stars is of interest from formation and detection perspectives. From population studies of larger planets detected by the radial velocity method, it was clear early on that host star metallicity[1] was related to the frequency at which these planets form (Gonzalez 1998; Heiter & Luck 2003a; Santos et al. 2004; Valenti & Fischer 2005), a trend that appears to decrease in strength with decreasing planet mass and/or radius (e.g., Sousa et al. 2008; Ghezzi et al. 2010; Schlaufman & Laughlin 2011; Buchhave et al. 2012; Wang & Fischer 2015; Buchhave & Latham 2015). Now, the most prevalent explanation of this trend is that it is evidence of the core accretion method of planet formation (e.g., Rice & Armitage 2003; Ida & Lin 2004; Alibert et al. 2011; Mordasini et al. 2012b), and that host star

[1]Usually parameterized by the number density of iron nuclei in a star's photosphere relative to the amount of hydrogen, normalized to these values in the Sun: [Fe/H], where [X/H]=log($N_X/N_H$) - log $(N_X/N_H)_{solar}$

metallicity is a proxy for the surface density of the solid material in a protoplanetary disk; higher solid surface densities facilitate the faster growth of the solid cores of larger planets, giving them more time to accrete gaseous envelopes.

In addition to the trend between host star [Fe/H] and the frequency of different types of planets, other relationships between stellar metallicity and planet properties have also come to light. For instance, Dawson & Murray-Clay (2013) used the evidence that giant planets orbiting [Fe/H]<0 stars generally have lower eccentricity orbits to suggest that planet-planet scattering is the dominant mechanism for inward migration of giant planets, since higher [Fe/H] systems are more likely to form multiple, closely spaced giant planets. Buchhave et al. (2014) analyzed ground-based optical spectra of *Kepler* Objects of Interest (KOIs) to measure spectroscopic metallicities and found three regimes of exoplanet sizes, split at $R_p \sim 1.7\ R_\oplus$ and $R_p \sim 3.9\ R_\oplus$, distinguished by different (increasing with $R_p$) host star metallicities. Using the same data but with a more rigorous statistical analysis, Schlaufman (2015) instead favored a single, continuous relationship between planet radius and stellar metallicity. Interestingly, recent results show that (i) 2-6 $R_\oplus$ planets with orbital periods from 1-10 days ("hot Neptunes") show an increase in host star [Fe/H] compared to typical planet-hosting stars, similar to that of hot Jupiter ($R_p \geq 10 R_\oplus$) planets (Dong et al. 2017), and (ii) at $<1$ day orbital periods, $\leq 2 R_\oplus$ planet host stars have significantly different metallicities than hot Jupiter host stars but similar metallicities as stars hosting $2-4 R_\oplus$ planets with 1-10 day periods (Winn et al. 2017). These studies exemplify the (evidently) intricate relationship between the metallicities of host stars and the sizes and orbital configurations of the planets that form around them.

We intend to further characterize this intricate relationship by investigating how planet orbital period is tied to host star metallicity. This is a topic that has recently been explored by several other studies.

Beaugé & Nesvorný (2013) examined both confirmed exoplanet systems and *Kepler*

candidate multi-planet systems to show (i) a lack of small ($R_p \lesssim 4\ R_\oplus$), short period ($P < 5$ days) planets around low metallicity (bulk [m/H] $< -0.2$ dex, from Buchhave et al. 2012) stars, and (ii) a dearth of 4-8 $R_\oplus$ planets at $P \leq 100$ days around low metallicity stars. At the time, trends also held in the planetary mass versus period plane; e.g., $M_p \sin i < 0.05 M_{\rm Jup}$ planets in short orbits were not found around [Fe/H] $< -0.2$ dex stars and planets between the masses of Neptune and Saturn with $P \leq 100$ days were mostly absent around [Fe/H] $< -0.2$ dex stars. The authors explained these observed trends with delayed formation and less planetary migration in metal-depleted protoplanetary disks. We note that the trends in Beaugé & Nesvorný (2013) are slightly reduced in significance when more up-to-date planet samples are considered (see Dawson et al. 2015, discussed below).

Similarly, Adibekyan et al. (2013) found from the HARPS GTO radial velocity survey (Mayor et al. 2003; Lo Curto et al. 2010; Santos et al. 2011) that $\sim$0.03 $M_{\rm Jup}$ to 4 $M_{\rm Jup}$ planets orbiting stars with [Fe/H] $< -0.1$ dex have longer periods than the same mass planets orbiting stars with [Fe/H] $> -0.1$ dex. Specifically, Adibekyan et al. find all $M_p \sin i < 0.03 M_{\rm Jup}$ planets orbiting [Fe/H] $> -0.1$ dex stars have periods $< 18$ days, and also suggest that smaller planets orbiting more metal-rich stars are more likely to migrate towards or form close to their host stars compared to planet orbiting more metal-poor stars.

Dawson et al. (2015) explored a theoretical framework motivated by these observational trends, combining analytical estimates for the formation of planetary embryos (that merge to form super-Earths and the cores of mini-Neptunes) with numerical simulations of atmospheric accretion in disks having varying solid surface densities. Interpreting their model predictions in the context of easily observed quantities (planet radius and host star metallicity), Dawson et al. find that disks with high solid surface density (metallicity) generate 2 $M_\oplus$ cores before the gas disk dissipates ($\sim$1 Myr), which enables the cores to more readily accrete significant atmospheres and thus increase their gas to rock fraction

($R_p$). Furthermore, these authors find a match with current observations – i.e., that metal-rich stars lack rock-dominated ($R_p < 1.5\ R_\oplus$) planets beyond $\sim$ 15 day periods – and suggest that this may indicate that embryo, and thus final core, masses grow faster at larger orbital distances in metal-rich versus metal-poor disks, thus producing gas-enveloped, larger $R_p$ planets.

Most recently, Mulders et al. (2016) used over 20,000 stars observed by both *Kepler* and LAMOST (Cui et al. 2012) to confirm that short period planets ($\lesssim$ 10 days) are preferentially found around more metal-rich stars ([Fe/H] $\simeq 0.15 \pm 0.05$ dex), whereas longer period planets orbit roughly solar metallicity ([Fe/H]$\sim$ 0) host stars. In the $P < 10$ day sample, it is the smallest radius planets ($< 1.7\ R_\oplus$) that have the largest host [Fe/H] contrast compared to their similarly-sized but longer period counterparts, with an occurrence-weighted $\Delta$[Fe/H] $\simeq\ 0.25 \pm 0.07$. Mulders et al. suggest that their results may be evidence that the inner edges of protoplanetary disks around more metal-rich stars are closer in than around more metal-depleted stars. The trend observed by Mulders et al. is in contrast to the assessment by Winn et al. (2017), who comment that their metallicities (from Petigura et al. 2017a, using HIRES/Keck data from the California *Kepler* Survey) of small planet host stars show no such period dependence. Differences in sample selection may influence the differences in the Mulders et al. vs. Winn et al. results.

Many of the works above use moderate to high resolution optical spectroscopy to derive host star parameters. Indeed, the original *Kepler* Input Catalog (KIC) was not intended for detailed studies of host star metallicity (Brown et al. 2011), which motivated numerous follow-up spectroscopic campaigns to better characterize KOIs (e.g., Bruntt et al. 2012; Buchhave et al. 2012, 2014; Everett et al. 2013; Dong et al. 2014; Brewer et al. 2016; Petigura et al. 2017a). In this work we present a study of host star [Fe/H] versus planetary orbital period using high resolution near infrared spectroscopy of KOIs taken by the Sloan Digital Sky Survey's Apache Point Observatory Galactic Evolution

Experiment (APOGEE, Majewski et al. 2017). In §2.2 we discuss the APOGEE stellar parameter derivation, and validate the [Fe/H] and $T_{\rm eff}$ values produced by APOGEE's automated stellar parameter pipeline (ASPCAP) by comparing its output to the results from several literature studies. §2.3 explains the data collection for our KOI sample. In §2.4 we present our analysis of the KOI planet and host star parameters, focusing on orbital period and [Fe/H], and conclude in §2.5 and §2.6 with the interpretation of our results and final conclusions, respectively.

## 2.2 Validating APOGEE Spectroscopic Parameters

All the data in this work were collected as part of APOGEE in the fourteenth Data Release (DR14, Abolfathi et al. 2018) of the third and fourth Sloan Digital Sky Survey (Eisenstein et al. 2011; Blanton et al. 2017). APOGEE utilizes a multi-object spectrograph (Wilson et al. 2010, 2012) mounted on the Sloan 2.5 m telescope (Gunn et al. 2006) to sample up to 300 sources simultaneously with high resolution ($R \sim 22,500$), high signal-to-noise ratio (SNR>100), $H$-Band (1.5–1.7 $\mu$m) spectroscopy. Details on the motivation and scope of the APOGEE survey are described in Majewski et al. (2017) and the targeting is described in Zasowski et al. (2013). All of the data from APOGEE is processed through automated reduction and stellar parameter pipelines (Nidever et al. 2015; Holtzman et al. 2015), and the spectroscopic parameters used for the stars in our sample are derived from the Automated Stellar Parameters and Chemical Abundances Pipeline (ASPCAP). We give a brief overview of ASPCAP here for convenience, but for details on the pipeline we refer the reader to García Pérez et al. (2016).

ASPCAP consists of two principle components: a `fortran90` optimization code (`ferre`, Allende Prieto et al. 2006[2]) that compares the observed APOGEE spectra to

[2]Available from `github.com/callendeprieto/ferre`

synthetic libraries, and a multifunctional IDL wrapper used for bookkeeping and reading and preparing the input APOGEE spectra. FERRE performs a $\chi^2$ minimization to find the best-fit set of atmospheric parameters (effective temperature, $T_{\rm eff}$; surface gravity, $\log g$; microturbulent velocity, $\xi_t$; and general solar-scaled metallicity, [M/H]) as well as C, N, and $\alpha$-element abundances from an interpolated library of synthetic ATLAS9 or MARCS model atmospheres. The atomic and molecular line list, gathered from the literature, has been updated regularly and for DR14 is described most recently in Shetrone et al. (2015) and Holtzman et al. (2018).

Once fundamental atmospheric parameters are found, ASPCAP extracts individual chemical abundances by fitting spectral windows optimized for each element. Iron has dozens of Fe I lines in the $H$-band, and Fe abundances are computed using ∼55 spectral windows. ASPCAP provides both raw and calibrated values for all of its spectroscopic parameters. Calibrated $T_{\rm eff}$ values are established using observations of globular and open clusters, and by requiring that there are no trends of abundances with $T_{\rm eff}$ in clusters (Holtzman et al. 2018). In this study, we adopt the DR14 calibrated $T_{\rm eff}$ and metallicity ([Fe/H]) values.

Because ASPCAP is optimized for red giants and not well tested for dwarfs, it is worthwhile to test the pipeline's performance against published spectroscopic studies of dwarfs. We select comparison studies focused on planet search targets, because the stellar samples are similar to those in our study (i.e., consisting mostly of FGK dwarfs). We compare first to four large surveys that derive stellar parameters using spectral synthesis (Bruntt et al. 2012; Buchhave et al. 2012; Huber et al. 2013; Brewer et al. 2016), each with enough stars to give a substantive comparison of ASPCAP's performance. These data and comparisons are shown in Table 2.1 and Figure 2.1, top panels. We also compare ASPCAP's performance to equivalent width analysis studies of detailed chemical abundances (Ghezzi et al. 2010; Adibekyan et al. 2012b; Nissen et al. 2014; Schuler et al. 2015) to better gauge the pipeline's performance. These data and comparisons are

shown in Table 2.2 and Figure 2.1, bottom panels. The equivalent width studies employ different methodologies, which allows us to compare ASPCAP against multiple strategies for deriving spectroscopic parameters. Summaries of these comparisons with ASPCAP's performance are given in §2.2.2 and §2.2.3.

### 2.2.1 ASPCAP Internal Errors on [Fe/H]

Each of the comparisons described below represents an estimate of the relative accuracy of the ASPCAP stellar parameter results, but not their precision. To first assess the internal error on the ASPCAP [Fe/H] values, we turn to the sample of stars in the solar-age open cluster M67 that was observed by APOGEE, with parameters derived in the same way as our sample of KOIs. From the astrometric survey of Yadav et al. (2008), we selected the stars that were observed by APOGEE having the highest M67 membership probability ($\geq$ 90%), and also had ASPCAP uncalibrated $\log g$ values[3] $\geq$ 4.0 and measured [Fe/H] values. From this sample of 76 stars, the [Fe/H] median is -0.021 dex, the mean is -0.018 dex, and the standard deviation is 0.073 dex. However, as described below (§2.4.1), in our analysis we include only high SNR (>100) spectra and exclude stars with $T_{\rm eff} < 4000$ K; performing the same cut to the M67 sample results in 46 stars with a [Fe/H] median of -0.016 dex, a mean of -0.011 dex, and a standard deviation of 0.053 dex.

M67 is known to be chemically inhomogeneous – Liu et al. (2016) found a metallicity difference of $\sim$ 0.05 dex between a solar twin and a solar analog in M67, as well an enhancement of $\sim$ 0.05 dex in neutron-capture elements in the solar analog versus the solar twin. Additionally, using SDSS-III DR12 APOGEE data, Bertran de Lis et al. (2016) found the spread in [O/Fe] in cool, low-gravity stars ($4000<T_{\rm eff}<4600$ K, $\log g < 3.8$) in M67 to be higher ($\sigma[{\rm O/Fe}]_{err} \sim 0.03$ dex) as compared to other solar or super-solar metallicity clusters NGC 6791 and NGC 6819, which show $\sigma[{\rm O/Fe}]_{err} \lesssim 0.01$

[3]The spectroscopic surface gravities for dwarfs in APOGEE DR14 are significantly lower than what is expected from stellar isochrones, and an acceptable calibration relation has not yet been developed by the ASPCAP team.

dex. However, only a handful of dwarf stars in NGC 6791 and NGC 6819 were targeted by APOGEE, and even fewer make our SNR cut. Thus we adopt the $\sigma$[Fe/H] value from M67, 0.053 dex, as a conservative error (since a significant part of the spread in [Fe/H] is likely intrinsic to the cluster) on the ASPCAP metallicities for the KOIs in our analysis henceforth.

### 2.2.2 Comparison to Literature – Spectral Synthesis Studies

Bruntt et al. (2012) utilized the analysis package VWA (Bruntt et al. 2010) to derive stellar parameters and elemental abundances for a sample of 93 G dwarfs in the *Kepler* field. Their data consist of high resolution ($R \approx 80,000$), and high SNR ($\sim 200 - 300$) optical spectra. The accuracy of their parameters resulted from adopting asteroseismic $\log g$'s, which they held fixed to derive the rest of their parameters. Bruntt et al. (2012) reported typical uncertainties in their $T_{\rm eff}$ and [Fe/H] determinations of 80 K and 0.07 dex, respectively. The overlapping sample with APOGEE contains 71 stars. The difference (ASPCAP−Other) in the effective temperature ($\Delta T_{\rm eff}$) and iron abundance ($\Delta$[Fe/H]) determinations between ASPCAP and Bruntt et al. (2012) have mean offsets and RMS scatter of $48 \pm 147$ K and $0.00 \pm 0.07$ dex.

Buchhave et al. (2012) systematically derived spectroscopic and stellar parameters for a sample of 152 planet-hosting stars (PHS's) with 1500 observations from multiple telescopes and spectrographs, having SNR $\geq 30$. The final stellar parameters were determined from the average of all the measurements of each particular star. For this study, the authors developed their own analysis package, SPC, designed to analyze spectra with low to modest SNR. The authors claimed typical uncertainties in the $T_{\rm eff}$ and [Fe/H] of 50 K and 0.08 dex respectively, which were derived as the scatter among their sample of measurements for each star. The overlap between Buchhave et al. (2012) and APOGEE is 75 stars. From these, the mean offset and RMS scatter for $\Delta T_{\rm eff}$ and $\Delta$[Fe/H] are

Table 2.1. Spectral Synthesis Study Comparison Parameters

| APOGEE ID | ASPCAP | | Bruntt et al. 2012 | | Bucchave et al. 2012 | | Huber et al. 2013 | | Brewer et al. 2016 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $T_{\rm eff}$ (K) | [Fe/H] | $T_{\rm eff}$ (K) | [Fe/H] | $T_{\rm eff}$ (K) | [m/H] | $T_{\rm eff}$ (K) | [Fe/H] | $T_{\rm eff}$ (K) | [Fe/H] |
| 2M19452396+4404359 | 6425 | 0.04 | 6344 | 0.01 | – | – | – | – | – | – |
| 2M19455565+4400329 | 4882 | 0.08 | – | – | 5082 | 0.03 | 5009 | -0.02 | – | – |
| 2M19480452+5024323 | 5927 | 0.10 | – | – | 6040 | 0.11 | – | – | 6008 | 0.18 |
| 2M19542140+4045024 | 6096 | -0.03 | – | – | – | – | 6081 | -0.03 | – | – |
| 2M20015142+4421140 | 6070 | -0.33 | 6114 | -0.36 | – | – | – | – | – | – |

Note. — This table is available in its entirety in a machine-readable form online. A portion is shown here for guidance regarding its form and content.

$43 \pm 147$ K and $-0.04 \pm 0.09$ dex, respectively.

Huber et al. (2013) produced a catalog of fundamental stellar parameters for 66 PHSs in the *Kepler* field. Huber et al. (2013) used a combination of SME (Valenti & Piskunov 1996) and SPC to obtain initial guesses of the $T_{\rm eff}$ , [Fe/H] and $\log g$. Stellar $\log g$'s were then fixed to asteroseismic solutions (derived using the initial $T_{\rm eff}$ and [Fe/H] from SPC and SME), and Huber et al. (2013) re-derived $T_{\rm eff}$ and [Fe/H] with the asteroseismic constraints in place. Like Buchhave et al. (2012), the Huber et al. (2013) data come from multiple telescopes, have modest SNR, and were analyzed in the same way for their initial $T_{\rm eff}$ and [Fe/H] measurements. The authors reported an average uncertainty of 82 K and 0.1 dex for $T_{\rm eff}$ and [Fe/H], respectively. All 66 stars in Huber et al. (2013) were observed with APOGEE. The mean offset and scatter for $\Delta T_{\rm eff}$ and $\Delta$[Fe/H] are $52 \pm 105$ K and $0.02 \pm 0.10$ dex, respectively.

Brewer et al. (2016) used SME (Valenti & Piskunov 1996) to provide a catalog of spectral and stellar properties for 1615 FGK dwarfs. All observations were taken with the HIRES spectrograph (Vogt et al. 1994) at the Keck I telescope with resolution $R \approx 70,000$, but vary in SNR (about 25% of their stars have SNR $< 100$, and the rest have SNR$\geq 100$). Brewer et al. (2016) note a strong dependence of their derived parameters with SNR. Of the Brewer et al. (2016) subsample that was also observed with APOGEE, only five stars have SNR $< 100$. These authors reported a typical uncertainty in $T_{\rm eff}$ of $\sim 25$ K and a typical [Fe/H] uncertainty between $\sim 0.01 - 0.04$ dex. Of the 1615 stars in their sample, 60 have also been observed by APOGEE. In this overlapping sample, we find a mean offset and scatter for $\Delta T_{\rm eff}$ and $\Delta$[Fe/H] of $82 \pm 126$ K and $0.06 \pm 0.10$ dex, respectively.

### 2.2.3 Comparison to Literature – Equivalent Width Studies

In addition to the large surveys described above, we compare ASPCAP's results

Table 2.2. Equivalent Width Study Comparison Parameters

| APOGEE ID | ASPCAP | | Adibekyan et al. 2012 | | Schuler et al. 2015 | | Nissen et al. 2014 | | Ghezzi et al. 2010 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $T_{\rm eff}$ (K) | [Fe/H] | $T_{\rm eff}$ (K) | [Fe/H] | $T_{\rm eff}$ (K) | [Fe/H] | $T_{\rm eff}$ (K) | [Fe/H] | $T_{\rm eff}$ (K) | [Fe/H] |
| 2M02360498+0653140 | 4781 | 0.01 | – | – | – | – | – | – | 4922 | -0.21 |
| 2M02515835+1122119 | 5809 | -0.70 | – | – | – | – | 5873 | -0.68 | – | – |
| 2M03402202−0313005 | 5996 | -0.72 | 5884 | -0.82 | – | – | 5859 | -0.86 | – | – |
| 2M13121982+1731016 | 6112 | -0.57 | – | – | – | – | 5956 | -0.72 | 5837 | -0.72 |
| 2M19134816+4014431 | 5875 | -0.17 | – | – | 5958 | -0.20 | – | – | – | – |

Note. — This table is available in its entirety in a machine-readable form online. A portion is shown here for guidance regarding its form and content.

to a few select studies that have computed stellar parameters and derived elemental abundances through equivalent width (EW) measurements. Specifically, we compare ASPCAP's results with Ghezzi et al. (2010), Adibekyan et al. (2012b), Nissen et al. (2014), and Schuler et al. (2015). In all of the following studies, stellar parameters were determined by adjusting the parameters until there was no correlation between the [Fe/H] values derived from Fe I lines and the lower excitation potential ($\chi$) of the lines, nor between [Fe I/H] and reduced EW [log (EW/$\lambda$)], and until there was agreement between abundances derived from Fe I and Fe II lines.

Ghezzi et al. (2010) obtained spectra of a sample of 117 PHS's from the Fiber-fed Extended Range Optical Spectrograph (FEROS; Kaufer et al. 1999). The setup chosen by Ghezzi et al. (2010) resulted in spectral coverage from 3560 to 9200 Å, and a nominal resolution of $R \approx 48,000$. The reported typical SNR per resolution element ranges from $\sim 200 - 500$. EWs were measured using the code ARES (Sousa et al. 2007) and stellar parameters were derived using the 2002 version of MOOG (Sneden 1973) assuming local thermodynamic equilibrium (LTE). The uncertainties for this sample were derived with the method outlined in Gonzalez (1998). Ghezzi et al. (2010) report typical $T_{\rm eff}$ uncertainties ranging from $\sim 30 - 70$ K and typical [Fe/H] uncertainties ranging from $\sim 0.02 - 0.05$ dex. For the four stars observed by APOGEE and Ghezzi et al. (2010), the mean offset and RMS scatter in $\Delta T_{\rm eff}$ between ASPCAP and Ghezzi et al. is $51 \pm 195$ K, and the mean offset and scatter in $\Delta$[Fe/H] is $0.07 \pm 0.14$ dex, indicating a fair agreement.

Adibekyan et al. (2012b) obtained a sample of 1111 FGK dwarfs from the HARPS (High Accuracy Radial velocity Planet Searcher) GTO (Guaranteed Time Observations) planet search program (Mayor et al. 2003; Lo Curto et al. 2010; Santos et al. 2011). The spectra taken from the HARPS spectrograph (Mayor et al. 2003) have spectral resolution $R \approx 110,000$ and a SNR ranging from $\sim 20 - 2000$, where 84% of the sample has SNR $\geq 100$. Their sample consists of dwarfs similar in $T_{\rm eff}$ to the Sun, the majority of which lie within 4500K to 6500K and with metallicities ranging from -1.39 to 0.55 dex. The

analysis was completed in a similar manner as Ghezzi et al. (2010), by assuming LTE, generating a grid of Kurucz model atmospheres (Kurucz 1993), and making use of ARES and MOOG. However, because Ghezzi et al. limited their line list to lines with $\log gf$ values measured in the lab, Adibekyan et al. (2012b)'s analysis benefits from a more extensive line list. The uncertainties from Adibekyan et al. (2012b) were determined by adding quadratically the uncertainties in the parameters of the atmospheric model and the scatter measured amongst the abundances of each individual line. Since HARPS is in the Southern Hemisphere, there are only eight stars that were also observed with APOGEE. The reported uncertainties were typically about 30K for $T_{\rm eff}$ and 0.03 dex in [Fe/H]. The mean offset and rms scatter of $\Delta T_{\rm eff}$ and $\Delta$[Fe/H] with respect to ASPCAP are $-6 \pm 92$ K and $-0.08 \pm 0.09$ dex, respectively.

Nissen et al. (2014) measured the C/O ratio in a sample of 66 sun-like stars, with $T_{\rm eff}$ ranging from 5400 K to 6400 K. The Nissen et al. (2014) data are from both HARPS and FEROS, with the same configurations as Adibekyan et al. (2012b) and Ghezzi et al. (2010). The sample of stars observed with HARPS has SNR $\gtrsim$ 300, while the FEROS-observed stars have a typical SNR $\sim$ 200. Nissen et al. (2014) derive $T_{\rm eff}$ in their HARPS-FEROS sample using photometric calibrations, and initially derive [Fe/H] values by interpolating within the plane-parallel model atmosphere MARCS grid (Gustafsson et al. 2008). They then derive more accurate [Fe/H] values by measuring the EWs of Fe II lines using the Uppsala program EQWIDTH. Nissen et al. report a typical uncertainty in [Fe/H] of 0.03 dex, and internal, one-sigma $T_{\rm eff}$ errors of 30 K. Though Nissen et al. do not explicitly describe how they derive their errors, they state their errors are drawn from uncertainties in their model atmosphere fits and equivalent width measurements. Comparing the eight stars in Nissen et al. that were also observed by APOGEE, the mean offset and rms scatter of $\Delta T_{\rm eff}$ and $\Delta$[Fe/H] are $8 \pm 230$ K and $-0.04 \pm 0.14$ dex, respectively.

Schuler et al. (2015) derived stellar parameters and elemental abundances for seven

PHS's identified by *Kepler*. The data were collected from HIRES on the Keck I telescope as part of the *Kepler* Follow-up Observing Program (KFOP, Gautier et al. 2007). The KFOP spectra have a spectral coverage of 3650-7950 Å, and a spectral resolution of $R \approx 50,000$. Schuler et al. only considered data with SNR $\geq 150$. Stellar parameters and abundances for this study were extracted using an LTE, curve-of-growth analysis. EWs were measured using the analysis package SPECTRE, and abundances were derived from the 2014 version of MOOG, along with synthetic fits to the data interpolated from the ATLAS9 Kurucz model atmosphere grid. Uncertainties in $T_{\rm eff}$ are reported as the difference between the adopted $T_{\rm eff}$ value and the value that results in a $1\sigma$ correlation between the [Fe/H] vs. $\chi$ and reduced EW relations. Schuler et al. report uncertainties in $T_{\rm eff}$ between 25 K and 45K, and uncertainties in [Fe/H] between 0.04 and 0.08 dex. In the Schuler et al. sample, all seven stars in the Schuler et al. sample were also observed with APOGEE, resulting in a mean offset and scatter between the two studies for $\Delta T_{\rm eff}$ and $\Delta$[Fe/H] of $110 \pm 119$ K and $-0.02 \pm 0.06$ dex, respectively.

Overall, after these various comparisons with multiple studies, we find that ASPCAP is accurate for $T_{\rm eff}$ and [Fe/H] within the scatter, and agree with these multiple optical studies that utilize different methodologies. We note that the mean offset in $T_{\rm eff}$ (57 K) versus the synthesis studies indicates that ASPCAP may be underreporting the $T_{\rm eff}$ compared to these other studies. However, we find almost no offsets in $T_{\rm eff}$ as compared to the studies of Nissen et al. (2014) and Adibekyan et al. (2012b). Furthermore, because -57 K is well within the RMS scatter for each of these studies, we do not consider it to be problematic for our purposes. Taking all these comparison studies into account, we find that the mean offset and RMS scatter in $\Delta$[Fe/H] (0.004 dex and 0.10 dex, respectively) are within the uncertainties required for this work.

Having validated the performance of ASPCAP, we now move on to the study of a particular subset of APOGEE data consisting of of repeated observations of *Kepler* objects of interest (KOIs) resulting in high SNR spectra.

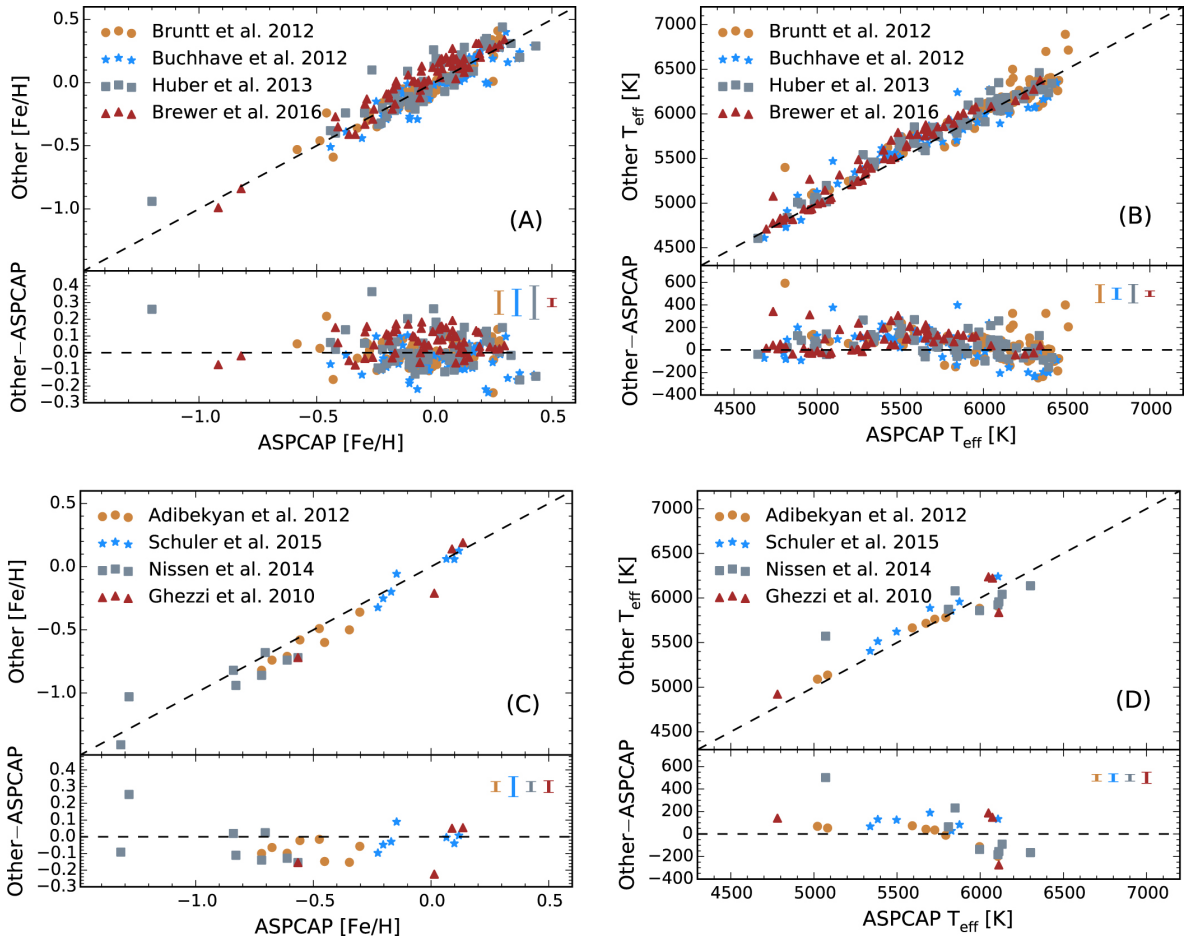


Fig. 2.1.— Comparisons of APOGEE [Fe/H] and $T_{\rm eff}$ measurements to values determined in the literature. Each plot shows the $T_{\rm eff}$ and [Fe/H] values from ASPCAP and the literature plotted against each other, as well as the difference (Other-ASPCAP). For the top panels, the dotted black line shows the one-to-one relationship, and on the bottom panels, the dotted black line shows the line of no difference. Typical errors from the literature studies are shown in the top right of each bottom panel. (A) ASPCAP's [Fe/H] determinations compared to measurements from synthesis studies. As a whole, the difference (Other−ASPCAP) shows good agreement with an RMS scatter of 0.09 dex, and a mean offset of 0.00 dex. (B) APOGEE $T_{\rm eff}$ measurements compared to those from synthesis studies. Though there is a slight bump around $T_{\rm eff} \sim 5500$ K, APOGEE shows excellent overall agreement with these surveys, with an RMS scatter of 129 K and mean offset of 57 K. (C) APOGEE's [Fe/H] determination compared to equivalent width studies in the literature. The difference in measured iron abundances by APOGEE and these studies show a mean difference of −0.05 dex and an RMS scatter of 0.11 dex. (D) APOGEE's $T_{\rm eff}$ determination compared to equivalent width studies in the literature. The difference in measured temperatures by APOGEE and these dedicated studies show a mean offset of 36 K and an RMS scatter of 166 K.

## 2.3 APOGEE KOI RV Sample

The primary goal of APOGEE, now in its second phase APOGEE-2 (Majewski et al. 2017), is to study the Milky Way through the radial velocities (RVs) and chemical abundances of as many as half a million stars, chosen to be primarily red giants across multiple stellar populations and Galactic regions. Additional science programs are also included in the survey, with one such program monitoring KOIs to search for false positives through RV variations (Fleming et al. 2015). APOGEE data reach an RV precision of $\sim$100 m s$^{-1}$ (Troup et al. 2016), allowing the search for eclipsing binaries and other grazing incidence geometries that may resemble transiting planets in the initial *Kepler* reduction pipeline. The APOGEE survey has observed $\sim$1500 KOIs with $\geq$18 epochs across eight APOGEE-2 fields (roughly the size of *Kepler* tiles), each KOI with a sufficient signal-to-noise ratio to get quality RVs at each epoch. As a result, the final "RV-normalized", *summed* spectra over all epochs are of very high SNR (typically a few hundred), which allows for the derivation of high-precision stellar parameters and elemental abundances for planet-hosting stars. The APOGEE targets were chosen with the goal of observing all possible "Confirmed" or "Candidate" KOIs with $H < 14$ in those five *Kepler* tiles. Some KOIs were excluded from the sample on the basis of unphysical impact parameters and planet radii consistent with stellar values. At the time of this work, APOGEE has observed $\sim$600 KOIs from the *Kepler* DR24 Q1-Q16 catalog (Mullally et al. 2015), orbiting $\sim$450 PHS, each with between 10 and 28 epochs at the time of this study.

Data concerning the orbital and planetary parameters for each KOI were gathered using the public NASA Exoplanet Archive, which provides the information in the form of interactive tables of confirmed and candidate planetary and stellar properties and includes a suite of integrated analysis tools (Akeson et al. 2017). Use of this archive allows us to exclude known false positives and ensures that we are using the most up-to-date KOI dispositions in the literature. The specific targets included in our analysis, vetted from

the ∼600 KOIs observed thus far by APOGEE-2, are described below (§2.4.1).

## 2.4 KOI Stellar Metallicity and Planet Period Relation

### 2.4.1 Selected Sample

All of the stars in our KOI sample were observed as part of the APOGEE KOI Goal Program (Fleming et al. 2015), as described above. Initially, that consists of 624 KOIs and 450 PHS's. To ensure the quality of the data, we restrict our analysis sample using a series of APOGEE flags and other constraints.[4] We first remove data with any of the STARFLAGS BAD_PIXELS, VERY_BRIGHT_NEIGHBOR, and LOW_SNR set. It is worth mentioning that a fraction of our sample falls on the high-persistence region of APOGEE's "blue chip". However, we decide to keep these data since persistence effects were shown to be minimized in DR14 (Holtzman et al. 2018). To exclude unreliable ASPCAP fits defined as values close to the edge of the model atmosphere grid, we also remove data with any of the following ASPCAPFLAGS: TEFF_BAD, LOGG_BAD, and METALS_BAD set. Because the focus of this study is [Fe/H], we remove KOIs with the PARAMFLAGS GRIDEDGE_BAD, CALRANGE_BAD, OTHER_BAD, and PARAM_FIXED flags set, with respect to the [Fe/H] parameter. In addition, we require that all of the summed APOGEE spectra in our sample have SNR $> 50$. Because the ASPCAP line list does not include FeH lines in DR14, which are important for modeling the metallicities of M dwarfs, we exclude all stars with $T_{\rm eff} < 4000\,{\rm K}$ (for a detailed discussion on this point see Souto et al. (2017)).

In addition to the cuts described above, our sample is further vetted based on the orbital periods and inferred planet radii of the KOIs. To avoid regions of parameter space associated with low survey completeness, we include only KOIs with orbital periods

[4]Descriptions of the DR14 APOGEE flags can be found at `http://www.sdss.org/dr14/algorithms/bitmasks/#APOGEE_TARGET1`

$P < 100$ days.[5] For multiple planet systems, only the planet with the shortest orbital period makes it into our analysis. In an effort to remove eclipsing binaries (EBs), we make several additional cuts. We restrict our sample to KOIs with inferred radii for the planet candidates $R_p < 20\ R_\oplus$. Planet candidates having radii larger than this limit are likely to be (EBs). To correct for false positives from our sample by removing eight known EBs identified in the literature and identify eight more likely binaries from visual inspection of spectra for which ASPCAP reported high $v \sin i$ values. In these cases, the reported $v \sin i$ values were a result of the combined spectra from the primary and companion stars. To remove more potential EBs, we identify stars with high RV variability, defined as the ratio of the scatter of the RV measurements to the error of the RV measurements, given by

$$\frac{\text{VSCATTER}}{\text{VERR\_MED}} > 17, \tag{2.1}$$

where VERR_MED is the median RV measurement error from all visits and VSCATTER is the RMS scatter of all the RV measurements. Because the RV measurement errors are often underreported in APOGEE (Troup et al. 2016), the adopted cutoff is the median value for the 16 known binaries in our sample ($\sim$17). After these cuts, and selecting only the inner-most planet for each system, our final sample consists of 282 KOIs, listed in Table 3.

The planet hosts in our sample are all FGK dwarfs with effective temperatures ranging from 4000 K – 6500 K. The [Fe/H] of our sample range from -0.6 – 0.4 dex. This parameter space is well covered by our tests of ASPCAP using literature comparisons (§2.2.2, §2.2.3). The spectral SNR in our KOI sample from APOGEE have a wide range; the inner 68% ranges in SNR from 70–280, while the asymmetric distribution peaks at SNR $\sim$ 140 with a tail to SNR $\gtrsim$ 500.

[5] For an estimate of survey completeness as a function of planet radius and orbital period, see Burke et al. (2015)

Table 2.3. Parameters of Selected Sample

| KOI | KIC | Period | Period Error | | Planet Radius | Planet Radius Error | | $K_p$ | APOGEE ID | $T_{\rm eff}$ | [Fe/H] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Upper Bound | Lower Bound | | Upper Bound | Lower Bound | | | | |
| | | (days) | (days) | (days) | ($R_\oplus$) | ($R_\oplus$) | ($R_\oplus$) | (mag) | | (K) | (dex) |
| K00041.02 | 6521045 | 6.89 | 2.28E-05 | -2.28E-05 | 1.30 | 0.08 | -0.07 | 11.20 | 2M19253263+4159249 | 5827 | 0.07 |
| K00049.01 | 9527334 | 8.31 | 4.21E-05 | -4.21E-05 | 2.74 | 0.43 | -0.13 | 13.70 | 2M19285977+4609535 | 5831 | -0.07 |
| K00084.01 | 2571238 | 9.29 | 3.83E-06 | -3.83E-06 | 2.10 | 0.26 | -0.10 | 11.90 | 2M19214099+3751064 | 5437 | -0.02 |
| K00100.01 | 4055765 | 9.97 | 1.22E-05 | -1.22E-05 | 16.89 | 2.15 | -4.29 | 12.60 | 2M19244270+3911581 | 6336 | -0.32 |
| K00103.01 | 2444412 | 14.91 | 1.28E-05 | -1.28E-05 | 2.62 | 0.33 | -0.17 | 12.59 | 2M19264400+3745057 | 5485 | 0.06 |
| K00119.01 | 9471974 | 49.18 | 2.48E-05 | -2.48E-05 | 8.20 | 0.50 | -0.55 | 12.65 | 2M19381420+4603443 | 5584 | 0.33 |
| K00156.02 | 10925104 | 5.19 | 8.92E-06 | -8.92E-06 | 1.0 | 0.08 | -0.07 | 13.74 | 2M19362914+4820582 | 4662 | 0.23 |

Note. — This table is available in its entirety in a machine-readable form online. A portion is shown here for guidance regarding its form and content.

### 2.4.2 Analysis & Results

As discussed in §2.1, a correlation between an exoplanet's orbital period and the metallicity of its host star could indicate that protoplanetary disks with higher solid surface density cause planets to migrate or form closer to their host stars (Beaugé & Nesvorný 2013; Adibekyan et al. 2013; Mulders et al. 2016). Alternatively or in addition, such a correlation could mean metal-rich disks spawn planet cores faster at larger orbital distances, facilitating the cores growing into gas-enveloped planets with larger $R_p$ and thus causing an absence of strictly rocky planets at longer periods around metal-rich stars (Dawson et al. 2015). Given the multiple interpretations and somewhat contradictory results regarding these host/planet properties (e.g., Winn et al. 2017), we want to assess the presence and strength of the [Fe/H]-$P$ correlation within the APOGEE KOI sample.

To assess first the correlation between host [Fe/H] and orbital period, we perform two different non-parametric tests, calculating Kendall's rank correlation coefficient, $\tau$ (Kendall 1938), and Spearman's rank correlation coefficient, $\rho$ (Spearman 1904). Kendall's $\tau$ coefficient is $\tau = -0.21$, with a $p$-value, $p_\tau = 1.40 \times 10^{-7}$ (the equivalent of a $5.1\sigma$ deviation from a normal distribution). Spearman's rank correlation coefficient is $\rho = -0.31$ with a $p$-value, $p_\rho = 1.67 \times 10^{-6}$ ($5.1\sigma$). These results indicate that as the host star [Fe/H] increases, the orbital periods of the planets around those stars decrease. To verify the robustness of these correlations, we perform a Monte Carlo simulation with $10^5$ sets of data. For every simulated dataset, we add a perturbation which is randomly drawn from a normal distribution with our adopted $1\sigma$ uncertainty of 0.053 dex (see §2.2.1), to the ASCAP-derived host [Fe/H] for each KOI. We then recalculate $\rho$ and $\tau$ for each simulated dataset. For Kendall's rank correlation, we recover the significance of this trend at a $4.83^{+0.31}_{-0.30}\sigma$ level, and for Spearman's rank correlation we recover a significance of $4.83^{+0.31}_{-0.31}\sigma$, where the errors represent the inner 68% of the posterior distribution.

To analyze further the correlation between orbital period on host [Fe/H], we adopt a

method similar to that employed by Buchhave et al. (2014). We generate $10^4$ test orbital periods equally separated in log space, spanning from the minimum to the maximum planetary orbital period in our sample. For each of these periods, $P_i$, we divide our KOI sample into two bins, one where the KOIs have orbital period $P > P_i$ and one with $P \leq P_i$. We then use a two-sample Kolmogorov-Smirnov (KS) test, as well as a $k$-Sample Anderson Darling (AD) test for redundancy, to determine the likelihood that the host star iron abundances from the "short" versus "long" period bins are drawn randomly from the same parent distribution. We find a critical period, $P_{\rm crit}$, where this likelihood is minimized (see Figure 2). In the case where the minimum $p$ value is equal among more than one period, it is because we are oversampling our period distribution. In this case, we take the mean. Within our dataset we find $P_{\rm crit} = 8.3$ days by the KS test and the AD test, with $p$ values of $p_{ks} = 5.0 \times 10^{-7}$ ($4.9\sigma$) and $p_{ad} = 8.6 \times 10^{-6}$ ($4.3\sigma$). To test the robustness of this critical period, we perform a Monte Carlo analysis and simulate $10^4$ sets of data, resampling as we did above, using the typical [Fe/H] uncertainty of 0.053 dex and the $P$ uncertainties as reported by the NASA Exoplanet Archive, which have a median of $4 \times 10^{-5}$ days. We recover the significance with both the KS test and AD test at the $4.5^{+0.4}_{-0.4}\sigma$ level. We find $P_{\rm crit}= 8.3^{+0.1}_{-4.1}$ days for both the KS test and AD test, which is consistent with our original findings. This method thus discovers two unique [Fe/H] populations within our dataset, one that is super-solar on average and contains planets orbiting closer to the host star, and one that is solar metallicity on average and contains planets that orbit farther from their host. That is, planets with orbital periods $P \leq 8.3$ days have statistically more metal-enriched hosts than planets with $P > 8.3$ days. This is the main finding of this Chapter. We note that performing the same tests on host star [Fe/H] and planetary semi-major axis ($a$), instead of period, produces similar results, with $a_{\rm crit} = 0.07$ AU.

In Figure 2, we also note two other, less significant dips at $P \sim 22$ days and at $P \sim 70$ days. The AD test found the shorter period dip at $P \sim 21$ days and the KS test found it

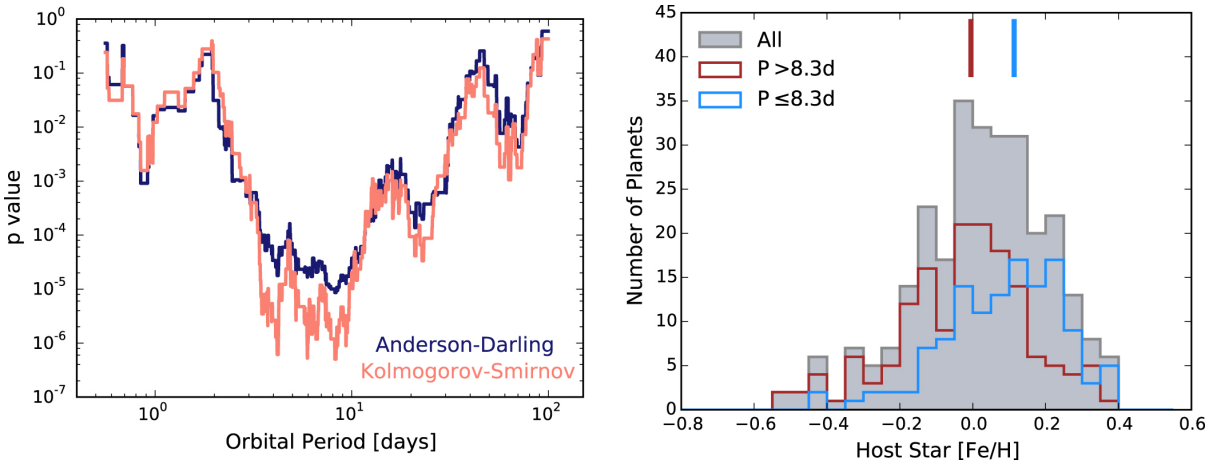


Fig. 2.2.— (Left) The $p$ values of the Kolmogorov-Smirnov and Anderson Darling tests for the probability that the [Fe/H] distributions of exoplanet candidates above and below the given orbital period are drawn from the same parent distribution. There is a statistically significant dip at $P = 8.2$ days in our sample, found by both an Anderson-Darling test, and a Kolmogorov-Smirnov test. (Right) Histogram of the host star metallicities of the long (red) and short (blue) period populations, split by $P_{\rm crit}$= 8.3 days. The combined distribution is shown in gray. The long period population peaks near solar metallicity while the short period population peaks above solar metallicity. The median Host [Fe/H] is shown by the tick marks for the long (red) and short (blue) period populations.

at $P \sim 23$ days. For the longer period dip, the minimum $p$-value found with the KS test is at $P = 63$ days, while the minimum found with the AD test is at $P = 71$ days. Running a Monte Carlo analysis with $10^4$ data sets, similar to the above but restricting our analysis within the range 15 days $< P <$ 45 days, we find the period that minimizes the $p$-value to be $P = 22.7^{+0.4}_{-1.5}$ days for both the AD and KS tests. However, the significance with which we recover this period is only $3.5^{+0.5}_{-0.5}\sigma$ for both the AD and KS tests. Because of the lowered significance, we are not comfortable claiming significance at this period within our study. However, there is theoretical motivation to support a critical period around 23 days (see §2.5.2), and it warrants further work. To test the third dip, we perform the same analysis restricting our test periods to $P > 45$ days, but the results are inherently less trustworthy considering the uneven sample sizes of KOIs with $P < 65$ days and $P > 65$ days. The period that minimizes the $p$-value for the AD and KS tests in this longest period range is $69^{+4}_{-9}$ days, with a significance of $2.8^{+0.4}_{-0.4}\sigma$ for each test. For the

same reasons as the $P \sim 23$ day dip, we do not consider this a significant minimum, and presently favor a two-population model split only at $P_{\rm crit}$= $8.3^{+0.1}_{-4.1}$ days.

The iron abundance for the short period population ($P \leq P_{\rm crit}$) is super solar, with a median [Fe/H] of 0.11 dex and a standard deviation of 0.17 dex. The long period population ($P > P_{\rm crit}$) is consistent with solar metallicity and has a standard deviation of 0.18 dex. To test whether the means of these two samples differ significantly, we perform a Mann-Whitney U-test (Mann & Whitney 1947) between these two populations. The Mann-Whitney U-test gives the probability that two separate populations have the same underlying mean. The test returns a $p$ value of $p_{mw} = 2.28 \times 10^{-7}$ ($5.0\sigma$) that the short and long period populations have the same mean metallicity. Thus, we can safely reject the null hypothesis that these distributions have the same mean, which is consistent with the results of the Kolmogorov-Smirnov and Anderson-Darling tests that the two [Fe/H] populations, separated by $P_{\rm crit}$, are sampled from different parent populations.

To further analyze the significance of the correlation in our sample, we follow the analysis of Mulders et al. (2016) and use the Nadaraya-Watson estimator (Nadaraya 1964; Watson 1964) to calculate how the mean iron abundance varies with orbital period (see Figure 3). The kernel regression of the mean metallicity, $\overline{\rm [Fe/H]}_{\rm KOI}$, as a function of orbital period is given by

$$\overline{\rm [Fe/H]}_{\rm KOI}(P) = \frac{\sum_{i=0}^{n} {\rm [Fe/H]}_i \, K(\log(P/P_i), \sigma)}{\sum_{i=0}^{n} K(\log(P/P_i), \sigma)}, \tag{2.2}$$

where the $n$ is the number of exoplanet candidates in the sample, ${\rm [Fe/H]}_i$ and $P_i$ are the host star metallicity and orbital period of each exoplanet candidate, respectively, and we use a log-normal kernel with constant bandwidth, $\sigma$, given by

$$K(\log x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-0.5(\log x/\sigma)^2}, \tag{2.3}$$

where $x$ is an arbitrary, dimensionless variable. In line with Mulders et al. (2016), we

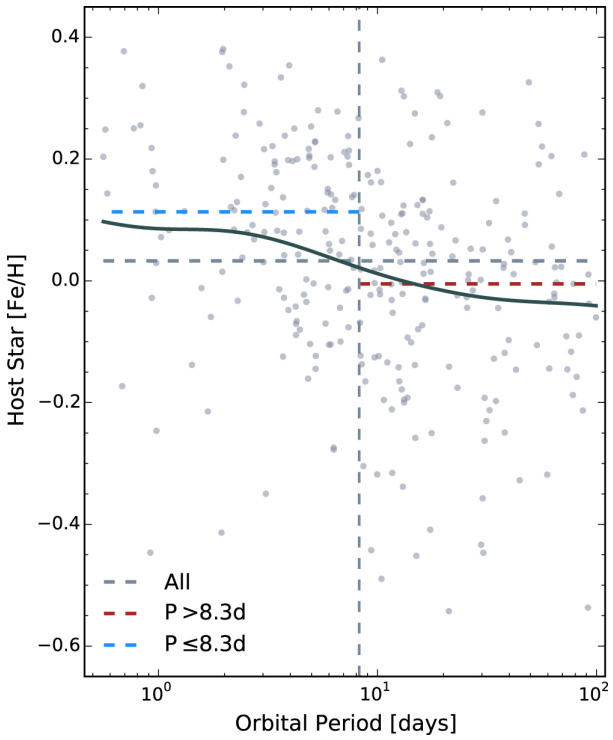


Fig. 2.3.— Host star metallicity as a function of exoplanet orbital period. The gray points are the KOIs in our sample. The vertical dashed line is at $P = 8.3$ days, which separates our short period and long period populations. The horizontal dashed lines show the median of the short period (blue), long period (red), and combined (gray) populations. The combined population has a median of 0.03 dex. The median metallicity of the short period population is super solar at 0.11 dex, while the median of the long period population is 0.00 dex. The solid gray line is the kernel regression of the mean metallicity. The kernel regression shows a steady decrease from a maximum of $\sim 0.10$ dex at the shortest periods, to a minimum of $\sim -0.04$ dex at the longest periods.

adopt a bandwidth of $\sigma = 0.29$.

The kernel regression of the mean metallicity is plotted over all the KOIs in our sample in Figure 3, along with the median [Fe/H] of the combined, short period, and long period samples. Using the kernel regression as a proxy for the mean [Fe/H], we find that the maximum mean metallicity ([Fe/H] $\sim 0.10$ dex) occurs at the shortest period in our sample, and the minimum mean metallicity in this trend ($\sim -0.04$ dex) occurs at the longest period in our sample. This difference of 0.14 dex is larger than the difference in the median metallicity between the long and short period sample by $\sim 0.03$ dex. However, the kernel regression clearly shows a decrease in mean [Fe/H] in our sample at

even the longest orbital periods.

## 2.5 Discussion

### 2.5.1 Metallicity-Period Correlation

Our results are consistent with those of Mulders et al. (2016), who find in their sample an increase of $0.15 \pm 0.05$ dex in host star metallicity for exoplanets orbiting at or interior to 10 days, as compared to longer-period planets. We find a statistically significant break in the KOI host star [Fe/H]-period distribution in our sample at 8.3 days, with shorter period planets orbiting stars with a median [Fe/H] of $0.11 \pm 0.17$ dex and longer period planets orbiting stars with a median [Fe/H] of $0.00 \pm 0.18$ dex. The Mann-Whitney and Kolmogorov-Smirnov probabilities in our data are $5.0\sigma$ and $4.9\sigma$, respectively, compared to Mulders et al. (2016) whose Mann-Whitney and Kolmogorov-Smirnov probabilities were comparable at $4.6\sigma$ and $4.3\sigma$, respectively. While our sample is significantly less than half the size ( 282 candidate or confirmed planets versus Mulders' 665), our data are measured from higher resolution, high SNR spectra, producing internal errors on [Fe/H] of only $\sim 0.053$ dex and typical offsets from literature values of $\sim 0.00 \pm 0.09$ dex. The Mulders et al. (2016) study used LAMOST [Fe/H] values, measured from $R \sim 2000$ optical spectra, typically with SNR $\leq 100$, and typical [Fe/H] internal errors (evaluated by way of repeat observations of some stars) of $\sim$0.055 dex and typical literatures offsets (using their calibrated [Fe/H] values) of $\sim -0.06 \pm 0.18$ dex. Thus, our study showcases an advantage APOGEE has over other surveys of similar scale. Even with our significantly smaller sample size compared to the Mulders et al. (2016) sample, we are able to recover the same correlation with greater confidence.

### Required Precision to Find the Trend

Given our smaller sample size, what is the minimum [Fe/H] precision required to be able to find the trend with orbital period that we do? To determine this precision, we replace the assumed [Fe/H] error (0.053 dex) with larger and larger errors until the resulting uncertainties on the significance of the Kolmogorov-Smirnov and Anderson Darling tests drop below $\sim 3\sigma$. With our well-vetted (e.g., removal of stars showing signs of binarity in their radial velocity variations) sample of high resolution, high SNR data, the greatest value that the mean [Fe/H] error can take on is $\sim$0.1 dex to still recover the observed trend with at least $3\sigma$ significance.

### Possible Mechanisms

What is the physical mechanism responsible for the observed trend between planetary orbital period and host star [Fe/H]? One possible explanation is that the dust sublimation radius (between 0.05 and 0.1 AU, or $\sim 4-12$ days around a solar mass star – Muzerolle et al. 2003, Eisner et al. 2005, Pinte et al. 2008, Min et al. 2011) in protoplanetary disks correlates with host star metallicity, and that it represents a semi-major axis cutoff inward of which no solids contribute to forming planets. If this were the case, we would expect a correlation between host star $T_{\rm eff}$ (as the dust sublimation radius is known to depend on the stellar luminosity and/or stellar mass), and orbital period. To test this correlation, we performed Kendall's $\tau$ and Spearman's $\rho$ correlation tests among our KOI sample. Because we are testing a potential semi-major axis cutoff, we include only the closest planets in the case of multiple-planet systems, as we do with the rest of this study. We find $\tau = 0.10$ with $p_\tau = 0.05$ ($1.7\sigma$) and $\rho = 0.14$ with $p_\rho = 0.06$ ($1.6\sigma$), which indicates that there is no significant correlation between $T_{\rm eff}$ and orbital period. To test further whether there is a difference in the host $T_{\rm eff}$ among the short-period and long-period subsamples, we performed a KS and AD test, similar to the above analysis. For the KS test, we calculate $p_{ks} = 0.2$ ($0.8\sigma$) and for the AD test we calculate $p_{ad} = 0.07$ ($1.3\sigma$).

Thus, both of these tests indicate that there is no significant difference in the temperature distributions of the host stars in the short-period and long-period samples. Figure 4, left, shows that while there is a slight decrease in [Fe/H] from cooler to hotter $T_{\rm eff}$ , both the shorter period/metal-rich and longer period/solar-metallicity subsamples decrease together with a roughly constant metallicity offset of $\sim$0.10 dex. Because the correlation between $T_{\rm eff}$ and [Fe/H] is the same for both subsamples, and the shorter period and longer period subsamples do not have significantly different temperature distributions (Figure 4, right), we do not believe this decrease in [Fe/H] at higher $T_{\rm eff}$ affects our conclusions. Thus, our data rule out dust sublimation effects as a possible explanation for the metallicity-orbital period correlation. Mulders et al. (2015b) also ruled out the hypothesis of the dust sublimation radius controlling the semi-major axis cutoff in planetary occurrence rates, although these authors were testing a dependence of semi-major axis on stellar mass, not metallicity.

A second possible explanation for the observed trend between planetary orbital period and host star [Fe/H] is that planets around higher metallicity stars migrate inward and are "trapped" closer to their host stars (e.g., Kuchner & Lecar 2002; Rein 2012; Plavchan & Bilinski 2013). This closer location could be related to the dust sublimation radius – for which we do not see evidence for as a factor in creating the P-[Fe/H] trend as described above – or the co-rotation radius, where gas is accreted onto the stellar surface. Usually the co-rotation radius is within the dust sublimation radius (Najita et al. 2007 and references therein), and the former also depends on the angular velocity of the star, which does not have a strong stellar mass dependence during the pre-main sequence stage (Bouvier 2013 and references therein). Thus it is difficult to assess the likelihood of this "planet trapping" explanation within the context of this work. However, as discussed below, small, short period planets around more metal-rich stars are generally rock-dominated, indicating accretion and migration mainly *within* the snow line. Thus their migration distance could not have been very far. The snow line is at a few AU around

a solar-mass star after ∼1 Myr, but this distance for any given disk will depend on the stellar mass, and also parameters like the heating mechanism(s), dust grain opacities, and the timing of planet formation (e.g., Kennedy & Kenyon 2008; Martin & Livio 2012; Mulders et al. 2015a; Xiao et al. 2017).

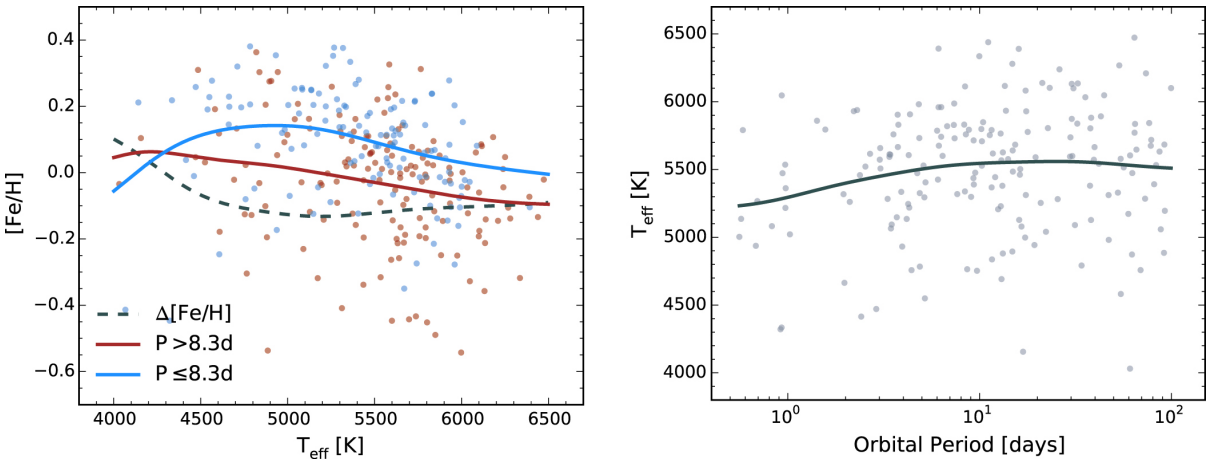


Fig. 2.4.— (Left) Kernel regression of the mean [Fe/H] as a function of PHS $T_{\rm eff}$ for the short-period (blue) and long-period (red) populations, with the difference of the two plotted as a gray dashed line. The upturn at cool temperatures is most likely a result of low-number statistics. Hotter than ∼4500 K, the two distributions show a constant offset of ∼ 0.10 dex, indicating no difference in host star $T_{\rm eff}$ between the two samples. (Right) Stellar $T_{\rm eff}$ versus planetary orbital period. The gray line shows the kernel regression of the mean $T_{\rm eff}$ . If the orbital period-[Fe/H] correlation is related to the dust sublimation radius, we would expect to see a statistically significant positive correlation in this plot, which we do not. Hence there is no evidence for a dust-sublimation effect.

Finally, we consider the possibility that the correlation between planet orbital period and stellar metallicity could be a byproduct of rocky planet ingestion by the star driven by inward migration (e.g., Meléndez et al. 2009; Ramírez et al. 2010, 2014; Schuler et al. 2015; Mack et al. 2014; Liu et al. 2014; Teske et al. 2015, 2016a,b; Bedell et al. 2017). For example, Mack et al. (2014) compared the elemental abundances of the twin stars in the planet-hosting wide binary system HD 20782/81. Both stars in that system host close-in planets within ∼0.2 AU, and Mack et al. (2014) found that both stars are significantly enriched in refractory elements. Those authors further used a model of planet ingestion to show that the enhanced refractory abundances were consistent with a scenario in which

the observed close-in planet pushed 10–20 $M_{\oplus}$ of rocky material (perhaps in the form of small rocky planetessimals) into the surface convection zone of the star. Presumably, this is a consequence of the inward migration of the observed close-in planet. Thus one possible explanation for the correlation we report here is that the planets in our sample that managed to migrate very close to their host stars ($P < 8.3$ d) were also more likely to shepherd other small rocky planets into the star, thus elevating its surface metallicity.

In theory, a testable prediction of this rocky material ingestion scenario would be a pattern of increasing stellar elemental abundances with elemental condensation temperature ($T_c$), as found by Mack et al. (2014) (and, originally in solar twins, Meléndez et al. 2009). However, we are currently unable to perform this test due to the uncertainties associated with dwarf stars abundances derived by ASPCAP in DR14; as noted in §2.2, ASPCAP is optimized for red giants and not well tested for dwarfs.[6] Moreover, while giant planet inward migration has been suggested as a mechanism for "pushing" refractory material onto the host star, we know of no studies suggesting the inward migration of small planets would be capable of creating a similar abundance signature in their host stars.

Interestingly, the scenario envisioned above could also potentially provide a natural explanation for the specific orbital period (8.3 d) dividing the metal-rich versus metal-poor samples. Previous work on the rotational evolution of pre–main-sequence stars has suggested a bimodal distribution of stellar rotation periods, the break between slow and rapid rotators occurring at ∼8 d (see, e.g., Choi & Herbst 1996, and references therein). The two groups of rotators have been interpreted by some authors as the result of different angular momentum histories, possibly due to rotational braking by those stars whose protoplanetary disks survive longer, draining angular momentum from the star via magnetic connection between star and disk at the co-rotation radius as noted

[6]We did attempt to make a preliminary assessment of the differences between the slopes of $T_c$-abundance trends for the short versus long period samples within our 300 stars with planets, but observed no difference between the slope distributions. If the short period planet host stars had ingested more refractory material, we would expect them to show more positive slopes.

above. If inward migration of planets, coupled with ingestion of rocky material by the star, is also related to the rapid dispersal of the protoplanetary disk, the result could be a natural division at ∼8 d of those stars that were more likely to ingest refractory elements (i.e., show elevated surface metallicity), and more likely to host a close-in planet with an orbital period shorter than ∼8 d.

If shorter period planets usually orbit more rapidly rotating stars, based on our observed planet $P$-[Fe/H]$_{star}$ trend we would predict that higher metallicity stars should show faster rotation and shorter disk lifetimes, on average. However, testing this prediction is complicated by the tendency for all stars to slow their rotation on Gyr timescales, regardless of their early rotational histories; we cannot know a priori the original rotation periods of the stars in our sample. To check whether their present-day stellar rotation periods show any trend with stellar metallicity or planetary orbital period, we cross-matched our sample with the McQuillan et al. (2013) and Walkowicz & Basri (2013) catalogs of *Kepler* stellar rotation periods, resulting in 82 stars in common. In this subsample we see no trends between stellar rotation period and stellar metallicity or planet orbital period. Moreover, near-infrared observations of young stars in clusters of various metallicites find that lower metallicity stars have *shorter* protoplanetary disk lifetimes (Yasui et al. 2009, 2010), and these observations are supported by models (Ercolano & Clarke 2010) and simulations (Nakatani et al. 2017) of protoplanetary disk evolution. This contradicts the prediction that shorter period planets orbiting more metal rich stars can be explained by shorter disk lifetimes (and by extension faster stellar rotation periods). Additionally, the idea of bimodal rotation rates among young stars has itself been disputed (see, e.g., Stassun et al. 1999).

Thus we conclude that, while it is still plausible that the main trend we have discovered here could be the result of ingestion by stars of rocky material due to the inward migration of planets, it does not appear to be a consequence of star-disk interaction in the context of stellar rotational evolution.

### 2.5.2 Planet Radius as a Third Dimension of the Correlation

Specifically, Mulders et al. (2016) find an occurrence-corrected $\Delta$[Fe/H] between planets interior and exterior to 10 days that varies with planetary radius, from $\Delta$[Fe/H]$= 0.25 \pm 0.07$ dex for $R_p < 1.7\ R_\oplus$ planets to $0.08 \pm 0.05$ dex for $1.7\ R_\oplus \leq R_p < 3.9\ R_\oplus$ planets to $0.10 \pm 0.12$ dex for $R_p \geq 3.9\ R_\oplus$ planets. We also find that the short period (metal-rich host star) planets in our sample are statistically smaller ($p_{mw} = 7.1\sigma, p_{ks} = 6.7\sigma$) than the planets at longer periods (around less metal-rich stars). The median, mean, and standard deviation of our metal-rich/short period planet population are 1.37, 2.01, and 1.91 $R_\oplus$, respectively, versus 2.29, 2.74, and 1.91 $R_\oplus$, respectively, in our solar-metallicity/long period planet population; typical errors on $R_p$ are 0.02 $R_\oplus$. Interestingly, the break in [Fe/H]-period space in our sample also appears to coincide with the reported "radius gap" around 1.8 $R_\oplus$ defined by Fulton et al. (2017) (see Figure 5). While metal-poor dwarf stars will generally have larger radii than metal-rich dwarf stars, potentially influencing any trends with $R_p$ and host star [Fe/H] (Gaidos & Mann 2013), the metallicity bias (the difference between [Fe/H] values of underlying population of stars versus those around which transiting planets are detected) for the stars in our sample is $\lesssim 0.02$ dex, below our measured precision.

Recently, Owen & Wu (2017) constructed a relatively simple analytical model, building on their previous numerical models (Owen & Wu 2013), of a low-mass planet – core with a gas-envelope – and how it changes with time under the influence of its host star flux. As their previous results (and those of Lopez & Fortney 2013 also showed), the radius distribution observed by Fulton et al. (2017) is matched by an "evaporation valley", such that smaller radius planets represent the bare cores of planets that have had their H/He envelopes photo-evaporated within the first 100 Myrs. Owen & Wu (2017) show specifically that the two peaks in $R_p$ observed by Fulton et al. (2017) arise because

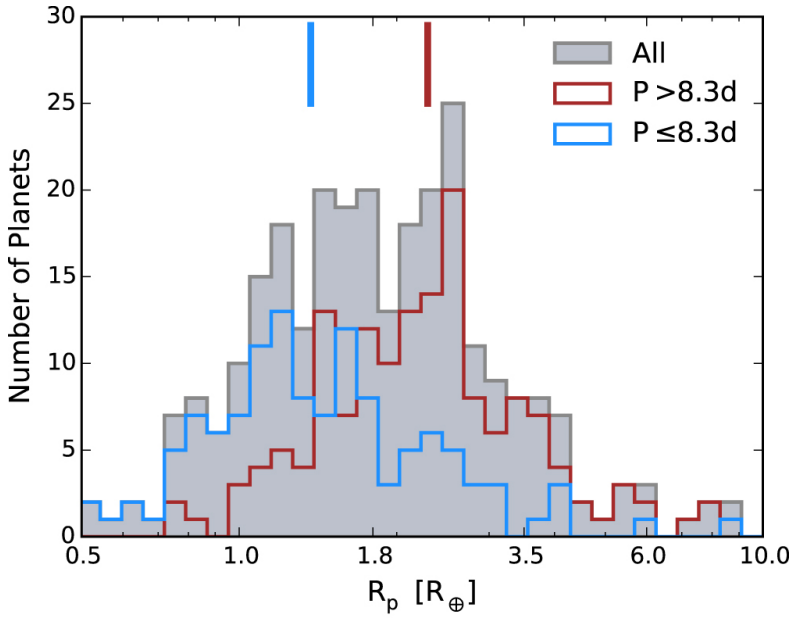


Fig. 2.5.— Histogram of KOI radii of the long (red) and short (blue) period population, split by $P_{\rm crit}$= 8.3 days. The combined distribution is shown in gray. These are the same colors corresponding to the metal-enriched (blue) and solar-metallicity (red) distributions in Figure 2, right panel. The tick marks at the top denote the median planet radius of the short-period (blue) and long period (red) populations.

the timescale for mass-loss is longest when the planet's radius has doubled in size due to an accreted volatile envelope of approximately a few percent the total planet mass. Their model is only able to reproduce the observed results when the planet envelopes are composed of primoridal H/He, *not* water, and when the cores are roughly Earth-like in composition ($\rho \sim 5.5$ g cm$^{-3}$). In summary, Owen & Wu (2017) show that the small, short period *Kepler* planets likely form from one parent population, with one average composition, and have a bimodal period-radius distribution due to envelope evaporation.

Almost simultaneous with Owen & Wu (2017), Jin & Mordasini (2017) produced work independent of Owen & Wu (2017) that compared theoretical models of planet formation, thermodynamical evolution, and atmospheric escape of rocky-cored versus icy-cored planets to the results of Fulton et al. (2017). Similarly, Jin & Mordasini (2017)'s goal was to understand better, in a statistical sense, how evaporation depends on planetary bulk composition. Jin & Mordasini (2017) suggest – assuming the radius gap is due to

atmospheric evaporation – that small, short period planets have mostly rocky cores made of silicates and iron, not mostly icy cores made of frozen $H_2O$, $CO_2$, $CH_4$ and/or $NH_3$. Since planets with mostly icy cores can only form *beyond* the snow line, this indicates that close-in low mass planets accreted mainly *within* the snow line, even considering migration (that is, migration must have been within the snow line). Classifying observed planets based on their ice mass fractions (derived from the mass and radius and an internal structure model, Mordasini et al. 2012a), as well as the planet's $R_p$ and semi-major axis $a$, the authors find eight categories of planets (see their Figures 6 and 7) that exhibit a clear compositional gradient with increasing planet radius. Interestingly, six of the eight planet categories requiring a rock-dominated composition ($R_p \lesssim 1.6\ R_\oplus$) are found within ∼0.09 AU (∼10 days around a solar mass star), and all eight of the rock-dominated planets are found within ∼0.17 AU (∼26 days around a solar mass star). We see a second, less significant dip at $P \sim 23$ days in Figure 2, perhaps corresponding to this second rocky planet orbital period limit.

Putting this all together, the following picture emerges: Most short period, small planets have rocky cores, but their size, volatile content, and thus density is (in a general sense) tied to their orbital period, and thus by this study, their host star metallicity. The trend we see predicts that planets with little to no volatiles, and thus the smallest $R_p$ values, should be in the closest orbits and thus around more metal-rich host stars. Jin & Mordasini (2017) caution that they do not see convincing evidence of a positive volatile content gradient with increasing semi-major axis, but that this is expected for evaporation. We note that Kolmogorov-Smirnov and Anderson Darling tests do not reveal a significant difference between the orbital periods of Jin & Mordasini 2017's 25 Type 1 and 3 planets versus their seven Type 6 planets (from their Table 1), but that with a larger number of small planets with well-constrained masses and radii, we can better test our prediction. Our results, as interpreted in the context of Jin & Mordasini's theoretical framework, suggest that in the hunt for small, rock-dominated planets with

little to no gaseous envelopes, one should be looking around more metal-rich stars.

## 2.6 Conclusions

In this work, we aim to characterize the intricate relationship between host star metallicity and planet orbital period, as it relates to the context of planet formation. We first demonstrate the veracity of ASPCAP's metallicities by comparing a sample of 221 FGK dwarfs in the APOGEE survey that also have quality parameters via optical spectroscopy in the literature. Then, using a sample of 282 short period ($P < 100$ days) *Kepler* exoplanets and exoplanet candidates observed by the APOGEE KOI goal program and the associated parameters derived from ASPCAP, we have characterized the correlation between planet orbital period and host star metallicity. In particular, we've found the following:

- There is a statistically significant correlation between host star [Fe/H] and planetary orbital period that is characterized by a critical period, $P_{\rm crit}= 8.3^{+0.1}_{-4.1}$ days, below which planets preferentially orbit more metal-rich stars. This corresponds to a semi-major axis of $\sim$0.07 AU for a solar mass star and is consistent with the drop in occurrence rate at $\sim$0.1 AU found by Mulders et al. (2015b).

- The minimum precision in [Fe/H] needed to see this trend within our carefully vetted sample is $\sim$0.1 dex. While this correlation has been seen in other studies (e.g. Mulders et al. 2016), the precision in APOGEE's abundance determinations allows us to find this correlation with higher confidence levels in a significantly smaller sample than what has been used for other planet host surveys of similar scale.

- Planets in the short-period/high-metallicity population have significantly smaller radii than the long period population ($p_{mw} = 7.1\sigma$). Based on previous work on

the "evaporation valley", this suggests that the population of planets around more metal-rich stars is mostly rocky and lacking substantial atmospheres, while the population of planets around more metal-poor stars have thicker atmospheres. Thus, to optimize the number of close-in, rocky exoplanets discovered around FGK dwarfs, transit surveys should prioritize super-solar metallicity stars.

- Based on the results of Jin & Mordasini (2017), we suspect that the critical period of 8.3 days may be tied to the bulk composition of the exoplanet population in a statistical sense. In addition to this period, we find some evidence for a second, less convincing critical period at $P \sim 23$ days, which may also correlate with the exoplanet population's composition. Although we do not currently believe this period is significant, its agreement with the results of Jin & Mordasini (2017) is intriguing enough to warrant further investigation.

- We hypothesize that there is some protoplanetary disk inner-radius with a metallicity-dependence at the time of planet formation that allows small, rocky planets to either form or migrate closer in to their host star in metal-rich conditions. Such an inner radius may be the dust-sublimation radius, but we would expect this radius and thus orbital period to correlate strongly with the host $T_{\rm eff}$ , and see no such correlation. The inner radius may instead be the gas co-rotation radius, but with our given observations it is hard to assess the likelihood of this explanation. Alternatively, the period-metallicity correlation that we observe may be the result of rocky planet ingestion, driven by inward planet migration. In this scenario, planets migrate inward and in the process shepherd rocky material (perhaps in the form of planetesimals) onto their host star, resulting in an increased surface metallicity. At this time the precision of APOGEE dwarf star abundances across a range of condensation temperatures preclude a robust test of this hypothesis.

APOGEE provides a valuable resource for characterizing exoplanet host stars from

the *Kepler* mission. In particular, studies of planetary architecture coupled with accurate metallicities, as presented here, can provide new constraints on planet formation that could not otherwise be obtained from smaller, more focused studies. In addition, APOGEE measures [Fe/H] to a level of precision that stands out from other spectroscopic surveys of similar scale, positioning APOGEE in a valuable area for future work in characterizing exoplanets and their host stars.

# Chapter 3

# The Influence of 10 Unique Chemical Species on the Distribution of Kepler Planets

## Summary

The chemical abundances of planet-hosting stars offer a glimpse into the composition of planet-forming environments. To further understand this connection, we make the first ever measurement of the correlation strength between planet occurrence and detailed chemical abundances for ten different elements (C, Mg, Al, Si, S, K, Ca, Mn, Fe, and Ni). We leverage data from the Apache Point Observatory Galactic Evolution Experiment (APOGEE) and *Gaia* to derive precise stellar parameters ($\sigma_{R_\star} \approx 2.2\%$, $\sigma_{M_\star} \approx 4.8\%$) for a sample of 1,018 *Kepler* Objects of Interest, which enables us to construct a sample of well-vetted *Kepler* planets with precisely measured radii ($\sigma_{R_p} \approx 3.2\%$). After controlling for biases in the *Kepler* planet detection pipeline and the selection function of the APOGEE survey sample, we characterize the relationship between planet occurrence and chemical abundance as the number density of nuclei of each element in a star's photosphere raised

to a power, $\beta$. The derived correlation strength, $\beta$, varies by planet type, but is consistent within our uncertainties across each of the ten elements. For hot planets ($P$ = 1-10 days), we find that an enhancement in any element of 0.1 dex corresponds to an increase in the occurrence rate of ≈20% for Super-Earths ($R_p$ = 1-1.9 $R_\oplus$) and ≈60% for Sub-Neptunes ($R_p$ = 1.9-4 $R_\oplus$). These trends are weaker overall for warm ($P$ = 10-100 days) planets of all sizes and for all elements, with the potential exception of Sub-Saturns ($R_p$ = 4-8 $R_\oplus$), whose occurrence is strongly correlated with enhanced metals ($\beta \approx 2.8$). Finally, we conclude this work with a caution to interpreting trends between planet occurrence and stellar age due to degeneracies caused by Galactic chemical evolution and make predictions for planet occurrence rates in nearby open clusters to facilitate demographics studies of young planetary systems.

## 3.1 Introduction

A clear host-star chemical influence on associated planets was recognized in early spectroscopic surveys primarily aimed at discovering planets through radial velocity (RV) variations, which found that stars hosting giant planets tend to have enhanced metallicities[1] (Gonzalez 1997; Heiter & Luck 2003b; Santos et al. 2004). More detailed population studies of RV-detected planets confirmed this trend between host star [Fe/H] and the frequency at which giant planets are found (Santos et al. 2004; Fischer & Valenti 2005), a trend that appears to decrease in significance with lower planet mass and/or radius (Sousa et al. 2008; Ghezzi et al. 2010; Schlaufman & Laughlin 2011; Buchhave et al. 2012; Wang & Fischer 2015; Ghezzi et al. 2018). This correlation is typically interpreted as evidence for the core accretion model of planet formation (e.g., Rice & Armitage 2003; Ida & Lin 2004; Alibert et al. 2011; Mordasini et al. 2012b; Maldonado et al. 2019), where

[1]In this study, we use metallicities and iron abundance interchangeably, where iron abundances are paramaterized by the number density of iron nuclei in a star's photosphere relative to the amount of hydrogen normalized to some zero-point, typically the Solar abundance: [Fe/H], where $[X/Y] \equiv \log(N_X/N_Y) - \log(N_X/N_Y)_0$.

host star metallicity is a proxy for the solid surface density of the protoplanetary disk; higher metallicities translate to more planet-forming material, which facilitates quick planetary core growth up to a critical mass of $\sim$10 $M_\oplus$, in turn allowing more time to accrete gaseous envelopes before gas dissipation in the protoplanetary disk.

The Planet-Metallicity Correlation (PMC) partly motivated large spectroscopic surveys of candidate and confirmed *Kepler* planet-hosting stars (e.g., Bruntt et al. 2012; Buchhave et al. 2012, 2014; Everett et al. 2013; Dong et al. 2014; Fleming et al. 2015; Brewer et al. 2016; Johnson et al. 2017). Within this population of close-in, transiting planets, more intricate relationships between stellar metallicity, planet radius, and orbital period have come to light. It is generally found that planets with larger radii have hosts with super-solar metallicity (Buchhave et al. 2014; Schlaufman 2015; Wang & Fischer 2015). This correlation appears strongest for large planets ($R_P \gtrsim 4\ R_\oplus$), and nearly disappears for the smallest planets ($R_p \lesssim 1.7\ R_\oplus$). While the PMC is weaker for small planets in general, that is not the case for small planets in short period ($P \lesssim 10$ days) orbits. The presence of such planets is positively correlated with metallicity, suggesting that an abundance of solids facilitates the growth and/or migration of small, close-in planets (Mulders et al. 2016; Wilson et al. 2018; Petigura et al. 2018; Narang et al. 2018). Thus, the amount of available solids in the protoplanetary disk seems to be a key variable in setting the planet mass, radius, *and* period distributions. While these works in particular demonstrated the intricate relationships between host-star chemistry and the formation/evolution of planetary systems, they also demonstrated the precision and resources needed to unveil such relationships.

While correlations of planetary architecture to bulk metallicity are well-established, some results indicate that these trends may be integrating over more detailed chemical relationships. For example, Adibekyan et al. (2012a) found that an increase in the abundance of certain $\alpha$-elements, such as Mg and Ti, increases the likelihood of planet occurrence. This work supported that of Brugamyer et al. (2011), who found that, beyond

the PMC, planet detection rates are positively correlated with enhanced Si abundances, but not with enhanced O abundances. Brugamyer et al. (2011) inferred from this that core accretion is driven by grain nucleation rather than icy mantle growth, and that $\alpha$-elements may drive the formation of planetesimals more efficiently than other elements. These investigations show the potential for detailed, multi-element stellar abundance studies to advance models of planet formation.

However, the data collection needed to study the relationships between planetary properties and the detailed chemical makeup of their host stars properly is particularly resource-intensive, as it requires high resolution, high signal-to-noise spectra of not only hundreds of planet-hosting stars, but also a significant fraction of the stars searched for planets (typically on the order of $10^{4-5}$ stars for *Kepler*). Because of this, an occurrence rate study with detailed chemical abundances has not been performed for the *Kepler* field, where much of our knowledge of the small planet population has originated.

The Apache Point Galactic Evolution Experiment (APOGEE; Majewski et al. 2017) provides a unique opportunity to perform such a study. APOGEE began in the third phase of the Sloan Digital Sky Survey (SDSS-III; Eisenstein et al. 2011), and is now in its second phase, APOGEE-2 as a part of SDSS-IV (Blanton et al. 2017). The APOGEE survey collects spectra with a multiplexed, high-resolution ($R \sim 22,500$), near-infrared ($\lambda \sim 1.5-1.7\,\mu$m) fiber-fed spectrograph (Wilson et al. 2012, 2019) mounted on the Sloan 2.5-meter telescope (Gunn et al. 2006) at Apache Point Observatory. The primary goal of APOGEE is to study the Milky Way through the RVs and chemical abundances of nearly 750,000 stars across multiple stellar populations and Galactic regions. Additional science programs are also included in the survey, with one such program monitoring stars with candidate planets from *Kepler* (*Kepler* Objects of Interest; KOIs) to search for false positives through RV variations (Fleming et al. 2015; Zasowski et al. 2017). This effort, the APOGEE-KOI Goal Program, has observed 1177 *Kepler* stars, with a median of 17 (mean: 17.7) epochs, as of the sixteenth Sloan data release (DR16; Ahumada et al. 2020;

Jönsson et al. 2020). Because of the large number of epochs, the combined, RV-aligned spectra are of high $S/N$ (median: 155, mean: 217), enabling precise derivations of stellar atmospheric parameters and chemical abundances.

In this Chapter, we utilize the data from the APOGEE-KOI program to explore the role of ten different chemical species (C, Mg, Al, Si, S, K, Ca, Mn, Fe, and Ni) in sculpting the population of *Kepler* planets. §3.2 describes our data, the derivation of stellar parameters for the KOIs in this study, and the resulting precision in planet radii for our sample. §3.3 describes the sample selection for measuring occurrence rates. §3.4 presents our results, enumerating chemical abundance trends present in the selected sample and our occurrence rate analyses. Finally, we end the chapter with a discussion on the implications of our results and reiterate our conclusions.

## 3.2 Data and Methods

### 3.2.1 The APOGEE-KOI Goal Program

The APOGEE-KOI Goal Program targets were chosen with the intention of observing all possible "Confirmed" or "Candidate" KOIs with $H < 14$ on six different *Kepler* tiles, one of which was observed as a pathfinder program in SDSS-III. One *Kepler* tile is roughly the size of the APOGEE footprint, thus allowing for a near one to one match between an APOGEE field and *Kepler* tile. Some KOIs were excluded from the sample on the basis of nonphysical impact parameters and putative planet radii consistent with stellar values. In total the DR16 APOGEE catalog contains observations for 1299 stars (totaling 1461 unique planet candidates without a "False Positive" disposition) in the *Kepler* Q1-Q17 DR24 KOI catalog (Mullally et al. 2015). Of the 1299 stars, 1177 are part of the APOGEE-KOI radial velocity survey and 122 stars were observed throughout the *Kepler* field as parts of other APOGEE programs (see e.g., Zasowski et al. 2013, 2017). In APOGEE DR16, six fields have been observed in total, labeled as K04, K06, K07,

K10, K16, and K21 (see Figure 3.1). Each field was selected on the basis of maximizing the number of available KOIs at the time of target selection. For three of the fields (K04, K06, and K07), KOIs were selected from the Q1-Q17 DR24 KOI catalog, while the other three fields (K10, K16, and K21) were queried from the NexSci Exoplanet Archive[2] immediately prior to the design of each field: 2014 March for K10, K21 and 2013 August for K16. These publicly available catalogs were dynamic, and therefore do not have a static or well-studied selection function. As a result, there are a number of KOIs that were discovered after sources were chosen for inclusion in the APOGEE-KOI program (these planet candidates are displayed as red dots in Figure 3.1). In §3.C.3, we account for biases that may arise from the exclusion of these planets in our analysis.

### 3.2.2 Stellar and Planetary Parameters

For each KOI observed in APOGEE, we re-derive fundamental stellar properties (e.g., $M_\star, R_\star$) and planet radii. The primary motivation for re-deriving stellar properties in our sample is to improve the precision of the planet radii by incorporating precise spectroscopic parameters derived from the high S/N, high resolution APOGEE spectra. This approach has the additional benefit of maintaining a uniform analysis in deriving properties for the planets in our sample so as not to add additional bias. While we only make use of the stellar radii in our analysis, we provide additional stellar properties for the sake of comparison and any future investigations.

#### Spectroscopic Parameters and Abundances: $T_{\rm eff}$, $\log g$, [Fe/H], [X/Fe]

The spectroscopic parameters in this work are adopted from APOGEE DR16 (Ahumada et al. 2020; Jönsson et al. 2020). All of the spectra from APOGEE are processed through automated data reduction pipelines (Nidever et al. 2015; Holtzman et al. 2018). The spectroscopic parameters used for stars in the APOGEE-KOI program are derived

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

from the Automated Stellar Parameters and Chemical Abundances Pipeline (ASPCAP; García Pérez et al. 2016). In DR16, ASPCAP consists of two components: a `fortran90` optimization code (`FERRE` [3];Allende Prieto et al. 2006) and an IDL wrapper used for book-keeping and preparing the input APOGEE spectra. `FERRE` performs a $\chi^2$ minimization across an interpolated library of synthetic stellar atmosphere models (e.g., Zamora et al. 2015), to find a best fit set of input parameters (effective temperature, $T_{\rm eff}$; bulk solar-scaled metallicity, [M/H]; surface gravity, $\log g$; microturbulent velocity, $\xi_t$; and C, N, and $\alpha$ abundances).

Once these best-fitting fundamental atmospheric parameters are found, ASPCAP fits individual spectral windows from a carefully curated linelist (Shetrone et al. 2015; Smith et al. 2021) optimized for each chemical element. In APOGEE DR16 both "raw" and calibrated spectroscopic parameters and abundance measurements are provided. $T_{\rm eff}$ is calibrated to reproduce the photometric values of González Hernández & Bonifacio (2009), $\log g$ in the case of dwarfs is calibrated using a combination of asteroseismic values and fits to isochrones. Calibrated abundances are zero-point shifted so that stars with solar [M/H] in the solar neighborhood have a mean [X/M]=0 (Jönsson et al. 2020). Unless otherwise stated, we use the calibrated parameters in this study. ASPCAP values of [X/Fe] are reported, which we change to [X/H] via the following equation, [X/H] $\equiv$ [X/Fe] + [Fe/H].

Abundance ratios for the ten chemical species in this study are defined in the same way as for [Fe/H], i.e., [X/Fe] $\equiv \log(N_X/N_{Fe}) - \log(N_X/N_{Fe})_0$. However, the chosen zero-point varies by chemical species and is not necessarily the corresponding Solar abundance (Jönsson et al. 2020). The APOGEE data products report two different values for carbon abundance ratios, one measured from atomic lines (`CI_FE` in the APOGEE DR16 data model) and one measured from molecular CO lines (`C_FE` in the APOGEE DR16 data model). For this work, we use the carbon abundance ratio as measured from atomic

[3]Available at https://github.com/callendeprieto/ferre

carbon lines, unless otherwise stated.

When deriving fundamental stellar properties (§3.2.2), we use the errors reported by ASPCAP for $T_{\rm eff}$, as comparisons in the literature have shown scatter consistent with these uncertainties (e.g., Wilson et al. 2018). However, the errors reported by ASPCAP are sometimes underestimated for $\log g$ and [Fe/H]. Therefore, when using these parameters to fit to evolutionary tracks in §3.2.2, we inflate the uncertainties on $\log g$ and [Fe/H]. We do this by multiplying all reported errors by a given value to define the median uncertainty. For [Fe/H], we inflate the errors so that the median uncertainty is 0.03 dex, a factor of 1.5× the median uncertainty determined from repeat observations of high $S/N$ spectra (Jönsson et al. 2020). We choose to inflate these errors because the typical uncertainty measured in Jönsson et al. (2020) was determined using a combined sample of giant and dwarf spectra, and ASPCAP generally measures more precise abundances for giant stars than for dwarf stars. The ASPCAP calibrated $\log g$ are systematically underestimated in FG dwarfs, forcing the fits to the evolutionary tracks to adopt models with systematically lower temperatures than the initial input measurements. To adjust for this, we inflated the ASPCAP $\log g$ uncertainties until the input and output temperatures showed no trend. In all, we inflated the $\log g$ uncertainties to have a median error of 0.15 dex, ∼1.8× larger than the ASPCAP reported uncertainties.

To ensure there are no systematic trends present in the ASPCAP abundances, we check for correlations with [X/Fe] and Teff. To test this, we select a sample of dwarf stars observed by APOGEE with high $S/N$ spectra. We start with the DR16 catalog, and remove all stars with log g $<$ 3.5, a distance, d $>$ 1 kpc, as measured from the geometric parallax in *Gaia* DR2 (Gaia Collaboration et al. 2018b; Bailer-Jones et al. 2018). In addition to these selection cuts designed to remove stars that are not broadly representative of our sample, we also apply a number of cuts designed to remove poor quality data. We remove stars with a spectrum $S/N < 100$, and stars with any of

the following ASPCAP or Star Flags set[4]: TEFF_BAD, LOGG_BAD, METALS_BAD, ALPHAFE_BAD, STAR_BAD, and VERY_CLOSE_NEIGHBOR.

With this sample of dwarf stars in APOGEE, we assume that there should be no trend in abundance-ratio with effective temperature. If a trend exists, it is more likely to indicate a problem in ASPCAP than an astrophysical source. Our goal is to identify a range of effective temperatures where the APOGEE abundance-ratio measurements are reliable and will not bias our inferences of the planet population. In general, we find two prominent features in the ASPCAP-derived abundance ratios at high and low $T_{\rm eff}$ range for ASPCAP that we consider to be systematic in nature and wish to avoid in our analysis (see Figure 3.2). At $T_{\rm eff} \lesssim 4700$ K there is a "hook" feature on the order of up to 0.1 dex, where the ASPCAP-derived abundances decrease dramatically then rise again, present for Mg, Si, and Al abundance ratios. We find this same feature in dwarfs in M67, which should all have the same abundance-ratios, leading us to conclude it is systematic in nature. On the hotter end, we find an increase in the abundance ratio at $T_{\rm eff} \gtrsim 6200$ K, which we believe is also a systematic trend. Thus, for this study we only use stars in the temperature range $4700\,{\rm K} < T_{\rm eff} < 6200\,{\rm K}$ for our occurrence rate analyses.

### Non-Spectroscopic Parameters: $\pi$, $Ks$, $E(B-V)$

For this study, we adopt the parallax, $\pi$, from *Gaia* DR2 (Gaia Collaboration et al. 2018b). We apply the global parallax systematic offset as derived by Zinn et al. (2019a), adding $\delta_\pi = 52.8 \pm 2.4\,\mu$as to the reported $\pi$ from *Gaia* DR2, and adding the uncertainty on the zero-point offset in quadrature with the reported $\sigma_\pi$. In conjunction with $\pi$, the stellar apparent magnitude sets a strict semi-empirical constraint on the stellar luminosity. To minimize the impact of dust extinction in our analysis we adopt the $Ks$-band magnitude from 2MASS (Skrutskie et al. 2006), as it is the longest wavelength

[4]for a description of these flags, see https://www.sdss. org/dr16/algorithms/bitmasks/

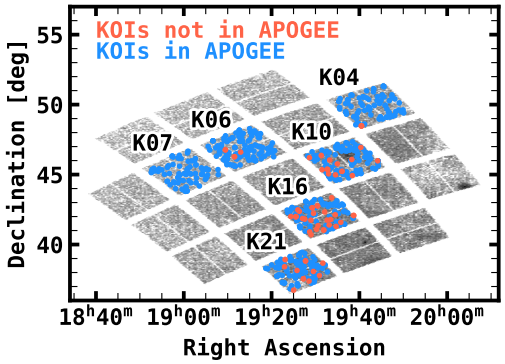


Fig. 3.1.— The right ascension and declination of stars in the APOGEE-KOI sample. The greyscale points show the density of stars in the *Kepler* stellar properties table at a particular sky coordinate, while the points show the DR24 KOIs observed (blue), and not observed (red) by the APOGEE-KOI program in a temperature range with reliable abundance-ratio measurements (see Figure 3.2). The name of each field is listed to the top left of the field.

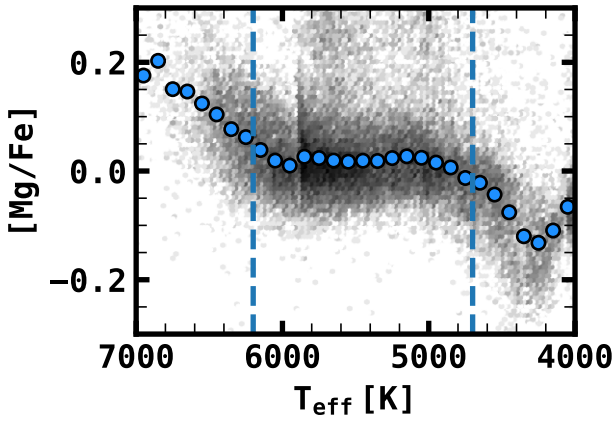


Fig. 3.2.— An example of a typical trend between abundance ratio and $T_{\rm eff}$ in the APOGEE dwarf sample. The gray-scale points show the density of stars with a given $T_{\rm eff}$ and [Mg/Fe] abundance, with darker shades representing more points. The blue points show the median abundance ratio in $T_{\rm eff}$ bins of 100K, and the dashed vertical lines show our adopted $T_{\rm eff}$ range for this study.

($\lambda \sim 2.2\,\mu$m) photometric band uniformly available for our sample.

To account for extinction from dust, we employ the 3D dust map from Green et al. (2019) which we access using the python package `dustmaps` (Green 2018). We add the uncertainty from the Green et al. (2019) three-dimensional dust map in quadrature with $\sigma_{E(B-V)} = 0.001$ mag to account for the typical uncertainties in the color excess ratios measured in Wang & Chen (2019) from which we adopt our reddening law.

### Fit to Stellar Evolutionary Tracks

To infer fundamental stellar parameters (e.g., $R_\star$, $M_\star$) for the stars in our sample we apply the python package `isofit`[5]. For the sake of brevity, we detail the methodology employed by the `isofit` package in the appendix (§3.A). In short, `isofit` compares observations to a grid of MESA Isochrones and Stellar Tracks (MIST) models (Dotter 2016; Choi et al. 2016) with masses ranging from 0.1 to 8.0 $M_\odot$, metallicities ranging from $-2$ to 0.5 dex, and evolutionary states ranging from the Zero-Age Main Sequence to the beginning of the White Dwarf Cooling track. After finding an initial best model, a Markov Chain Monte Carlo (MCMC) analysis is applied to estimate the credible ranges for each parameter.

For each host star in our initial planet candidate sample, we run `isofit` with the following observable quantities and associated uncertainties: $\pi$, $Ks$, $E(B-V)$, $T_{\rm eff}$, $\log g$, and [Fe/H]. We instantiate the MCMC sampling using 30 walkers, with 350 steps and 200 burn-in steps. While modest, we find that this returns posterior distributions in stellar mass and radius that are consistent with the distributions returned after convergence[6], and these settings significantly reduce our computational load. We report the stellar parameters as the median for each parameter in the posterior distribution and the upper

[5]Available at https://github.com/robertfwilson/isofit

[6]This is true for stars on the main sequence, and for parameters that are well constrained, such as stellar radius and luminosity. These settings do not typically return an adequate posterior distribution for other parameters, such as age, or in parameter spaces where degeneracies are likely, such as near the base of the Red Giant Branch.

and lower limits as the 84th and 16th percentile of the posterior, respectively. In all, we derive fundamental stellar parameters for 1,018 stars (281 stars did not have reliable ASPCAP solutions). The stellar parameters derived from `isofit` are given in Table 3.1.

### Accuracy and Precision of Stellar Properties

We assume the larger of the absolute value between the median and upper or lower limits to be a reliable metric for the precision of the stellar parameters inferred in our sample. These uncertainties are displayed in Figure 3.3. For stellar radius, we find a mean uncertainty of $\sigma_{R_\star} = 2.5\%$ and median uncertainty of $\sigma_{R_\star} = 2.2\%$. This error is largely limited by the uncertainty in $T_{\rm eff}$ and $Ks$. It is more difficult to say what sets the minimum uncertainty in $M_\star$, given that there are several inputs that are correlated. In all, we find both the median and mean uncertainty to be 4.8%. However, we caution that for some stars our reported uncertainty in $M_\star$ is likely underestimated. There are likely grid effects that prevent the walkers from exploring the full range of parameter space in $M_\star$, especially for stars with $\sigma_{M_\star} \lesssim 3\%$. We also note once again for emphasis that the reported uncertainties in stellar mass do not take model uncertainties into account, and are entirely model-dependent. While $\log g$ does offer a semi-empirical mass constraint when combined with the inferred radius, which only depends on the bolometric correction as a model-dependent constraint, it is not as limiting in our case where we inflate the $\log g$ uncertainties to have a median of 0.15 dex. To this end, comparing the masses derived with different sets of model grids are likely to reveal larger uncertainties in the inferred mass, but such an exercise is outside the scope of this work.

To judge the accuracy of the stellar parameters in our sample, we compare the results from `isofit` to the parameters derived in Berger et al. (2020b), which has a measured mass and radius for each star in our sample. Berger et al. (2020b) derived masses and radii for ∼186,000 stars in the *Kepler* field by comparing photometric effective temperatures, *Gaia* parallaxes, and 2MASS $Ks$-band magnitudes to a custom set of MIST model grids,

Table 3.1: Derived Properties for 1,018 KOIs in APOGEE.

| Column | Column Label | Column Description |
|---|---|---|
| 1 | KIC | *Kepler* Input Catalog Identification Number |
| 2 | APOGEE_ID | The APOGEE Star Identification |
| 3 | Teff | effective temperature of the star in K |
| 4 | Teff_e | 16th percentile of derived posterior in Teff |
| 5 | Teff_E | 84th percentile of derived posterior in Teff |
| 6 | logg | logarithm of the surface gravity of the star in $cm/s^2$ |
| 7 | logg_e | 16th percentile of derived posterior in logg |
| 8 | logg_E | 84th percentile of derived posterior in logg |
| 9 | feh | metallicity of the star, [Fe/H] |
| 10 | feh_e | 16th percentile of derived posterior in feh |
| 11 | feh_E | 84th percentile of derived posterior in feh |
| 12 | mass | mass of the star in $M_{\odot}$ |
| 13 | mass_e | 16th percentile of derived posterior in mass |
| 14 | mass_E | 84th percentile of derived posterior in mass |
| 15 | radius | radius of the star in $R_{\odot}$ |
| 16 | radius_e | 16th percentile of derived posterior in radius |
| 17 | radius_E | 84th percentile of derived posterior in radius |
| 18 | logL | logarithm of the bolometric luminosity of the star in $L_{\odot}$ |
| 19 | logL_e | 16th percentile of derived posterior in logL |
| 20 | logL_E | 84th percentile of derived posterior in logL |
| 21 | density | density of the star in $\rho_{\odot}$ |
| 22 | density_e | 16th percentile of derived posterior in density |
| 23 | density_E | 84th percentile of derived posterior in density |
| 27 | distance | distance of the star in pc |
| 28 | distance_e | 16th percentile of derived posterior in distance |
| 29 | distance_E | 84th percentile of derived posterior in distance |
| 30 | ebv | the reddening of the star in units of $E(B-V)$ |
| 31 | ebv_e | 16th percentile of derived posterior in ebv |
| 32 | ebv_E | 84th percentile of derived posterior in ebv |

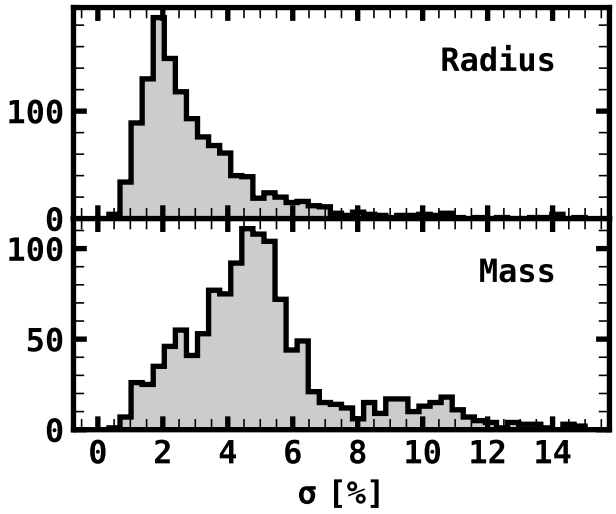


Fig. 3.3.— The relative errors of the stellar radius (top) and mass (bottom) in the APOGEE-KOI sample derived by `isofit`. The mean and median stellar radius uncertainties are 2.5% and 2.2%, respectively. The mean and median uncertainties on the stellar mass are 4.8%.

and spectroscopic [Fe/H] where applicable. For stars with no spectroscopic [Fe/H], the authors assumed a thin disk metallicity prior. These comparisons are highlighted in Figure 3.4.

We find overall agreement consistent with our reported uncertainties. The mean difference in radii, calculated as $(R_\star - R_{\rm B20})/R_\star$, gives a mean and scatter of $-0.7 \pm 3.4\%$, where $R_{\rm B20}$ is the radii inferred by Berger et al. (2020b). This is well within the combined uncertainties defined in our sample and in Berger et al. (2020b). However, there are some systematic differences. While there is generally excellent agreement in $R_\star$, the radii in the APOGEE sample are systematically lower by as much as $\sim$5% for lower-mass stars ($\lesssim 0.75 M_\odot$). This may be caused by the use of slightly different model grids. Berger et al. (2020b) adjust their model grids for stars with $M_\star \lesssim 0.75 M_\odot$ by adopting empirical relations from Mann et al. (2015, 2019). However, while we still report stellar parameters for all the KOIs observed by APOGEE, only a subset of those stars (largely FGK dwarfs, see §3.3.3) are actually used for the occurrence rate analyses in the later sections of this

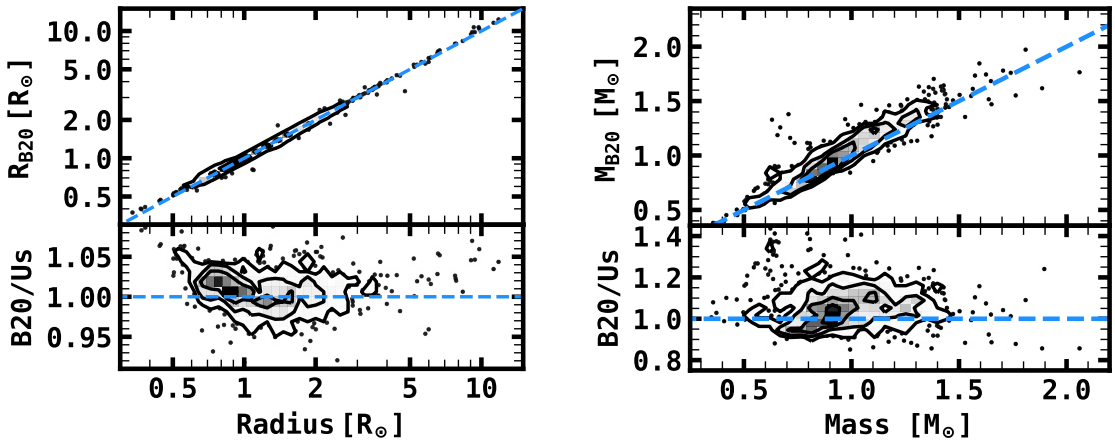


Fig. 3.4.— Comparison of the fundamental stellar properties derived in this work versus the stellar properties derived by Berger et al. (2020b) for the same stars (B20). The dashed blue lines in each case represent the one-to-one agreement between the two samples. *Left:* Comparisons of the stellar radii derived in this work. Overall there is excellent agreement, with scatter in the ratio of radii of 3.4%, and an average offset of $< 1\%$. *Right:* Comparisons of the stellar masses derived in this work and in Berger et al. (2020b). They agree overall within the scatter, but have an offset of $\approx 6\%$, in that the APOGEE sample has a lower mass on average.

work. For this reason, there is no strong motivation to make adjustments to our model grid for low mass stars.

Performing the same comparison for $M_\star$, we find the mean and scatter of $(M_\star - M_{\rm B20})/M_\star = -6.1 \pm 8.1\%$, where $M_{B20}$ is the mass derived in Berger et al. (2020b). While there is a somewhat significant offset, it is still within the reported scatter for the comparison. However, this offset is larger than our reported uncertainties ($\sigma_{M_\star} \approx 5\%$) in $M_\star$, but as mentioned above, $\sigma_{M_\star}$ is likely underestimated for a fraction of stars in our sample. This offset is most likely due to a difference in the $T_{\rm eff}$ of the two samples. We find that the effective temperatures between our sample and those of Berger et al. (2020b) have $T_{\rm eff} - T_{\rm eff,\,B20} = -78 \pm 193$ K. This lower temperature may explain the differences in the inferred stellar masses between the two samples. However, this difference is mostly for stars with effective temperatures from 5000-6000 K. The difference in effective temperature is minimal for stars with $T_{\rm eff} \lesssim 5000$ K, so although this may be

the cause of the mass difference in the APOGEE-KOI sample and the B20 sample, this is likely not the reason for the systematic differences in radii for low mass stars.

In addition to the comparisons with Berger et al. (2020b), we check our stellar radii against those inferred from high-resolution spectroscopy (Martinez et al. 2019, see Figure 3.5). Martinez et al. (2019) derived atmospheric parameters from the archival spectra in the CKS sample by measuring equivalent widths for a carefully curated sample of Fe I and Fe II lines, following the methodology outlined in Ghezzi et al. (2010, 2018). The Martinez et al. (2019) sample is a more fair comparison for the purposes of checking the accuracy of the stellar radii in our sample due to the combination of spectroscopic $T_{\rm eff}$, $\log g$, and [Fe/H] used as opposed to Berger et al. (2020b) who used photometric temperatures and inflated errors on [Fe/H]. We find relatively good agreement, with $(R_\star - R_{\rm M19})/R_\star = -1.1 \pm 1.4\%$, where $R_{M19}$ is the radii from Martinez et al. (2019). Thus, although there is an offset, the radii derived in Martinez et al. (2019) largely agree with those derived here, and the difference is within systematic uncertainties of $\approx$2% for radii derived from *Gaia* DR2 parallaxes (Zinn et al. 2019b). The difference between our radii and those derived in Martinez et al. (2019) can likely be traced to differences in the effective temperature between the two samples. On average, we find $T_{\rm eff} - T_{\rm eff,\,M19} = -108 \pm 171$ K. This difference explains our smaller inferred radii (see Figure 3.5).

The comparisons in this section to Berger et al. (2020b) and Martinez et al. (2019) imply that ASPCAP is underestimating the effective temperature for stars in the APOGEE-KOI sample. This difference in temperature scale drives differences of $\sim$1% and $\sim$6% in the inferred radii and masses, respectively. A similar conclusion was found in Serenelli et al. (2017), who found that for dwarf and subgiant stars the ASPCAP temperature scale was underestimated compared to photometric temperature scales by $\sim$200 K. The difference in effective temperature between the Berger et al. (2020b) sample and the APOGEE sample is unsurprising because Berger et al. (2020b) calibrated their temperatures using stars with interferometric measurements from Boyajian et al. (2013) and

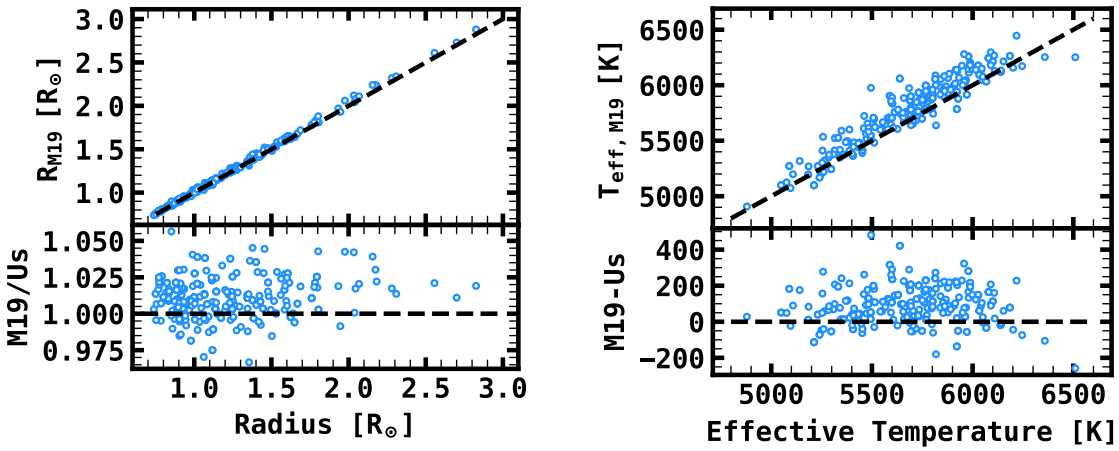


Fig. 3.5.— Comparison of the fundamental stellar properties derived in this work versus the stellar properties derived by Martinez et al. (2019) for the same stars (M19). In each panel, the dashed black line denotes agreement. *Left:* Comparison of the stellar radii. We find relative agreement, with an average offset and scatter of 1.1±1.3% in the ratio of the radii. *Right:* Comparison of the effective temperatures derived by ASPCAP and the effective temperatures from M19. There is a mean offset and scatter of $T_{\rm eff} - T_{\rm eff,\,M19} = -108 \pm 171$ K between the two samples. The systematically lower $T_{\rm eff}$ in ASPCAP is the likely reason for the systematic offset in stellar radii.

Huang et al. (2015), which are in agreement with the Serenelli et al. (2017) photometric temperature scale that was systematically higher than the ASPCAP temperature scale.

The question remains as to the origin of the difference between the ASPCAP and photometric temperature scales. One possible insight is that the APOGEE temperature scale is calibrated to color-$T_{\rm eff}$ relations derived via the Infrared Flux Method (IRFM) by González Hernández & Bonifacio (2009), while the Berger et al. (2020b) calibrate their $T_{\rm eff}$-color relation using effective temperatures derived from stars with directly-measured diameters via interferometry (Boyajian et al. 2013; Huang et al. 2015). The IRFM method of inferring $T_{\rm eff}$, depending on one's choice of zero-point calibration, can have systematic uncertainties of ∼100 K. Such a difference may be at the root of the discrepancy in the ASPCAP effective temperature scale.

### Planet Radii

We derive each of the planet radii using the reported transit depth in the DR24 KOI catalog (Mullally et al. 2015). We apply the simple relationship,

$$R_p = R_\star \sqrt{\delta_{tr}}$$

to calculate the planet radii in our sample, where $\delta_{tr}$ is the measured transit depth. The uncertainty in planet radius for our catalog is found by propagating the errors on $R_\star$ with the uncertainties from the *Kepler* DR24 transit depth measurement. The resulting planet radii in our sample have a median uncertainty of $\sigma_{R_p}/R_p = 3.2\%$ (mean: 3.4%).

## 3.3 Sample Selection and Planet Classes

For this study we define three individual samples that we introduce here before describing them in detail below. The first sample is the stellar planet-search sample, $\mathcal{S}$. $\mathcal{S}$ is the parent sample of stars that may have been observed by the APOGEE-KOI program. This translates to the *Kepler* field stars within the APOGEE footprint that are then down-selected based on our scientific goals. The second sample is $\mathcal{C}$, or the control sample, which is a subset of $\mathcal{S}$. Because we don't have detailed chemical abundances for each star in $\mathcal{S}$, $\mathcal{C}$ acts as a proxy from which we can infer the bulk properties (i.e., abundance-ratio distributions) of $\mathcal{S}$. The final sample is the vetted planet sample, $\mathcal{P}$. $\mathcal{P}$ is the sample of planets whose host stars were observed by the APOGEE-KOI Goal Program that is then further vetted to remove False Positives and ensure a well-characterized sample of planet candidates. Each of these samples are displayed in Figure 3.6, along with their metallicity distributions.

### 3.3.1 $\mathcal{S}$: Stellar Planet Search Sample

To select the appropriate planet search sample, $\mathcal{S}$, we start from the catalog of stars in Berger et al. (2020b). We downsample this table to replicate the selection function of the APOGEE-KOI survey. These cuts are listed below.

1. *Brightness Cut*, $H < 14$: This is the brightness limit in the APOGEE-KOI planet sample, chosen because it is the limit for which a one-hour integration with APOGEE yields a S/N $\gtrsim$ 10, i.e., sufficient to derive reliable radial velocities. We apply this cut to each star in the field sample.

2. *APOGEE Field Cut*: $100'' < d < 1.5\,^\circ$, where $d$ is the angular distance from the center of the nearest APOGEE-KOI field. The upper limit of $1.5\,^\circ$ represents the limit placed by the Sloan 2.5-meter telescope's field of view, and $100''$ is an instrumental limit derived from a central post that obscures targets in the center of the plate design (Owen et al. 1994; Zasowski et al. 2017).

At this point, it is important to note that the individual fields for the APOGEE-KOI program were chosen to maximize the number of observable KOIs per field. If each *Kepler* tile is expected to have the same number of KOIs, the choice to maximize the number of targets in the APOGEE-KOI program may introduce a bias leading us to overestimate the planet occurrence rate. However, it is more likely that the planet yield per field is driven by a combination of the number of stars per field where transiting planets are detectable, which would favor the fields closer to the Galactic mid plane, and the quality of the light curves in the particular field, which would be diminished by crowding and favor fields farther from the Galactic mid plane. Both of these effects are accounted for in our occurrence rate methodology either directly (e.g., the number of planet-search stars) or indirectly (e.g., the expected $S/N$ for a transiting planet with a given period and radius). Therefore, we believe that the choice of observed fields does not impart a significant bias that is not already accounted for in our methodology.

We applied a further series of criteria to ensure that our sample is well suited to the ASPCAP analysis and completeness model we employ in §3.C.3, and to remove stars that are evolved or likely to be a member of a binary system. To select this sample, we make use of the stellar properties derived by Berger et al. (2018, 2020b) to apply the following cuts:

1. *Effective Temperature Cut,* $4700\,\mathrm{K} < T_{\mathrm{eff}} < 6200\,\mathrm{K}$: We remove stars outside the temperature range well-suited to the ASPCAP analysis. Late K dwarfs show non-physical trends in abundance ratio, and F-type stars have increased abundances compared to the Solar neighborhood reported by ASPCAP (see Figure 3.2).

2. *Maximum Transit Duration Cut,* $t_{\mathrm{dur,\,max}} < 15\,\mathrm{hr}$: Because the *Kepler* Transiting Planet Search module (TPS; Twicken et al. 2016) doesn't include transit durations, $t_{\mathrm{dur}} > 15$ hr, we remove stars that can reasonably include such long duration transits from our planet-search sample. This criterion is logically analogous to removing evolved stars from the planet search sample. This is typical in *Kepler* occurrence rate studies, usually as a recommendation to removing stars with large radius, such as $R_\star \gtrsim 1.25 R_\odot$, when applying empirical measurements of the *Kepler* pipeline detection efficiency (Christiansen et al. 2015, 2016; Christiansen 2017; Burke & Catanzarite 2017a). To determine such stars, we employ the following approximation for the transit duration of a planet assuming a circular orbit and impact parameter of $b = 0$, with a given period, $P$,

$$t_{\mathrm{dur}} \approx 1.426\,\mathrm{hr} \left(\frac{\rho_\star}{\rho_\odot}\right)^{-1/3} \left(\frac{P}{\mathrm{days}}\right)^{1/3}, \tag{3.1}$$

where $\rho_\star$ is the mean density of the star. Finally, $t_{\mathrm{dur,\,max}}$ is obtained by setting $P = 300$ days. The motivation behind setting a limit of 300 days is to avoid regions of parameter space where planets would have fewer transits and as a result may introduce a higher rate of false alarms in our sample, which for this work we assume

is negligible.

3. *Astrometric Noise Cut, $RUWE < 1.2$*: We utilize the Renormalized Unit Weight Error ($RUWE$) from *Gaia* DR2 provided in Berger et al. (2020b) to remove stars that are likely to show signs of multiplicity. The $RUWE$ parameter is a combination of goodness of fit metrics that quantifies deviations of a given star's sky motion from a 5-parameter astrometric solution. Single stars are expected to show a Gaussian distribution centered at $RUWE = 1$, which suggests that sources with $RUWE$ significantly greater than that expected from a Gaussian distribution are likely to have companions that induce detectable centroid offsets in the *Gaia* DR2 astrometric pipeline. We choose $RUWE < 1.2$ as our cutoff, as motivated by Bryson et al. (2020), to be the limit above which we would reliably expect stars to be binaries.

4. *Likely Binary Cut*, `BinFlag` $\neq 1$ `or` 3: We remove stars that are likely to be binaries, as determined by Berger et al. (2018). Berger et al. (2018) use `BinFlag`=1 or `BinFlag`=3 to denote a star likely to be a binary due to its inferred radius. We do not remove stars with `BinFlag`=2, which are stars likely to be binaries as determined from high-resolution AO or speckle imaging, because those data are only available for a small subset of the planet search sample, and removing such stars is likely to create a bias.

After applying these cuts we are left with 20,625 stars in $\mathcal{S}$. This defines our planet-search sample, with stars that have typical masses ranging from 0.7-1.3 $M_{\odot}$, and distances ranging from 100-2000 pc.

### 3.3.2 $\mathcal{C}$: APOGEE-*Kepler* "Control" Sample

In addition to the KOIs that were observed in the APOGEE-KOI program, a number of stars were chosen to fill the APOGEE plates as a control sample for the purpose of

comparing the chemistry of stars with and without detected transiting planets. The control sample was chosen to reflect the bulk properties of the KOI sample by matching the joint distributions of effective temperatures, $H$-band magnitudes, and $\log g$ from the *Kepler* Input Catalog (KIC; Brown et al. 2011). It is from this sample of stars that we construct $\mathcal{C}$.

At this point, we want to emphasize the purpose of $\mathcal{C}$. $\mathcal{C}$ is used solely to infer the abundance distributions of $\mathcal{S}$. Therefore, there are two requirements needed to ensure that $\mathcal{C}$ is representative of the abundances of $\mathcal{S}$. First, it must broadly reflect the Galactic coordinates, distances, masses, and ages of the stars in $\mathcal{S}$, properties that are known to correlate with chemical abundance distributions (see e.g., Hayden et al. 2015). The second criterion is that there must not be systematic differences that would bias the ASPCAP analysis. For example, differences in $S/N$, $T_{\rm eff}$, and $\log g$ may all lead to systematic offsets in the derived abundances that could lead one to conclude there are differences in the underlying distributions when that is not truly the case.

Because $\mathcal{C}$ already reflects $\mathcal{S}$ in terms of Galactic coordinates, distances, and $H$ mag (and therefore $S/N$) by its very construction, we only need to apply the cuts that ensure the stars in $\mathcal{C}$ are amenable to the ASPCAP analysis, and that they reflect the ages and masses of the stars of interest. Therefore, we apply the *Maximum Transit Duration Cut* and the *Effective Temperature Cut*, because differences in the distribution of stellar densities (and therefore $\log g$) can be indicators of age differences, and differences in effective temperature are most likely to lead to systematic offsets in the derived abundances. After these two cuts, we are left with 72 stars in $\mathcal{C}$.

### 3.3.3 $\mathcal{P}$: Vetted Planet Sample

To ensure that we have a high purity planet sample, we apply an additional series of cuts to the planet candidates designed to remove False Positive detections, remove planets where the transit depth, and therefore planet radius measurement, may not be accurate,

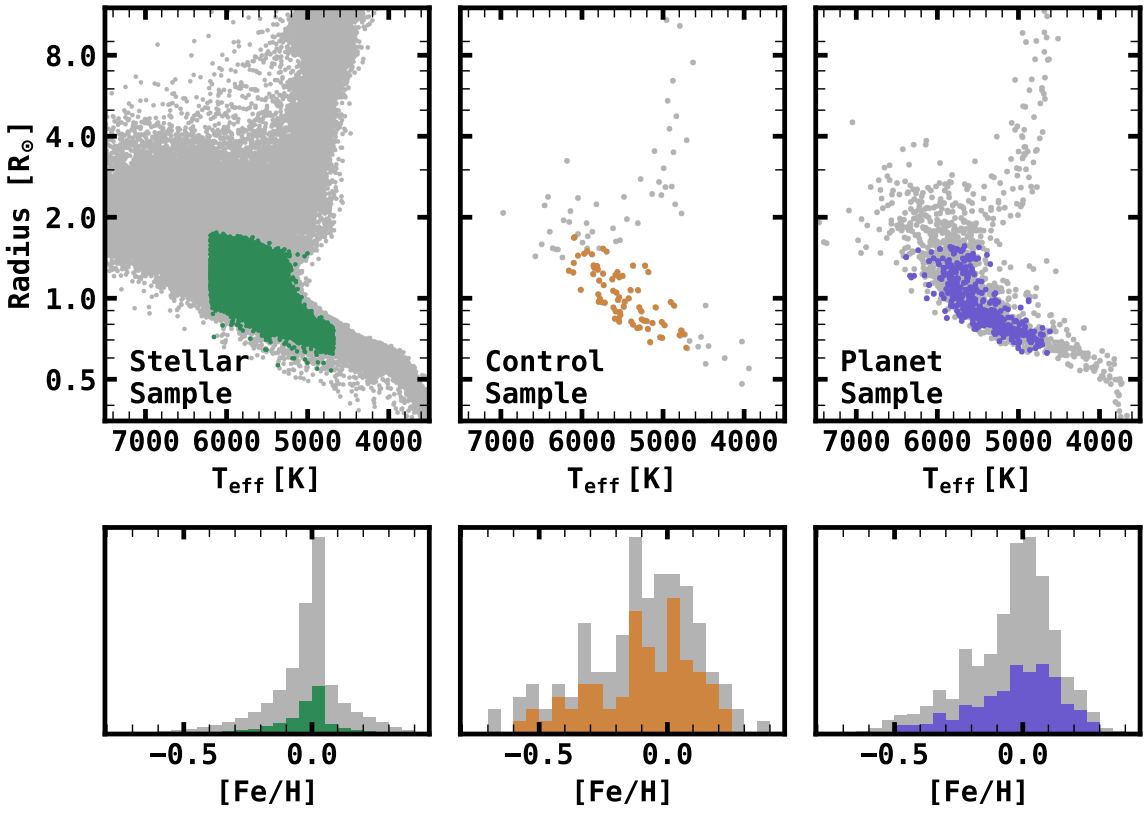


Fig. 3.6.— The three samples considered in this study. The effective temperature and radii of the stars in each sample are shown along the top row, and the metallicity distribution function for each sample is shown along the bottom row. The metallicity distributions are scaled to arbitrary units. *Left*: *Kepler* field stars with parameters derived in Berger et al. (2020b). The stars cut from $\mathcal{S}$ are shown in grey, and those included in $\mathcal{S}$ are shown in green. The metallicities for the stars in $\mathcal{S}$ are heterogeneous, or assumed to be solar, and thus are not as reliable for this study. *Center*: The stars in the Control sample (grey), and the subset of these stars included in $\mathcal{C}$ (tan). *Right*: All the stars in the APOGEE-KOI program (gray) and the stars included in $\mathcal{P}$ (purple).

and to restrict our sample to the parameter space well-defined by our completeness correction model (§3.C.3). We define and motivate each of these cuts below.

1. *ASPCAP Solution Cut:* First, we remove planet candidates whose host stars do not have a reliable ASPCAP solution. This cut was already implicitly made when adopting the stellar and planetary radii, but we repeat it here for emphasis. Because we are interested in measuring planet occurrence rates and their change with chemical abundances, we restrict our sample to stars for which the ASPCAP pipeline

has derived a reliable solution to the spectroscopic fit. Spectra that do not have such a fit will not have derived abundances and are therefore not appropriate to include in our analysis. We correct for this bias in §3.C.3.

2. *Reliability Cut*: To remove as many contaminants from $\mathcal{P}$, we remove all planet candidates with a False Positive disposition in the DR24 KOI catalog.

3. *Impact Parameter Cut, $b < 0.9$*: We remove all planet candidates with impact parameter, $b > 0.9$, as measured in the DR24 KOI catalog. Modeling transits with large impact parameters leads to greater uncertainties in the transit depth and therefore planet radius of the sample. Thus, we remove planet candidates with large impact parameters to ensure that we have a sample of planets with well-measured radii.

4. *Planet Radius Cut, $R_p < 23\,R_\oplus$*: We place an upper limit on the radius of a planet candidate in our sample of $23\,R_\oplus$ ($2.1\ R_{\rm Jup}$), which is consistent with the radius of the largest confirmed transiting exoplanet currently known, HAT-P-67b (Zhou et al. 2017). While inflated Hot Jupiters are known to have radii as large as $\sim 2R_{\rm Jup}$, most objects with radii larger than $2R_{\rm Jup}$ are likely very low-mass stars.

5. *Excess RV Variability Cut, $\epsilon_{RV} < 4.4$*: To remove EBs and eclipsing brown dwarfs from $\mathcal{P}$, we define a metric for excess RV variability, $\epsilon_{RV}$, as

$$\epsilon_{RV} \equiv MAD(RV)/\sigma_{RV}\ , \tag{3.2}$$

where $MAD(RV)$ is $1.4826\times$ the median absolute deviation of the individual RV measurements, and $\sigma_{RV}$ is the median RV uncertainty for all epochs. To estimate $\sigma_{RV}$, we add the reported RV uncertainty for each visit in quadrature with $\sigma_{RV,\,\rm min} = 72\ {\rm m\,s^{-1}}$, which has been noted as a reliable lower limit on the relative RV error for high S/N observations in DR16, where the reported error may be

underestimated (Price-Whelan et al. 2020). Given the varying brightness of our targets, the RV uncertainties are highly correlated with the single epoch spectrum $S/N$. As a result, a flat cut in the scatter of the RV measurements could remove bonafide planet candidates with dim host stars, while missing astrophysical False Positives around bright host stars. $\epsilon_{RV}$, therefore, gives a more accurate assessment of whether a given star is RV-variable than a flat cut in the scatter of the RV measurements. We decide on $\epsilon_{RV} = 4.4$ because that is equal to the median plus twice the median absolute deviation in our sample. APOGEE RV observations in the KOI sample are capable of placing upper limits into the planetary mass regime, typically between 1-10 $M_{jup}$, depending on the orbital period of the transiting planet, spectrum $S/N$ at each epoch, and mass of the host star. Therefore, by removing all stars with significant RV variability in our sample, we in turn remove any contaminating eclipsing binaries. APOGEE's RV precision is not quite effective enough to detect planetary mass companions without detailed modeling, so our metric for RV variability is not likely to remove any real planets, such as hot Jupiters. We justify this briefly in §3.4.2.

After these cuts we are left with 504 total planet candidates in $\mathcal{P}$. The breakdown of these planets by field is shown in Table 3.2. The radius and period characteristics of these candidates are shown in Figure 3.7. There are a number of trends evident in this figure. For instance, the radius gap (Fulton et al. 2017) is clear in both the top and bottom panels of our figure, as well as a slope in orbital period in the gap measured by previous authors (Fulton & Petigura 2018; Martinez et al. 2019); these two features qualitatively validate the precision and accuracy of the radii in $\mathcal{P}$.

### 3.3.4 Adopted Planet Classes

We divide the planets in $\mathcal{P}$ into multiple classes based on their orbital period and radius, as many previous studies have shown metallicity correlations that depend on these

| Field | $\alpha$(h:m:s) | $\delta$(d:m:s) | $n_{pl}$ | $n_\star$ | $F_\star$ |
|---|---|---|---|---|---|
| K04 | 19:42:47 | +49:54:07 | 70 | 3303 | 0.16 |
| K06 | 19:13:39 | +46:52:30 | 86 | 2947 | 0.143 |
| K07 | 19:00:17 | +45:12:47 | 69 | 2686 | 0.13 |
| K10 | 19:36:30 | +46:00:18 | 101 | 3966 | 0.192 |
| K16 | 19:31:05 | +42:05:24 | 82 | 4211 | 0.204 |
| K21 | 19:26:13 | +38:09:36 | 96 | 3512 | 0.17 |
| **All** | **N/A** | **N/A** | **504** | **20,625** | **1.00** |

Table 3.2: The coordinates, number of stars in $\mathcal{S}$, number of planets in $\mathcal{P}$, and fraction of stars in $\mathcal{S}$ for each APOGEE-KOI field.

properties. The adopted planet size classes are motivated partially by empirical and theoretical boundaries where applicable, and partially by conventions in the literature, as explained below. For the planet size classes, we define the following:

1. *Sub-Earths*, $R_p < 1R_\oplus$: The number of planets in this class suffers particularly severely from low survey completeness, and for that reason these planets are drastically skewed toward lower orbital periods. Because of this, we don't consider these planets when measuring occurrence rates, and are hesitant to draw major conclusions when comparing the abundances of their host stars to those of stars in $\mathcal{C}$. There are 57 Sub-Earths in $\mathcal{P}$.

2. *Super-Earths*, 1.0 $R_\oplus \leq R_p < 1.9R_\oplus$: Super-Earths are defined as planets larger than earth, with an upper limit set by the minimum of the radius gap in the planet radius distribution between 1-4 $R_\oplus$ in our sample (Figure 3.7). The 1.9 $R_\oplus$ boundary we find between Super-Earths and Sub-Neptunes is slightly different than that found by Fulton et al. (2017), and closer to the boundary found by Martinez et al. (2019). There are 198 Super-Earths in $\mathcal{P}$.

3. *Sub-Neptunes*, 1.9 $R_\oplus \leq R_p < 4R_\oplus$: The lower boundary is driven by the radius gap as discussed above. The upper boundary is placed as the limit where the occurrence of Sub-Neptunes tends to zero. While a more precise physically-motivated boundary is not clear, we choose $4R_\oplus$ as an upper limit to be consistent with conventions

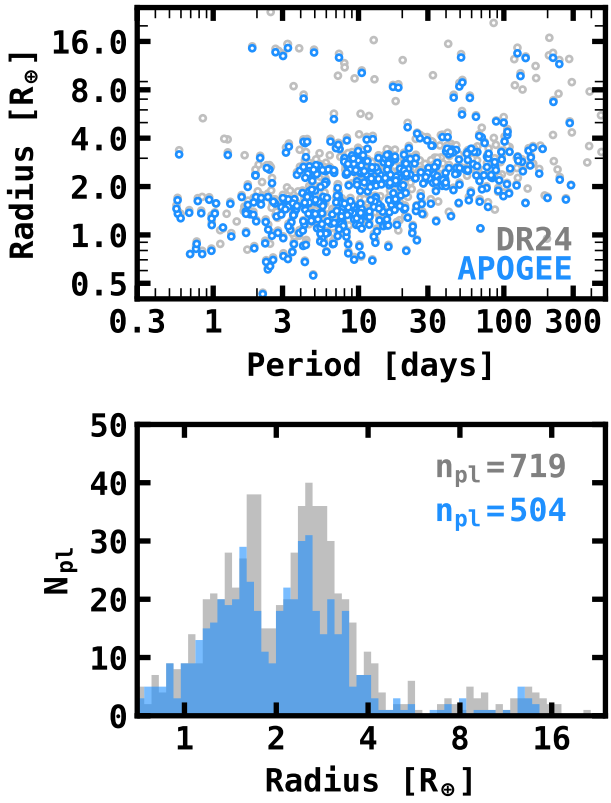


Fig. 3.7.— The planets in $\mathcal{P}$, plotted with all the DR24 planet candidates that have a host in $\mathcal{S}$. *Top*: The planet radius and orbital period of all planets in $\mathcal{P}$. The grey points show all the planets from the DR24 KOI catalog with a host star in $\mathcal{S}$. *Bottom*: The radius distribution of the planets in $\mathcal{P}$. The grey histogram shows the radii of all the planets in DR24 with a host in $\mathcal{S}$, while the blue shows those in $\mathcal{P}$. The primary reasons for exclusion in $\mathcal{P}$ are RV variability, a poor solution from ASPCAP, or pre-DR24 target selection.

in the literature. There are 236 Sub-Neptunes in $\mathcal{P}$.

4. *Sub-Saturns*, $4\ R_\oplus \leq R_p < 8_\oplus$: The lower radius boundary for Sub-Saturns is given by the decrease in Sub-Neptune occurrence rates described above, and the upper limit is driven by the approximate radius at which planets are typically $\gtrsim 100 M_\oplus$ (Petigura et al. 2017b). There are 11 Sub-Saturns in $\mathcal{P}$.

5. *Jupiters*, $8\ R_\oplus \leq R_p < 23\ R_\oplus$: The radius range for Jupiter-sized planets is given by the upper boundary for Sub-Saturns, and by the upper limit placed by the largest

known confirmed planet, as mentioned in §3.3.3. There are 17 Jupiters in $\mathcal{P}$.

In addition to these size classes, we also define three different period boundaries for planets of differing orbital separations (i.e., orbital period).

1. *Hot*, $P \leq 10$ days[7]: There is a well-documented break in the occurrence rate of planets with respect to orbital period, showing two different regimes above and below $P \sim 10$ days (Youdin 2011; Howard et al. 2012; Mulders et al. 2015c). There are 235 hot planets in $\mathcal{P}$.

2. *Warm*, $10 < P \leq 100$ days: The boundary for warm planets is given by the lower bound on hot planets, and on the upper end where completeness becomes an issue for Super-Earths. This range of orbital periods is also consistently used in the literature, so we adopt it as well for ease of comparison. There are 239 warm planets in $\mathcal{P}$.

3. *Cool*, $100 < P \leq 300$ days: We define this period range as our cool sample. The number of planets in this range suffers severely from decreased *Kepler* survey efficiency, and only contains 30 planets in $\mathcal{P}$. In addition, studying the population of *Kepler* planets with $P \gtrsim 300$ days requires a careful approach to modeling the *Kepler* False Alarm rate, which we assume to be negligible (Bryson et al. 2020).

We refer to these classes often throughout the rest of this work.

### 3.3.5 Assessment of Differences Between Host Star Abundances and the Field

In this section we examine whether there are any clear correlations with planet type and host chemical abundance. We also make more detailed comparisons between the

[7]Note: For the occurrence rate analyses, our definition of hot planets doesn't include planets with $P < 1$ day, due to the lack of injections used to test the *Kepler* pipeline completeness at these short periods (see §3.C.3 and Figure 3.20).

abundances of $\mathcal{C}$ and $\mathcal{P}$. The chemical abundances of both $\mathcal{P}$ and $\mathcal{C}$ are shown in Figure 3.8. For this section, we rely on the abundance ratios to Fe, [X/Fe], because there is a clear offset in [Fe/H] between $\mathcal{C}$ and $\mathcal{P}$ visible in Figure 3.8, where stars in $\mathcal{C}$ are more metal-poor on average. This is a well-known property of the stars with known transiting planets when compared to the general population of stars (i.e., regardless of whether they are known to host planets or not) in the *Kepler* field. Because of this difference, using [X/H] as a metric is almost certainly guaranteed to reproduce the [Fe/H] differences already known, and our goal is to search for new differences.

After defining the planet size and orbital period classes above, the first natural ques-

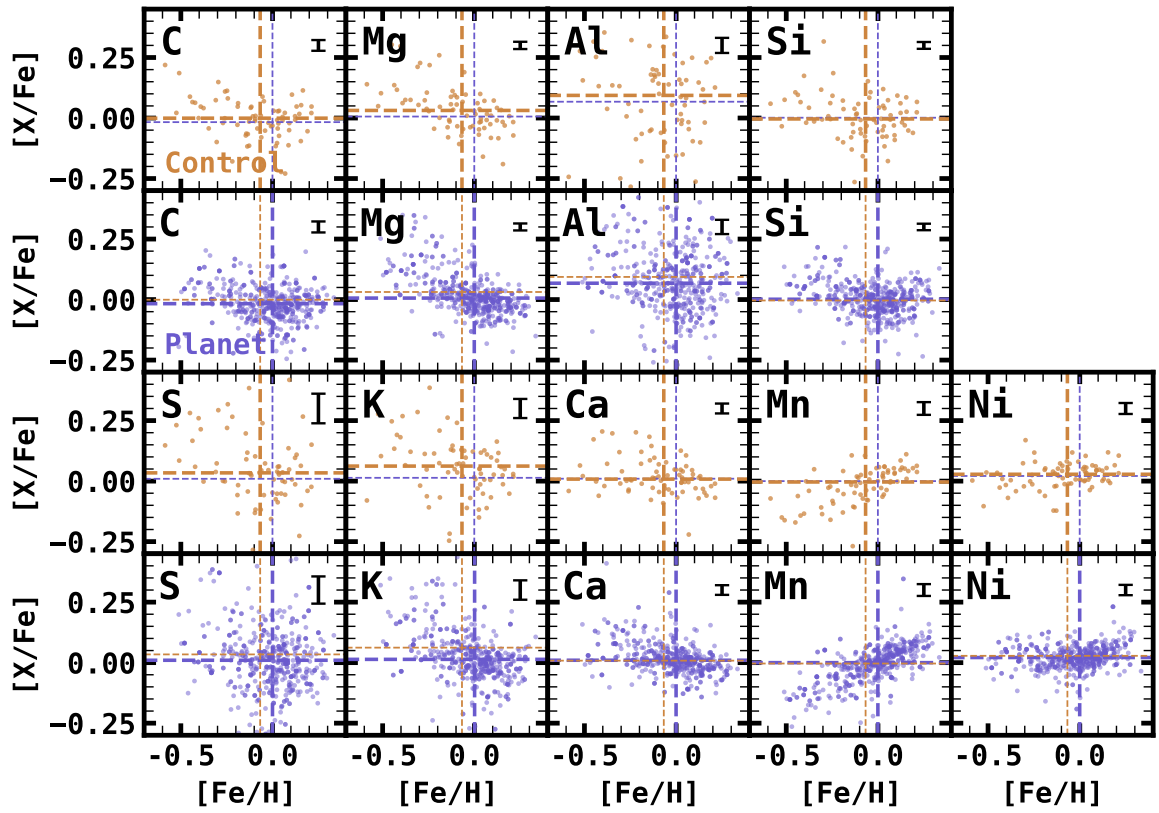


Fig. 3.8.— Chemical abundances for the planet host (purple) and control (tan) samples. The chemical abundance displayed is shown in the upper left corner of each panel. The median error ($\pm 1\sigma$) for each abundance is shown by the black error bar in the top right corner of each panel, and the dashed lines indicate the median abundances for the planet host sample (purple) and the control sample (tan).

tion is whether hosts of differing planet classes tend toward specific abundance patterns. Therefore, to detect any differences in the distribution of the host star abundances and the abundances of general stars in the field, we apply four unique statistical tests, considering a result significant if the $p$-value for the statistic is $<0.005$. The results of these tests are shown in Table 3.7, and for the sake of brevity they are discussed further in the Appendix (3.B). In short, we find no new credible differences, according to these tests, between the chemistry of stars in $\mathcal{C}$ and those in $\mathcal{P}$ that are not easily explained by already known trends between planet properties and the metallicities of their host stars (Santos et al. 2004; Valenti & Fischer 2005; Ghezzi et al. 2010, 2018; Buchhave et al. 2014; Schlaufman 2015; Mulders et al. 2016; Wilson et al. 2018; Petigura et al. 2018; Narang et al. 2018). The median and median absolute deviations of abundance distributions for these samples are listed in Table 3.3.

| [$X_i$/Fe] | $\mathcal{C}$ | $\mathcal{P}$ |
|---|---|---|
| Fe[a] | -0.068±0.183 | -0.001 ± 0.161 |
| C | -0.001±0.071 | -0.017 ± 0.060 |
| Mg | 0.031±0.082 | 0.006 ± 0.059 |
| Al | 0.094±0.201 | 0.067 ± 0.125 |
| Si | -0.004±0.090 | 0.001 ± 0.058 |
| S | 0.034±0.125 | 0.010 ± 0.100 |
| K | 0.062±0.096 | 0.014 ± 0.075 |
| Ca | 0.008±0.059 | 0.010 ± 0.046 |
| Mn | -0.004±0.077 | -0.000 ± 0.073 |
| Ni | 0.028±0.044 | 0.021 ± 0.042 |

Table 3.3: The median and median absolute deviation of each abundance distribution in $\mathcal{C}$ and $\mathcal{P}$. [a]For iron, the abundance is reported with respect to Hydrogen, [Fe/H]

## 3.4 Results

### 3.4.1 Abundance Trends with Planet Period and Radius

In this section, we test whether there are any correlations between the host star abundances and planet properties. While these correlations can reveal important trends, it is important to note that the trends discussed in this section do not take completeness or detection biases into account. When appropriate, we mention when we believe an effect may be a result of a lack of completeness. A more thorough investigation would include correcting for biases in the *Kepler* and APOGEE-KOI surveys, which is performed in §3.4.2.

In Figures 3.9 and 3.10 we plot the mean and variance of the abundance distributions for different planet radius and planet period bins. As in the literature, we recover an anti-correlation [Fe/H] of the host star and the planet orbital period. We also recover a positive correlation between the planet radius and the host star [Fe/H]. Within these broader correlations, there are a few interesting results. For instance, while there is a general anti-correlation between planet orbital period and host star [Fe/H], there is an increase in the average metallicity distribution at $P \sim 30$ days. This slight increase is apparent in Figure 3 of Petigura et al. (2018) as well, though to a lesser extent. This feature is also pointed out in Wilson et al. (2018) as a possible transition period at $P \sim 23$ days. While the exact cause of this bump is not well-constrained by this work, we hypothesize that it is due to an increase in the relative number of Sub-Saturns at these orbital periods. Because the presence of Sub-Saturn planets are positively correlated with enhanced metallicity, and they also have increasing occurrence rate at warm orbital periods.

We also see a number of interesting trends between planet radius and host star [Fe/H]. For one, we confirm the claim made by several authors (Buchhave et al. 2014; Schlaufman

2015; Wang & Fischer 2015; Ghezzi et al. 2018; Petigura et al. 2018) that larger radius planets are positively correlated with host star [Fe/H]. Digging deeper we also find a few other interesting results. For instance, there is an apparent increase in the metallicity of Sub-Earths. However, as cautioned, these planets suffer from low completeness, and are heavily skewed toward shorter periods. Thus, this bump can be explained by the stellar metallicity planet orbital period trend discussed above. Another interesting trend we find is that Sub-Neptunes with larger radii ($R_p \sim 3 - 4R_\oplus$) have host stars with enhanced [Fe/H] compared to smaller Sub-Neptunes ($R_p \sim 1.9 - 3R_\oplus$). This is predicted by the theory of atmospheric loss via core-heating (Gupta & Schlichting 2019, 2020). The radius of Sub-Neptunes are expected to increase with metallicity, $Z$, via the relation $d \log R_p / d \log Z \sim 0.1$.

To test for significant trends in our sample, we calculate the Spearman $\rho$ rank correlation coefficient between the iron normalized abundances for the planet hosts in our sample and the logarithm of the radii and periods of the planets in our sample. The results of these statistical tests are shown in Table 3.4. As with the tests in the previous section, we consider a result significant if the $p$-value is $<0.005$. In this vein we uncover a few statistically significant correlations. The most clear correlations we recover are correlations with planet radius and [Mn/Fe], [Mg/Fe], and [S/Fe]. Perhaps unsurprisingly, the correlations with [Mn/Fe] is positive, and the correlation with [Mg/Fe] is negative, meaning that they are most likely influenced by statistically strong correlations with [Fe/H]. We can see from Figure 3.8, in fact, [Mn/Fe] displays strong correlations with [Fe/H], and [Mg/Fe] shows a strong anti-correlation.

However, the origin of the positive trend with [S/Fe] is less clear. [S/Fe] does not display the same correlation as [Mn/Fe]. Interestingly, [S/Fe] is the only abundance (aside from [Mn/Fe] for the same reasons as above) that is significantly correlated with planet period as well ($p = 2.2 \times 10^{-5}$). Even more interesting, these correlations cannot be explained by already known trends with [Fe/H]. If that were the case, [S/Fe] would be

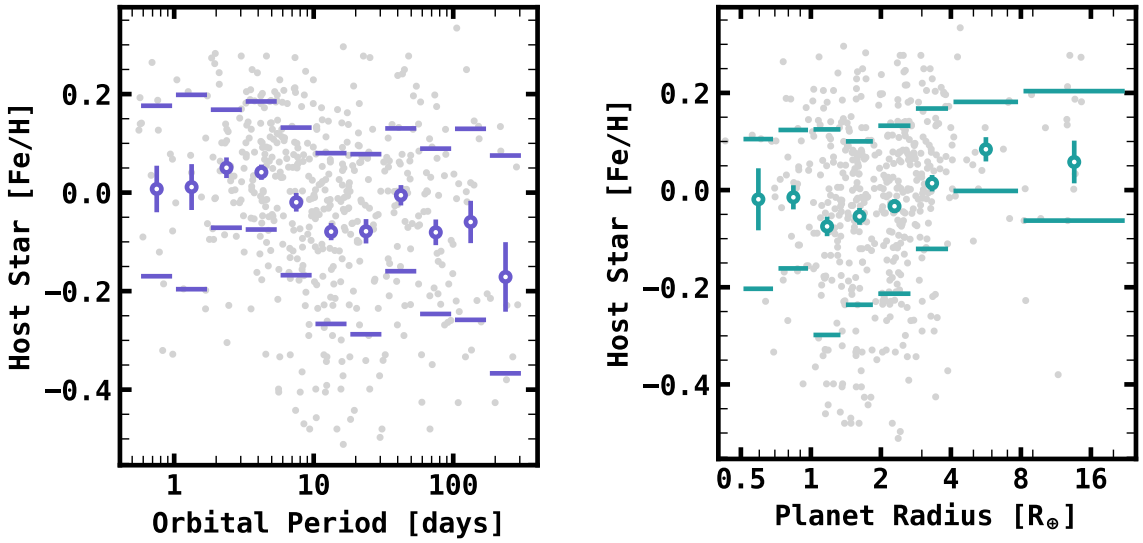


Fig. 3.9.— *Left:* The average metallicity for host stars of planets in given orbital period bins. The circular points show the average metallicity, while the horizontal lines show the 68% confidence interval on the metallicity distribution. We recover the same planet period–stellar metallicity anti-correlation reported in previous literature (e.g., Mulders et al. 2016; Wilson et al. 2018) *Right:* The average host star metallicity for planets of differing radius bins. The bins are chosen to reflect the size classes, Sub-Earths, Super-Earths, Sub-Neptunes, Sub-Saturns, and Jupiters. The Sub-Earth, Super-Earth and Sub-Neptune classes are split into two radius bins each. We find similar relations as in the literature, that there is a constant increase in the average host metallicity for planets with radii ranging from $\sim$1-8$R_{\oplus}$. The increased metallicity for the Sub-Earth planets is likely a result of the bias against detecting sub-Earths at $P \gtrsim 10$ days.

expected to show a correlation with either planet period or radius and then must show an anti-correlation in the other, as with [Fe/H]. However, [S/Fe] shows a strong positive correlation with both planet radius and planet period. It's not clear that the source of the trend with S abundance-ratios is astrophysical, but it warrants further investigation.

### 3.4.2 Planet Occurrence as a Function of Chemical Abundance

In this section we calculate the occurrence rates of planets as a function of $P$, $R_p$, and $[X/H]$. We fit a parametric model to describe the general trends of the planetary distribu-

tion function (PLDF) and their dependence on these properties. This analysis represents an improvement from the analysis in §3.3.5, as we are now accounting for the selection functions of *Kepler* and APOGEE; thus the conclusions we draw about the PLDF from this analysis should be independent of observational biases.

We employ a common strategy to measure the PLDF that has been used in previous studies: the number of planets per star (NPPS) is calculated over a grid of $P$ and $R_p$, utilizing the inverse detection efficiency method and a maximum likelihood approach (e.g., Youdin 2011; Fressin et al. 2013; Burke et al. 2015; Mulders et al. 2015c, 2018; Petigura et al. 2018). We give a brief description of our completeness model below, but refer the reader to the Appendix (§3.C) for details on our methodology.

### Completeness Model

In this subsection we give a brief description of our completeness model, $\eta(\boldsymbol{x}, \boldsymbol{z})$, where $\boldsymbol{x}$ are planet properties and $\boldsymbol{z}$ are stellar properties, but refer the reader to the appendix for details (§3.C.3). Our approach varies slightly from most *Kepler* occurrence rate studies, because we also need to correct for biases inherent in the follow-up program. In other words, inclusion in $\mathcal{P}$ is dependent on more than membership in $\mathcal{S}$ and a detected planet candidate in *Kepler*. There are additional biases imposed by the APOGEE selection

| [X/Fe] | $n_{pl}$ | $\rho_P$ | $p_P$ | $\rho_{R_p}$ | $p_{R_p}$ |
|---|---|---|---|---|---|
| Ni | 504 | -0.003 | 0.95 | 0.117 | 0.0087 |
| Si | 504 | -0.001 | 0.99 | -0.028 | 0.53 |
| Mg | 504 | 0.084 | 0.058 | -0.155 | 0.00049 |
| C | 504 | 0.044 | 0.33 | -0.002 | 0.96 |
| Al | 500 | 0.049 | 0.28 | 0.045 | 0.31 |
| Ca | 504 | 0.068 | 0.13 | -0.032 | 0.48 |
| Mn | 504 | -0.133 | 0.0028 | 0.163 | 0.00023 |
| S | 503 | 0.164 | 0.00022 | 0.139 | 0.0018 |
| K | 500 | 0.101 | 0.023 | -0.060 | 0.18 |

Table 3.4: The results of the Spearman $\rho$ rank coefficient to test for correlations between abundance ratios to iron and $\log P$ and $\log R_p$.

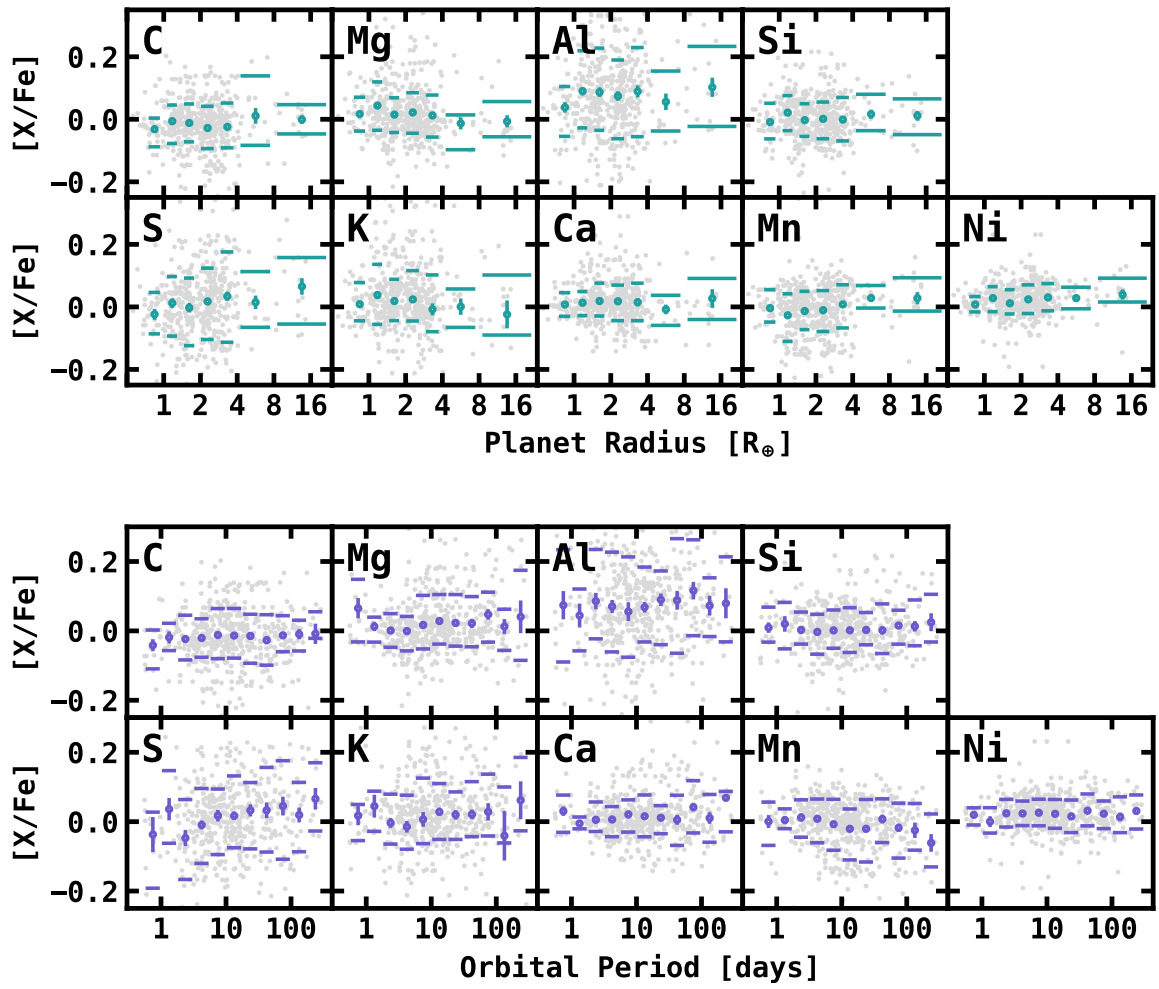


Fig. 3.10.— *Top:* Trends with planet radius and abundance ratios to iron. Just like with Figure 3.9, the points represent the means of each bin, with error bars representing the error on the mean [X/Fe] from bootstrapping. The horizontal lines show the 16th and 84th percentiles of the distribution in each bin to display the variance. We detect significant positive correlations between [Mn/Fe], and [S/Fe] vs. $R_p$. *Bottom*: The distribution of host star abundance ratios to iron as a function of planet period. We detect a statistically significant positive correlations between [S/Fe] and $P$. Such a correlation cannot be explained with well known trends in [Fe/H].

function, instrumental setup, and spectroscopic analysis pipeline that must be considered.

In total we account for four unique biases for a planet candidate to be included in $\mathcal{P}$:

1. The geometric probability that a planet with a randomly oriented orbital plane transits its host star ($p_{\rm tra}$)

2. The probability that a transiting planet is detected by *Kepler* ($p_{\rm det}$),

3. The probability that a detected planet candidate was observed in the APOGEE-KOI program ($p_{\rm apo}$)

4. The probability that ASPCAP doesn't fail to produce reliable atmospheric parameters for the host star ($1 - p_{\rm fail}$).

Assuming that each of the four terms above are independent, we calculate the total average survey efficiency for each field as the product of each term, given by

$$\langle \eta \rangle = \frac{1}{n_\star} \sum_i^{n_\star} p_{\rm tra,i} \times p_{\rm det,i} \times p_{\rm apo,i} \times (1 - p_{\rm fail,i}) \ , \tag{3.3}$$

where $\langle \eta \rangle$ is the average survey efficiency across $\mathcal{S}$. The mean survey efficiency for each field is shown in Figure 3.11. By marginalizing over all the stars in $\mathcal{S}$ in this way, we've removed stellar properties from our expression for survey efficiency, so that $\eta = \eta(\boldsymbol{x}) = \eta(P, R_p)$. This relies on an implicit assumption that chemical abundances are not correlated with survey efficiency.

### Occurrence Rates in the $P$-$R_p$ Plane

We first calculate the occurrence rate of planets in the $P$-$R_p$ plane, making use of the completeness model in §3.C.3. Because we are not applying any stellar properties (i.e., abundances) for these calculations, we calculate the occurrence rates as described in §3.C.1 and §3.C.2 for equally spaced bins in $\log P$ and $\log R_p$.

We first divide the $P$-$R_p$ plane into logarithmic bins of $\Delta \log P \times \Delta \log R_p = 0.25$ dex $\times$ 0.15 dex, and we plot these occurrence rates in Figure 3.12. Each bin is shaded in accordance with its occurrence rate, and annotated with our measured occurrence rate and error, or with an upper limit on the occurrence rate in the case that a planet was not detected in that bin. For compactness, the error on the occurrence rate is taken to be

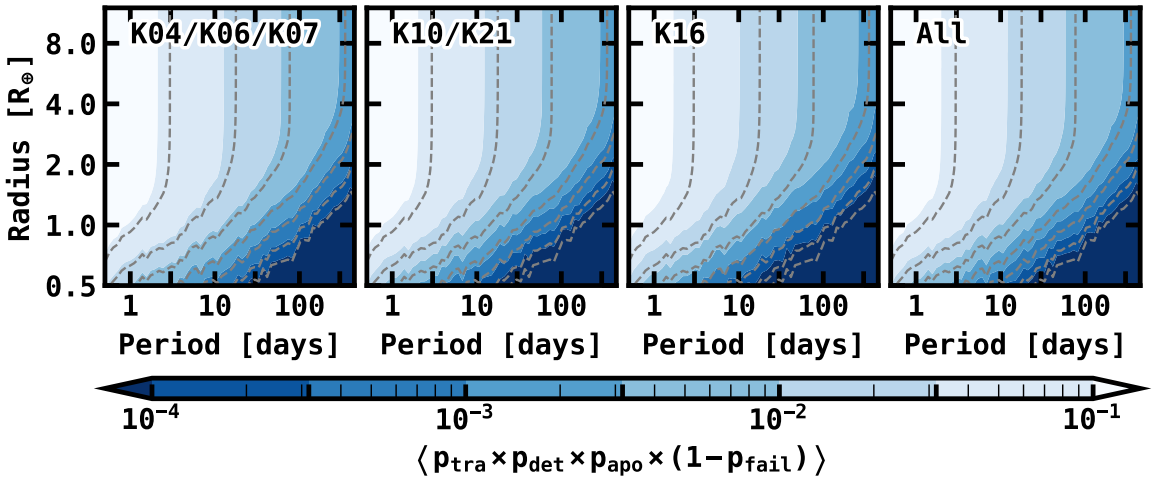


Fig. 3.11.— The mean completeness model, $\eta = p_{\rm tra} \times p_{\rm det} \times p_{\rm apo} \times (1 - p_{\rm fail})$, for each APOGEE field in $\mathcal{S}$, and the combined model from all fields. The blue filled contours give the survey efficiency in the $P$-$R_p$ plane, representing the probability that a given planet orbiting a star in $\mathcal{S}$ is in $\mathcal{P}$. The lightest shade shows where $\eta > 0.1$, while the darkest shade represents survey efficiencies of $\eta < 10^{-4}$. The gray dashed lines are the corresponding contours for the *Kepler* DR24 pipeline efficiency and are shown for comparison to highlight the effects of the APOGEE-KOI program selection function. The panels representing the APOGEE fields are organized from least to most divergent from the DR24 pipeline efficiency with the combined survey efficiency on the far right.

half of the 68% confidence interval around the measured value, which is why some of the errors imply a range of uncertainty with a negative occurrence rate, which is unphysical. Bins that do not have any annotations represent regions with low completeness where our derived upper limit is not restricting.

We use the same bins as in Petigura et al. (2018) for the sake of comparison, and we find that our results are qualitatively similar. For instance, we both find that the most abundant planet types are warm Sub-Neptunes, warm Super-Earths, and then cool Jupiters, in that order. For Jupiters, we find a sharp rise in occurrence rate for $P > 100$ days. This trend is present in the CKS sample as well, and has been noted in RV surveys (Cumming et al. 2008). This rise in occurrence rate is thought to be correlated with the water ice line at $\sim$1 au, leading to the facilitation of more massive planetary cores. There is also an island of relatively high occurrence for hot Jupiters centered

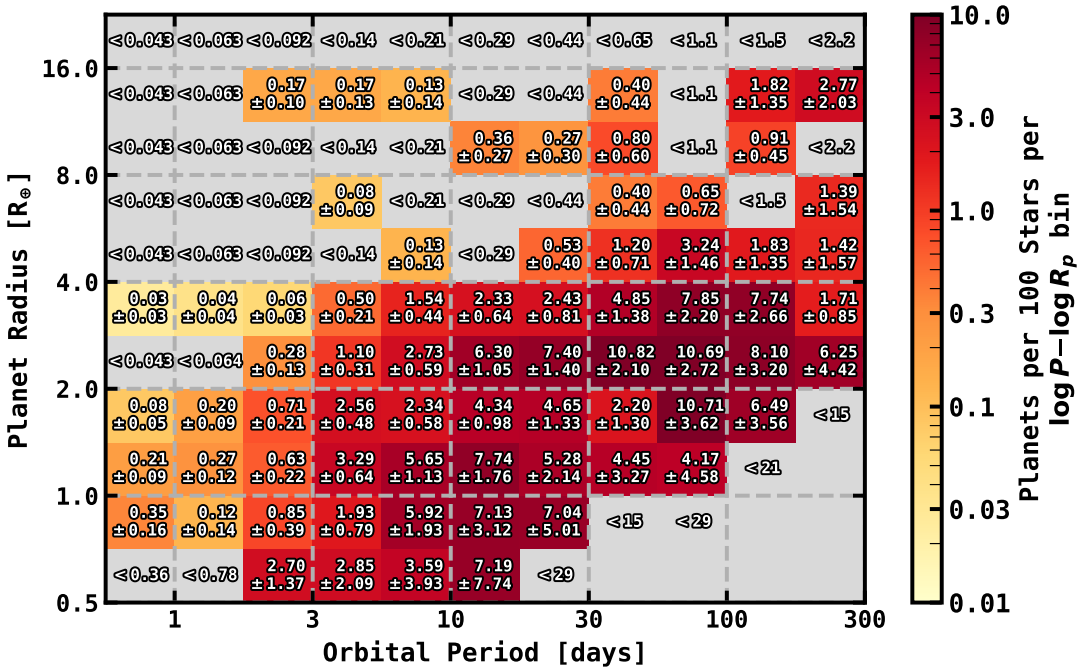


Fig. 3.12.— The planet occurrence rate in the $P$-$R_p$ plane. We divide the planet into bins of size $\Delta \log P \times \Delta \log R_p = 0.25 \times 0.15$ dex. The color shows the measured occurrence rate in the bin of interest on a logarithmic scale. The gray bins do not have any detected planets. The numbers in each cell shows the occurrence rate in units of number of planets per 100 stars. The uncertainty shown is taken to half of the 68% confidence interval range on the occurrence rate. Bins without detected planets have the upper limit displayed. Bins with no detected planets and no upper limit listed are areas of low completeness where an upper limit is not constraining.

on $P \approx 3$ days. From our data alone, it's not clear if this is a statistically significant increase centered at $P \approx 3$ days, or if it is simply a result of declining occurrence rates below $P \approx 100$ days. However, this increase in occurrence rates is also found in the California Planet Search program (Cumming et al. 2008), the CKS survey (Petigura et al. 2018), and other *Kepler* occurrence rates studies (e.g., Hsu et al. 2019), lending credibility to its existence. Overall, we find a similar occurrence rate for hot Jupiters of $f = 0.48^{+0.16}_{-0.24}$ planets per 100 stars, compared to the CKS team's measurement of $f \approx 0.57$ planets per 100 stars. This agreement in the occurrence rate of hot Jupiters bolsters our claim from §3.3.3 that removing RV variable sources does not remove a significant fraction of hot Jupiters.

However, we find a few key differences with previous studies as well. For instance, the occurrence of Sub-Neptunes and Super-Earths is nearly twice as high in some of the bins as compared to that found by CKS survey. One explanation for this apparent difference in the occurrence rates of small planets is simply a systematic difference in the planet radii. For instance, this work typically has more precisely-measured planet radii due to the inclusion of *Gaia* parallaxes in our analysis, which could cause certain bins in the $P$-$R_p$ plane to appear to have higher occurrence simply due to sharper features in the occurrence rate distribution. The bins themselves were also chosen arbitrarily, so increased occurrence for a given bin can appear inflated due to the choice of bin edges. To more accurately judge this potential difference, we calculate occurrence rates in arbitrarily small bins in the $P$-$R_p$ plane, then convolve these occurrence rates with a two-dimensional Gaussian kernel of size $\Delta \log P \times \Delta \log R_p = 0.25$ dex $\times$ 0.1 dex. The occurrence rates as a result of this smoothing are shown in Figure 3.13. This figure gives a more intuitive understanding of the occurrence rate of planets in the $P$-$R_p$ plane, and avoids the effects of binning that may misrepresent the PLDF. We find that our occurrence rates indeed are slightly larger than in Petigura et al. (2018) at the peak of the warm Sub-Neptune and Super-Earth distributions. This difference may be due to the APOGEE-KOI selection function, which selects a higher fraction of lower mass stars due to its magnitude cut in the near-infrared where such stars are brighter, rather than on the optical $Kp$ magnitude. Lower mass stars are known to have increased occurrence rates for small ($R_p \lesssim 4R_\oplus$) planets (Mulders et al. 2015c).

One feature present in our occurrence rate distribution is the radius gap (Fulton et al. 2017), with a notable dependence on the location of the gap with orbital period. This trend was uncovered by an independent analysis of the CKS spectra performed by Martinez et al. (2019). We find excellent agreement between the slope they found and the planet occurrence rate distribution in our sample. This slope in the radius gap is shown as a dashed black line in Figure 3.13.

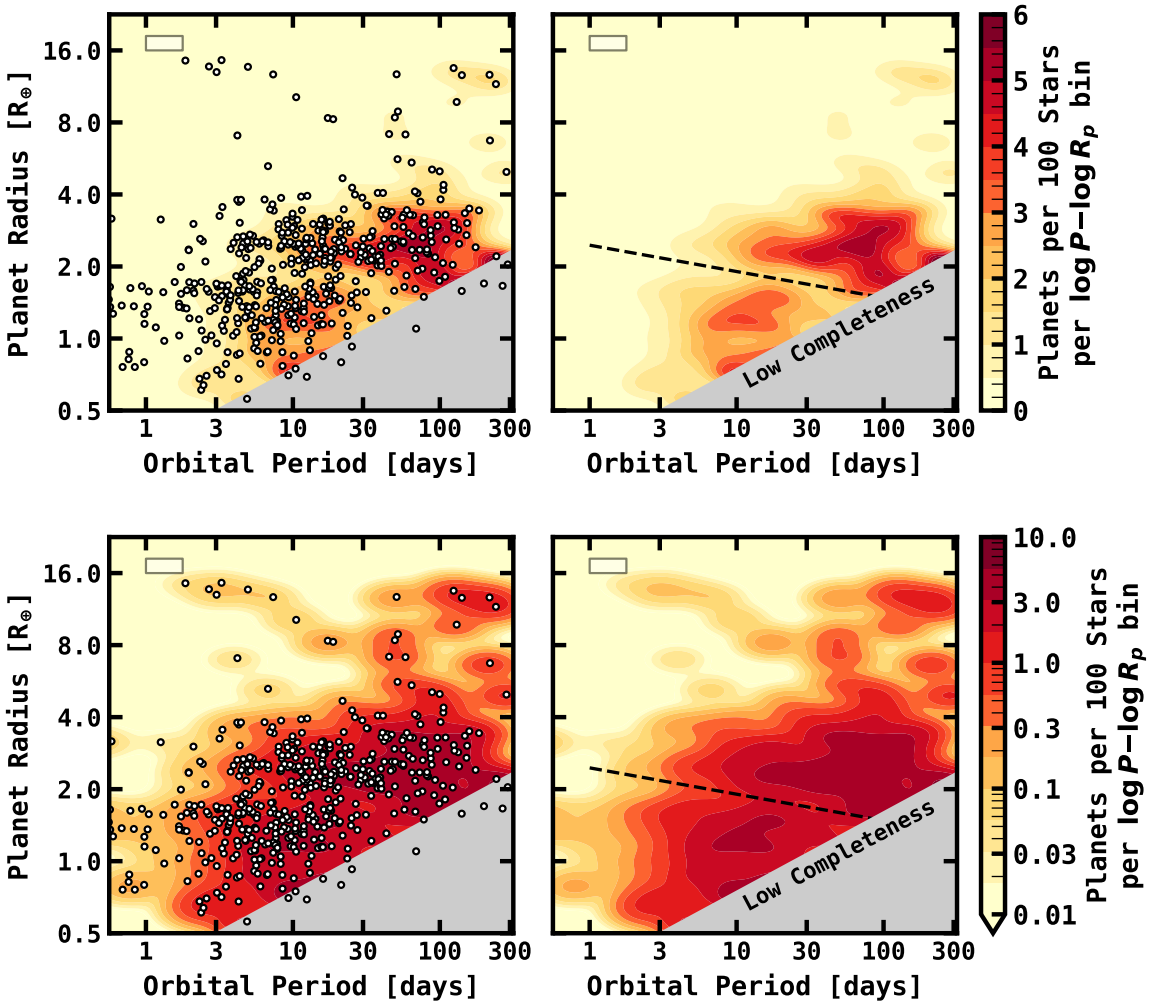


Fig. 3.13.— The planet occurrence rate in the $P$-$R_p$ plane. For each row of figures, the left panel shows the planets in $\mathcal{P}$ as white points, and the filled contours show the derived occurrence rate, with darker shades representing higher occurrence rates. The top row displays the occurrence rates on a linear scale, and the bottom row displays occurrence rates on a log scale, where darker shades of red indicate higher occurrence and lighter shades of yellow indicate lower occurrence. A box in the upper left corner of each panel shows the FWHM of the Gaussian kernel used to calculate the contours for this figure. The gray region denotes areas of low survey completeness. The dashed black line shows the location and slope of the radius gap measured by Martinez et al. (2019).

We also find that the occurrence rate of Sub-Neptunes and Super-Earths as a function of orbital period can be well described by a distribution of the form,

$$f_P = CP^{\alpha}(1 - e^{-(P/P0)^{\gamma}}) \ , \tag{3.4}$$

which effectively acts as a power law distribution, with a break at $P = P_0$. At $P \ll P_0$, the distribution acts as a power law with $f_P \propto P^{\alpha+\gamma}$, and at $P \gg P_0$, the distribution acts as a power law of the form, $f_P \propto P^{\alpha}$. We fit the differential occurrence rate of small planets with respect to period using this functional form for both Sub-Neptunes and Super-Earths. We use bin sizes of $\Delta \log P = 0.005$ dex, and initialize the MCMC routine with 50 walkers, 5000 total steps, and 1000 burn-in steps. Sub-Saturns and Jupiters are not well described by this functional form. The fits are displayed in Figure 3.14.

For Super-Earths, we find a transition period of $P_0 = 6.3^{+1.7}_{-1.3}$ days, and for Sub-Neptunes we find a transition period of $P_0 = 11.2^{+3.9}_{-2.5}$ days. This is consistent with the theory of photoevaporation (Owen & Wu 2013, 2017), as planets at shorter orbital periods are subject to higher incident FUV and XUV flux, and are thus more subject to atmospheric stripping. As a result, one would expect the occurrence rate of Sub-Neptunes to drop before the occurrence rate of Super-Earths. Super-Earths and Sub-Neptunes have a consistently steep rise in occurrence at short orbital period, with $\gamma = 2.1^{+0.3}_{-0.2}$ for Sub-Neptunes and $\gamma = 1.9^{+0.2}_{-0.2}$ for Super-Earths. At longer orbital periods, Sub-Neptunes level off in occurrence rate with $\alpha = 0.08^{+0.14}_{-0.17}$ consistent with no change, and Super-Earths may have a slight decrease in occurrence rate at longer orbital periods with $\alpha = -0.11^{+0.13}_{-0.14}$, though these are also consistent with no change. These parameters are all consistent with those measured by Petigura et al. (2018).

In addition, the transition period we measure for Super-Earths, is in agreement with the transition period found in Wilson et al. (2018), $P_0 = 8.3^{+0.1}_{-4.3}$ days, who analyzed planets of all size classes. In Wilson et al. (2018) the transition period was measured by

finding the period in which the metallicity distributions of host stars with their innermost detected planet above and below the transition period are the most statistically different.

### Occurrence Rates with $P$, $R_p$, and [X/H]

To test the significance with which each element correlates with planet occurrence, we fit a parametric function of the form

$$f_{X,P} = CP^{\alpha}10^{\beta X} \quad , \tag{3.5}$$

where $X = [X/H]$, using the bootstrapping monte-carlo method described in §3.C.4. This is an extension of the model used by Petigura et al. (2018), who modeled the correlation between planet occurrence rates and metallicity. The abundance term in the above equation is equivalent to a power law relationship with the number density of atoms in the star's photosphere,

$$f_{X,P} \propto \left(\frac{n_X}{n_H}\right)^{\beta} \quad , \tag{3.6}$$

where $n_X$ is the number density of atoms of element $X$, and $n_H$ is the number density of hydrogen atoms in a star's photosphere. With this relationship in mind, a value of $\beta > 0$ would indicate a correlation between the number of planets and the presence of that particular element, while a value of $\beta < 0$ would indicate an anti-correlation between the planet occurrence rate and the number density of atoms of that particular element.

If we naively assume that that the abundance ratios in the stellar photosphere are the same as the abundance ratios of the protoplanetary disk in the first $\sim 1-10$ Myrs during planet formation before the disk disperses, then a non-zero differential occurrence rate density between two independent elements may indicate that the presence of one element more efficiently facilitates or suppresses planet formation compared to the presence of the other element. Such a result may indicate the composition of dust grains that grow

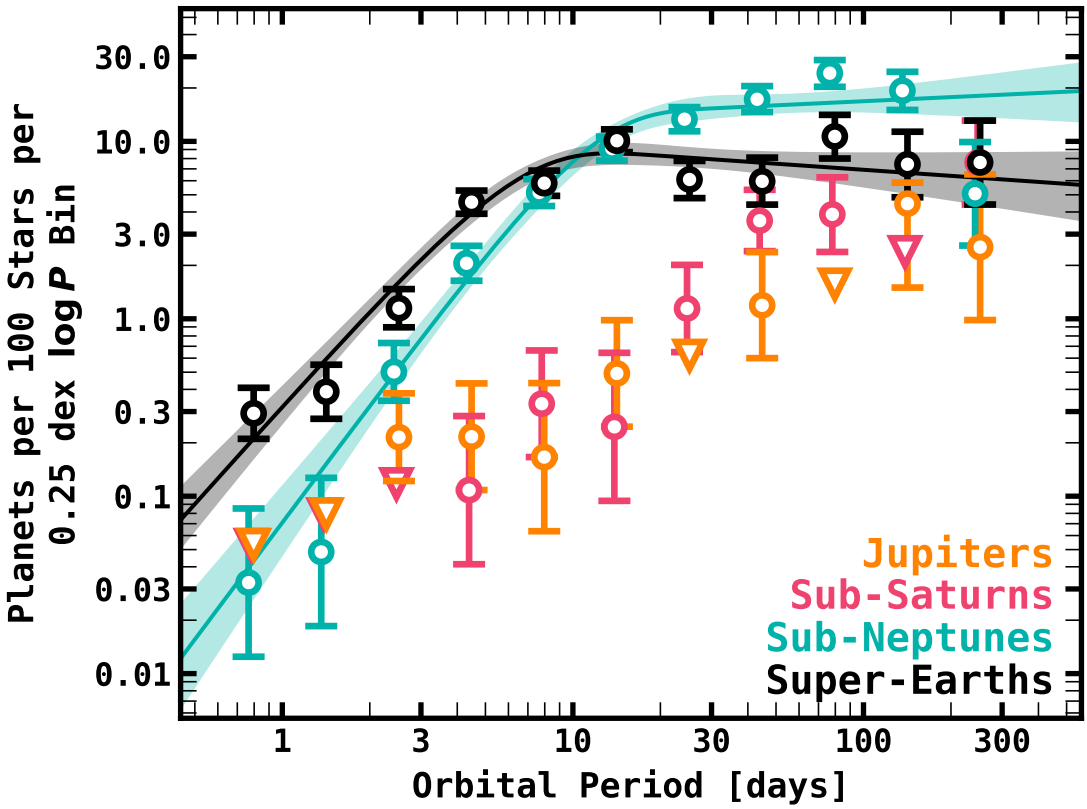


Fig. 3.14.— The number of planets per star (multiplied by 100) for a given orbital period bin and planet size class. The colors denote the planet size class. The circular points denote the occurrence rates while the triangular points denote upper limits. We fit the Sub-Neptune and Super-Earth occurrence rates with a function of the form, $f_P \propto P^{\alpha}(1 - e^{-(P/P0)^{\gamma}})$. The lines show the adopted best fit solution, and the shaded regions denote the $1\sigma$ confidence interval of credible models. The occurrence rates shown in this figure are displayed over substantially larger bin sizes than those used to fit the model but are displayed here to guide the eye.

to planetesimals more efficiently, or gaseous molecules that are preferentially accreted.

As discussed in the appendix (§3.C.4), the conclusions stemming from this analysis are limited by uncertainties in the stellar abundance distribution function, $\mathcal{F}_\star(X)$, rather than the Poisson error. In other words, the low number of stars in $\mathcal{C}$ are the dominant source of uncertainty in deriving $\beta$. For each planet size and period class we observe, we find consistent values for the period dependence, $\alpha$, across each elements in a given planet size and period class. We also find no correlation between $\alpha$ and $\beta$, for any element $X$ and any planet period period or size class in the posterior distributions. The results of these parametric fits are listed in Table 3.8, and plotted in Figures 3.15 and 3.16.

For the hot period class of planets, we find a positive correlation with all abundances and planet size classes, except Sub-Saturns that we are not able to constrain due to the low number of detections. The fits and range of credible models for all the hot planets are plotted in Figures 3.15 and 3.16. Because the model is two-dimensional, we integrate over the period dependence and only display the dependence on the elemental abundances.

The hot Jupiters in our sample are poorly constrained, but still consistent with $\beta > 0$, with $\beta$ ranging from $\beta = +11.3^{+6.4}_{-4.8}$ for Si, to $\beta = +4.2^{+2.7}_{-2.1}$ for Ni. All of these values are consistent at the $1\sigma$ level, but not well constrained. For hot Super-Earths, the element number density correlation ranges from $\beta = +0.50^{+0.48}_{-0.46}$ at the lowest for K, and $\beta = +1.16^{+0.47}_{-0.43}$ for C at the highest. These values are consistent at the $1\sigma$ level, and there are no clear differences between each different element.

For hot Sub-Neptunes, the correlation coefficients, $\beta$, are all $> 0$ and mostly consistent across all elements. However, we do find hints at variation among different chemical species. The correlation strengths range from $\beta = 2.98^{+0.85}_{-0.78}$ for Ca to $\beta = 1.26^{+0.51}_{-0.47}$ for Al (a $\sim 2\sigma$ discrepancy). However, we are hesitant to trust these differences due to potential non-LTE effects that may bias the Al abundance ratios, which are calculated in ASPCAP assuming LTE. The dependence on other elements range between these extremes. Future studies are needed to determine if credible variations exist. Such a difference may give

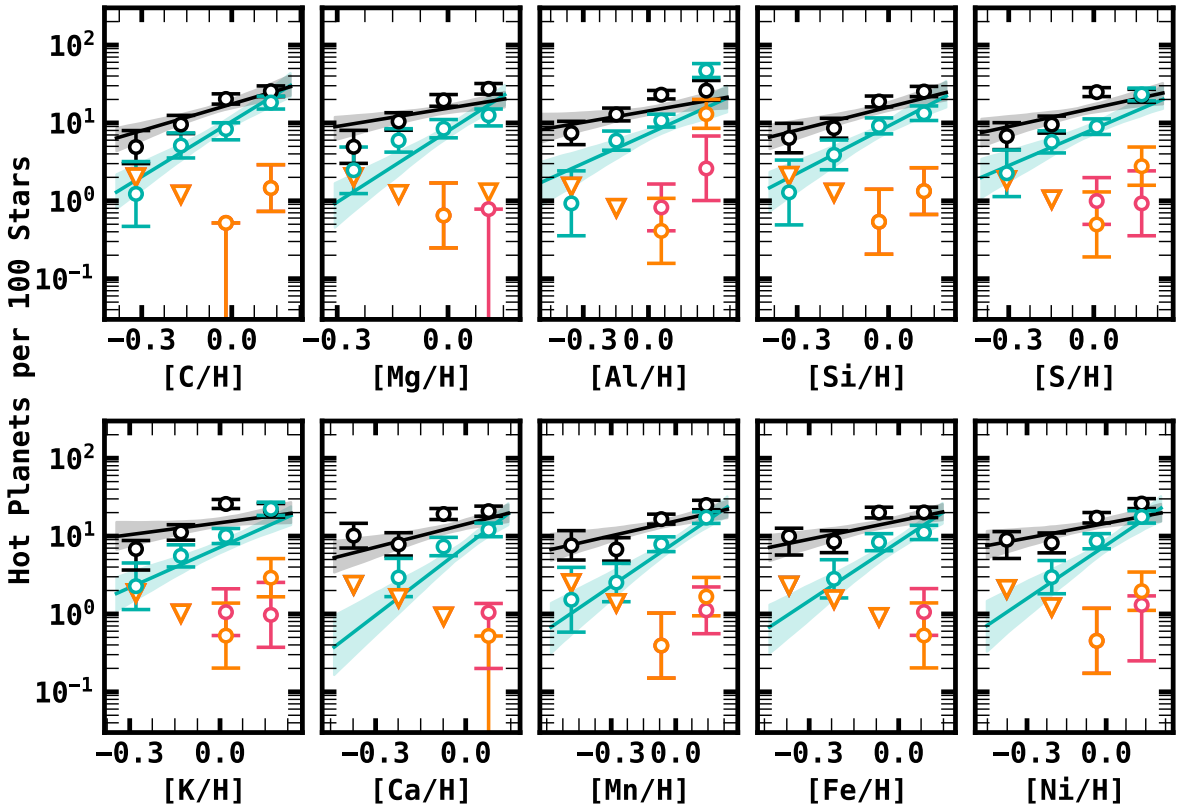


Fig. 3.15.— The occurrence rate of hot planets ($P = 1-10$ days) as a function of chemical abundances for ten different chemical elements. The colors represent planets of different size classes (Jupiters: Orange, Sub-Saturns: Pink, Sub-Neptunes: Teal, Super-Earths: Black). The points show the number of planets per 100 stars per bins equally spaced across the inner 90% of the abundance distribution for each element. The triangles show upper limits (90th percentile) on the planet occurrence rate, and the lines and shaded regions show our best fit and $1\sigma$ uncertainty to a power law distribution of the form, $f_{X,P} \propto P^{\alpha}\, 10^{\beta X}$, where we've integrated over the period dependence. Models are not shown for Sub-Saturns and Jupiters when the fit is poorly constrained. We emphasize once again that the occurrence rates and upper limits displayed in this figure are for larger bin sizes than those used to fit the power law distribution, and are included to guide the eye.

rise to important mechanisms in the formation or evolution of hot Sub-Neptune systems.

For warm planets, the correlation strength is reduced compared to the corresponding strength for hot planets for all planet size classes except possibly Sub-Saturns considering we were unable to constrain the correlation strength for hot Sub-Saturns. For warm Super-Earths, we find a tentative anti-correlation of $\beta \sim -0.6$ for most elements, though

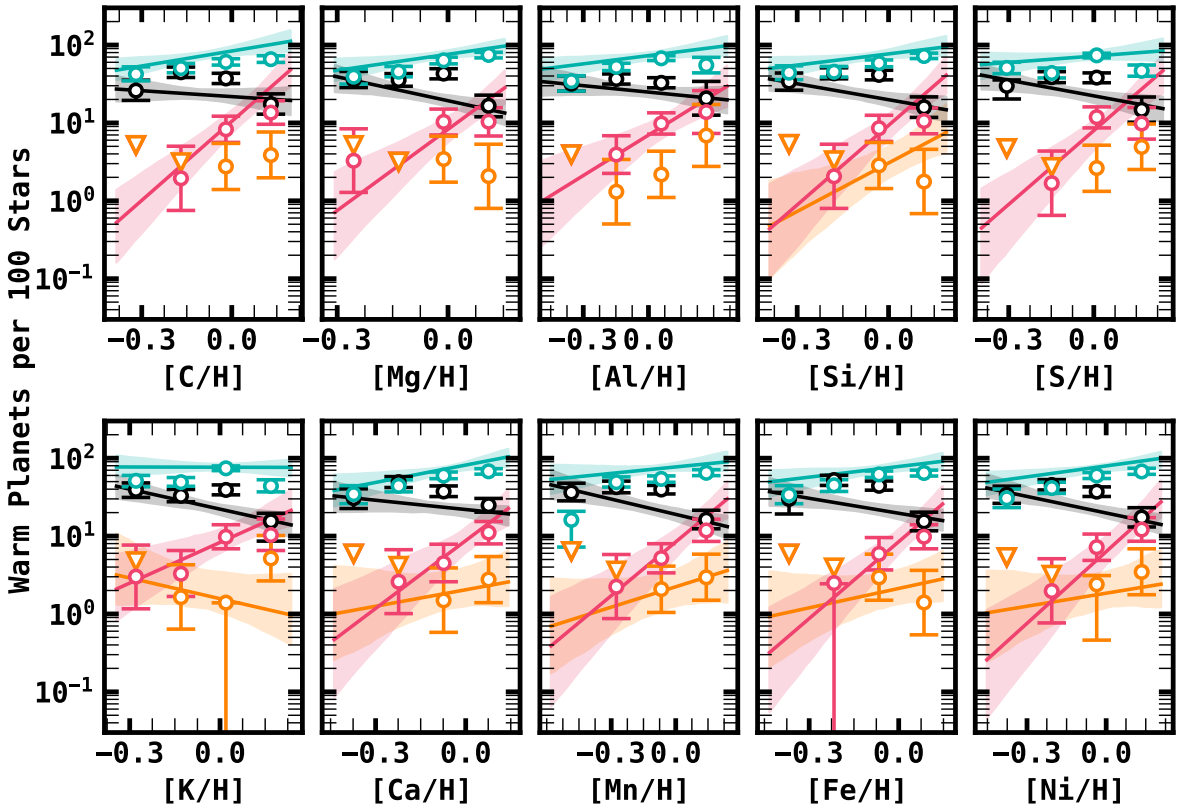


Fig. 3.16.— The same as Figure 3.15, but for warm planets ($P = 10 - 100$ days). We see overall weaker trends for each element and planet size class, with the possible exception of Sub-Saturns given the low number of detections at short periods.

they are also consistent with no correlation. Therefore, we do not make any claims about the dependence of the warm Super-Earth occurrence rate and the abundance of any chemical species. The abundance of Sub-Neptunes gives the opposite result, and we find that the there is a slight correlation, with $\beta \sim +0.4$ across all elements, but with similarly-sized errors such that we are unable to make a claim that the occurrence of warm Sub-Neptunes is positively correlated with the abundance of any particular chemical species. Warm Jupiters also have this same result, with $\beta$ ranging from $-0.9$ to $+2.1$ and errors ranging from 1 to 1.8 dex. Although our uncertainties are larger for each of these different chemical trends, these values are all consistent with the [Fe/H] dependence found by Petigura et al. (2018).

The Sub-Saturns are the only planet size class that have a measurable correlation between planet occurrence and chemical abundances with $P = 10 - 100$ days. For Sub-Saturns we measure a range of correlations from $\beta = 1.7^{+1.1}_{-0.9}$ for K, to $\beta = 3.5^{+1.6}_{-1.4}$ for Mg. These values are all still consistent across the ten elements within our uncertainties. Our measured correlation for Fe ($\beta = 3.3^{+1.5}_{-1.3}$) is larger than, but still is consistent with Petigura et al. (2018) who reported $\beta = 2.1^{+0.7}_{-0.7}$ for warm Sub-Saturns.

One trend we've noticed is that the magnitude of the strength of the correlation for Mn ($|\beta|$), is lower than for Fe in most period and planet size classes, though not significantly enough to claim a distinction. The lower value for Mn may be particularly surprising considering that [Mn/H] has the strongest correlation with [Fe/H] of all the abundances, one might expect that this effect be enhanced. However, considering it is always within our uncertainties, we infer that it is just an interesting coincidence at this time, and no further reflection is required.

## 3.5 Discussion

### 3.5.1 Variations in Correlation Strength Between Different Chemical Species

In this work we've made the first measurement of the dependence of the planet occurrence as a function of detailed chemical abundances in the *Kepler* field. The measured $\beta$ values are shown in Figures 3.17 and 3.18 for each element and planet size class. We are unable to confidently detect any differences in $\beta$ for different chemical species within a given planet size and period class. This lack of difference may be due to one of, or a combination of three effects. First, the lack of difference may be intrinsic (i.e., the enhancement/depletion of all elements are equally correlated with planet occurrence); second, our null result may be due to our uncertainties, which are limited by uncertainties

in $\mathcal{F}_\star$ for Super-Earths and Sub-Neptunes, and by the lack of detections in $\mathcal{P}$ for Sub-Saturns and Jupiters, or third, we are unable to detect differences in this dataset due to degeneracies caused by the lack of unique stellar populations probed in the *Kepler* field. I.e., the stars in the *Kepler* field have abundance ratios that are highly correlated for each element.

Interpreting such a variation, were it to exist, would put credible constraints on theories of planet formation. For instance, the increase in [Fe/H] for large planet hosts is often cited as evidence for core accretion. Such an enhancement has been determined for the [Fe/H] of such planet hosts, but has not been readily identified for other abundances, though there are previous results from RV surveys (Brugamyer et al. 2011; Adibekyan et al. 2012a). Following a similar line of logic, the relative increase in $\beta$ of differing chemical species for large planets, may indicate what elements most efficiently form planetary cores. They may also give insights into the location that certain planets form. For example, if the occurrence rate of hot Jupiters are more positively correlated with a volatile element such as C, an element likely to be in gaseous form at close orbital separations (Lodders 2003), one may infer that the core of such planets formed at greater orbital distance where those elements were contained in solid form (i.e., exterior to the respective molecule's ice line) before migrating interior to the respective molecules ice line (Öberg et al. 2011; Marboeuf et al. 2014), though these inferences can be complicated by effects such as cosmic ray ionisation (Eistrup et al. 2018).

Though the difference is not statistically significant, $\beta$ has a large range for Sub-Neptunes, ranging from ≈1.3-1.6 for Al and S to ≈3.0 for Ca and Mg, a $\approx 1.5\sigma$ difference for S, and $\approx 1.9\sigma$ difference for Al, if taken at face value. A variation in the strength of the correlation with hot Sub-Neptunes, were it to exist, may arise from differences in the planet's atmospheric opacity. In the core-powered mass loss framework, Sub-Neptunes with higher metallicity host stars are assumed to have higher metallicity atmospheres, which leads to longer atmospheric-loss timescales (Gupta & Schlichting 2020). By this

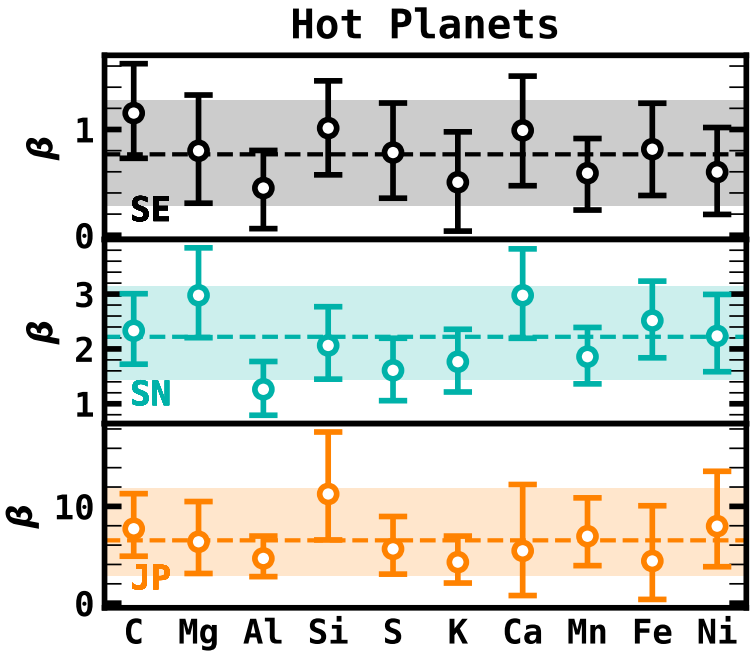


Fig. 3.17.— The derived $\beta$ and uncertainties for each chemical species, and each planet size class with periods ranging between 1-10 days. The colors show the planet size class for which $\beta$ was derived (Black: Super-Earths, Teal: Sub-Neptunes, Orange: Jupiters). The dashed line shows the mean across all elements, and the shaded region shows the inner 68% of the posteriors to the fits performed across all elements in §3.4.2. In order of increasing planet size class, the averages are $\beta_{\rm avg} = 0.8$, 2.2, and 6.5, though the Jupiters in our sample have a range of $\beta \sim$ 3-12.

logic, if we naively assume that the atmospheres of Sub-Neptunes have the same (or similar) abundance ratios for heavy elements as their host stars, one may infer that the occurrence of Sub-Neptunes should be higher around stars with an enhancement in elements that increase atmospheric loss timescales. While such a correlation between the metal content of a planet's atmosphere and the abundance of such elements in the hosts star's photosphere is not confirmed for multiple elements or detailed abundance ratios, the bulk metallicity of the planet's host star does correlate with the bulk metallicity of a planet's atmosphere for giant planets (i.e., Sub-Saturns and Jupiters), particularly at lower masses (Thorngren et al. 2016).

Therefore, it is not unreasonable to think that the enhancement of heavy elements

in a star may correlate with a similar enhancement in the atmospheres of planets such as sub-Neptunes that orbit such stars. If this is indeed the case, then a difference in $\beta$ between, e.g., S and Ca may arise from differences in the contributed opacity from S and Ca atoms in the planet's atmosphere. In this scenario, the lower $\beta$ value for S abundances doesn't increase the atmospheric loss timescale as much as Ca, resulting in a weaker correlation between the occurrence rate of Sub-Neptunes and the number density of S atoms. Thus, these longer timescales would result in higher occurrence rates for Sub-Neptunes, everything else being equal. In this way, measuring the dependence of planet occurrence rates on host star chemical abundances may provide an interesting route in probing the physics of exoplanet atmospheres.

Another potential cause for such a difference in $\beta$ between different elements can be due to the density of the planetary core. If it is assumed that the mineralogical makeup of planetesimals dictates the planet's interior structure, and planetesimals' mineralogical makeup may be inferred from stellar abundances (Dorn et al. 2017a,b; Hinkel & Unterborn 2018), then one expectation would be that the abundance of elements that result in a more dense core would be more likely to prevent atmospheric stripping. Such a trend may be observable as a stronger $\beta$ in Sub-Neptunes for elemental ratios that result in more dense cores.

### 3.5.2 Disentangling the Effects of Stellar Age, Mass, and Galactic Chemical Evolution

For each of ten chemical species in this study, we calculated the occurrence rate density, $f_{X,P}$ and compared it to the same quantity for nine other elements. However, while an important first step, a simple one-to-one comparison may not necessarily be appropriate to positively identify the influence of one particular element over another, as chemical abundances for all of these elements are correlated, due to Galactic chemical evolution. To test the dependence of each different chemical species on the planet occurrence rate

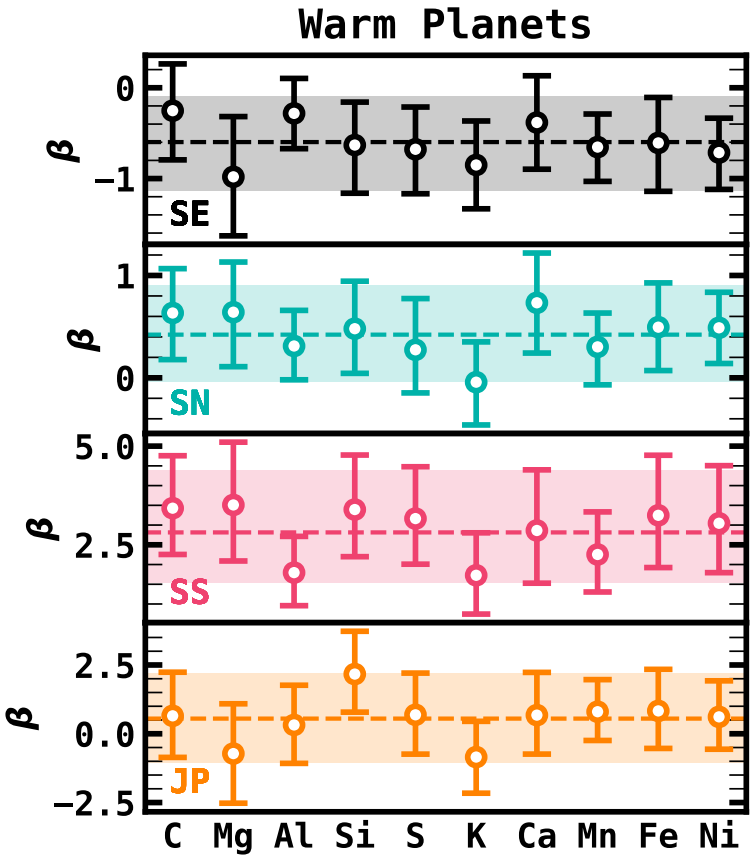


Fig. 3.18.— The derived $\beta$ and uncertainties for each chemical species, and each planet size class with $10 < P < 100$ days. The legend is the same as in Figure 3.17, with the planet size classes represented by differing colors (Black: Super-Earths, Teal: Sub-Neptunes, Pink: Sub-Saturns, Orange: Jupiters). In order of increasing planet size class, the averages are $\beta_{\rm avg}$ = -0.6, 0.4, 2.8, and 0.6.

separately from the known effects of enhanced bulk metallicity, it is insufficient to simply measure the quantity, $f_X$. It is equally insufficient to measure $f_{X/Z} = df/d[\mathrm{X/Fe}]$, as the chemical abundance trends with [X/Fe] and [Fe/H] are often not linear, and vary element by element based on a complicated function of star formation history, radial migration, and nucleosynthetic yields (e.g., Wyse 1995; McWilliam 1997; Sellwood & Binney 2002; Hayden et al. 2014; Nidever et al. 2015).

Determining the importance of unique elements in facilitating planet occurrence rates would provide an important constraint, but in practice can be very difficult to disentangle

from a number of other confounding variables. For instance, while we tend to find positive trends with elemental abundances and planet occurrence, it's not possible to say definitely whether those trends are due to one specific element, a combination of several, or simply due to the culmination of all elements heavier than H and He. In one sense it is unsurprising that we find very little variation between $\beta$ for different elements, because the abundance ratios for each of these elements are strongly correlated with one another. Thus, disentangling such effects will rely on either more precise observations, a much larger sample where subtle differences can be detected, or targeted planet-search surveys across multiple different stellar populations with unique chemical abundance patterns, such as in the thick disk or the halo.

Another important issue is the relative trends with chemical abundances and stellar age and mass. Because lower metallicity stars in the thin disk were formed before the enrichment of the interstellar medium, such stars may skew toward older ages and lower masses. Disentangling these effects is particularly challenging, given that credible trends with planet occurrence and stellar mass have been unequivocally uncovered in the literature (e.g., Mulders et al. 2015c; Dressing & Charbonneau 2015; Fulton & Petigura 2018; Ghezzi et al. 2018), and estimates of stellar age are becoming more precise due to surveys such as *Gaia*. For these reasons, when interpreting trends between age and planet properties, it is imperative that host star chemistry is taken into account. In short, stellar mass, age, and composition are all strong confounding variables with one another.

### Demonstration of an Age-Metallicity Degeneracy in Planet Demographics

There have been a number of claims relating to the demographics of planets and stellar age. For instance, Berger et al. (2020a) found that the relative fraction of Super-Earths to Sub-Neptunes is lower for young (<1 Gyr) stars than for the old (>1 Gyr) stars. Berger

et al. (2020a) inferred from this that there is ∼Gyr evolution in the atmospheric-loss timescale for stars near the radius gap, as predicted by core-powered mass loss (Gupta & Schlichting 2019, 2020).

While Berger et al. (2020a) cite age and long-term planetary evolution as a cause for a decrease in the frequency of Sub-Neptune planets, in this study we find that a dramatic decrease in the frequency of Sub-Neptunes can be attributed to even a small depletion of heavy elements. To test whether this relative decrease in the number of Sub-Neptunes can be explained by a difference in metallicity and subsequent change in occurrence rate between the "Old" and "Young" samples, we cross-matched our sample of all the KOIs observed in APOGEE with the "Young" (<1 Gyr) and "Old" (>1 Gyr) sample of planets from Berger et al. (2020a). In total, there are 25 and 23 planets with hosts in the "Old" and "Young" samples, respectively, with [Fe/H] measured by APOGEE. We then calculate the metallicity distribution function for each sample using a Gaussian kernel density estimate with a bandwidth chosen by Scott's rule (Scott 2010). The distribution functions are shown in Figure 3.19. The "Young" subsample is slightly skewed toward higher metallicities compared to the "Old" subsample.

Using the measured metallicity distribution functions to compute the expected occurrence rates, we find that the expected occurrence in the young sample is $\approx 1.4\times$ higher for Sub-Neptunes and $\approx 1.1\times$ higher for Super-Earths with $P$ = 1-10 days. For comparison, Berger et al. (2020a) found that $N_{SupEarth}/N_{SubNep}$ was $0.61 \pm 0.09$ and $1.00 \pm 0.10$ for the "Young" and "Old" samples, respectively. Thus, Berger et al. (2020a) found that $N_{SupEarth}/N_{SubNep}$ was decreased by a factor of $0.61 \pm 0.12$ from the "Young" to "Old" samples. Under the naive assumptions that $f_{SupEarth}/f_{SubNep} \propto N_{SupEarth}/N_{SubNep}$ for short periods, and that the Berger et al. (2020a) KOI samples are skewed toward $P < 10$ days, we would expect $N_{SupEarth}/N_{SubNep}$ to decrease by a factor of $0.80^{+0.11}_{-0.12}$. These values have only a $\approx 1.7\sigma$ discrepancy, though this is before correcting for detection biases which would lower the number of Super-earths. Thus, provided that the

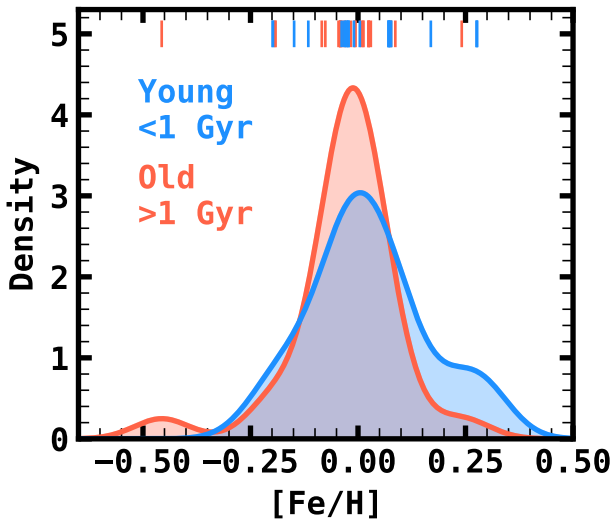


Fig. 3.19.— The metallicity distribution functions of the planet hosts in the"Young" and "Old" samples from Berger et al. (2020a) that were observed by APOGEE. The metallicities for each planet host are displayed as vertical ticks near the top of the figure.

inferred metallicity distribution functions of the "Old" and "Young" planets observed by APOGEE is representative of the metallicity distribution functions in the sample used by Berger et al. (2020a), then the full change in the relative number of Super-Earths to Sub-Neptunes may partially be explained by slight metallicity differences between the "Old" and "Young" samples.

It is worth noting that Berger et al. (2020a) were extremely careful to use known spectroscopic metallicities to control for differences when constructing their "Young" and "Old" samples. However, these metallicities came from a heterogeneous catalog, and different spectroscopic pipelines often have systematic differences as large as $\sim$0.1 dex. Thus, with precisely and homogeneously measured metallicities from APOGEE, we are able to detect a slight difference that may bias results such as these.

We intend the above exercise to be a demonstration rather than a repudiation of the conclusions inferred in Berger et al. (2020a). In reality, the logic behind inferring a metallicity distribution of a field sample by measuring the host star metallicity distribution of a planet sample is in conflict with the premise of this study, and should not be trusted.

Instead, this exercise is intended to demonstrate how even small metallicity differences may bias an inferred planetary distribution function, and therefore motivate the need for high-resolution, high-$S/N$ spectroscopic surveys to provide uniform metallicities and chemical abundances for a significant fraction of stars, so that biases from such effects can be adequately controlled.

### Age-Metallicity Degeneracies in the Population of Hot Jupiters

There are also claims of demographic trends for the population of hot Jupiters with age. For instance, by comparing the Galactic dynamics of field stars against the dynamics of hot Jupiter host stars, Hamer & Schlaufman (2019) claimed that hot Jupiters are destroyed by tides while their hosts are on the main sequence. This study found that hot Jupiter host stars are kinematically cold compared to the field, implying they are younger on average. However, Hamer & Schlaufman (2019) didn't have the necessary data to take into account the strong correlation with metallicity and the occurrence of hot Jupiters, and the correlation between metallicity and Galactic kinematics. It has been shown by a number of investigations that populations of metal-enhanced stars have lower Galactic velocity dispersion (e.g., Anguiano et al. 2018). On the other hand, one can make a similar claim about the age of hot Jupiter host stars and their correlation with metallicity. Perhaps the Planet Metallicity Correlation can be partially explained by a correlation with stellar age.

One potentially keystone open cluster for understanding the interplay between age and chemistry on the demographics of large planets is NGC 6791. NGC 6791 is a uniquely old cluster at $7.0 \pm 2.5$ Gyr (Netopil et al. 2016), with an enhanced metallicity, [Fe/H] = +0.35 (Cunha et al. 2015; Donor et al. 2020), and a high relative alpha abundance of [$\alpha$/Fe] = +0.1 (Linden et al. 2017). An occurrence rate study within this cluster could show the relative importance of tides and enhanced heavy metals. If tides (i.e., correlations with age) are the dominant force in shaping the hot Jupiter population, then

one should find an occurrence rate more similar to the field. If the dominant correlation is with enhanced abundances, the hot Jupiter population should be dramatically larger than in the field. While NGC 6791 was observed in the *Kepler* field, the main sequence turn off is very dim with a *Gaia* magnitude of $G \sim 17$. Thus, measuring a reliable planet occurrence rate may prove difficult in practice.

### Predictions for Planet Occurrence Rates in Nearby Open Clusters

An ideally perfect experiment to disentangle correlations between age and abundance ratios would be to measure the planet occurrence rate in several different open clusters with the same age but significantly varying abundance patterns, or vice versa. Such an experiment is likely not possible due to the lack of existence of such clusters with a wide enough abundance spread, and with enough stars. However, nearby open clusters still provide opportunities for measuring changes in planet demographics with age.

To date, there have been a few such studies targeting young transiting planets with *K2* (Howell et al. 2014) which observed open clusters such as the Hyades (Mann et al. 2016, 2018; Vanderburg et al. 2018), the Pleiades (Gaidos et al. 2017), and Praesepe (Rizzuto et al. 2018), among others. However, while there have been uniform searches for transiting planets in such clusters (Rizzuto et al. 2017), occurrence rates and completeness corrections have proven difficult due to effects such as crowding and the presence of strong correlated noise in the light curves of young stars. In an attempt to facilitate planet demographics studies for such clusters, we make predictions for planet occurrence rates for a number of nearby open clusters given their metallicity. Deviations in the actual planet occurrence rate from our predictions may then be attributed to age effects. Though some of the clusters mentioned below are scarcely populated, measurements of the planet occurrence rate may be approachable if measured in the aggregate (i.e., the expected planet occurrence rate from stars across multiple clusters). An alternate approach may be to measure occurrence rates with RV surveys, which do not have as

harsh of a geometric bias as transit surveys but come with other complications such as enhanced stellar activity in young stars.

We make predictions for the occurrence rate of hot planets ($f_{\rm hot}$) for the Pleiades, Praesepe, the Hyades, Ruprecht 147, M67, NGC 188, and NGC 6791. We make predictions for hot planets because they are more likely to be discovered with TESS (Ricker et al. 2015) and *K2*, and because hot planets have the strongest correlations with enhanced heavy elements. These predictions are listed in Table 3.5. For each of these clusters, we adopt the ages and distances of the clusters derived from *Gaia* DR2 photometry (Gaia Collaboration et al. 2018a), and the metallicities from APOGEE DR16 (Donor et al. 2020) where available, and from a homogenized catalog (Netopil et al. 2016) when not observed by APOGEE. We assume that the stars in the cluster all have equivalent metallicity and derive the expected occurrence rate using the posterior distributions of the fits to Equation 3.5 with [Fe/H] performed in §3.4.2.

There are two important caveats to these predictions. First, these predictions only hold if comparing a collection of planet-search stars with the same mass distribution as $\mathcal{S}$. Second, these predictions only hold true if a power law is an accurate parameterization of the true shape function over the metallicities of interest. In fact, there is some evidence that small planet occurrence may plateau at host star metallicities greater than ∼0.2 dex (Zhu 2019). In addition to these limitations, predictions for clusters with [Fe/H] below ∼-0.5 dex and greater than ∼0.2 dex rely on extrapolation and may be suspect as a result.

These predicted occurrence rates are meant to serve as a benchmark with which to compare age trends in the planet population. While it is tempting to explain differences between the population of planets in a cluster against the field population by invoking the age of the cluster or even the cluster environment, we show here that the metallicity of a cluster is a strong confounding variable, and alone could be responsible for a 100% increase in the planet occurrence rate for as small a difference as ∼0.1-0.2 dex. In this

way, accounting for metallicity effects in planet occurrence rates is a crucial step in understanding the difference between the populations of young planets in clusters, and older planets in the field (or in older clusters).

### Diffusion as Yet Another Confounding Variable in Age-Metallicity Correlations

One more important consideration when interpreting the correlations between planet occurrence and stellar abundances is the role of atomic diffusion. Atomic diffusion acts to deplete the surface abundances of certain elements by as much as $\sim$0.15 dex, depending on stellar mass and age (Souto et al. 2018, 2019). These processes complicate the interpretation of results such as those from this work, and general metallicity-planet occurrence rate trends because surface metal abundances are lower than the abundances of the nebula from which the star and planetary system formed. The relative depletion of surface abundance is a complicated function of stellar age, mass, and chemical species, complicating matters further. Ideally, one would correct for diffusion to estimate the initial abundances of the nebula in interpreting the role of specific elements in shaping the planetary distribution function, or estimating, e.g., initial abundances of planetary atmospheres assuming they are similar to the nebular composition. However, such a correction relies on accurate estimates of stellar masses and ages, and an accurate and precise model of the effects of diffusion.

## 3.6 Conclusion

In this Chapter we investigate the trends in the distribution of *Kepler* planets with the chemical abundances of their host stars as measured for stars in the APOGEE-KOI program (Fleming et al. 2015). Leveraging precise atmospheric parameters as measured by high S/N, high resolution near-infrared spectra, we derive precise planetary radii

($\sigma_{R_p} \approx 3.2\%$) for 504 planets. Using this sample of planet hosts, along with a control sample representative of the planet-search sample, we measure the abundance distribution functions for the *Kepler* field stars and derive planet occurrence rates as a function of abundance ratios for C, Mg, Al, Si, S, K, Ca, Mn, Fe, and Ni. In general, we find that the enhancement of any of these ten elements correlates with increased occurrence rates, and the strength of the correlation between planet occurrence rate abundance ratio is consistent across these ten elements.

At $P < 10$ days, we find that an enhancement of 0.1 dex in any of the ten elements in this study results in a ≈20% increase in the occurrence of Super-Earths and a ≈60% increase in the occurrence of Sub-Neptunes. The strength of these correlations are weaker for planets with $P = 10 - 100$ days, and we can only confidently confirm a positive correlation with the occurrence rate of Sub-Saturns and the enhancement of elements heavier than H and He in this period regime. Finally, we conclude this work with a discussion on the implications of such differences in the dependence of planet occurrence rates on the abundances of particular chemical elements, as well as a caution to the interpretation of trends in planet demographics in the context of Galactic chemical evolution and the ages and masses of planet hosting stars.

## 3.A Appendix A: Description of `Isofit`

`isofit` makes use of the `DFInterpolator` from the `isochrones` package (Morton 2015) to interpolate between a grid of MESA Isochrones and Stellar Tracks (MIST) models (Dotter 2016; Choi et al. 2016). The parent grid is defined from the MIST grid of models with solar-scaled alpha abundances and rotation and interpolated in initial [Fe/H] ($[\mathrm{Fe/H}]_{\mathrm{init}}$), initial mass ($M_{\mathrm{init}}$), and Equivalent Evolutionary Phase ($EEP$)[8]. The points in the parent grid for each of these parameters are $0.1 \leq M_{\mathrm{init}}/M_{\odot} \leq 8$ in steps of 0.02, $-2 < [\mathrm{Fe/H}]_{\mathrm{init}} < 0.5$ in steps of 0.05 dex, and $202 \leq EEP \leq 1710$, in steps of 1.

[8]For a detailed description of the $EEP$ parameter, see Dotter (2016).

The range in $EEP$ roughly represents each step in a stellar evolutionary track from the Zero-Age Main Sequence to the beginning of the White Dwarf cooling track. In total, the parent grid contains ∼15 million valid models.

To infer model parameters for a given set of observations, in this case $\theta_i = \{T_{\rm eff}, \log g, [{\rm Fe/H}], \pi, Ks, E(B-V)\}$, `isofit` computes the likelihood for the model input parameters, $x_i = \{M_{\rm init}, [{\rm Fe/H}]_{\rm init}, EEP, d, E(B-V)\}$ and derives an integrated posterior distribution over all likelihoods and priors (e.g., Serenelli et al. 2013; Huber et al. 2017). More specifically, the posterior probability is given by

$$p(\boldsymbol{x}|\boldsymbol{\theta}) \propto p(\boldsymbol{x})p(\boldsymbol{\theta}|\boldsymbol{x}) \tag{3.7a}$$

$$\propto p(\boldsymbol{x}) \prod_i \exp\left[-\frac{(\theta_i - \theta_i(\boldsymbol{x}))^2}{2\sigma_{\theta,i}^2}\right] \tag{3.7b}$$

where $\sigma_{\theta,i}$ are the Gaussian errors on the measurement $\theta_i$, and $\theta_i(\boldsymbol{x})$ are the inferred model parameters for input vector $\boldsymbol{x}$. The likelihood for the inferred model parallaxes is given by

$$p(\pi|d) \propto \exp\left[-\frac{1}{2\sigma_\pi^2}\left(\pi - \frac{1}{d}\right)^2\right] \tag{3.8}$$

where $d$ is the model distance used to derive apparent magnitudes. For apparent magnitudes, in this case $Ks$, `isofit` calculates the inferred model apparent magnitude using the MIST grid of bolometric corrections, $BC_m$, the inferred model distance modulus, $\mu = 5\log d - 5$, and the inferred model bolometric magnitude, $M_{\rm bol}$,

$$m = M_{\rm bol} - BC_m + \mu + A_m \tag{3.9}$$

where $A_m$ is the extinction in band $m$, calculated from $E(B-V)$ and the extinction law from Wang & Chen (2019).

For model output parameters that do not have associated observations, we assume a

flat prior. The exception for this is distance, which has a decreasing density prior with a length scale, $l = 1350\,\mathrm{pc}$ (as in, e.g., Bailer-Jones 2015; Huber et al. 2017), given by

$$p(d) \propto \frac{d^2}{2l^3} e^{d/l} \,. \tag{3.10}$$

Finally, we take the natural log of each term, and sum them together to get the log-likelihood estimate for a given set of input parameters, $\boldsymbol{x}_i$.

To find the initial best fit model, `isofit` calculates the log-likelihood for all the models of an initial course grid, interpolated from the parent grid with steps of 0.05 $M_\odot$, 5 $EEP$, and 0.1 dex in [Fe/H], that agree with the spectroscopic parameters, $T_{\rm eff}$, $\log g$, and [Fe/H] within $\pm 5\sigma$. Then, `isofit` calculates a fine grid around the course grid point that returns the maximum log-likelihood, and repeats the process but with finer step sizes of 0.01 $M_\odot$, 0.5 $EEP$, and 0.01 dex in [Fe/H]. We then instantiate an MCMC routine (`emcee`;Foreman-Mackey et al. 2013) with a Gaussian ball centered around the model parameters that return the largest log-likelihood with the observed $\pi^{-1}$ and $E(B-V)$.

## 3.B Appendix B: Results of Statistical Tests Comparing $\mathcal{C}$ and $\mathcal{P}$ Abundances

We first test for normality in each distribution using the Shapiro-Wilkes test for normality. In this case, we find that the abundance distributions in $\mathcal{C}$ are only consistent with a normal distribution in the case of [Fe/H], [Si/Fe], and [C/Fe], i.e., for all other abundances the $p$-value was sufficiently low that we reject the null hypothesis that the data were pulled from a normal distribution. This lack of normality motivates us to adopt non-parametric tests: the Kolmogorov-Smirnov (KS) test, the Mann-Whitney U-test (MW), and the Brown-Forsythe test (BF). The KS test is used to measure any difference between two cumulative distribution functions. Because of this the KS test is applicable in a variety

of situations, but in general is not very sensitive. Therefore, we also apply the MW test, which tests for differences in the means of two samples, and the BF test which tests for differences in the variances of the two samples. For each subsample of planet type, we apply these three tests against the abundance distributions in $\mathcal{C}$. We do not conduct tests on sub-samples where $n_{pl} < 10$ to avoid erroneous conclusions caused by small number statistics.

We find a few statistically significant differences from this methodology. First, we find that the Fe abundances for each planet size class is enhanced compared to the field, except in the case of Super-Earths, which do not meet our significance criterion. The only other statistically significant difference between $\mathcal{P}$ and $\mathcal{C}$ distributions are with [K/Fe]. However, [K/Fe] is highly correlated with [Fe/H], so it's most likely that this result is only tracing differences in the Fe abundances already described.

Perhaps not surprisingly, we find that the planet subgroup with the lowest MW p-value, $p_{mw}$, are Jupiters. The presence of large planets have been shown numerous times to correlate with enhanced metallicity (Santos et al. 2004; Valenti & Fischer 2005; Ghezzi et al. 2010, 2018; Buchhave et al. 2014; Schlaufman 2015; Wang & Fischer 2015). While all Super-Earths in general are not significantly Fe-enhanced compared to the field ($p_{mw} = 0.071$), hot Super-Earth hosts are Fe-enhanced. In fact, we find that all hot planet hosts are iron-enhanced compared to $\mathcal{C}$, a result that agrees with the literature, and has been pointed out by a number of authors (Mulders et al. 2016; Wilson et al. 2018; Petigura et al. 2018; Narang et al. 2018). Hot planets also seem to be correlated with [K/Fe] and [S/Fe], in that they have a significantly different variance. This is likely due to hot planet hosts having [Fe/H] distributions skewed above solar, and as a result there are no hosts with high K and S abundances relative to iron. In other words, the decreased variance in [K/Fe] and [S/Fe] planet hosts is driven by lower variance in Fe abundances. Thus, these are also likely due to trends with Fe.

Finally, the last subgroup we find that is significantly different compared to the field

is the population of Warm Sub-Neptune hosts. These stars have a significantly different variance compared to $\mathcal{C}$. A difference in variance is also present in the populations of all Sub-Neptunes, hot planets, and hot Sub-Neptunes. This agrees with the interpretation from Petigura et al. (2018) that stars with greater metallicities are capable of a wider diversity of planets. In other words, the abundance distribution of the field sample should have larger variance compared to hosts of a specific planet classes due to the fact that some planet classes are only present around higher metallicity host stars.

## 3.C Appendix C: Occurrence Rate Methodology

### 3.C.1 Formalism and Definitions

Our methodology treats the detection of a transiting planet as an independent random process, i.e., as a Poisson process. We use NPPS as our definition of planet occurrence, $f$. As a note, this is not equivalent to the quantity of the Fraction of Stars With Planets (FSWP) that is often used as a definition of planet occurrence. For a transit survey, a measurement of FSWP requires detailed modeling of multiplicity, mutual inclinations, and other effects that are outside the scope of this work. We instead default to NPPS, which is blind to these properties. In the interest of comparing to other works (e.g., Petigura et al. 2018), we often report our occurrence rates in units of number of planets per 100 stars.

For a given star with properties $\boldsymbol{z}$, the probability of hosting a planet with properties $\boldsymbol{x}$ can be expressed as

$$df = \frac{\partial f(\boldsymbol{x}, \boldsymbol{z})}{\partial \boldsymbol{x}} d\boldsymbol{x} \quad , \tag{3.11}$$

where integrating over the planet properties, $\boldsymbol{x}$, gives $f(\boldsymbol{z})$, the average number of planets for a star with properties $\boldsymbol{z}$. In this work, $\boldsymbol{x}$ is some combination of $\log R_p$ and $\log P$,

and $\boldsymbol{z}$ is the abundance of some chemical species. We typically adopt $Z$ as our symbol for metallicity, and use $X$ to refer to an arbitrary chemical element. For compactness we adopt the following notation for a partial derivative of $f$ with respect to an arbitrary variable $x_1$ and $x_2$,

$$f_{x_1} \equiv \frac{\partial f}{\partial \log x_1} \; ; \; f_{x_1,x_2} \equiv \frac{\partial^2 f}{\partial \log x_1 \, \partial \log x_2} \quad . \tag{3.12}$$

This is similarly defined for chemical abundances as,

$$f_X \equiv \frac{\partial f}{\partial [X/H]} \; ; \; f_Z \equiv \frac{\partial f}{\partial [Fe/H]} \quad , \tag{3.13}$$

where chemical abundance ratios are always defined with respect to hydrogen. Note, this is a change from §3.3.5 where we were searching for new trends independent of [Fe/H]. For the remainder of this study, we wish to compare the strength of the correlation with the enhancement of each chemical element with planet occurrence. Thus, we adopt [X/H] to express each element on a similar scale. We express the differential distribution for NPPS as

$$f_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{z}) d\boldsymbol{x} \equiv C g(\boldsymbol{x}, \boldsymbol{z}; \theta) \tag{3.14}$$

where $g(\boldsymbol{x}, \boldsymbol{z}; \theta)$ is a shape function (i.e., some parametric prescription used to describe the PLDF) that depends on planet and/or stellar properties with shape parameters $\theta$. A functional form for $g(\boldsymbol{x}, \boldsymbol{z})$ must be assumed, with as many shape parameters, $\theta_i$, as necessary.

The total number of planets orbiting $n_\star$ stars (indexed by $i$) is then

$$n_{pl} = C \sum_i^{n_\star} \int g(\boldsymbol{x}, \boldsymbol{z_i}; \theta) d\boldsymbol{x} \tag{3.15a}$$

$$= n_\star C \int \mathcal{F}_\star(\boldsymbol{z}) g(\boldsymbol{x}, \boldsymbol{z}; \theta) d\boldsymbol{x} d\boldsymbol{z} \tag{3.15b}$$

where the integration takes place over some range of planet properties. In equation 3.15b, the sum over all stars is replaced by an integral over the probability distribution of stellar properties, $\mathcal{F}_\star(\boldsymbol{z})$. $\mathcal{F}_\star$ is normalized so that $\int \mathcal{F}_\star(\boldsymbol{z}) d\boldsymbol{z} = 1$. In practice, the summation over the known properties of each planet search star is preferable, but in principle an accurate measurement of $\mathcal{F}_\star(\boldsymbol{z})$ gives an equivalent result.

In this work, we calculate our occurrence rates for bins with some combination of [X/H], [Fe/H], $\log P$, and/or $\log R_p$. The width of a bin is given by $\Delta\boldsymbol{x}$, where $\Delta\boldsymbol{x} = \prod_i \Delta x_i$, where $i$ indexes over the dimensions of the bin. The occurrence within a bin, $f_{\rm bin}$, depends on the number of independent trials, $n_{\rm trial}$, that yield a detected planet, and the survey efficiency, $\eta$, which may depend on both stellar and planetary properties. We compute $n_{\rm trial}$ as,

$$n_{{\rm trial},\, j} = \sum_i^{n_\star} \eta(\boldsymbol{x}_j, \boldsymbol{z}_i) \tag{3.16}$$

$$= n_\star \left\langle \eta(\boldsymbol{x}_j) \right\rangle \tag{3.17}$$

where $\langle . \rangle$ denotes the arithmetic mean. For a given survey efficiency, $\eta(\boldsymbol{x}, \boldsymbol{z})$, $\eta$, also includes the number of false positives in a given sample. However, because we have removed RV variable sources from $\mathcal{P}$, the false positive rate from astrophysical sources is negligible. We also ignore false alarms from instrumental effects. While incorporating such false alarms is important for deriving robust occurrence rates in principle, the actual false alarm rate in *Kepler* is negligible for planets with $P \lesssim 300$ days (Mullally et al. 2016).

Following the examples of Bowler et al. (2015) and Petigura et al. (2018), we assume that the planet occurrence is log-uniform within a given bin of size $\Delta \boldsymbol{x}_{i,j}$, which should be reasonable at small enough bin sizes. In this case, $n_{\rm trial}$ for a bin can be expressed as,

$$n_{\rm trial} = \frac{n_\star}{\Delta \boldsymbol{x}_{i,j}} \int \langle \eta(x_{i,j}) \rangle \, d\boldsymbol{x} \quad . \tag{3.18}$$

Thus, for a given cell with $n_{\rm trial}$ trials and $n_{\rm pl}$ detected planets, the likelihood of $f_{\rm bin}$ can be described by a binomial distribution of the form,

$$P(f_{\rm bin}|n_{\rm pl}, n_{\rm trial}) = P(n_{\rm pl}|f_{\rm bin}, n_{\rm trial}) \tag{3.19}$$

$$= C f_{\rm bin}^{n_{\rm pl}} (1 - f_{\rm bin})^{n_{\rm nd}} \tag{3.20}$$

where $n_{\rm nd} = n_{\rm trial} - n_{\rm pl}$ is the number of non-detections, and $C$ is a normalization constant that takes the form

$$C = \frac{(n_{\rm trial} + 1)\,\Gamma(n_{\rm trial} + 1)}{\Gamma(n_{\rm pl} + 1)\,\Gamma(n_{\rm nd} + 1)} \tag{3.21}$$

When analyzing occurrence rates as a function of stellar properties, we bin the planet and stellar properties in bins bounded by $[\log P_1, \log P_2]$, $[\log R_{p,1}, \log R_{p,2}]$, and $[X_1, X_2]$. In this way, we calculate $n_{\rm trial}$ via equation 3.18, multiplied by the fraction of stars, $F_\star$, with abundances between $[X_1, X_2]$,

$$F_\star = \int_{X_1}^{X_2} \mathcal{F}_\star(X) dX \quad . \tag{3.22}$$

In the case where there are no detected planets in a given bin, we estimate an upper limit on the occurrence rate for that bin by numerically solving for the integral,

$$\int_0^{f_{\rm bin}} P(f|n_{pl}, 0) df = 90\% \quad . \tag{3.23}$$

We note that in practice, we only use these upper limits for display purposes, and don't directly incorporate them into our analysis.

### 3.C.2 Parametric Fits to the Differential Occurrence Rate Distributions

We wish to express the strength of the correlation between a star's chemical composition and the occurrence of various types of planets. We do this via a parametric relation, often with a power law in this work, to gauge the strength of the correlation. We note here that a power-law prescription for describing the differential occurrence rate does not necessarily reflect the true shape function of the occurrence rate. However, such a prescription can give a precise estimate of the average strength with which the differential occurrence rate relies on the underlying abundance. So, while a power-law fit of this form gives a precise estimate of correlation strength, this prescription may not be appropriate to robustly predict $f$ for a given stellar sample.

To this end, we can estimate the differential occurrence rate for a bin of size $\Delta\boldsymbol{x}$ via the following relation,

$$Cg(\boldsymbol{x}, \boldsymbol{z}; \theta) = \frac{f_{\rm bin}}{\Delta \boldsymbol{x}} \quad . \tag{3.24}$$

To find the best fit shape parameters, $\theta$, we maximize the log-likelihood of the function, described for a given bin, $i$, by

$$\ln L_i = n_{pl,i} \ln Cg\Delta\boldsymbol{x} + n_{nd,i} \ln(1 - Cg\Delta\boldsymbol{x}) \tag{3.25}$$

where each cell is an independent constraint on $Cg\Delta\boldsymbol{x}$. Therefore, by maximizing the

combined log-likelihood over all bins, indexed by $i$,

$$\ln L = \sum_i \ln L_i \quad , \tag{3.26}$$

we find the best fit shape parameters, $\theta$. To zero in on the best fit shape parameters, we apply arbitrarily small bin sizes. As a result a number of these bins have few or no detected planets. Because the log-likelihood function we apply incorporates non-detections, this methodology is stable even to few detected planets. This approach has two advantages. The first is that our assumption that a bin is log-uniform has more merit in smaller bins, and the second is that for a small enough bin the errors on the occurrence rate will be dominated by Poisson statistics rather than uncertainties in $\mathcal{F}_\star$. We expand upon this assumption in §3.C.4. After finding the best-fit parameters, $\theta$, that maximize the likelihood function, we then apply an MCMC routine[9] initialized at those parameters to explore the range of credible models.

### 3.C.3 Completeness Corrections

In this subsection we describe our completeness model, $\eta(\boldsymbol{x}, \boldsymbol{z})$. Our approach varies slightly from most previous *Kepler* occurrence rate studies, because we also correct for biases inherent in the follow-up program as well. Most *Kepler* occurrence rate studies (this one included) select some subsample of planet-search stars from the *Kepler* Input Catalog (Brown et al. 2011), and then compute a detection efficiency model from the *Kepler* pipeline that is marginalized over their planet-search sample. However, this strategy alone is not adequate because we have additional biases that are not quantified by the *Kepler* detection pipeline. In other words, inclusion in $\mathcal{P}$ is dependent on more than membership in $\mathcal{S}$ and a detected planet candidate in *Kepler*. There are additional biases imposed by limitations in the APOGEE selection function, instrumental setup,

[9] as implemented in the python package `emcee` (Foreman-Mackey et al. 2013)

and spectroscopic analysis pipeline that must be considered.

We consider four unique biases for a planet candidate to be included in $\mathcal{P}$: first, the planet has to transit its host star; second, the transiting planet must be included in the DR24 KOI catalog; third, the planet's host star must have been observed by APOGEE; and fourth, ASPCAP must have returned reliable abundance and spectroscopic parameters for the host star. We take each of these criteria as their own independent process, so that we can model the total completeness as the product of the probabilities that a planet candidate passes each step. Thus, our completeness model, $\eta$, can be described by four terms: the geometric probability that a planet with a randomly oriented orbital plane transits its host star ($p_{\rm tra}$), the probability that a transiting planet is detected by *Kepler*($p_{det}$), the probability that a candidate was observed in the APOGEE-KOI program ($p_{\rm apo}$), and the probability that ASPCAP doesn't fail to produce reliable atmospheric parameters for the host star ($1 - p_{\rm fail}$). We go into more detail for each of these terms below before presenting the combined, average survey efficiency.

### Transit Probability ($p_{\rm tra}$)

The probability that a given planet transits depends only on the geometry of the orbit. We make the assumption that the inclination of all orbits follows an isotropic distribution. Under these assumptions, the probability for a planet to transit in our sample is simply,

$$p_{\rm tra} = \frac{0.9 R_\star}{a(1-e^2)} \tag{3.27}$$

where we set $e = 0$ for simplicity, include a factor of 0.9 to account for our cut on impact parameter, and $a$ is calculated from $M_\star$ and $P$ using Kepler's third law.

### Kepler Pipeline Detection Efficiency ($p_{\rm det}$)

In this section we give our model for the *Kepler* DR24 pipeline completeness. For a planet to be detected, it must have a high enough $S/N$ to be detected, and it must pass multiple levels of vetting. In place of transit $S/N$, the *Kepler* pipeline utilizes the Multiple Event Statistic (MES). The MES is a measure of the null hypothesis that a *Kepler* light curve does not have a transit signal at a given epoch ($t_0$), duration ($t_{\rm dur}$), and period ($P$). Under the assumption of white noise, the MES distribution is Gaussian with a mean of zero and variance of one. Under the alternative hypothesis however, i.e., that there is a transit signal, the mean of the MES distribution is shifted by a constant proportional to the $S/N$ of the transit. *Kepler* defines a threshold of MES $> 7.1$ for a detected signal.

It is common to parameterize the *Kepler* pipeline completeness with a one-dimensional model in expected MES from a putative transiting planet ($\rm MES_{exp}$). However, this is not an appropriate model for the DR24 pipeline completeness, because an introduced $\chi^2$ metric used to veto false alarms severely reduced the completeness for planets with $P > 40$ days (Christiansen et al. 2016). Therefore, we parameterize *Kepler*'s pipeline efficiency in two dimensions, with $\rm MES_{exp}$ and $P$. Rather than utilizing an analytical model, we take a purely empirical approach to assess $p_{\rm rec}(P, \rm MES_{exp})$ (as in, e.g., Petigura et al. 2013; Dressing & Charbonneau 2015).

We apply the results of the Monte Carlo injection and recovery tests executed by Christiansen et al. (2016). To ensure that the light curves are representative of the light curves from $\mathcal{S}$, we remove the injection results from stars that are inconsistent with stars in $\mathcal{S}$. We applied the following cuts to the sample of stars with injected signals, $R_\star < 2R_\odot$, $4700\,{\rm K} < {\rm T_{eff}} < 6200\,{\rm K}$, and $RUWE < 1.2$. We also attempted to limit the collection of transit recoveries to stars in APOGEE fields, but we found no differences in $p_{\rm rec}$ between stars in and out of APOGEE fields so elected to use stars from the entire *Kepler* field to improve our statistics. In total, this resulted in 94,141 individual injection

and recovery tests.

To measure $p_{\rm rec}(P, {\rm MES}_{\rm exp})$, we defined a grid in $P$ and ${\rm MES}_{\rm exp}$ and measured the fraction of recovered injections within each bin. The injections were performed using a uniform prior in $P$. However, we are interested in assessing the completeness in logarithmic bins. To account for this, we define linear bins from $P = 0.25 - 10$ days, in steps of 0.25 days, and then 75 logarithmically spaced bins from 10-500 days. We binned the ${\rm MES}_{\rm exp}$ of each putative signal in steps of 0.5 from 0-20, where the recovery fraction for ${\rm MES}_{\rm exp} > 20$ is assumed to be constant. The resulting grid is shown in the top panel of Figure 3.20. Due to few injections at $P \lesssim 10$ days, and at large MES, we replace any binned points that have $< 3$ injections, ${\rm MES}_{\rm exp} > 13$, and $P < 40$ days with $p_{\rm rec} = 0.997$, the expected pipeline efficiency at arbitrarily large MES. We then convolve the $p_{\rm rec}$ grid with a Gaussian kernel having a width of the bin size in each dimension to smooth over any artificial features and interpolate over points with no injections, such as those at arbitrarily low ${\rm MES}_{\rm exp}$ or very short $P$. The resulting grid that we apply is shown in the bottom panel of Figure 3.20.

For each star in $\mathcal{S}$, we calculate $p_{\rm rec}$ over the $P$-$R_p$ grid. We calculate ${\rm MES}_{\rm exp}$ for every combination of $P, R_p$ for each star in $\mathcal{S}$ and interpolate on the grid defined above to measure $p_{\rm rec}(P, R_p)$. We calculate ${\rm MES}_{\rm exp}$ as

$$ {\rm MES}_{\rm exp} = \left(\frac{R_p}{R_\star}\right)^2 \frac{1}{\sigma_{cdpp}(t_{\rm dur})} \sqrt{\frac{T_{obs}}{P}} \tag{3.28} $$

where $T_{obs}$ is the observation baseline and $\sigma_{cdpp}(t_{\rm dur})$ is the combined differential photometric precision (Christiansen et al. 2012) on the timescale of the transit duration, $t_{\rm dur}$. We calculate $t_{\rm dur}$ using equation 3.1. The *Kepler* data products contain measurements of $\sigma_{cdpp}$ on timescales of 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, which are the transit durations searched by *Kepler*'s TPS module. We interpolate between the values provided at these timesales to estimate $\sigma_{CDPP}$ for any arbitrary $t_{\rm dur}$.

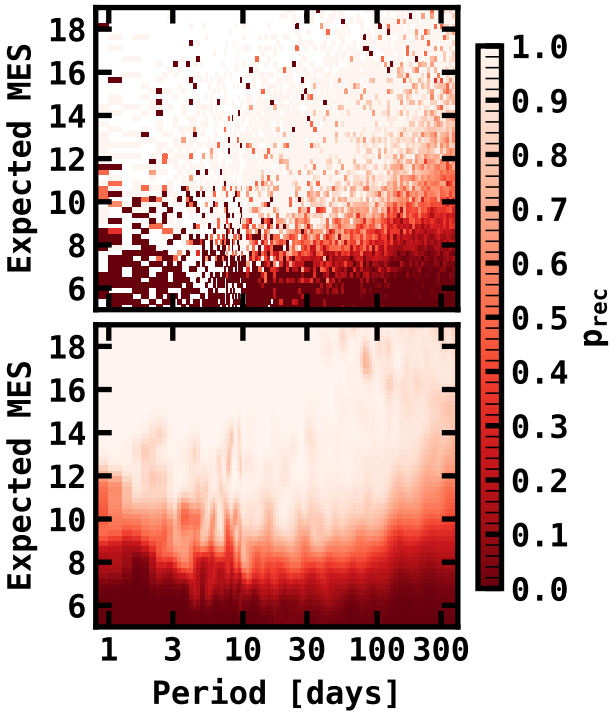


Fig. 3.20.— The DR24 *Kepler* pipeline detection efficiency. The shading shows the probability of the *Kepler* TPS module to recover a transit signal ($p_{\rm rec}$) with an expected Multiple Event Statistic ($\rm MES_{exp}$) and Period ($P$). Darker shades of red represent a lower recovery fraction, and lighter shades represent a higher recovery fraction. *Top*: The recovery probability ($p_{\rm rec}$) of the *Kepler* pipeline from the injection and recovery experiments in Christiansen et al. (2016). White spaces denote bins with no data. *Bottom*: The interpolated, and smoothed grid of $p_{\rm rec}$ that we apply for our completeness corrections.

The values for $T_{obs}$ and $\sigma_{cdpp}$ were taken from the DR24 stellar properties table hosted on the NExScI Exoplanet Archive[10].

One last requirement in the *Kepler* pipeline is that the planet candidate must have at least three transits in the data. We quantify this probability by the window function, $p_{\rm win}$, as estimated in Burke et al. (2015),

$$p_{\rm win} = 1 - (1 - f_d)^M - M f_d (1 - f_d)^{M-1} - \frac{M(M-1)}{2} f_d^2 (1 - f_d)^{M-2} \,, \tag{3.29}$$

[10]https://exoplanetarchive.ipac.caltech.edu (Akeson et al. 2017)

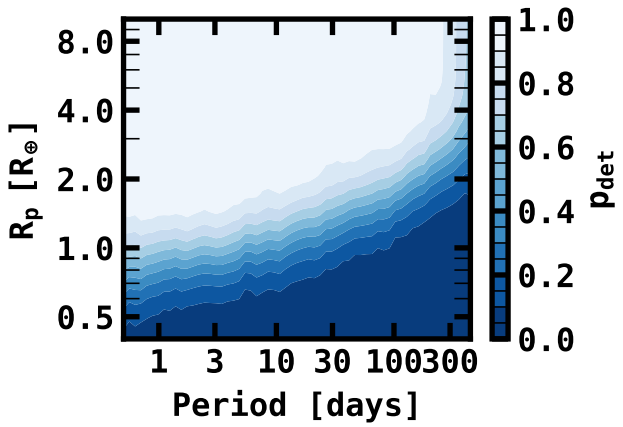


Fig. 3.21.— The average DR24 *Kepler* TPS pipeline detection efficiency for the stars in $\mathcal{S}$. Darker shades of blue represent a lower detection fraction, and lighter shades represent a higher detection fraction.

where $f_d$ is the duty cycle of the observations and $M = T_{obs}/P$. For planets with $P < 300$ days, this probability is typically $> 0.98$, so this is not a strong source of bias in our sample, but we include it for posterity.

The joint probability of *Kepler* detecting a transiting planet, then is given by the product $p_{\rm det} = p_{\rm rec} \times p_{\rm win}$. This full *Kepler* pipeline sensitivity is shown in Figure 3.21.

### Pre-DR24 Selection Bias and Probability of Fiber Collisions ($p_{\rm apo}$)

The probability that a KOI was included in the APOGEE program depends on the APOGEE field. For fields K04, K06, and K07, this bias is dominated by the rejection of APOGEE targets based on fiber collisions. A "collision" occurs when two fibers, if placed on the plate, would be separated by less than the size of the protective ferrule around each fiber. For the APOGEE-N spectrograph the collision radius corresponds to $71.5''$ (Zasowski et al. 2013, 2017). In the case of fiber collisions, one target is assigned priority, and the other is removed from the target list. We don't explicitly cut stars out of $\mathcal{S}$ for this purpose because this quantity is dynamic, and priority is assigned on a target by target basis, making it difficult to apply to a field sample. Instead we make

a simplifying assumption to correct for this bias in assuming that all KOIs are equally likely not to be observed due to conflicts with other APOGEE-KOI targets. This is not strictly true, as planet candidates with low-mass host stars were given priority over competing planet candidates. However, because there are so few instances where this happens (<3%) and because we remove M dwarfs from $\mathcal{S}$, this is unlikely to lead to a noticeable bias in stellar host properties compared to the field. This bias is measured by selecting all the planet candidates without a "False Positive" disposition in DR24 with a host in $\mathcal{S}$ for a given field, and comparing that to the targets in the APOGE-KOI program in that field. Note, this is not the same quantity as the number of planets in $\mathcal{P}$ in that field, because we reject some candidates based on observations from APOGEE. For fields K04, K06, and K07 only five KOIs from DR24 were not observed, resulting in a flat probability of $p_{\rm apo} = 0.988$. This is the dominating bias for these fields.

For fields K10, K21, and K16, the most prolific bias is the selection of KOIs before a static catalog was available from *Kepler*. We consider this bias in conjunction with the fiber collision bias described above, as it largely dominates, and because there is not a catalog of "expected" planets with which to compare. This is not only due to the nature of the KOI catalogs at the time of target selection, but also improvements in the *Kepler* pipeline used to reject both false positives, such as the detection of photocenter offsets and significant secondary eclipses, as well as false alarms caused by instrumental systematics such as sudden pixel sensitivity dropouts, rolling bands, and abrupt changes in the photometric noise profile between quarters (Coughlin 2015; Mullally et al. 2016).

The targets for K16 were chosen as part of SDSS-III (Eisenstein et al. 2011), and used as a pathfinder for the APOGEE-KOI program. This field observed 163 planet candidates with $H < 14$ known at the time, in 2013 August. Of those 163 candidates, 153 had a "Candidate" status, 4 were confirmed, and 6 were not yet dispositioned (Fleming et al. 2015). The DR24 catalog, however, has 166 planet candidates in $\mathcal{S}$ with either a "Candidate" or "Confirmed" disposition. The APOGEE-KOI program observed 126

(75.9%) of these planet candidates. This discrepancy is due largely to two effects: improved vetting that removed a significant fraction of the 163 original planet candidates and improvements to the *Kepler* pipeline supplemented with additional data that allowed for the discovery of planet candidates with lower S/N transits. To take this bias into account, we measure $p_{\rm apo}$ by taking the ratio of DR24 candidates observed by APOGEE to the total number of DR24 candidates as a function of transit S/N, and model the increasing fraction with a modified gamma cumulative distribution function of the form,

$$p(S/N) = \frac{a}{d^b \Gamma(b)} \int_0^{S/N} (\xi - c)^{b-1} e^{-(\xi - c)/d} d\xi \quad . \tag{3.30}$$

Measuring the fraction of observed planet candidates in S/N bins of 1.0, we find the best fit parameters to be $a = 0.86$, $b = 6.0$, $c = 3.9$, and $d = 1.0$ (See Figure 3.22). This fit implies that at high $S/N$, only 86% of the *Kepler* planet candidates from August 2013 would survive the more detailed vetting procedures introduced in the DR24 pipeline (Coughlin 2015; Mullally et al. 2016). This bias is applied across the $P$-$R_p$ grid with the expected MES in place of the transit S/N in calculating $\langle \eta \rangle$. This fit is displayed in Figure 3.22. Perhaps surprisingly, we didn't see any significant trend in either $P$ or $R_p$ alone.

We apply the same analysis jointly to fields K10 and K21 (as targets between these two fields were selected from the same KOI catalog) which were observed as part of the SDSS-IV bright time extension program (Beaton et al., in prep). The DR24 catalog contains 318 planet candidates in fields K10 and K21 with either a "Candidate" or "Confirmed" disposition. APOGEE observed 130 out of 148 (87.8%) of these candidates in field K21 and 155 out of 170 (91.2%) of these candidates in field K10. Combining these two fields, and fitting a modified gamma cumulative distribution function as for Field K21, we find the best fit parameters of $a = 0.98$, $b = 0.22$, $c = 7.5$ , $d = 1.7$. This fit implies that at high $S/N$, 98% of the *Kepler* planet candidates detected at the time

of the APOGEE-KOI survey target selection would survive the more detailed vetting procedures introduced in the DR24 pipeline. Therefore, the dominant factor in the discrepancy between the number of planet candidates in these fields and those observed in APOGEE is dominated by improvements to the *Kepler* pipeline that resulted in the detection of lower S/N transits. We multiply this correction over all stars in fields K10 and K21 when calculating $\langle \eta \rangle$. See Figure 3.22.

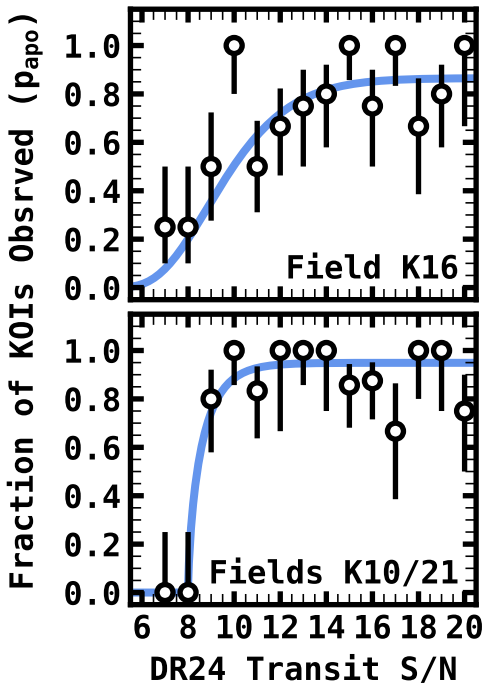


Fig. 3.22.— The probability that a particular planet candidate is observed in the three fields with early target selection in the APOGEE-KOI survey. There is a bias incurred from selecting planet candidates from a pre-DR24, non-static catalog, as well as avoiding APOGEE fiber collisions. The points show the fraction of DR24 planet candidates observed in APOGEE at a given transit S/N for field K16 (top) and fields K10 and K21 (bottom). The error bars are derived assuming a binomial distribution. The blue lines show our adopted models to correct for this bias.

### Probability of ASPCAP Failure ($p_{\rm fail}$)

Although the typical star in our sample has a spectrum with $S/N > 100$, our sample still includes a non-negligible fraction of dim stars near the $S/N$ limit of ASPCAP's capabilities. Because of these low $S/N$ sources, there is a non-negligible fraction of planet candidates ($\sim$10%) that would otherwise be in $\mathcal{P}$ that are not included in our analysis. To account for this bias, we assume the dominant reason for an ASPCAP failure is due to this low $S/N$ effect, although some spectra may fail due to other reasons (e.g., stray light from a bright companion). To account for this bias we model the failure probability as a function of $H$ because the number of cadences are designed to derive reliable orbital solutions, and therefore are not set by any stellar parameters or observable quantities that should bias the results of the occurrence rates.

To measure the failure probability, $p_{\rm fail}$, we take a similar approach to measuring $p_{\rm apo}$. We calculate the fraction of stars for a given $H$ bin of 0.2 magnitudes that have an ASPCAP-derived best fit solution for the input parameters (i.e., $T_{\rm eff}$, $\log g$, [M/H], $\xi_t$, [C/M], [N/M], and [$\alpha$/M]). We then fit the fraction of stars without an ASPCAP solution using a simple modified power law of the form,

$$p_{\rm fail}(H) = c + e^{a(H-b)} \tag{3.31}$$

where $H$ is the magnitude of the bin. We find the best fit parameters to be $a = 1.70$, $b = 14.45$, and $c = 0.02$. The best fit model and fractions are displayed in Figure 3.23. These fits imply that the failure rate due to low $S/N$ should be $\sim$1 at $H = 14.45$, and that the failure rate for reasons other than low $S/N$ is 2%, which is fairly insignificant compared to the other uncertainties considered in this study.

#### $\eta$: The Combined Survey Efficiency

We calculate the total average survey efficiency for each field as the product of each term described above, given the form

$$\langle \eta \rangle = \frac{1}{n_\star} \sum_i^{n_\star} p_{\rm tra,i} \times p_{\rm det,i} \times p_{\rm apo,i} \times (1 - p_{\rm fail,i}) \ , \tag{3.32}$$

where $\langle \eta \rangle$ is the average survey detection efficiency. The mean survey efficiency for each field is shown in Figure 3.11. In this way, i.e., by marginalizing over all the stars in $\mathcal{S}$, we've removed stellar properties from our expression for survey efficiency, so that $\eta = \eta(P, R_p)$. This relies on an implicit assumption that chemical abundances are not correlated with survey efficiency. Finally, for each $\log P$-$\log R_p$ bin, we calculate the combined survey efficiency for all the stars in the APOGEE-KOI program as a weighted sum of the efficiency for each APOGEE-KOI field indexed by $i$,

$$\langle \eta \rangle = \sum_i F_{\star,i} \times \langle \eta \rangle_i \ , \tag{3.33}$$

where $F_{\star,i}$ is the fraction of stars in $\mathcal{S}$ that are in field $i$ (see Table 3.2).

### 3.C.4 Errors Due to Uncertainties in $\mathcal{F}_\star(X)$

We do not have measured chemical abundances for all the stars in $\mathcal{S}$, so we do not know precisely how many stars from $\mathcal{S}$ are in a given metallicity bin. However, with knowledge of the distribution function over each abundance, $\mathcal{F}_\star(X)$, we can estimate this number by integrating the distribution over the bin and multiplying by the total number of stars in $\mathcal{S}$ (see Equation 3.15b).

To define $\mathcal{F}_\star$, we fit the abundance distributions in $\mathcal{S}$ using the measured abundances in $\mathcal{C}$ with a Gaussian kernel density estimator (KDE). The choice of bandwidth for the KDE is non-trivial, as it may impart significant bias if overestimated and introduce

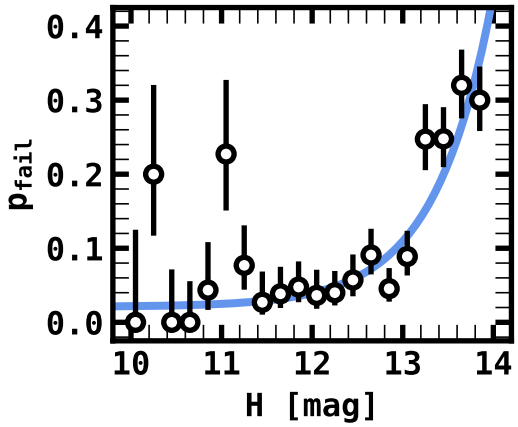


Fig. 3.23.— The ASPCAP failure rate as a function of $H$ magnitude. The data points are the fraction of stars observed in the APOGEE-KOI program where ASPCAP did not derive a solution in a given magnitude bin, with error bars assumed from a binomial distribution. The blue line shows our fitted model for the failure rate.

variance if underestimated. To select the optimal bandwidth, we fit a Gaussian KDE for a large sample of stars in the entire APOGEE database. The intent is to define a realistic distribution function from this sample, draw a number of stars equal to those in $\mathcal{C}$, and compare how well a given bandwidth recreates the defined distribution. We select this sample of stars such that it should broadly reflect our assumptions about the true abundance distributions in $\mathcal{S}$.

We remove stars from the APOGEE DR16 catalog with $\log g < 3.5$, $T_{\rm eff} < 4000$ K, $T_{\rm eff} > 6500$ K, $\pi/\sigma_\pi < 10$ in *Gaia* DR2, and a distance $> 1$ kpc as reported by Bailer-Jones et al. (2018). In addition to these sample selection cuts, we also apply a number of cuts designed to remove stars with poor quality measurements. We remove stars with $S/N < 50$ and any of the following ASPCAP or Star Flags set[11]: `TEFF_BAD`, `LOGG_BAD`, `METALS_BAD`, `ALPHAFE_BAD`, `STAR_BAD`, `SN_BAD`, and `VERY_CLOSE_NEIGHBOR`. This leaves ∼111,000 stars to define the parent sample.

From this APOGEE dwarf star sample we fit a KDE to the [Fe/H] measurements, and use this as our ground truth metallicity distribution. We then randomly sample 72

[11]for a description of these flags, see `https://www.sdss.org/dr16/algorithms/bitmasks/`

measurements from the KDE (chosen to match the number of stars in $\mathcal{C}$), and test each bandwidth from 0.01 – 0.30 dex, in intervals of 0.005 dex using a Kfolds cross-validation procedure with ten folds. We repeat this experiment 1000 times, and take the mean of the 1000 iterations to be the optimal model. This same experiment is run for each of the ten elements, resulting in our choices of bandwidth for each element given in Table 3.6. We display the resultant KDEs for each abundance in Figure 3.24. It's worth noting at this point that we are not assuming that $\mathcal{S}$ has the same abundance distribution as the APOGEE dwarf sample. Rather, we are assuming that optimizing our model selection for fitting the APOGEE dwarf sample will also optimize the model selection for fitting the abundance distributions of $\mathcal{C}$.

To calculate the planet occurrence rate, we rely on knowing the quantity $n_{\rm trial}$. In most occurrence rate studies, the uncertainty on this quantity is ignored, as uncertainties in the derived occurrence rates are typically dominated by Poisson error (e.g., Youdin 2011; Fressin et al. 2013; Petigura et al. 2013; Burke et al. 2015; Mulders et al. 2015c; Petigura et al. 2018; Narang et al. 2018). However, because the number of stars in $\mathcal{C}$ is small compared to $\mathcal{S}$, our strategy of extrapolating abundance distributions from $\mathcal{C}$ leads to a significant uncertainty in $\mathcal{F}_\star$, and therefore $n_{\rm trial}$ for a given abundance bin as compared with other studies. To estimate these errors and whether they are significant in relation to the Poisson error in our data, we again use the APOGEE dwarf sample described above. We perform a similar experiment to the one used to determine the optimal bandwidth, i.e., randomly drawing 72 measurements from the defined KDE for each APOGEE dwarf star abundance. We then fit a Gaussian KDE to the randomly drawn measurements using the optimal bandwidth determined above, and measure the difference in $F_\star$ as inferred from equation 3.22 between the defined KDE and the experiment KDE. We repeat this procedure 1,000 times to derive the typical uncertainty, $\sigma_{F_\star}$, and the bias (i.e., mean offset from the true value), $\delta_{F_\star}$, in $F_\star$ for each abundance bin. Performing this experiment, we draw two interesting conclusions. First, $\delta_{F_\star}$ is negligible for bins within the inner 90th

percentile of the abundance distributions. Therefore, when fitting the occurrence rate distributions, we omit abundances outside of this range. Secondly, the typical error on $\sigma_{F_\star}$ is ∼20-30%, and increases to ∼50-60% as you move away from the median of the distribution. This error is relatively small compared to the Poisson errors of planet size classes with few ($\lesssim$ 10) detected planets. Thus, this uncertainty is important to take into account for the occurrence rates of Sub-Neptunes and Super-Earths, but is less constraining when deriving occurrence rates for Sub-Saturns and Jupiters in our sample.

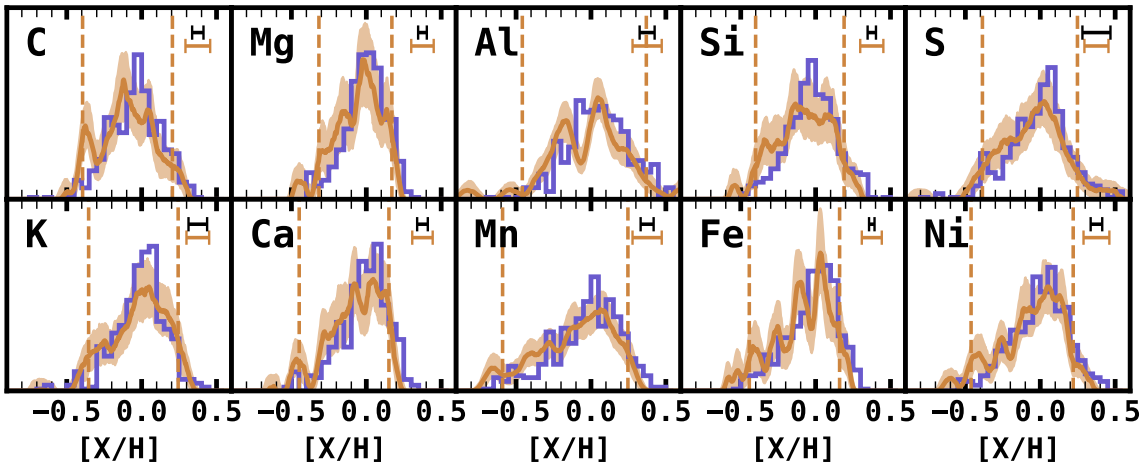


Fig. 3.24.— Our measurements for $\mathcal{F}_\star(X_i)$, applying a Gaussian KDE. The tan line shows our derived KDE, with the shaded region showing the $1\sigma$ region of credible models obtained from Monte Carlo sampling and bootstrapping. The abundance ratio distributions in $\mathcal{P}$ are shown in purple for comparison. The bandwidth used to model the abundance ratio distributions for $\mathcal{C}$ are represented by the tan error bar in the upper corner, and the vertical dashed lines mark off the inner 90th percentiles used in the occurrence rate analyses. The black error bar shows the median abundance uncertainty. The element is noted in the top left of each panel.

To account for this uncertainty, we modify our fitting procedure when deriving occurrence rates that depend on abundances. The procedure outlined in §3.C.2 is repeated 100 times. In each iteration we resample the abundances in $\mathcal{C}$ with replacement, adding an offset randomly drawn from a Gaussian distribution with the width set by the error reported in ASPCAP. The collection of posterior distribution from each of the 100 independent MCMC routines are then used to determine the range of credible models. This

bootstrapping routine is only performed when determining occurrence rates as a function of abundances where we rely on $\mathcal{C}$ to measure $\mathcal{F}_\star$. The range of credible models for $\mathcal{F}_\star$, as determined from the Monte Carlo bootstrapping routine, are shown in Figure 3.24. These estimates can also be interpreted as the abundance-ratio distribution functions for each chemical species in $\mathcal{S}$, providing the first such inferences for field dwarfs observed by *Kepler*.

Table 3.7:: Significance testing for the abundances between each planet subsample and $\mathcal{C}$. [a]Note: For Fe, we use [Fe/H].

| [X/Fe][a] | $P$-class | $R_p$-class | $n_{pl}$ | $\langle$[X/Fe]$\rangle$ | Norm? | $p_{ks}$ | $p_{mw}$ | $p_{bf}$ | Sig? |
|---|---|---|---|---|---|---|---|---|---|
| Fe | All | All | 504 | -0.001±0.161 | No | 0.038 | 0.0042 | 0.15 | Yes |
| Fe | All | SE | 198 | -0.032±0.172 | No | 0.54 | 0.14 | 0.53 | No |
| Fe | All | SN | 236 | 0.006±0.154 | No | 0.0076 | 0.00098 | 0.055 | Yes |
| Fe | All | SS | 11 | 0.062±0.088 | Yes | 0.022 | 0.004 | 0.015 | Yes |
| Fe | All | JP | 17 | 0.128±0.186 | Yes | 0.011 | 0.0016 | 0.6 | Yes |
| Fe | Hot | All | 235 | 0.044±0.142 | No | 0.00065 | 6.1e-06 | 0.018 | Yes |
| Fe | Hot | SE | 129 | -0.009±0.148 | No | 0.022 | 0.005 | 0.13 | Yes |
| Fe | Hot | SN | 64 | 0.102±0.132 | Yes | 7.9e-06 | 2.3e-08 | 0.00064 | Yes |
| Fe | Hot | SS | 2 | – | – | – | – | | |
| Fe | Hot | JP | 6 | – | – | – | – | | |
| Fe | Warm | All | 239 | -0.034±0.161 | No | 0.57 | 0.18 | 0.17 | No |
| Fe | Warm | SE | 65 | -0.134±0.177 | Yes | 0.16 | 0.075 | 0.66 | No |
| Fe | Warm | SN | 153 | -0.010±0.131 | No | 0.066 | 0.037 | 0.044 | No |
| Fe | Warm | SS | 7 | – | – | – | – | | |
| Fe | Warm | JP | 6 | – | – | – | – | | |
| Fe | Cool | All | 30 | -0.039±0.264 | Yes | 0.31 | 0.44 | 0.34 | No |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Fe | Cool | SE | 4 | – | – | – | – | | |
| Fe | Cool | SN | 19 | -0.205±0.190 | Yes | 0.21 | 0.42 | 0.78 | No |
| Fe | Cool | SS | 2 | – | – | – | – | | |
| Fe | Cool | JP | 5 | – | – | – | – | | |
| Ni | All | All | 504 | 0.021±0.042 | No | 0.65 | 0.4 | 0.21 | No |
| Ni | All | SE | 198 | 0.019±0.039 | No | 0.45 | 0.23 | 0.13 | No |
| Ni | All | SN | 236 | 0.021±0.045 | No | 0.93 | 0.44 | 0.58 | No |
| Ni | All | SS | 11 | 0.049±0.031 | Yes | 0.34 | 0.1 | 0.17 | No |
| Ni | All | JP | 17 | 0.047±0.027 | Yes | 0.12 | 0.059 | 1.0 | No |
| Ni | Hot | All | 235 | 0.022±0.038 | No | 0.59 | 0.42 | 0.097 | No |
| Ni | Hot | SE | 129 | 0.022±0.037 | No | 0.55 | 0.3 | 0.088 | No |
| Ni | Hot | SN | 64 | 0.030±0.043 | No | 0.3 | 0.14 | 0.65 | No |
| Ni | Hot | SS | 2 | – | – | – | – | | |
| Ni | Hot | JP | 6 | – | – | – | – | | |
| Ni | Warm | All | 239 | 0.021±0.044 | Yes | 0.84 | 0.41 | 0.41 | No |
| Ni | Warm | SE | 65 | 0.019±0.043 | Yes | 0.73 | 0.27 | 0.49 | No |
| Ni | Warm | SN | 153 | 0.017±0.049 | Yes | 0.82 | 0.39 | 0.61 | No |
| Ni | Warm | SS | 7 | – | – | – | – | | |
| Ni | Warm | JP | 6 | – | – | – | – | | |
| Ni | Cool | All | 30 | 0.012±0.056 | Yes | 0.98 | 0.43 | 0.99 | No |
| Ni | Cool | SE | 4 | – | – | – | – | | |
| Ni | Cool | SN | 19 | 0.009±0.054 | Yes | 0.94 | 0.44 | 0.46 | No |
| Ni | Cool | SS | 2 | – | – | – | – | | |
| Ni | Cool | JP | 5 | – | – | – | – | | |
| Si | All | All | 504 | 0.001±0.058 | No | 0.15 | 0.42 | 0.00055 | Yes |
| Si | All | SE | 198 | 0.011±0.055 | No | 0.21 | 0.37 | 0.0026 | Yes |
| Si | All | SN | 236 | -0.002±0.057 | No | 0.19 | 0.29 | 0.0034 | Yes |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Si | All | SS | 11 | 0.004±0.061 | Yes | 0.33 | 0.24 | 0.14 | No |
| Si | All | JP | 17 | 0.020±0.059 | Yes | 0.84 | 0.39 | 0.19 | No |
| Si | Hot | All | 235 | 0.002±0.056 | Yes | 0.14 | 0.4 | 0.0012 | Yes |
| Si | Hot | SE | 129 | 0.012±0.050 | No | 0.18 | 0.36 | 0.0028 | Yes |
| Si | Hot | SN | 64 | -0.003±0.063 | Yes | 0.29 | 0.15 | 0.033 | No |
| Si | Hot | SS | 2 | – | – | – | – | | |
| Si | Hot | JP | 6 | – | – | – | – | | |
| Si | Warm | All | 239 | 0.001±0.059 | No | 0.22 | 0.4 | 0.0023 | Yes |
| Si | Warm | SE | 65 | -0.001±0.066 | Yes | 0.76 | 0.44 | 0.11 | No |
| Si | Warm | SN | 153 | -0.001±0.055 | No | 0.15 | 0.31 | 0.0052 | No |
| Si | Warm | SS | 7 | – | – | – | – | | |
| Si | Warm | JP | 6 | – | – | – | – | | |
| Si | Cool | All | 30 | 0.004±0.056 | Yes | 0.38 | 0.31 | 0.17 | No |
| Si | Cool | SE | 4 | – | – | – | – | | |
| Si | Cool | SN | 19 | 0.004±0.059 | Yes | 0.31 | 0.16 | 0.4 | No |
| Si | Cool | SS | 2 | – | – | – | – | | |
| Si | Cool | JP | 5 | – | – | – | – | | |
| Mg | All | All | 504 | 0.006±0.059 | No | 0.024 | 0.051 | 0.11 | No |
| Mg | All | SE | 198 | 0.019±0.065 | No | 0.16 | 0.22 | 0.16 | No |
| Mg | All | SN | 236 | -0.002±0.059 | No | 0.019 | 0.024 | 0.15 | No |
| Mg | All | SS | 11 | -0.009±0.047 | Yes | 0.067 | 0.074 | 0.55 | No |
| Mg | All | JP | 17 | -0.034±0.039 | Yes | 0.022 | 0.018 | 0.22 | No |
| Mg | Hot | All | 235 | 0.002±0.052 | No | 0.0054 | 0.012 | 0.03 | No |
| Mg | Hot | SE | 129 | 0.015±0.062 | No | 0.047 | 0.12 | 0.14 | No |
| Mg | Hot | SN | 64 | -0.016±0.049 | No | 0.0017 | 0.00072 | 0.012 | Yes |
| Mg | Hot | SS | 2 | – | – | – | – | | |
| Mg | Hot | JP | 6 | – | – | – | – | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Mg | Warm | All | 239 | 0.013±0.067 | No | 0.17 | 0.22 | 0.19 | No |
| Mg | Warm | SE | 65 | 0.023±0.073 | No | 0.77 | 0.48 | 0.36 | No |
| Mg | Warm | SN | 153 | 0.006±0.063 | No | 0.11 | 0.13 | 0.21 | No |
| Mg | Warm | SS | 7 | – | – | – | – | | |
| Mg | Warm | JP | 6 | – | – | – | – | | |
| Mg | Cool | All | 30 | -0.003±0.078 | Yes | 0.47 | 0.13 | 0.58 | No |
| Mg | Cool | SE | 4 | – | – | – | – | | |
| Mg | Cool | SN | 19 | -0.001±0.079 | Yes | 0.75 | 0.28 | 0.51 | No |
| Mg | Cool | SS | 2 | – | – | – | – | | |
| Mg | Cool | JP | 5 | – | – | – | – | | |
| C | All | All | 504 | -0.017±0.060 | No | 0.1 | 0.047 | 0.13 | No |
| C | All | SE | 198 | -0.011±0.054 | Yes | 0.27 | 0.18 | 0.044 | No |
| C | All | SN | 236 | -0.027±0.067 | Yes | 0.053 | 0.012 | 0.44 | No |
| C | All | SS | 11 | 0.062±0.126 | Yes | 0.15 | 0.13 | 0.29 | No |
| C | All | JP | 17 | 0.013±0.052 | Yes | 0.82 | 0.44 | 0.16 | No |
| C | Hot | All | 235 | -0.021±0.060 | Yes | 0.091 | 0.036 | 0.1 | No |
| C | Hot | SE | 129 | -0.014±0.057 | Yes | 0.31 | 0.14 | 0.064 | No |
| C | Hot | SN | 64 | -0.030±0.073 | Yes | 0.045 | 0.011 | 0.59 | No |
| C | Hot | SS | 2 | – | – | – | – | | |
| C | Hot | JP | 6 | – | – | – | – | | |
| C | Warm | All | 239 | -0.015±0.066 | Yes | 0.22 | 0.07 | 0.29 | No |
| C | Warm | SE | 65 | -0.010±0.051 | Yes | 0.6 | 0.32 | 0.19 | No |
| C | Warm | SN | 153 | -0.027±0.067 | Yes | 0.092 | 0.023 | 0.47 | No |
| C | Warm | SS | 7 | – | – | – | – | | |
| C | Warm | JP | 6 | – | – | – | – | | |
| C | Cool | All | 30 | 0.001±0.044 | Yes | 0.67 | 0.38 | 0.19 | No |
| C | Cool | SE | 4 | – | – | – | – | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| C | Cool | SN | 19 | -0.018±0.053 | Yes | 0.75 | 0.29 | 0.6 | No |
| C | Cool | SS | 2 | – | – | – | – | | |
| C | Cool | JP | 5 | – | – | – | – | | |
| Al | All | All | 500 | 0.068±0.124 | No | 0.072 | 0.31 | 0.058 | No |
| Al | All | SE | 198 | 0.075±0.126 | Yes | 0.21 | 0.48 | 0.014 | No |
| Al | All | SN | 234 | 0.071±0.129 | No | 0.13 | 0.31 | 0.38 | No |
| Al | All | SS | 11 | -0.004±0.100 | Yes | 0.6 | 0.28 | 0.14 | No |
| Al | All | JP | 15 | 0.083±0.132 | Yes | 0.28 | 0.38 | 0.12 | No |
| Al | Hot | All | 234 | 0.065±0.134 | No | 0.11 | 0.28 | 0.23 | No |
| Al | Hot | SE | 129 | 0.075±0.133 | Yes | 0.34 | 0.48 | 0.07 | No |
| Al | Hot | SN | 63 | 0.062±0.155 | No | 0.52 | 0.19 | 0.66 | No |
| Al | Hot | SS | 2 | – | – | – | – | | |
| Al | Hot | JP | 6 | – | – | – | – | | |
| Al | Warm | All | 236 | 0.074±0.116 | No | 0.12 | 0.39 | 0.019 | No |
| Al | Warm | SE | 65 | 0.067±0.102 | No | 0.19 | 0.47 | 0.013 | No |
| Al | Warm | SN | 152 | 0.081±0.139 | No | 0.26 | 0.43 | 0.11 | No |
| Al | Warm | SS | 7 | – | – | – | – | | |
| Al | Warm | JP | 4 | – | – | – | – | | |
| Al | Cool | All | 30 | 0.056±0.086 | Yes | 0.12 | 0.32 | 0.031 | No |
| Al | Cool | SE | 4 | – | – | – | – | | |
| Al | Cool | SN | 19 | 0.044±0.053 | Yes | 0.063 | 0.3 | 0.048 | No |
| Al | Cool | SS | 2 | – | – | – | – | | |
| Al | Cool | JP | 5 | – | – | – | – | | |
| Ca | All | All | 504 | 0.010±0.046 | No | 0.7 | 0.49 | 0.12 | No |
| Ca | All | SE | 198 | 0.013±0.039 | No | 0.36 | 0.36 | 0.013 | No |
| Ca | All | SN | 236 | 0.010±0.050 | No | 0.85 | 0.46 | 0.47 | No |
| Ca | All | SS | 11 | 0.003±0.024 | Yes | 0.18 | 0.15 | 0.41 | No |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Ca | All | JP | 17 | 0.003±0.046 | No | 0.94 | 0.38 | 0.49 | No |
| Ca | Hot | All | 235 | 0.006±0.039 | No | 0.37 | 0.38 | 0.025 | No |
| Ca | Hot | SE | 129 | 0.011±0.037 | No | 0.53 | 0.43 | 0.0048 | Yes |
| Ca | Hot | SN | 64 | 0.003±0.043 | No | 0.45 | 0.3 | 0.54 | No |
| Ca | Hot | SS | 2 | – | – | – | – | | |
| Ca | Hot | JP | 6 | – | – | – | – | | |
| Ca | Warm | All | 239 | 0.011±0.049 | No | 0.83 | 0.41 | 0.3 | No |
| Ca | Warm | SE | 65 | 0.016±0.043 | No | 0.45 | 0.17 | 0.27 | No |
| Ca | Warm | SN | 153 | 0.011±0.049 | No | 0.93 | 0.46 | 0.39 | No |
| Ca | Warm | SS | 7 | – | – | – | – | | |
| Ca | Warm | JP | 6 | – | – | – | – | | |
| Ca | Cool | All | 30 | 0.012±0.084 | No | 0.9 | 0.44 | 0.56 | No |
| Ca | Cool | SE | 4 | – | – | – | – | | |
| Ca | Cool | SN | 19 | 0.026±0.079 | Yes | 0.77 | 0.47 | 0.81 | No |
| Ca | Cool | SS | 2 | – | – | – | – | | |
| Ca | Cool | JP | 5 | – | – | – | – | | |
| Mn | All | All | 504 | -0.000±0.073 | No | 0.42 | 0.17 | 0.027 | No |
| Mn | All | SE | 198 | -0.008±0.071 | No | 0.63 | 0.34 | 0.13 | No |
| Mn | All | SN | 236 | 0.003±0.073 | No | 0.37 | 0.084 | 0.11 | No |
| Mn | All | SS | 11 | 0.034±0.043 | Yes | 0.021 | 0.0055 | 0.2 | No |
| Mn | All | JP | 17 | 0.018±0.069 | Yes | 0.061 | 0.016 | 0.54 | No |
| Mn | Hot | All | 235 | 0.009±0.061 | No | 0.042 | 0.022 | 0.023 | No |
| Mn | Hot | SE | 129 | 0.000±0.066 | No | 0.34 | 0.28 | 0.12 | No |
| Mn | Hot | SN | 64 | 0.031±0.069 | Yes | 0.03 | 0.0014 | 0.13 | Yes |
| Mn | Hot | SS | 2 | – | – | – | – | | |
| Mn | Hot | JP | 6 | – | – | – | – | | |
| Mn | Warm | All | 239 | -0.007±0.071 | No | 0.84 | 0.47 | 0.097 | No |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Mn | Warm | SE | 65 | -0.022±0.060 | No | 0.091 | 0.044 | 0.44 | No |
| Mn | Warm | SN | 153 | -0.004±0.077 | No | 0.76 | 0.26 | 0.14 | No |
| Mn | Warm | SS | 7 | – | – | – | – | | |
| Mn | Warm | JP | 6 | – | – | – | – | | |
| Mn | Cool | All | 30 | -0.019±0.071 | No | 0.6 | 0.22 | 0.87 | No |
| Mn | Cool | SE | 4 | – | – | – | – | | |
| Mn | Cool | SN | 19 | -0.055±0.113 | Yes | 0.41 | 0.26 | 0.97 | No |
| Mn | Cool | SS | 2 | – | – | – | – | | |
| Mn | Cool | JP | 5 | – | – | – | – | | |
| S | All | All | 503 | 0.010±0.100 | No | 0.25 | 0.053 | 0.16 | No |
| S | All | SE | 197 | 0.003±0.091 | No | 0.06 | 0.02 | 0.33 | No |
| S | All | SN | 236 | 0.026±0.112 | No | 0.63 | 0.19 | 0.26 | No |
| S | All | SS | 11 | 0.009±0.071 | Yes | 0.87 | 0.32 | 0.18 | No |
| S | All | JP | 17 | 0.068±0.132 | Yes | 0.74 | 0.24 | 0.38 | No |
| S | Hot | All | 234 | 0.002±0.097 | No | 0.035 | 0.0075 | 0.25 | No |
| S | Hot | SE | 128 | 0.003±0.089 | No | 0.11 | 0.025 | 0.38 | No |
| S | Hot | SN | 64 | -0.001±0.135 | No | 0.23 | 0.031 | 0.85 | No |
| S | Hot | SS | 2 | – | – | – | – | | |
| S | Hot | JP | 6 | – | – | – | – | | |
| S | Warm | All | 239 | 0.019±0.099 | No | 0.6 | 0.22 | 0.18 | No |
| S | Warm | SE | 65 | -0.001±0.104 | No | 0.13 | 0.053 | 0.55 | No |
| S | Warm | SN | 153 | 0.030±0.108 | No | 0.65 | 0.35 | 0.15 | No |
| S | Warm | SS | 7 | – | – | – | – | | |
| S | Warm | JP | 6 | – | – | – | – | | |
| S | Cool | All | 30 | 0.028±0.103 | Yes | 0.65 | 0.42 | 0.086 | No |
| S | Cool | SE | 4 | – | – | – | – | | |
| S | Cool | SN | 19 | 0.056±0.145 | Yes | 0.83 | 0.47 | 0.34 | No |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| S | Cool | SS | 2 | – | – | – | – | | |
| S | Cool | JP | 5 | – | – | – | – | | |
| K | All | All | 500 | 0.014±0.074 | No | 0.00024 | 0.00025 | 0.22 | Yes |
| K | All | SE | 196 | 0.020±0.074 | No | 0.0024 | 0.003 | 0.018 | Yes |
| K | All | SN | 236 | 0.013±0.076 | No | 0.00049 | 0.00037 | 0.81 | Yes |
| K | All | SS | 11 | 0.022±0.042 | Yes | 0.067 | 0.076 | 0.35 | No |
| K | All | JP | 17 | 0.005±0.077 | No | 0.067 | 0.021 | 0.92 | No |
| K | Hot | All | 233 | 0.011±0.067 | No | 5.2e-05 | 7.1e-05 | 0.25 | Yes |
| K | Hot | SE | 129 | 0.021±0.068 | No | 0.00082 | 0.001 | 0.033 | Yes |
| K | Hot | SN | 64 | 0.001±0.078 | No | 0.00032 | 0.00046 | 0.44 | Yes |
| K | Hot | SS | 2 | – | – | – | – | | |
| K | Hot | JP | 6 | – | – | – | – | | |
| K | Warm | All | 237 | 0.015±0.080 | No | 0.0032 | 0.002 | 0.043 | Yes |
| K | Warm | SE | 63 | 0.014±0.081 | No | 0.039 | 0.049 | 0.088 | No |
| K | Warm | SN | 153 | 0.016±0.080 | No | 0.0048 | 0.0017 | 0.1 | Yes |
| K | Warm | SS | 7 | – | – | – | – | | |
| K | Warm | JP | 6 | – | – | – | – | | |
| K | Cool | All | 30 | 0.038±0.108 | No | 0.27 | 0.11 | 0.25 | No |
| K | Cool | SE | 4 | – | – | – | – | | |
| K | Cool | SN | 19 | 0.022±0.093 | No | 0.21 | 0.087 | 0.2 | No |
| K | Cool | SS | 2 | – | – | – | – | | |
| K | Cool | JP | 5 | – | – | – | – | | |

Table 3.8:: Best fit parameters for the planet occurrence . rate functions in equation 3.5

| $R_p$-Class | $P$-class | [X/H] | $\log C$ | $\alpha$ | $\beta$ |
|---|---|---|---|---|---|
| SE | hot | C | $-1.67^{+0.11}_{-0.11}$ | $+1.46^{+0.15}_{-0.15}$ | $+1.16^{+0.47}_{-0.43}$ |
| SE | hot | Mg | $-1.75^{+0.11}_{-0.11}$ | $+1.48^{+0.16}_{-0.16}$ | $+0.80^{+0.53}_{-0.50}$ |
| SE | hot | Al | $-1.78^{+0.11}_{-0.11}$ | $+1.49^{+0.16}_{-0.15}$ | $+0.45^{+0.36}_{-0.38}$ |
| SE | hot | Si | $-1.69^{+0.11}_{-0.11}$ | $+1.43^{+0.16}_{-0.15}$ | $+1.01^{+0.45}_{-0.44}$ |
| SE | hot | S | $-1.67^{+0.11}_{-0.11}$ | $+1.38^{+0.15}_{-0.15}$ | $+0.78^{+0.47}_{-0.43}$ |
| SE | hot | K | $-1.74^{+0.10}_{-0.11}$ | $+1.45^{+0.15}_{-0.15}$ | $+0.50^{+0.48}_{-0.46}$ |
| SE | hot | Ca | $-1.72^{+0.11}_{-0.12}$ | $+1.40^{+0.17}_{-0.16}$ | $+0.99^{+0.51}_{-0.52}$ |
| SE | hot | Mn | $-1.69^{+0.10}_{-0.11}$ | $+1.41^{+0.15}_{-0.15}$ | $+0.59^{+0.33}_{-0.35}$ |
| SE | hot | Fe | $-1.71^{+0.11}_{-0.12}$ | $+1.44^{+0.16}_{-0.16}$ | $+0.81^{+0.43}_{-0.44}$ |
| SE | hot | Ni | $-1.73^{+0.11}_{-0.11}$ | $+1.43^{+0.16}_{-0.15}$ | $+0.60^{+0.42}_{-0.40}$ |
| SE | warm | C | $+0.07^{+0.32}_{-0.31}$ | $-0.50^{+0.23}_{-0.24}$ | $-0.25^{+0.52}_{-0.54}$ |
| SE | warm | Mg | $-0.14^{+0.32}_{-0.32}$ | $-0.40^{+0.24}_{-0.24}$ | $-0.98^{+0.66}_{-0.65}$ |
| SE | warm | Al | $-0.05^{+0.30}_{-0.29}$ | $-0.38^{+0.22}_{-0.23}$ | $-0.28^{+0.38}_{-0.39}$ |
| SE | warm | Si | $+0.03^{+0.32}_{-0.31}$ | $-0.50^{+0.23}_{-0.24}$ | $-0.63^{+0.47}_{-0.53}$ |
| SE | warm | S | $+0.03^{+0.31}_{-0.31}$ | $-0.48^{+0.23}_{-0.24}$ | $-0.68^{+0.47}_{-0.49}$ |
| SE | warm | K | $-0.14^{+0.31}_{-0.31}$ | $-0.35^{+0.23}_{-0.24}$ | $-0.85^{+0.48}_{-0.48}$ |
| SE | warm | Ca | $-0.07^{+0.32}_{-0.32}$ | $-0.42^{+0.23}_{-0.24}$ | $-0.38^{+0.51}_{-0.51}$ |
| SE | warm | Mn | $-0.05^{+0.32}_{-0.31}$ | $-0.46^{+0.23}_{-0.24}$ | $-0.65^{+0.36}_{-0.38}$ |
| SE | warm | Fe | $-0.02^{+0.32}_{-0.31}$ | $-0.47^{+0.23}_{-0.24}$ | $-0.61^{+0.50}_{-0.53}$ |
| SE | warm | Ni | $+0.04^{+0.31}_{-0.31}$ | $-0.51^{+0.23}_{-0.24}$ | $-0.71^{+0.38}_{-0.41}$ |
| SN | hot | C | $-2.60^{+0.21}_{-0.23}$ | $+2.35^{+0.28}_{-0.27}$ | $+2.33^{+0.67}_{-0.61}$ |
| SN | hot | Mg | $-2.86^{+0.23}_{-0.26}$ | $+2.51^{+0.31}_{-0.29}$ | $+2.98^{+0.87}_{-0.77}$ |
| SN | hot | Al | $-2.80^{+0.22}_{-0.24}$ | $+2.40^{+0.29}_{-0.28}$ | $+1.26^{+0.51}_{-0.47}$ |
| SN | hot | Si | $-2.61^{+0.21}_{-0.23}$ | $+2.30^{+0.28}_{-0.27}$ | $+2.06^{+0.71}_{-0.61}$ |
| SN | hot | S | $-2.58^{+0.20}_{-0.22}$ | $+2.22^{+0.28}_{-0.26}$ | $+1.61^{+0.59}_{-0.55}$ |

| | | | | | |
|---|---|---|---|---|---|
| SN | hot | K | $-2.62^{+0.20}_{-0.21}$ | $+2.17^{+0.27}_{-0.26}$ | $+1.77^{+0.59}_{-0.55}$ |
| SN | hot | Ca | $-2.95^{+0.26}_{-0.28}$ | $+2.59^{+0.34}_{-0.32}$ | $+2.98^{+0.85}_{-0.78}$ |
| SN | hot | Mn | $-2.81^{+0.23}_{-0.25}$ | $+2.49^{+0.31}_{-0.29}$ | $+1.86^{+0.53}_{-0.50}$ |
| SN | hot | Fe | $-2.79^{+0.24}_{-0.26}$ | $+2.45^{+0.32}_{-0.30}$ | $+2.52^{+0.72}_{-0.68}$ |
| SN | hot | Ni | $-2.84^{+0.23}_{-0.26}$ | $+2.48^{+0.31}_{-0.29}$ | $+2.24^{+0.76}_{-0.65}$ |
| SN | warm | C | $-0.75^{+0.19}_{-0.20}$ | $+0.44^{+0.13}_{-0.13}$ | $+0.63^{+0.43}_{-0.46}$ |
| SN | warm | Mg | $-0.77^{+0.19}_{-0.19}$ | $+0.43^{+0.13}_{-0.13}$ | $+0.64^{+0.49}_{-0.53}$ |
| SN | warm | Al | $-0.83^{+0.19}_{-0.19}$ | $+0.45^{+0.13}_{-0.13}$ | $+0.31^{+0.35}_{-0.33}$ |
| SN | warm | Si | $-0.75^{+0.19}_{-0.19}$ | $+0.41^{+0.13}_{-0.13}$ | $+0.48^{+0.46}_{-0.44}$ |
| SN | warm | S | $-0.71^{+0.19}_{-0.20}$ | $+0.37^{+0.13}_{-0.13}$ | $+0.27^{+0.50}_{-0.42}$ |
| SN | warm | K | $-0.80^{+0.19}_{-0.19}$ | $+0.44^{+0.13}_{-0.13}$ | $-0.04^{+0.39}_{-0.42}$ |
| SN | warm | Ca | $-0.73^{+0.19}_{-0.19}$ | $+0.41^{+0.13}_{-0.13}$ | $+0.73^{+0.49}_{-0.49}$ |
| SN | warm | Mn | $-0.81^{+0.19}_{-0.20}$ | $+0.45^{+0.13}_{-0.13}$ | $+0.30^{+0.33}_{-0.37}$ |
| SN | warm | Fe | $-0.77^{+0.19}_{-0.20}$ | $+0.43^{+0.13}_{-0.13}$ | $+0.49^{+0.43}_{-0.42}$ |
| SN | warm | Ni | $-0.81^{+0.19}_{-0.19}$ | $+0.45^{+0.13}_{-0.13}$ | $+0.49^{+0.35}_{-0.35}$ |
| SS | hot | C | – | – | – |
| SS | hot | Mg | – | – | – |
| SS | hot | Al | – | – | – |
| SS | hot | Si | – | – | – |
| SS | hot | S | – | – | – |
| SS | hot | K | – | – | – |
| SS | hot | Ca | – | – | – |
| SS | hot | Mn | – | – | – |
| SS | hot | Fe | – | – | – |
| SS | hot | Ni | – | – | – |
| SS | warm | C | $-4.84^{+0.96}_{-1.11}$ | $+2.29^{+0.62}_{-0.55}$ | $+3.43^{+1.33}_{-1.17}$ |
| SS | warm | Mg | $-4.98^{+0.96}_{-1.10}$ | $+2.30^{+0.62}_{-0.55}$ | $+3.51^{+1.59}_{-1.42}$ |

| | | | | | |
|---|---|---|---|---|---|
| SS | warm | Al | $-5.05^{+0.97}_{-1.11}$ | $+2.30^{+0.62}_{-0.56}$ | $+1.79^{+0.92}_{-0.84}$ |
| SS | warm | Si | $-4.90^{+0.96}_{-1.10}$ | $+2.30^{+0.61}_{-0.55}$ | $+3.39^{+1.38}_{-1.19}$ |
| SS | warm | S | $-4.46^{+0.87}_{-0.99}$ | $+1.99^{+0.56}_{-0.51}$ | $+3.16^{+1.31}_{-1.16}$ |
| SS | warm | K | $-4.53^{+0.87}_{-0.98}$ | $+2.08^{+0.56}_{-0.51}$ | $+1.73^{+1.08}_{-0.98}$ |
| SS | warm | Ca | $-4.77^{+0.98}_{-1.13}$ | $+2.21^{+0.64}_{-0.57}$ | $+2.86^{+1.54}_{-1.34}$ |
| SS | warm | Mn | $-4.92^{+1.00}_{-1.16}$ | $+2.27^{+0.65}_{-0.58}$ | $+2.25^{+1.08}_{-0.95}$ |
| SS | warm | Fe | $-4.91^{+1.00}_{-1.16}$ | $+2.28^{+0.65}_{-0.58}$ | $+3.25^{+1.52}_{-1.32}$ |
| SS | warm | Ni | $-5.01^{+1.01}_{-1.16}$ | $+2.28^{+0.65}_{-0.59}$ | $+3.05^{+1.46}_{-1.25}$ |
| JP | hot | C | $-3.68^{+0.64}_{-0.80}$ | $+1.47^{+0.88}_{-0.81}$ | $+7.69^{+3.62}_{-2.84}$ |
| JP | hot | Mg | $-3.67^{+0.66}_{-0.86}$ | $+1.34^{+1.01}_{-0.93}$ | $+6.34^{+4.16}_{-3.27}$ |
| JP | hot | Al | $-3.45^{+0.59}_{-0.75}$ | $+0.52^{+0.79}_{-0.82}$ | $+4.62^{+2.31}_{-1.88}$ |
| JP | hot | Si | $-4.66^{+0.98}_{-1.27}$ | $+1.76^{+1.09}_{-0.95}$ | $+11.28^{+6.40}_{-4.75}$ |
| JP | hot | S | $-3.50^{+0.63}_{-0.81}$ | $+0.83^{+0.93}_{-0.92}$ | $+5.61^{+3.35}_{-2.61}$ |
| JP | hot | K | $-3.65^{+0.61}_{-0.76}$ | $+1.46^{+0.88}_{-0.80}$ | $+4.24^{+2.69}_{-2.16}$ |
| JP | hot | Ca | $-3.77^{+0.90}_{-1.29}$ | $+0.17^{+1.74}_{-1.94}$ | $+5.41^{+6.85}_{-4.62}$ |
| JP | hot | Mn | $-3.83^{+0.74}_{-0.95}$ | $+0.99^{+0.94}_{-0.91}$ | $+6.91^{+3.97}_{-3.04}$ |
| JP | hot | Fe | $-3.67^{+0.86}_{-1.23}$ | $+0.15^{+1.72}_{-1.92}$ | $+4.36^{+5.71}_{-3.97}$ |
| JP | hot | Ni | $-3.91^{+0.86}_{-1.16}$ | $+0.53^{+1.15}_{-1.19}$ | $+7.96^{+5.67}_{-4.19}$ |
| JP | warm | C | $-2.71^{+1.16}_{-1.25}$ | $+0.65^{+0.79}_{-0.81}$ | $+0.65^{+1.59}_{-1.51}$ |
| JP | warm | Mg | $-2.82^{+1.15}_{-1.25}$ | $+0.65^{+0.79}_{-0.80}$ | $-0.72^{+1.81}_{-1.80}$ |
| JP | warm | Al | $-1.47^{+1.46}_{-1.40}$ | $-0.41^{+1.00}_{-1.16}$ | $+0.32^{+1.44}_{-1.40}$ |
| JP | warm | Si | $-2.21^{+0.93}_{-0.97}$ | $+0.45^{+0.63}_{-0.65}$ | $+2.17^{+1.56}_{-1.38}$ |
| JP | warm | S | $-2.72^{+1.14}_{-1.24}$ | $+0.65^{+0.78}_{-0.79}$ | $+0.69^{+1.52}_{-1.43}$ |
| JP | warm | K | $-1.93^{+1.03}_{-1.05}$ | $+0.13^{+0.71}_{-0.76}$ | $-0.85^{+1.30}_{-1.31}$ |
| JP | warm | Ca | $-1.85^{+1.03}_{-1.05}$ | $+0.13^{+0.71}_{-0.76}$ | $+0.68^{+1.55}_{-1.42}$ |
| JP | warm | Mn | $-1.84^{+1.03}_{-1.06}$ | $+0.12^{+0.71}_{-0.77}$ | $+0.80^{+1.17}_{-1.05}$ |
| JP | warm | Fe | $-1.84^{+1.03}_{-1.05}$ | $+0.12^{+0.71}_{-0.76}$ | $+0.83^{+1.51}_{-1.36}$ |

| | | | | | |
|---|---|---|---|---|---|
| JP | warm | Ni | $-1.88^{+1.04}_{-1.05}$ | $+0.12^{+0.71}_{-0.77}$ | $+0.61^{+1.31}_{-1.18}$ |

| Cluster | Age | [Fe/H] | Distance | $f_{\rm hot}$ | $f_{\rm hot}$ |
|---|---|---|---|---|---|
| | Gyr | dex | pc | $1-1.9\ R_\oplus$ | $1.9-4\ R_\oplus$ |
| Pleiades | 0.1 | $-0.01$ | 136 | $15.4^{+2.2}_{-1.9}$ | $7.7^{+1.5}_{-1.3}$ |
| Praesepe | 0.7 | $+0.16$ | 47 | $21.2^{+6.2}_{-4.8}$ | $20.7^{+7.7}_{-5.7}$ |
| Hyades | 0.8 | $+0.13$ | 186 | $20.1^{+5.2}_{-4.1}$ | $17.4^{+5.6}_{-4.2}$ |
| Ruprecht 147 | 2.0 | $+0.12$ | 310 | $19.7^{+4.9}_{-3.9}$ | $16.4^{+5.0}_{-3.9}$ |
| M67 | 3.5 | $+0.03$ | 880 | $16.6^{+2.8}_{-2.3}$ | $9.7^{+2.0}_{-1.7}$ |
| NGC 188 | 5.5 | $+0.11$ | 1990 | $19.3^{+4.6}_{-3.7}$ | $15.5^{+4.5}_{-3.5}$ |
| NGC 6791 | 7.0 | $+0.35$ | 2300 | $30^{+16}_{-11}$ | $63^{+51}_{-28}$ |
| ***Kepler* Field Stars** | – | – | – | $8.1^{+0.7}_{-0.8}$ | $4.7^{+0.6}_{-0.6}$ |

Table 3.5: Predicted occurrence rates, $f_{\rm hot}$, of hot planets ($P = 1-10$ days) for a few nearby open clusters in the absence of long-term planetary evolution. $f_{\rm hot}$ is given in units of Number of Planets per 100 Stars for each size class. Our model is extrapolated for NGC 6791, so the uncertainties for this cluster are quite large. The occurrence rates for the *Kepler* field are from this study.

| $X$ | $\sigma_{X/H}$ |
|---|---|
| Fe | 0.09 |
| Ni | 0.12 |
| Si | 0.11 |
| Mg | 0.10 |
| C | 0.12 |
| Al | 0.13 |
| Ca | 0.09 |
| Mn | 0.14 |
| S | 0.11 |
| K | 0.11 |

Table 3.6: The bandwidth adopted for the Gaussian kernel used to estimate the distribution for each elemental abundance in $\mathcal{C}$.

# Chapter 4

# TraSH in the Kepler Field. Improved Properties for ∼163,000 Stars and a Transit Detection Pipeline Designed to Minimize Bias with Spectral Type

## Summary

Despite being arguably the most successful planet-hunting mission to date, the *Kepler* mission suffers from a bias against detecting Transiting planets with Subgiant Hosts (TraSH). This bias derives primarily from the choice of transit templates searched in the light curves, which is optimized for finding planets around Sun-like stars. However, TraSH plays a crucial role in our understanding of planet age demographics, due to the precision in age-dating subgiant stars afforded via isochrone modeling. To first identify subgiant stars in the *Kepler* field, we apply parallaxes from *Gaia* DR2 to broadband

photometry to derive fundamental stellar properties (e.g., age, $M_\star$, $R_\star$, $\rho_\star$) for ∼163,000 stars in the *Kepler* field. We derive these properties with careful attention to homogeneity so as to avoid bias and release them in the CAtolog of Transit SearcH Input Parameters (`CaTShIP`). Finally, to reddress the bias against TraSH in the original *Kepler* mission, we present `TraSH-DUMP` (TRAnsiting planets with Subgiant Hosts – Detection with an Unbiased Matched filter Pipeline), a custom transiting planet detection pipeline which utilizes an optimized set of transit templates based on the properties of the star being searched. Leveraging the stellar characterization from the `CaTShIP`, we demonstrate the performance of `TraSH-DUMP` and consider the biases that would arise from a transit search aimed at detecting planets with evolved hosts.

## 4.1 Introduction

The *Kepler* mission (Borucki et al. 2010; Borucki 2016) has discovered thousands of exoplanets enabling population studies of large samples of planets spanning a wide range of stellar host types (Howard et al. 2012; Fressin et al. 2013; Mulders et al. 2015b; Dressing & Charbonneau 2013, 2015). However, while this remarkable database has been exploited to explore planet hosts on the main sequence (MS), there is still a distinct lack of knowledge regarding planets with post-MS host stars. Such planetary systems are expected to vary substantially from their MS counterparts (Veras 2016, and references therein).

One important consideration in understanding the population of planets with evolved host stars is tides. Tidal evolution is expected to destroy close-in giant planets as the host star evolves (Villaver & Livio 2009; Kunitomo et al. 2011; Schlaufman & Winn 2013; Villaver et al. 2014; Sun et al. 2018). However, the timescale at which inspiral occurs is debated, and theoretical predictions are extremely sensitive to the interior structure of the host star (Weinberg et al. 2017). While some studies suggest stars are

destroyed as early as the MS (Hamer & Schlaufman 2019), other studies have found that the occurrence of hot Jupiters is consistent from the MS to the base of the RGB, implying that tidal disruption is negligible until the star evolves onto the RGB (Grunblatt et al. 2018). One proposed explanation for this apparent discrepancy is a population of transient planets with moderate eccentricities and larger initial orbital separations that, with rapidly decaying orbits of their own, "replace" the population of hot Jupiters destroyed on the MS (Villaver et al. 2014; Grunblatt et al. 2019). No matter the actual situation, it is clear that post-MS stellar evolution brings on rapid changes in planetary architectures. It is therefore imperative for constraining these various models that not only the occurrence rate of planets at short periods be measured, but the rates for planets at all accessible orbital distances, and across all stellar ages, masses, and evolutionary states.

In addition to understanding the role of tides and stellar evolution in reshaping planetary architectures, stellar age-dating is a necessary tool for understanding planetary system evolution that is driven by the long-term processes in the planetary systems themselves rather than their host star's evolution. Recently, Berger et al. (2020a) argued for such evolution by comparing the ratio of the number of mini-Neptunes to the number of super-Earths in a population of old (>1 Gyr) and young (<1 Gyr) host stars. The authors found that the fraction $N_{SupEarth}/N_{SubNep}$ was lower for the younger sample of host stars, and inferred from this that sub-Neptunes lose their atmospheres over ∼Gyr timescales. David et al. (2021) came to a similar conclusion using both isochrone-derived ages from the California-*Kepler*-Survey (CKS; Johnson et al. 2017; Petigura et al. 2017a) as well as ages inferred from gyrochronology. David et al. (2021) argued that the small ($R_p \sim 1\text{-}4\,R_\oplus$) planet size distribution changes over the course of several ∼Gyr, finding evidence that the occurrence of planets within the radius gap (Fulton et al. 2017) is lower for younger stars. From this, they interpreted that large Super-Earths are the result of atmospheric loss that takes place over several Gyr. These studies provide motivation for

precise stellar age dating to understand the dependence of exoplanet demographics on age.

Understanding the population of TRAnsiting planets with Subgiant Hosts (TraSH) is critical in this endeavor because subgiants are uniquely suited to precision age dating due to their rapid evolution. However, as a result of the scientific imperative to detecting small planets around Sun-like stars, the detection algorithm applied by the *Kepler* team only searched for transits with durations up to 15 hours (Thompson et al. 2018), leading to unreliable completeness estimates for stars that can reasonably host planets with transit durations >15 hours, including subgiants and RGB stars (Burke & Catanzarite 2017b).

Transiting planets around such stars are likely to have durations in excess of 15 hours for periods as low as $P \sim$ 100-300 days, well within the detectable parameter space for the *Kepler* mission. However, only a handful of planets with transit durations longer than 15 hours have been confirmed in the literature. Whether these planets have not been discovered due to the biases discussed, or because of an intrinsic rarity, is not known. However, in the limit in which the transit signal to noise is sufficiently large, there should be more detected planets with long durations due to an increased transit probability, $R_\star/a$. There may be other systematic biases other than the duration of transits searched, such as corruption of long duration events by *Kepler*'s systematics removal pipeline. To fully understand whether the low number of detected planets is due to an inherent decrease in occurrence rates, a bias caused by the decreases planet-star radius contrast, or simply a bias in the *Kepler* light curve processing and transit search algorithms, a dedicated study with uniquely-suited light curve processing and transit detection algorithms is needed.

We present the first systematic exploration of the TraSH in *Kepler* program. The present study is part of an effort aimed at mapping the demographics of planets as a function of stellar mass and age by constructing an independent planet catalog with

completeness estimates for stars across the entirety of the HR diagram. We present the foundations for those goals here. Using an entirely homogeneous analysis, we construct a catalog of the most precise stellar ages, masses, radii, and other fundamental parameters for ∼163,000 stars observed by the *Kepler* mission (§4.2). In addition to the most precise stellar masses and radii to date, we are able to resolve multiple stages of MS evolution for high mass ($M_\star \gtrsim 1.2 M_\odot$) stars in our catalog, enabling inferences of the stellar interior structure needed to fully test predictions of tidal decay theory. We also introduce a new detection algorithm, `TraSH-DUMP`, designed to search for planets orbiting stars in any region of the HR diagram with minimal bias, unlike most transiting planet searches which suffer from 2nd order detection biases arising from template mismatch (§4.3). Finally we conclude this work by discussing the implications of these findings and other considerations needed to measure planet occurrence rates across a diverse range of stellar masses, ages, and evolutionary states (§4.4 & §4.5).

## 4.2 `CaTShIP`: The CAtalog of Transit SearcH Input Parameters

The goal of the CAtolog of Transit SearcH Input Parameters (`CaTShIP`) is to (1) measure precise fundamental stellar properties to aid the search of transiting planets, (2) reduce uncertainties in the properties of transiting planets found in the *Kepler* data, (3) determine the relative ages/evolutionary states of the stars in the *Kepler* field, and (4) derive stellar ages, masses, radii, etc. with meticulous attention to homogeneity so as to avoid biases that may result from the design of surveys outside our control.

For example, one potential bias for our science goals that we avoid is removing stars with known close ($\lesssim 1''$) companions detected with high-resolution imaging surveys such as Robo-AO (Ziegler et al. 2018). The vast majority of the targets in these surveys are planet candidates, creating a bias where field stars will have a higher binary fraction

than planet candidates, effecting our planet occurrence statistics. To avoid biases such as these, all input data we use is from surveys with homogeneous data across each star in our sample, with a possible exception of metallicity information when available.

In addition to the above goals, we wish to optimize the precision of the fundamental parameters derived in the `CaTShIP`, rather than the accuracy. In other words, our science goals are more readily accomplished with a better understanding of the degree to which we can differentiate stellar properties ($M_\star, R_\star$, etc.) from one star to the next as opposed to pinning down an absolute value for these properties. To this goal, the uncertainties reported in the `CaTShIP` are largely model-dependent and are not appropriate for calibrating such models. To adequately accomplish this precision, we use as few input parameters to fit to our stellar model grid as possible, for fear that any one poor estimate could deviate away from the precision granted from carefully selected photometric bands.

### 4.2.1 Stellar Sample

To select a field sample of stars with good photometry and reliable parallaxes, we first cross match the point sources detected by Everett et al. (2012) with the 2MASS point source catalog. For this cross-match, and those listed below, we take the best point source within $1''$. We require each star to have reliable photometric measurements in each band for all the stars in this cross-match. Therefore, we remove all sources that are not detected in $U$, $B$, and $V$ from Everett et al. (2012), and we remove all sources from the 2MASS point source catalog with a quality flag other than "AAA". At this point, we then cross match this catalog with the *Kepler* stellar properties table (Mathur et al. 2017) and remove any stars that have a poorly-constrained parallax in *Gaia* EDR3, which we define as $\pi/\sigma_\pi < 20$. The final resulting field catalog contains 164,211 stars.

### 4.2.2 Input Measurements

We use five parameters to fit each star to a stellar evolutionary track: $T_{\rm eff}$, $\pi$, $E(B-V)$, $Ks$, and [Fe/H]. In the following sections, we explain the origin of each measurement and any systematic corrections we apply to either improve accuracy or homogenize our measurements across each star in our catalog.

#### Parallax from *Gaia* EDR3

We use the derived $\pi$ from *Gaia* EDR3 (Gaia Collaboration et al. 2021). The parallax constrains the distance to the star which, combined with extinction corrections, $T_{\rm eff}$, and $Ks$ magnitude, constrains the stellar radius and luminosity, arguably the most important quantities to measure, along with $T_{\rm eff}$, to reliably infer a star's evolutionary state.

#### Photometry from 2MASS Corrected for Blended Sources

In conjunction with $\pi$, a star's apparent magnitude constrains the luminosity semi-empirically (with only bolometric corrections acting as model-dependent values). However, the apparent magnitude can be biased by dust extinction. To minimize this effect in our analysis, we use the $Ks$-band magnitude from the 2MASS survey (Skrutskie et al. 2006), because it is the longest wavelength ($\lambda \sim 2.2\,\mu$m) uniformly available for each star in our sample.

However, 2MASS has an effective resolution of $\sim$3.5″, which is $\sim$2× larger than the effective resolution of $\sim$1.5″ in the catalog from Everett et al. (2012). This is likely to lead to blended sources, and therefore cause stars to appear redder than they are because the $Ks$-band flux will be artificially inflated, but not the $B$-band flux. Therefore, instead of using the $Ks$-band magnitudes directly from 2MASS, we instead use the $Ks$-band magnitudes from Berger et al. (2020b), who applied a correction for stars with a *Gaia*-detected source within 4″. The correction assumed that the flux contamination

from the secondary source seen by *Gaia* was entirely within the 2MASS aperture at separations below 1.25″, and entirely outside the aperture at separations >4″, with a linear relationship for the flux contamination between these two extremes. Given *Gaia*'s typical resolution of ∼0.5-1″, the effective resolution of the corrected $Ks$ magnitudes adopted in this work should be consistent with the effective resolution of the catalog in Everett et al. (2012).

### Dust Correction and Choice of Extinction Law

We employ the 3D dust map from Green et al. (2019) to derive the extinction for each star in our sample, which we access using the the python package `dustmaps` (Green 2018). To derive the extinction in each band we assume the reddening law from Wang & Chen (2019) and convert the arbitrary reddening units, $E$, from Green et al. (2019) to $E(G_{BP} - G_{RP})$ by converting to $E(B-V)$ via their relation, $E(B-V) = 0.996\,E$ and then applying the relations listed in Table 3 of Wang & Chen (2019) to calculate extinction values. For these extinctions we assume an error at least as large as 0.002 mag, given the typical uncertainties for the conversions derived in Wang & Chen (2019). Because Wang & Chen (2019) don't report color excess ratios for the $U$-band, we apply their expression for the excess color ratio, $A_\lambda/A_V$ using $\lambda_{\rm eff} = 0.3656\,\mu$m to derive the extinction for this band.

### Effective Temperature from $B - Ks$ Color

To measure stellar effective temperatures, we develop a relation using synthetic photometry from stellar evolutionary tracks described in §2.3. We define each model as one point in $B-Ks$ and $T_{\rm eff}$ space. In color bins of 0.05 mag, we calculate the weighted mean of the temperature for all models in that bin. The weights are calculated from the metallicity of each model as a Gaussian with mean of 0.0 dex and standard deviation of 0.2 dex. We choose to bias our temperature calibration in this way because it is representative of

the metallicity distribution in the *Kepler* field (Dong et al. 2014), and our model grids are asymmetric with respect to metallicity, so without the metallicity weights we may be biased toward lower-metallicity models. Because we don't expect many low-metallicity stars, we choose to dilute their influence in our relation. We fit these points to a 10th order polynomial to define the conversion from $(B-Ks)$ to effective temperature. To calculate the relative errors in this scheme, we take the standard deviation of the residuals of the models and the polynomial function in 0.05 magnitude bins. We then fit the scatter to a polynomial as a function of $B-Ks$ to derive a function for the intrinsic color-dependent uncertainties in our temperature relation. These relations are displayed in Figure 4.1.

The adopted error on the $T_{\rm eff}$ relation is taken to be the larger of either the scatter in the MIST models, as described, or a minimum of $\sim$2%. This minimum is given by the systematic uncertainty when measuring interferometric angular diameters (e.g., von Braun & Boyajian 2017).

### Data-Driven Photometric Metallicities Trained on LAMOST DR5

Uncertainties in metallicity are well known to lead to degeneracies when attempting to infer stellar ages and masses from stellar isochrones. Therefore, we make a concerted effort to include a measured metallicity for each star in our sample, without any particular assumptions about the stellar population to which the star belongs (e.g., the thin disk versus the thick disk). There are some instances where this is not possible, and we are thus forced to assume a thin disk prior, as the *Kepler* field is dominated by such stars.

In addition to our need for as many stars as possible to have a measured metallicity, we also require our stars to be uniform in metallicity, so we avoid using metallicities from multiple spectroscopic pipelines, which often have systematic differences of $\gtrsim$0.1 dex (Furlan et al. 2018). Thus, by including multiple sources for metallicities in a sample such as ours, we would in turn have to account for such systematic differences from

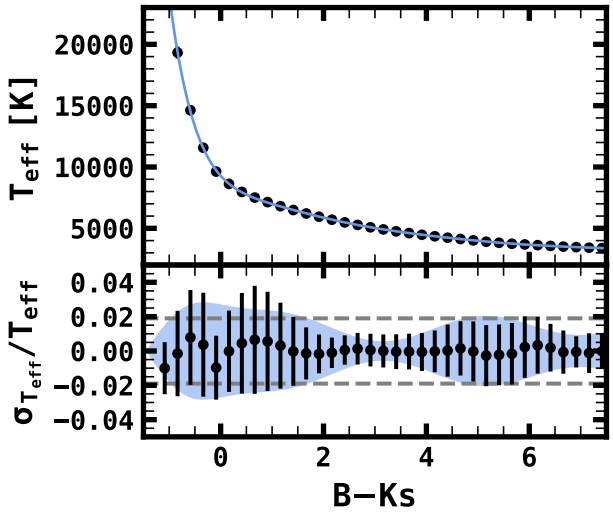


Fig. 4.1.— Calibration used to fit synthetic photometry to a model atmosphere effective temperature. *Top:* A 10th order polynomial fit to the [Fe/H]-weighted mean effective temperature for $B-Ks$ bins of size 0.05 mag, shown as black points. *Bottom:* The error expected from applying this relationship. The shaded blue region shows the expected uncertainty, and grey dashed lines show 1.9% uncertainty. The black points show the residuals from the fit. Only one in ten points are shown for clarity.

pipeline to pipeline by inflating the errors for each metallicity, which in turn negates the improvement in precision gained by adopting spectroscopic metallicities in the first place.

In an attempt to solve the problem of adopting metallicities for a significant sample of stars while also maintaining a uniform analysis and preserving the precision afforded by spectroscopic metallicities, we develop a methodology for inferring photometric metallicities. We choose one survey from which we adopt spectroscopic metallicities, LAMOST DR5 (Cui et al. 2012). LAMOST was chosen because of the LAMOST-*Kepler* program, which observed a large fraction of stars with no bias toward color or spectral type (De Cat et al. 2015; Zong et al. 2020). We then wish to train a model from which we can infer the metallicities of stars that were not observed by LAMOST. For this purpose we define a model for transforming a set of observed photometric values to inferred metallicities. We use $(U-B)_0$, $(B-Ks)$, and $M_{Ks}$ as inputs to our model, and map these inputs to a given metallicity. We choose the $(U-B)_0$ color because line blanketing in the $U$-band has

historically been an efficient method to infer photometric metallicities, albeit with worse precision as compared to spectroscopy. The effects of line blaketing on stellar SEDs is well known to vary by luminosity class and spectral type, so we also include $(B-Ks)_0$, and $M_{Ks}$ to act as proxies for temperature and luminosity, respectively, to allow our model to accurately disentangle the mapping from $(U-B)_0$ to [Fe/H].

Using these three inputs, we define a model using a Support Vector Regressor (SVR). An SVR works similarly to a Support Vector Machine (SVM), which is a simple classification algorithm. SVM identifies a hyperplane or line in multidimensional space that separates two or more classes of objects from one another, and then adopts that boundary to classify future objects. SVR works similarly, but instead of using a hyperplane to classify objects, it calculates a residual between the hyperplane and an input point in multi-dimensional space to define a continuous mapping from one domain to another, in our case, from $\theta_i = \{(U-B)_{0,i}, (B-Ks)_{0,i}, M_{Ks,i}\}$ to [Fe/H].

To train our SVR, we cross-match our catalog with that of LAMOST DR5, totaling 56,583 sources in common. To construct an adequate training sample, we downsample this cross-matched catalog to remove sources with poor photometry or stars with spectral types for which LAMOST metallicities are poorly constrained. Thus we remove stars with colors outside of the ranges $-0.5 < (B-Ks)_0 < 7$ and $-0.5 < (U-B)_0 < 2.2$. We also remove stars with [Fe/H] outside of the range used by the model grid in `isofit`, $-2 <$ [Fe/H] $< 0.5$. Finally, we remove stars with $\pi/\sigma_\pi < 10$ and stars with a spectrum $S/N < 50$ in the $i$-band. After these cuts we are left with 49,285 stars.

Finally, we randomly assign the LAMOST stars with good photometry and spectroscopic metallicities to a training sample and a test sample of equal sizes and fit the training sample of stars. To test the accuracy and precision of the SVR, we then use the trained model to predict the metallicities of the test sample. Our trained SVR is able to predict the metallicities of the test sample with minimal bias in the input parameters, and with relatively good precision considering the metallicities are derived from photom-

etry, see Figure 4.2. In the residuals for the test sample we find a median of 0.001 dex, in the sense that the predictions are slightly overestimated on average, with a median absolute deviation of 0.115 dex. The standard deviation of the measurements is higher, at 0.141 dex, implying there are some outliers. There are no trends with the residual in the test sample for any of the input parameters, implying that any biases in the sample are at least consistent with those of the LAMOST dataset.

One potential bias in the photometric metallicities is that they tend to be closer to solar than their actual measured value. This bias is easy to understand. Because the *Kepler* field naturally has more stars closer to solar metallicity, the gradient in the number of stars as a function of metallicity ($dn/dZ$) is always increasing toward solar. Thus, the average number of stars contaminating a given metallicity bin will always be higher in the direction of solar. Thus, while globally there appears not to be a bias, there is a tendency for the SVR to infer higher metallicities for stars with subsolar metallicities in LAMOST, and lower metallicities for stars with supersolar metallicities in LAMOST. In other words, assuming a Gaussian error for each measured metallicity, there are always more contaminating stars from metallicity bins closer to solar than away from solar, artificially bringing the mean metallicity of stars within a particular bin of observables closer to solar. However, this is not a worrisome bias. In fact, in some sense this is analogous to adopting a thin disk prior on top of the photometrically-derived metallicities, which still represents a drastic improvement from no measured metallicities.

We re-train the SVR on the set of 49,285 LAMOST stars originally split into test and training samples, and assume that the resulting bias and uncertainties are the same as when using only the training sample. We then use the SVR to derive photometric metallicities for 106,322 stars. For stars with a metallicity measured by LAMOST, we adopt whichever of the photometric metallicity or the spectroscopic metallicity has a smaller uncertainty, resulting in 55,148 stars with a spectroscopic metallicity. Many of these stars did not make the spectrum $S/N$ cut to be considered in the training

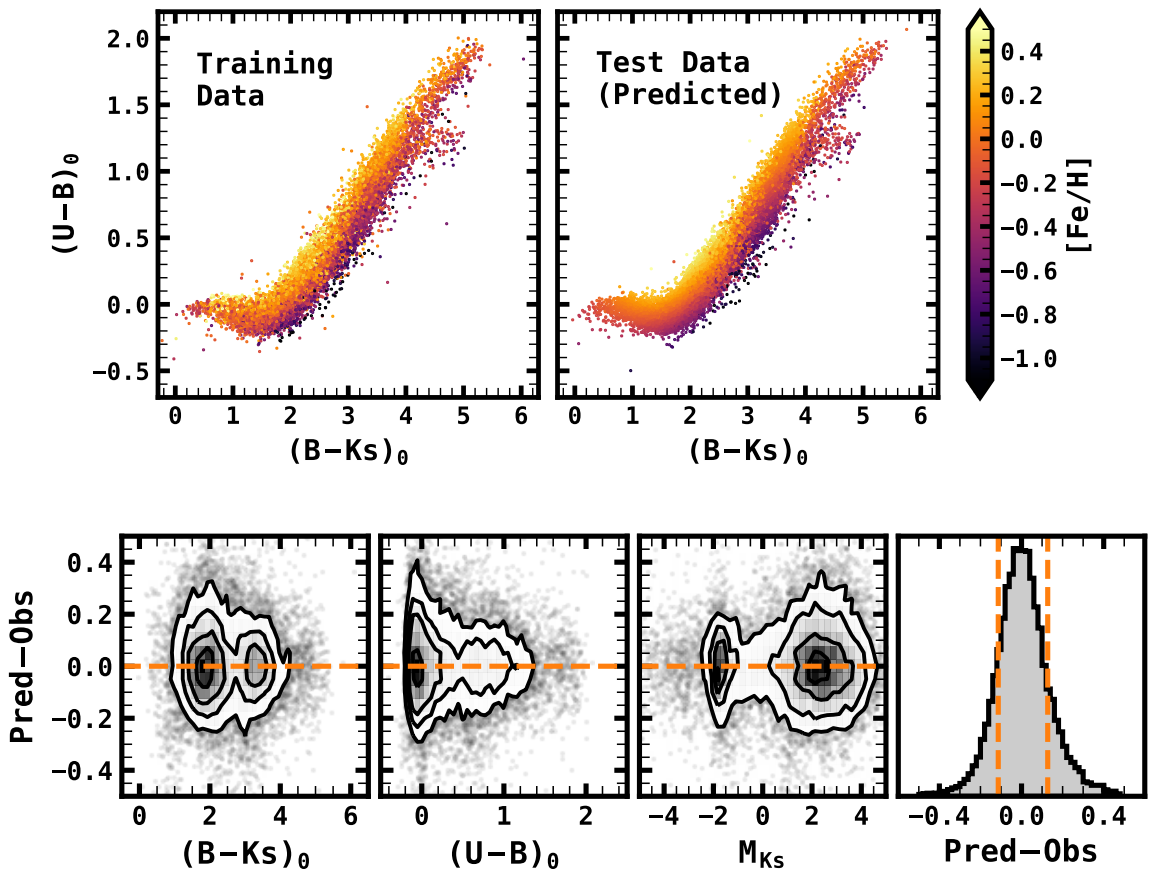


Fig. 4.2.— The results of the Support Vector Regression analysis used to infer photometric metallicities for stars in the CaTShIP. *Top Row:* The Color-Color diagram of the training (left) and test (right) samples. The color of the training data represents the metallicities as measured by LAMOST, while the color of the test data represents the metallicities predicted by the Support Vector Regression. *Bottom Row:* The residuals between the predicted (Pred) metallicities and the metallicities observed by LAMOST (Obs) in the test sample, as a function of the three inputs to the Support Vector Regressor. There are no trends with any of the input parameters. The far right panel shows the distribution of residuals from the test data, where the dashed orange lines denote the 16th and 84th percentile.

sample, but still had a more precise metallicity than what we would have derived from photometry. The remaining 2,741 stars are outside the color/temperature range needed to derive reliable photometric metallicities, so we adopt a thin disk prior of solar with a spread of 0.2 dex. This sample of stars is dominated by M dwarfs and stars at the tip of the RGB. We include the photometric metallicity for every star in our input catalog,

but note which metallicity was adopted for deriving stellar properties in the `FeH_source` column.

### 4.2.3 Fit to Stellar Evolutionary Tracks

To infer fundamental stellar parameters for the stars in our sample (e.g., $R_\star$, $M_\star$, age) we apply the python package `isofit`[1] (Wilson et al., *Submitted*). `isofit` compares observations to a grid of MESA Isochrones and Stellar Tracks (MIST) models (Dotter 2016; Choi et al. 2016) with masses ranging from 0.1 to 8.0 $M_\odot$, metallicities ranging from $-2$ to 0.5 dex, and evolutionary states ranging from the Zero-Age Main Sequence (ZAMS) to the beginning of the White Dwarf Cooling track. After finding an initial best model, a Markov Chain Monte Carlo (MCMC) analysis is applied to estimate the credible ranges for each parameter.

For each star in our sample, we run `isofit` with the following observable quantities and associated uncertainties: $\pi$, $Ks$, $E(B-V)$, $T_{\rm eff}$, and [Fe/H]. We instantiate the MCMC sampling using 50 walkers, with 1000 steps and 500 burn-in steps. We report the stellar parameters as the median for each parameter in the posterior distribution and the upper and lower limits as the 84th and 16th percentile of the posterior, respectively. In all, we derive stellar parameters for 163,164 stars.

### 4.2.4 The CaTShIP HR Diagram

In this section we discuss the results of our evolutionary track modeling and the features that are naturally present as a result of stellar evolution. In many places we also compare our results to those of Berger et al. (2020b), "B20" hereafter. Most of the differences between our methodology and that of B20 can be traced back to one of two sources, our choice to include uniformly-derived photometric (and spectroscopic) metallicities, or the set of basis vectors used to define the model grids in our respective fitting routines. In

[1]Available at https://github.com/robertfwilson/isofit

some cases, we also benefit from the use of more precise parallaxes from *Gaia* EDR3, though in most cases the error budget is dominated by uncertainties in $T_{\rm eff}$ and $Ks$-band magnitude.

The model grid from B20 is defined in three dimensions – age, mass, and metallicity – while our grid is defined by Equivalent Evolutionary Phase (EEP), mass, and metallicity. The EEP parameter is a uniform basis that is designed to adequately represent each phase in stellar evolution by the same number of points and to have a representative point across models of different masses that represent similar phases in stellar evolution. For example, EEP=202 corresponds to the ZAMS for stars of all masses. Thus, a set of EEP, mass, and metallicity corresponds to one unique model, as does the set of age, mass, and metallicity. The motivation behind interpolating a grid in EEP, rather than age, is to ensure that all phases of stellar evolution are adequately represented and shortly-lived phases are not improperly ignored. That being said, the model grid used by B20 was spaced finely enough to avoid such problems. In this case, the primary difference arises mainly from the difference in timescale between each grid point. The methodology of B20 of interpolating the grid in age acts similarly to adopting a linearly-sampled prior on age, while for us adopting the EEP as our basis results in posterior distributions more analogous to sampling logarithmically in age.

The resulting HR diagram from our modeling is shown in Figure 4.3, along with the HR diagram for the same stars analyzed by B20 for comparison. For these figures we've removed all stars with an inferred age >14 Gyr, as that is not possible given the age of the universe. In turn, this also removes most of the poorly-fit stars in our sample. The resulting catalog has 143,392 stars. Comparing this catalog to the corresponding stars from B20 shows few differences.

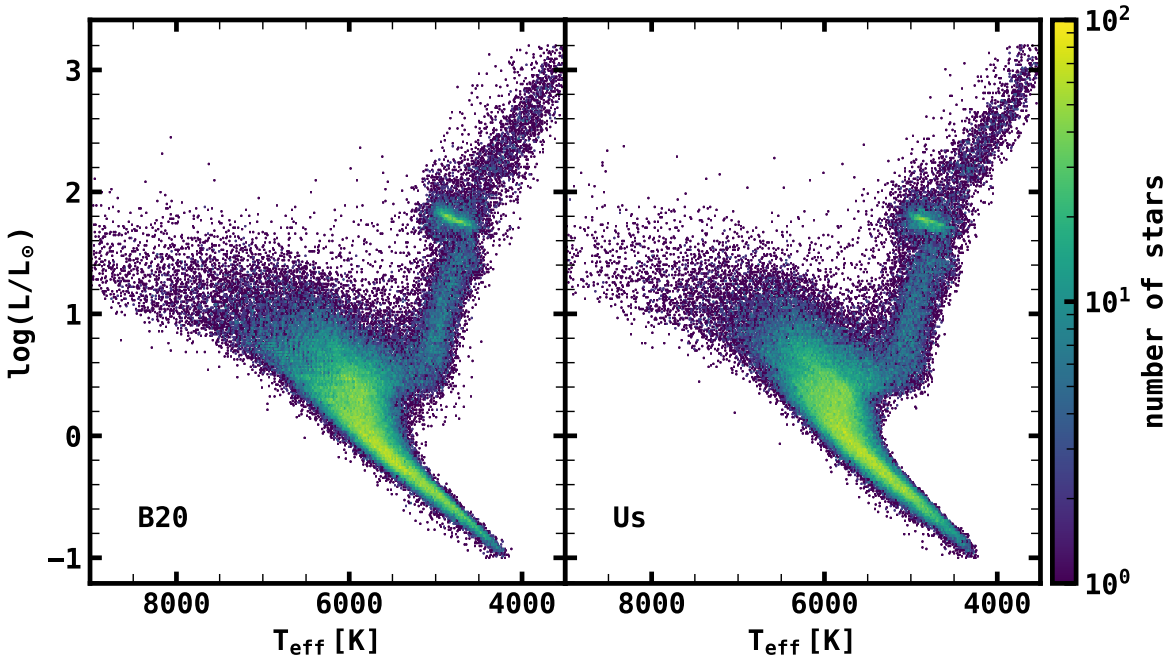


Fig. 4.3.— The temperature-luminosity distribution of ∼150,000 stars in our sample (right) and the corresponding stars from B20 (left). Stars with inferred ages >14 Gyr in our analysis are not included.

### Main Sequence

One difference that is noticeable between the two HR diagrams in Figure 4.3 is a broader main sequence in the `CaTShIP`. The reason for the increased scatter in our sample is the use of metallicities. Lower metallicity stars on the main sequence have a lower luminosity than higher metallicity main sequence stars of the same $T_{\rm eff}$. These metallicity differences lead to luminosity differences as large as 50-100% for a star with Solar effective temperatures and a metallicity difference near 0.1 dex. Thus, the scatter in the luminosity for stars on the main sequence reflects the metallicity distribution function of the `CaTShIP`. We expand upon this difference in the next section.

### Red Giant Branch

Another feature that is prominent between the two samples is the Red Clump (RC), which is the overdensity of stars at $\log L/L_{\odot} \approx 1.8$ and $T_{\rm eff} \approx 4500$ K caused by stable

core He burning (See Figure 4.4). The RC in our sample has slightly less scatter in luminosity than that in B20, though there is very little difference overall in the $T_{\rm eff}$ or luminosity distributions. Just below the RC is a slight over density at $T_{\rm eff}$ ≈4600 K and $\log L/L_{\odot} \approx 1.4$. This is the RGB bump. This slight overdensity results because a star's ascent up the RGB is briefly stalled as the H-shell burning passes through the convective H inhomogeneity envelope (Thomas 1967; Iben 1968). There is another overdensity above the RC at $T_{\rm eff} \approx 4400$ K and $\log L/L_{\odot} \approx 2.2$. This slight overdensity is the Asymptotic Giant Branch (AGB) bump. This results from an analogous pause in evolution to the RGB bump.

### 4.2.5 The CaTShIP Mass-Luminosity Diagram

In addition to the HR diagrams between the B20 sample and the CaTShIP, we also display a Mass-Luminosity diagram from the two samples in Figure 4.5. In some sense, this may be easier to interpret than the HR diagram, as stars evolve vertically in this diagram.

There are features of the stellar evolution models present in the CaTShIP sample, though there are a few systematics present in this figure as well. For instance, toward the tip of the RGB, there appears to be another point at $\log L/L_{\odot} \approx 2.6$ above which the mass distribution function increases sharply. This is a systematic effect in the CaTShIP. The most luminous giants in our sample do not have well measured metallicities because they are too red to be adequately inferred by our photometric metallicity relation, and so we assume a thin disk-like metallicity prior, which though it's a default for nearby stars may not be a reasonable representation of high luminosity stars.

#### Main Sequence

This difference in the main sequence is much more apparent between the two samples in Figure 4.5. In the B20 sample there are three distinct lines of stellar overdensities that represent distinct phases in stellar evolution. The line lowest in $\log L/L_{\odot}$ is the

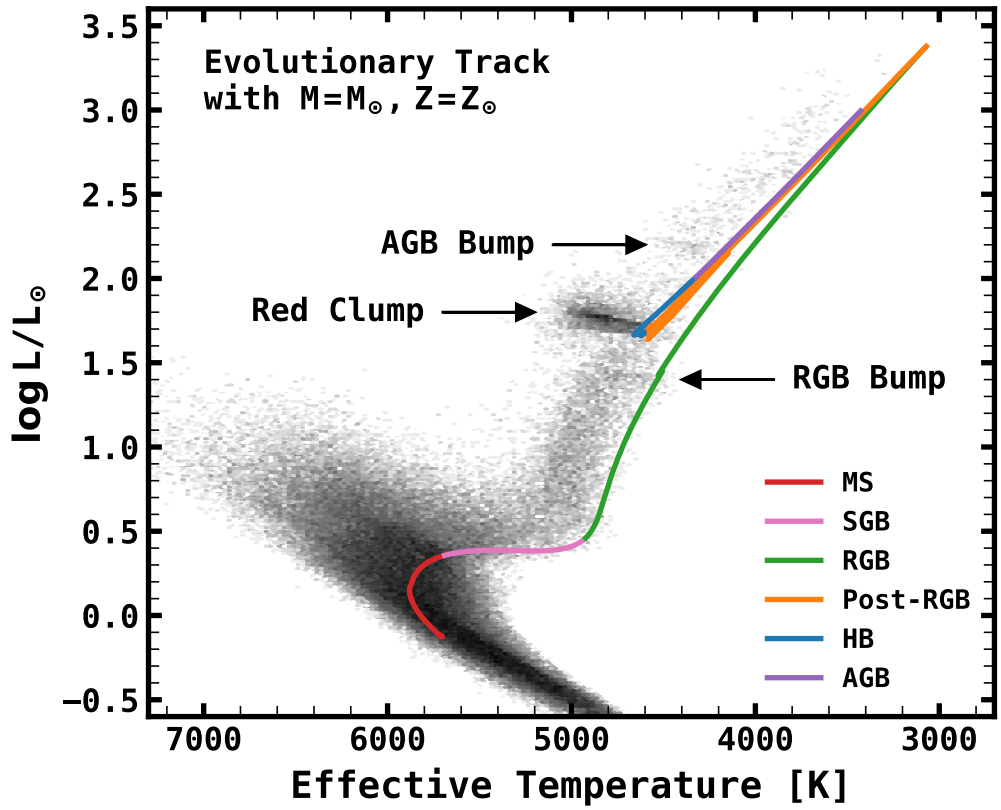


Fig. 4.4.— The Red Giant Branch as seen by the `CaTShIP`. The solid line denotes the evolutionary track undergone by a solar metallicity solar mass star from the Zero-Age Main Sequence to the end of the Asymptotic Giant Branch (AGB), with the color of the line denoting the evolutionary states, (SGB: Subgiant Branch, RGB: Red Giant Branch, HB: Horizontal Branch) . The density of points is shown in grey-scale with darker shades representing a higher density. The RGB and AGB bumps are denoted by arrows, as is the Red Clump, also referred to as the Horizontal Branch.

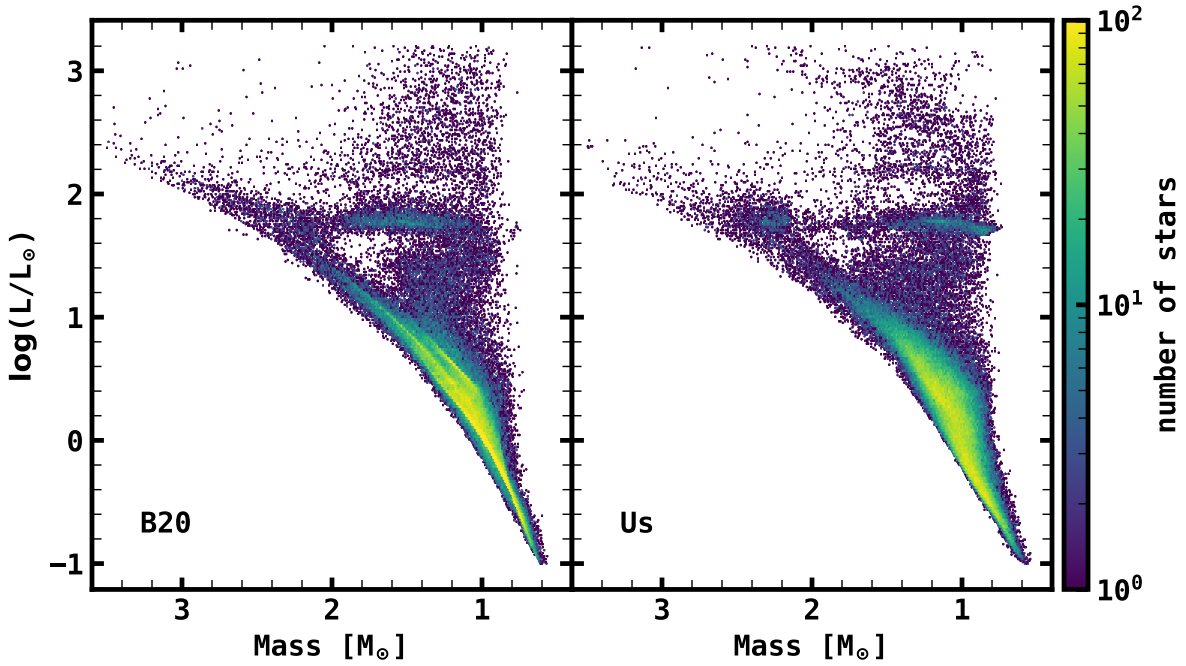


Fig. 4.5.— The mass-luminosity diagram for the stars in B20 (left) and the same stars in the `CaTShIP` (right). These samples are identical to those shown in Figure 4.3. The color represents the logarithmic density of stars at a given mass and luminosity.

main sequence, the line ranging from a $M \sim 1.2M_{\odot}$, and $\log L/L_{\odot} \sim 0.4$ to $M \sim 2M_{\odot}$, and $\log L/L_{\odot} \sim 1.5$, is the terminal age main sequence (TAMS), and the line just above that is the subgiant branch. These three stages in evolution appear strikingly in the B20 sample. However, they are not clear in our sample. This may be due to either differences in luminosity caused by our inclusion of a homogeneous metallicity scale, or may be due to larger mass uncertainties. Inspecting the mass-luminosity diagram in detail for differing metallicities, it becomes clear that our sample does in fact display the same overdensities shown by the B20 sample, but only when controlling for metallicity differences (see Figure 4.6).

In fact, when separating the sample within metallicity bins with widths of 0.1 dex, we are able to identify not just the three overdensities apparent in the B20 sample, but a fourth overdensity where the TAMS splits into two distinct overdensities at $M \gtrsim 1.2M_{\odot}$. The overdensity at lower luminosity is the Intermediate age main sequence (IAMS), while

the overdensity at higher luminosity is the TAMS. These two overdensities are present due to the "convective hook", which occurs in higher mass stars that have a convective core near the end of their main sequence lifetime. The B20 sample was not able to distinguish between these two phases of stellar evolution. However, it's worth noting that the fact that these overdensities appear in our sample is not evidence for their existence, as the model which includes these physical processes was used to infer the luminosity of the stars in our sample. In other words, while we can precisely determine the evolutionary phase of high-mass ($\gtrsim 1.2 M_{\odot}$) stars on the main sequence, this precision is model-dependent, and therefore cannot be used to validate our choice of models.

#### Red Clump and Secondary Clump

Other than the Main Sequence, another major differences we see between the sample in B20 and the `CaTShIP` are the masses of the stars located in the RC and the secondary clump. Compared to B20, we find a larger fraction of stars at the secondary clump. Such a difference may be due to the accuracy afforded to us by metallicities. On the giant branch, the stellar temperature is a strong function of both stellar mass and metallicity, where more massive stars and lower metallicity stars tend to be hotter. This is a problem when determining ages, because with photometry alone it is often not possible to break the degeneracy between mass and age in this area of parameter space. This is true on most parts of the RGB as well. However, with the photometric metallicities used to construct the `CaTShIP`, we can partially break this degeneracy and determine accurate masses, and as a result, ages. However, some degeneracy still remains.

### 4.2.6 Precision and Accuracy of Stellar Parameters

We take the larger of the absolute value between the median and upper/lower limits (i.e., the range of the inner 68th percentile) to be a reliable metric for the precision of the stellar properties inferred in our sample. The uncertainties for a few of the parameters

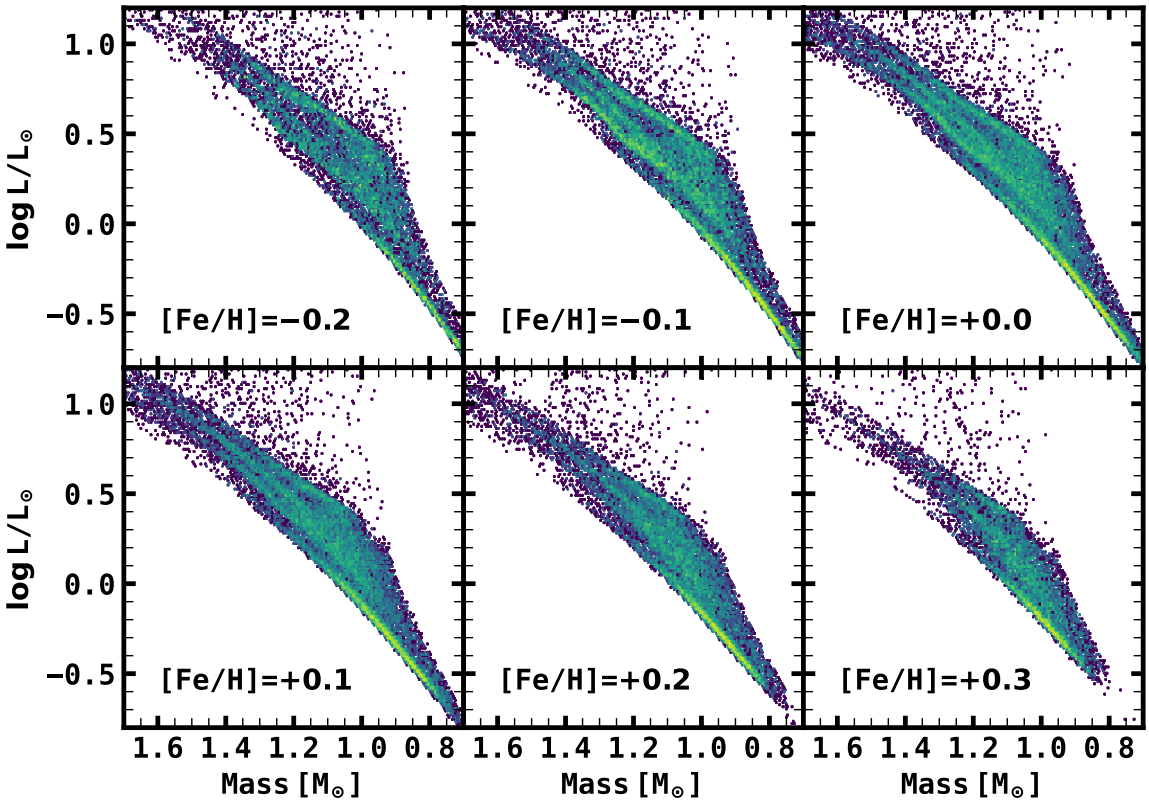


Fig. 4.6.— The Mass-Luminosity diagram for the `CaTShIP` sample split into metallicity bins with widths of 0.1 dex. The center metallicity is shown in the bottom left of each panel. The colors show the density of points, with yellow/green representing more stars and purple representing fewer stars. The over-densities in each panel suggest that our methodology is able to distinguish multiple phases of stellar evolution on the main sequence.

are displayed in Figure 4.7. Also shown in Figure 4.7 are the distributions for mass, radius, and age of the `CaTShIP` sample.

Overall we find generally the same distributions as those in B20, but we tend to have better precision. The radius distribution for our sample is nearly identical with a broad distribution at ≈1-1.5 $R_{\odot}$, with a decrease toward larger radii with the exception of a peak at ≈11$R_{\odot}$. This peak is due to the RC, which we've noted above. The location of this peak is slightly lower than that from B20, likely due to the $\lesssim$1$M_{\odot}$ mass stars near the RC which congregate at lower luminosity and radius.

We find that our distribution in mass has less variance than the B20 mass distribution,

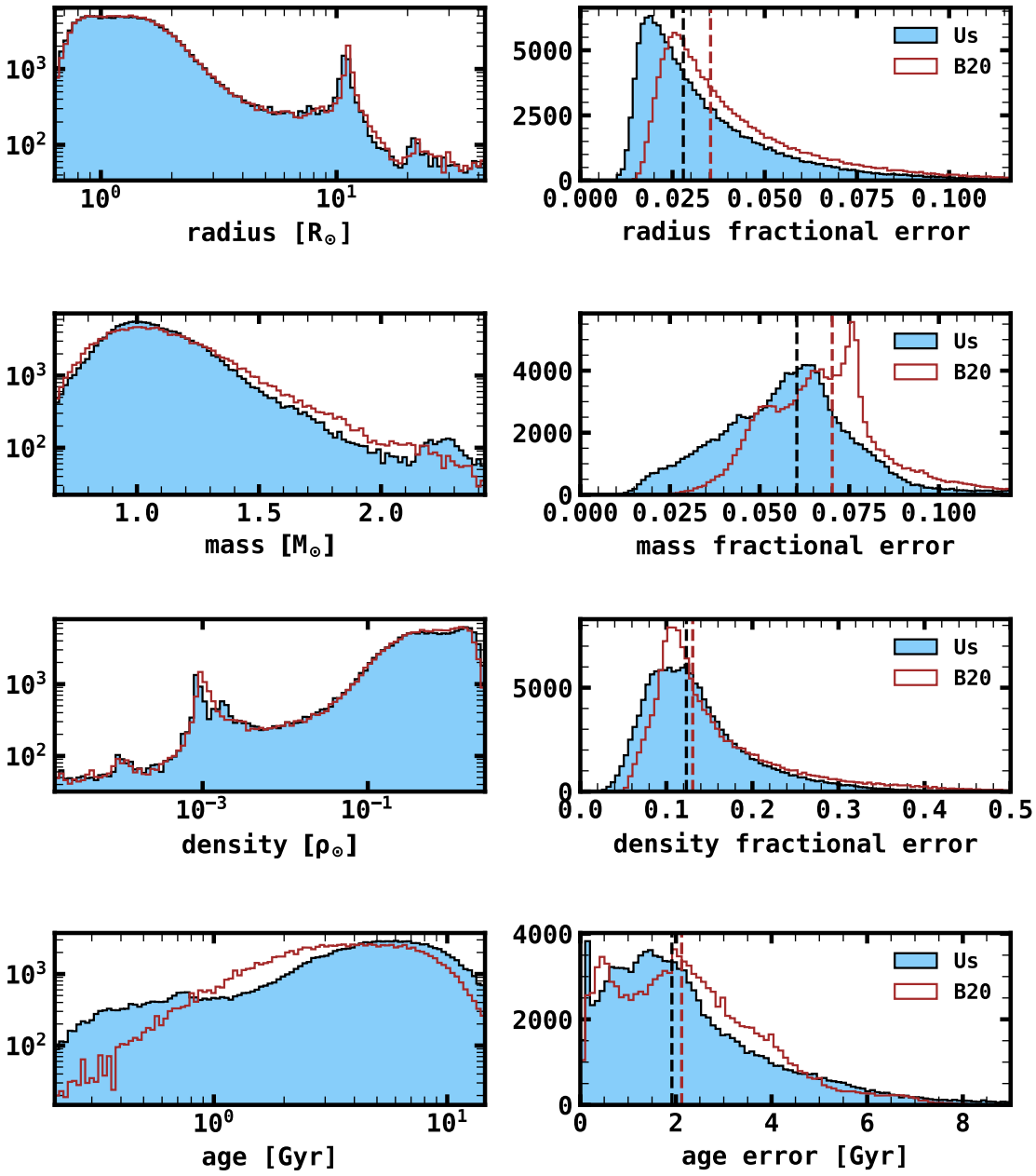


Fig. 4.7.— The distribution of stellar parameters for the stars in our sample and their associated distribution of uncertainties. The blue histogram shows the stellar parameters and uncertainties from this work, while the red histogram is that of B20 for the same sample of stars, and is shown for comparison. The blacked dashed line denotes the median in our sample, and the red dashed line denotes the median from the B20 sample. For each of these parameters, we generally find that our sample has more or less the same distribution as in B20, but with better precision.

but peaks in the same place as in B20, at $\approx 1 M_\odot$, with a nearly identical median of $1.33 M_\odot$ versus $1.34 M_\odot$ for B20. The apparently lower width of our distribution compared to B20 can likely be attributed to the higher precision masses inferred from our data. This is most apparent in the mass uncertainty distributions. Our sample has a median uncertainty of 6%, while that from B20 has a median uncertainty of 7%. These differences do not seem particularly large, though there are some notable features in the B20 uncertainty distributions that are not present in our sample, such as the large peak at an uncertainty of ≈7.5%. In the B20 sample that peak corresponds to main sequence stars without a measured metallicity. Thus, we've mitigated this peak in our own sample with our use of photometric metallicities. The `CaTShIP` also has slightly better mass estimates for stars on the RGB due to our better metallicities, but these still remain the most uncertain parameters in our sample because of degeneracies with mass. This is particularly true for cool RGB stars.

The density distributions are nearly identical to that of B20, and while our sample has a slightly better precision on average, the fractional uncertainty only decreases from ≈13% to ≈12% from B20 to the `CaTShIP`. The stellar mean density is one of the more important parameters for our science goals, as it is used to determine the grid of transit durations used in the `TraSH-DUMP` search. Because transits can have a shorter duration if transiting at higher impact parameters or due to eccentric orbits, the motivation for the star's mean density to be precise is not as strong as the motivation for the density to be accurate.

Finally, the ages in the `CaTShIP` are slightly more precise than for B20. We have a median uncertainty of 1.9 Gyr, while the B20 sample has a median uncertainty of 2.1 Gyr. These uncertainties are for all the stars with informative ages, so we are not including low mass stars in this statement. One particularly interesting feature in our histogram is the peak at ≈100-200 Myr. This feature is in fact real, and is not a systematic uncertainty. This peak corresponds to high-mass ($\gtrsim 2 M_\odot$) stars that have not yet evolved onto the

giant branch. Because such stars do not have long lifetimes, they can be age-dated very precisely via isochrone- and evolutionary track-modelling, such as this. However, the lifetime of such stars depends steeply on the mass, and their positions on the HR diagram are degenerate with metallicity. We are able to break this degeneracy with our photometric metallicities and in a few cases infer an age with a precision of ∼10-100 Myr. However, one caveat to this is that the ages and masses are only accurate if the metallicities from the training sample (i.e., LAMOST spectra sample) are accurate, which may be questionable for hotter ($\gtrsim$ 7000K) stars due to fewer overall metal lines in the spectra of such stars, but should be accurate for high-mass stars in the cooler subgiant phase or near the Hertzsprung gap.

### Comparison to Asteroseismology

To determine the accuracy, rather than just the precision, of our stellar parameters, we cross-match the `CaTShIP` sample with the APOKASC dwarf and subgiant catalog (Serenelli et al. 2017). The APOKASC catalog is an effort that combines spectroscopic temperatures and metallicities from the Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2015; Majewski et al. 2017) with global seismic parameters, $\Delta\nu$ and $\nu_{\rm max}$, measured from *Kepler* light curves (Pinsonneault et al. 2018). The global seismic parameters set extremely strict constraints on the stellar density and surface gravity through the asteroseismic scaling relations, (Ulrich 1986; Brown 1991; Kjeldsen & Bedding 1995), which Serenelli et al. (2017) incorporate directly into a grid of stellar evolutionary tracks via the same Grid-Based modelling approach upon which `isofit` is based. The APOKASC catalog has typical statistical uncertainties in mass and radius of ∼4.7% and ∼2.7%, respectively. The `CaTShIP` contains 212 stars in common with the APOKASC dwarf star catalog. The other 213 stars from the APOKASC dwarf star catalog were too bright to be included in this analysis, due to saturation in the Everett et al. (2012) photometry. The comparisons for our catalog with that of the

APOKASC catalog for radius, mass, density, and age are shown in Figure 4.8.

There are a few notable differences between the input parameters in our sample versus those used in the APOKASC catalog. For instance, we find a systematic offset of 0.06 dex in [Fe/H], in that the `CaTShIP`'s input parameters are slightly more metal rich, with a scatter of 0.09 dex, typical of our uncertainties. This difference is likely to generate differences in the inferred ages and densities of the APOKASC catalog and the `CaTShIP`.

Comparing our catalog to the APOKASC catalog, we find good agreement in the inferred stellar parameters within the combined uncertainties of our sample and the APOKASC sample. For radii, perhaps our most precisely inferred quantity, we find an agreement within $\sim$3%, with a systematic offset of $\sim$2%. In general, the `CaTShIP` tends toward better precision for stars at larger distances. This is likely due to the $S/N$ of the oscillation modes in the *Kepler* light curves decreasing for dimmer stars. In addition to the combined uncertainties for the radii of stars in the `CaTShIP` and the APOKASC catalogs, there is a known systematic offset between radii inferred from asteroseismology and from *Gaia* parallaxes of $\sim$2% (Zinn et al. 2019a). This explains the median offset between our radii and those of the APOKASC in Figure 4.8.

For mass, we again find good agreement between that of our sample and the APOKASC sample. In general, we tend to infer slightly smaller masses by $\sim$3%, but the overall scatter between our sample and the APOKASC catalog is mostly within our combined uncertainties of $\sim$4-6%. The scatter is slightly larger overall, which could be attributed to differences in our adopted temperature scale or metallicity scale. Because we trained our metallicities on LAMOST data, and APOKASC adopted their metallicities from APOGEE (Majewski et al. 2015), there are likely to be some systematic differences.

The largest differences between our sample and that of APOKASC are in the densities. We tend to infer densities that are $\sim$10% smaller than those of APOKASC. This offset is close to the typical scatter of the combined uncertainties between the `CaTShIP` dwarfs and subgiants ($\approx$7%) and the APOKASC catalog ($\sim$ 4%), of $\approx$8%. However, the offset

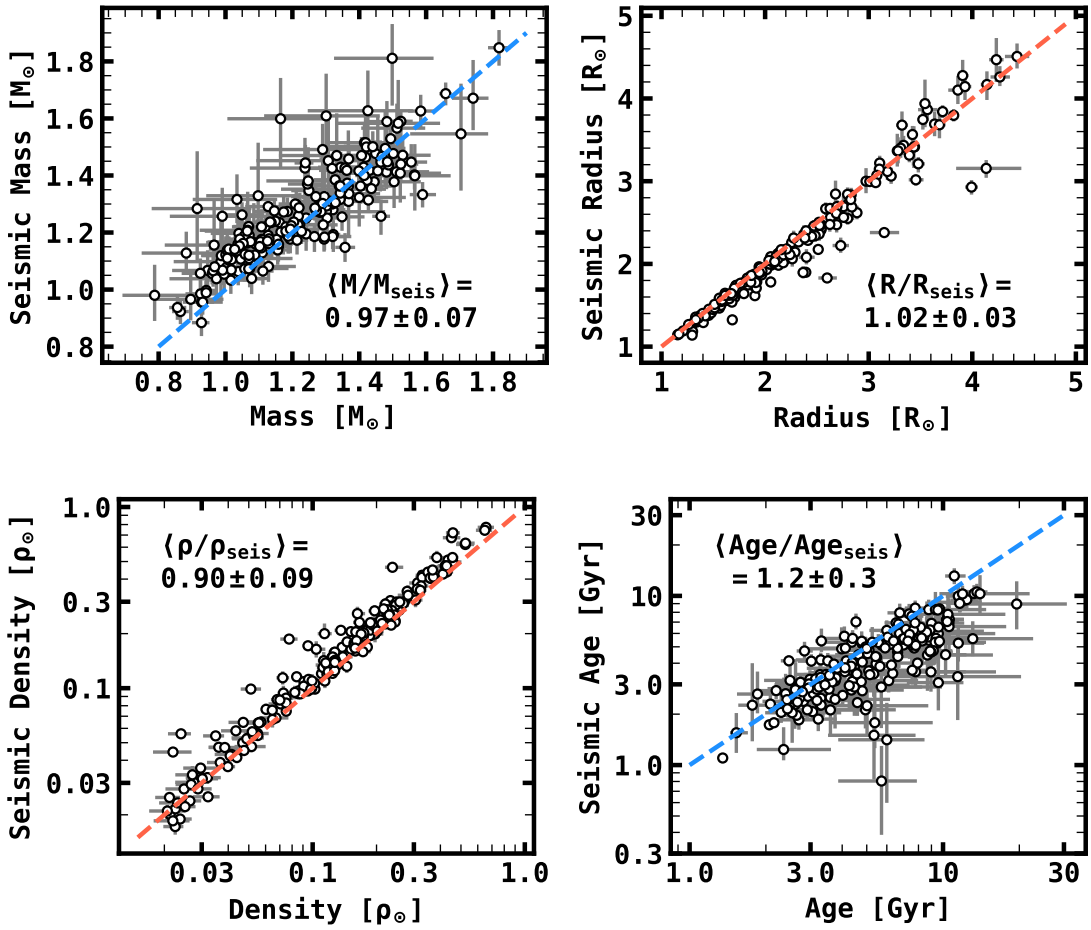


Fig. 4.8.— Comparison of the derived properties in the `CaTShIP` to 212 dwarfs and subgiants with parameters derived from asteroseismology. For each panel, the dashed line indicates unity, and the median and median absolute deviation are displayed. In most cases, the differences between our values and those in Serenelli et al. (2017) agree within the combined uncertainties.

is still somewhat significant, and likely derives mostly from the strong dependence on the stellar radius, which is known to have systematic uncertainties. The densities tend to agree better for older stars (e.g., subgiants and stars near the TAMS), and disagree more for younger stars. This is expected, as isochrones have weaker age and mass constraints for stars on the main sequence as compared to slightly evolved stars.

Finally, comparing the ages between these two samples reveals a large scatter, which is expected due to the difficulty in inferring ages. As mentioned, the type of modelling

done here has trouble distinguishing ages for stars on the main sequence, as there are only slight changes in luminosity and temperature for a given mass with differing ages. Therefore, on average, our modelling may give estimates for age that are overestimated for stars on the main sequence. We find that effect here, as we tend to estimate ages that are ∼20% greater than those inferred using global asteroseismic parameters, where one is able to adopt stricter density constraints. The difference in metallicity between the two samples is also likely to play a role in overestimating ages in the `CaTShIP` relative to the APOKASC. However, the ages between our sample are mostly in agreement within the combined uncertainties of the `CaTShIP` and the APOKASC catalog of ∼20%. The offset is likely due to the larger radii and lower density estimates for our sample, on average.

In all, we find a few systematic differences between our sample and the APOKASC catalog. Most of the differences can be traced back to systematic differences either in the radii/luminosity inferred from geometric parallaxes versus asteroseismic parameters, or differences in the metallicities in the `CaTShIP` and the APOKASC. However, overall the inferred parameters from these two catalogs agree relatively well with only a few outliers.

### Properties for Stars in NGC 6811

As another test of the accuracy in our sample, we compare ages derived from stars likely to be members of the open cluster NGC 6811. NGC 6811 has a canonical age of $1.0 \pm 0.1$ Gyr and metallicity estimates ranging from -0.05 to 0.05 dex (Frinchaboy et al. 2013; Molenda-Żakowicz et al. 2013; Sandquist et al. 2016; Bossini et al. 2019). The metallicity range of this cluster makes it a prime candidate to test our age estimates, given that they are near to solar metallicity, as are the majority of the stars in the `CaTShIP`. To select probable members of NGC 6811, we first query the `CaTShIP` for stars within half of a degree from the center of the cluster given in Cantat-Gaudin et al. (2018). From this sample, we then remove stars that disagree from the mean proper motion of the cluster in either right ascension or declination by more than $2\sigma$. After these cuts we are left

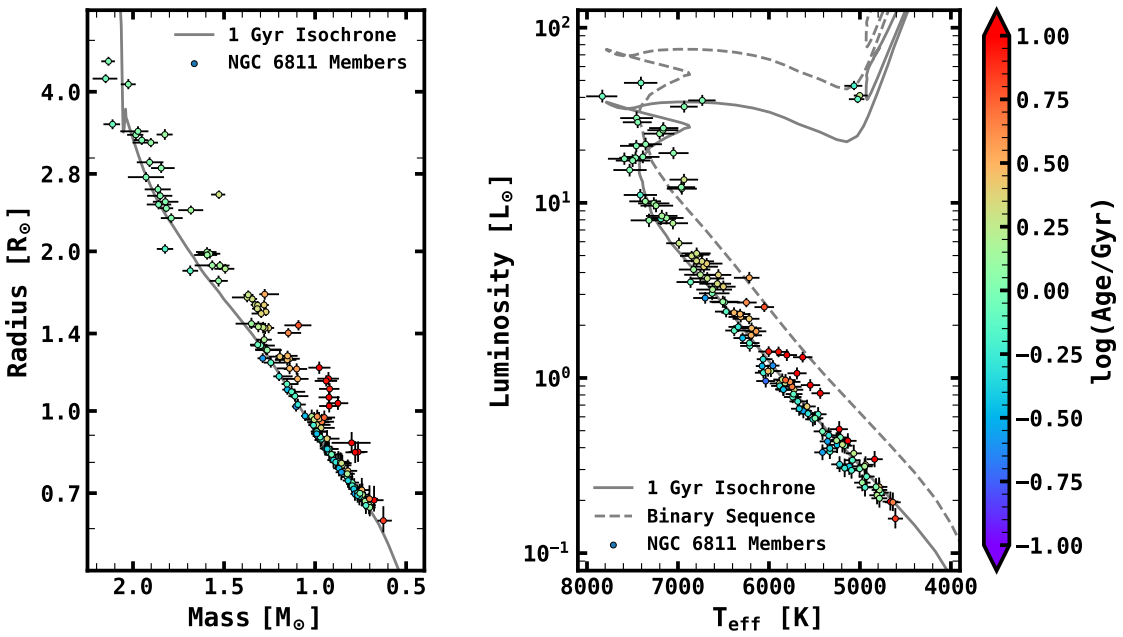


Fig. 4.9.— Comparisons with our derived ages and the canonical age, 1 Gyr, for stars in the open cluster NGC 6811. In each figure, members of the open cluster NGC 6811 are colored by our inferred age, and plotted against a 1 Gyr isochrone with [Fe/H]=0.05. *Left*: The derived mass and radius of each star in the cluster. The gray line shows the expected mass and radius from a 1 Gyr isochrone. *Right*: A Hertzsprung-Russel diagram of the NGC 6811 cluster members in our sample. Note the relatively good age agreement at higher masses and disagreement for stars with higher and lower derived luminosity than the displayed isochrone. In general, the stars represented by green points have ages in relative agreement with the canonical age of the cluster.

with 116 stars that are likely to be members of NGC 6811 (see Figure 4.9).

From these stars, we find that our metallicity estimates tend to agree with the properties of the cluster in the literature. For these stars, we inferred a median metallicity of -0.05 dex and a median absolute deviation of 0.09 dex, well within the expectations for this cluster. The derived parameters tend to agree with what is expected from isochrones as well. We show this agreement in Figure 4.9. The right panel shows the H-R diagram of the probable cluster members, and the left panel shows the mass and radii of the probable cluster members. The points are colored by the logarithm of their derived ages in the `CaTShIP`, where green represents an age of 1 Gyr, and therefore good agreement with the literature ages. In each of these figures, the gray line denotes the expected values for

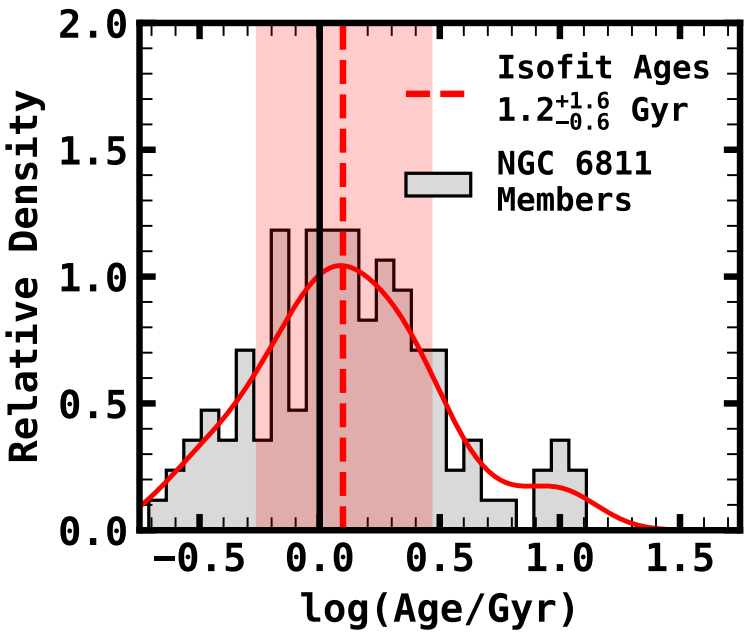


Fig. 4.10.— The distribution of stellar ages measured for stars in NGC 6811. The gray histogram shows the individual stars in the sample, the solid red line shows the kernel density estimate, the dashed red line shows the median of the inferred ages, and the shaded region shows the inner 68% of the inferred ages. The solid black line shows the canonical age of NGC 6811 of ≈1 Gyr.

an isochrone of 1 Gyr and solar metallicity.

The stars with the biggest deviations in age also have the largest deviations from the expected luminosity and radius given the sample. The reason for these disagreements is likely due to contamination from unresolved binary stars. Although we use $Ks$-band photometry that was corrected for contamination from secondary stars, many are still likely not to have been noticed, and as a result artificially increase the apparent brightness of the stars. In our case, the increase in apparent brightness leads to a larger inferred radius and luminosity, which results in a larger inferred age. This represents the primary impact that unresolved binary stars have on the derived parameters for stars in our sample. They generally lead to significantly overestimated ages.

In addition to the stars with larger discrepancies in age in the `CaTShIP`, there are a few stars with increased differences in mass and radius, particularly at ∼1.5-2 solar

masses. These stars are likely different due to our metallicities being less reliable at hotter temperatures and bluer colors. This is mostly due to the fewer number of stars in the LAMOST training sample, but it may also be impacted by the LAMOST spectra yielding less precise or accurate metallicities at these temperatures. Though we note that while there are a few differences in mass, radius, and luminosity, the ages for these stars are still in good agreement, though that may be in part due to the short MS lifetimes of such stars.

Despite some of the differences between our inferred values and the expected values from the literature for stars in NGC 6811, we find that the majority of stars in this sample in fact do have reliable ages. The distribution of ages mostly agree with the canonical age of the cluster, and has a median and spread of $1.2^{+1.6}_{-0.6}$ Gyr (see Figure 4.10). Therefore, despite the outliers at inferred ages of $\sim$10 Gyr caused by unresolved binaries, we are still able to trust the ages of our stars approximately within a factor of 2, which is roughly consistent with our estimated uncertainties of $\sim$50% for stars in the NGC6811 sample. We also notice a similar trend in age compared to our comparison with ages in the APOKASC catalog, that we infer ages that are relatively older by $\approx$20%. Also, as seen in Figure 4.9, and as expected, the ages tend to be most accurate for more massive stars, especially those with $M \gtrsim 1.5 M_{\odot}$.

## 4.3 TraSH-DUMP: An Unbiased Transit Detection Pipeline

Now that we have generated a catalog of reliable stellar parameters, we turn to the question of transit detection. The inspiration for this pipeline came from the large fraction of subgiants in the *Kepler* field ($\approx$21%) that were revealed by *Gaia* DR2 (Berger et al. 2018) and the difficulty in estimating occurrence rates for evolved stars using the *Kepler* pipeline due to the limited grid of transit durations searched, particularly at

long periods. To address that difficulty, we've written and implemented a detection algorithm that is informed by the stellar properties of the star being searched. In this way, we intend to avoid second-order detection biases that arise from differences in limb-darkening coefficients, and biases that arise from searching over an inadequate grid of transit templates.

While some of these biases can be mitigated with a careful approach to constructing a transit template, there are still difficulties that arise from detecting transiting planets orbiting evolved stars that are unavoidable, such as a significant increase in the radius ratio contrast between the putative planet and evolved star, and stellar variability on timescales similar to the durations being searched.

### 4.3.1 The Philosophy Behind TraSH-DUMP

The goal of our new code, TRAnsiting planets with Subgiant Hosts – Detection with an Unbiased Matched filter Pipeline (`TraSH-DUMP`), is to increase the detection sensitivity of a transit search by utilizing known stellar properties of the host star. At the heart of `TraSH-DUMP` is a wavelet-based adaptive matched filter, based on the algorithm in Jenkins (2002), but implemented with a customized set of transit templates for each light curve in the planet-search sample. For each star, limb-darkening coefficients are determined from the star's surface gravity and effective temperature, and the transit durations searched are determined from the star's density, allowing us to include longer duration transit templates in our search for planets orbiting low density stars (e.g., subgiants) and include short duration templates in our search for planets orbiting high density stars (e.g., M dwarfs).

In the process of customizing the transit templates used for each star, `TraSH-DUMP` both maximizes the sensitivity to low S/N transits and inherently reduces the bias present in most traditional transiting planet searches caused by template mismatch (Burke & Catanzarite 2017b). Template mismatch is a second order effect that decreases the

detection efficiency for transit shapes that are different from the template used for the search. In the *Kepler* detection pipeline, which is optimized for Sun-like stars, this effect results in a nearly 20% drop in detection efficiency for planets orbiting cool dwarfs, and is magnified in the presence of strong red noise (Burke & Catanzarite 2017b; Christiansen 2017; Christiansen et al. 2020). By correcting for this bias in the detection algorithm, the detection efficiency is boosted and the completeness corrections needed to calculate occurrence rates across multiple stellar populations from the `TraSH-DUMP` planet catalog is simplified.

### 4.3.2 The `TraSH-DUMP` Algorithm

In this section we walk through the `TraSH-DUMP` algorithm, and explain our implementation of the wavelet-based adaptive matched filter in detail. We begin this section with the signal detection theory upon which `TraSH-DUMP` is built, and then go into the details of our implementation, including the light curve pre-processing steps and identification of statistically significant signals. This process is highlighted in Figure 4.11

#### Theory Behind the Matched-Filter

The theory behind our approach is discussed in detail by Jenkins (2002), but we provide an abridged discussion here for the convenience of the reader.

In a given time series with equally spaced cadences, $x(n)$ (where $n$ denotes the index of the time sample), we wish to detect a transit signal, $s(n)$, in the presence of Gaussian white noise, $w(n)$. In this case, if one presumes that $x(n) = w(n) + s(n)$, a matched filter can be constructed of the form,

$$l = \frac{x^T R^{-1} s}{\sqrt{s^T R^{-1} s}} = \frac{(R^{-1/2} x)^T (R^{-1/2} s)}{\sqrt{(R^{-1/2} s)^T (R^{-1/2} s)}} \quad , \tag{4.1}$$

where $l$ is a detection metric, $R$ is the auto-correlation matrix of $w(n)$, and $T$ is the

transpose operator. In this case, $l$ has the property that when no signal is present it is a number randomly drawn from a Gaussian distribution with a mean of zero and variance of one. In the case of a signal, however, $l$ is randomly drawn from a Gaussian with a variance of one and a mean equal to $\sqrt{s^T R^{-1} s}$, which is the strength of the signal. However, $R$ is not generally known beforehand, so there must exist some way to estimate the noise autocorrelation matrix, $R$.

In the presence of stationary white noise, the detection metric can be calculated in the frequency domain, where the noise autocorrelation matrix is estimated via a power spectrum, $P(\omega)$, typically from a periodogram. In these cases, Equation 4.1 can be recast as,

$$l = \frac{\tilde{x} \cdot \tilde{s}}{|\tilde{s}|} \tag{4.2}$$

$$= \int_{-\pi}^{\pi} \frac{X(\omega) S^*(\omega)}{P(\omega)} d\omega \Bigg/ \sqrt{\int_{-\pi}^{\pi} \frac{S(\omega) S^*(\omega)}{P(\omega)} d\omega} \tag{4.3}$$

where $\tilde{x}$ and $\tilde{s}$ are the whitened time-series and signal, respectively, $X(\omega)$ and $S(\omega)$ are the Fourier transforms of the time-series and signal, respectively, and the asterisk denotes the complex conjugate. In practice, this is not a computationally efficient approach because this relationship implies the need for calculating a power spectrum at every single possible transit location in the time series.

Thus, the need for a wavelet transform is born from the requirement that the time-series must be whitened adaptively, i.e., that the detection algorithm takes into account changes in the amplitude of the noise with respect to time. A wavelet is a natural choice for this application as it is a representation of a waveform in both time and frequency, and is analogous to a Fourier transform that can be localized in time. Within this framework, an overcomplete discrete wavelet transform acts as a series of high-pass and low-pass filters, which measure the frequency response of a time series in $\log_2(N) = M$ individual channels, where $N$ is the number of data points, and each channel, indexed by

$i = 0, 1, 2, ..., M$, represents the response of the time-series, $x(n)$, to a specific filter. The filter for band $i$ is symmetric in $\log f$ and has a central frequency, $f_i = 2^{-(i+1/2)}$ times the Nyquist frequency, $f_{\rm nyquist}$, with a range of $2^{-(i+1)}$ to $2^{-i}$ times $f_{\rm nyquist}$, at each time, $n$. Thus, the wavelet transform results in $M$ individual time-series, which we denote as $\hat{x}_i(n), i = 0, 1, .., M$, with each band preserving the frequency structure of filter $i$.

Finally, applying the properties of the overcomplete wavelet transform and the matched filter from Equation 4.2, we measure the detection statistic at each time, $l(n)$, also known as the Single Event Statistic (SES), by doubly-whitening the time-series from each band-pass, and correlating the results with the overcomplete discrete wavelet transform of the transit signal, $\hat{s}_i(n)$,

$$\mathrm{SES}(n) = \frac{\mathbb{N}(n)}{\sqrt{\mathbb{D}(n)}} \tag{4.4}$$

$$= \frac{\sum_{i=1}^{M} 2^{-\min(i,M-1)} [\hat{x}_i(n)\sigma_i^{-2}(n)] * \hat{s}_i(n)}{\sqrt{\sum_{i=1}^{M} 2^{-\min(i,M-1)} \sigma_i^{-2}(n) * \hat{s}_i^2(n)}} \ , \tag{4.5}$$

where $\sigma_i(n)$ are the whitening coefficients in band $i$ for each point in time, $n$, and $*$ is a correlation operator. The whitening coefficients are estimated via a decimated median absolute deviation (MAD) filter with a window of $8\times$ the transit duration being searched.

Finally, we can take advantage of the periodic nature of the transit signal to combine the SES at multiple different locations into a multiple event statistic (MES), which we use as the detection metric for `TraSH-DUMP`. For a light curve with non-varying exposure time $t_{\rm exp}$, we test for transits at a given period, $P$, and time of mid-transit, $t_0 = n_0 + j \times t_{\rm exp}$, where $j = 0, 1, 2, ..., P/t_{\rm exp}$ and $n_0$ is the first point in the time-series, by binning the SES into the Multiple Event Statistic (MES) via

$$\mathrm{MES}(t_0, P) = \sum_{n \in A} \mathbb{N}(n) \Big/ \sqrt{\sum_{n \in A} \mathbb{D}(n)} \ , \tag{4.6}$$

where $A$ is the set of all cadences observed at times, $t$, that satisfy the condition $|t - (t_0 +$

$m \times P)| < t_{\rm exp}/2$ for some integer, $m$. In this way we define the MES as our detection statistic.

### Pre-Search Light Curve Processing

Before running `TraSH-DUMP` for a particular target, we remove bad data points flagged by the *Kepler* mission, and then the light curve from each sector is detrended with a time-windowed biweight kernel, implemented through the python package `wotan` (Hippke et al. 2019). The window for the biweight kernel is chosen to be at least 3× the maximum transit duration searched, typically ∼3-4 days. This approach was shown to be the most effective at preserving transit signals in the data while removing stellar variability. Each sector is detrended individually, with the edges of the data gaps removed. The length of the data removed near the edges are the data within half of the window used plus an additional 0.5 days if the remaining edge data deviate by $>2.5\sigma$ from the median of the detrended time-series.

After detrending, we apply an algorithm to remove outliers. The algorithm works by computing a running measure of the MAD and removes points outisde some threshold, typically defined as 4-5× the MAD at that location. Positive flux outliers have two deleterious results on planet detection. The first effect is that positive outliers may contribute to overestimating the local noise profile and as a result artificially reduce sensitivity to transiting events. This is not a huge concern in `TraSH-DUMP` due to our use of the MAD to estimate the whitening coefficients of the time series. The second effect is more consequential. The wavelet transform across a discontinuous outlier is likely to lead to "ringing" behavior, similar to the ring-down caused by discontinuities in a Fourier transform, which contaminates the power spectrum of the wavelet decomposition at lower frequencies and can lead to false alarms or non-Gaussian MES distributions. For removing negative flux outliers, a concern is removing data within a transit. To avoid this, we only remove points identified as a negative outlier if there are no adjacent

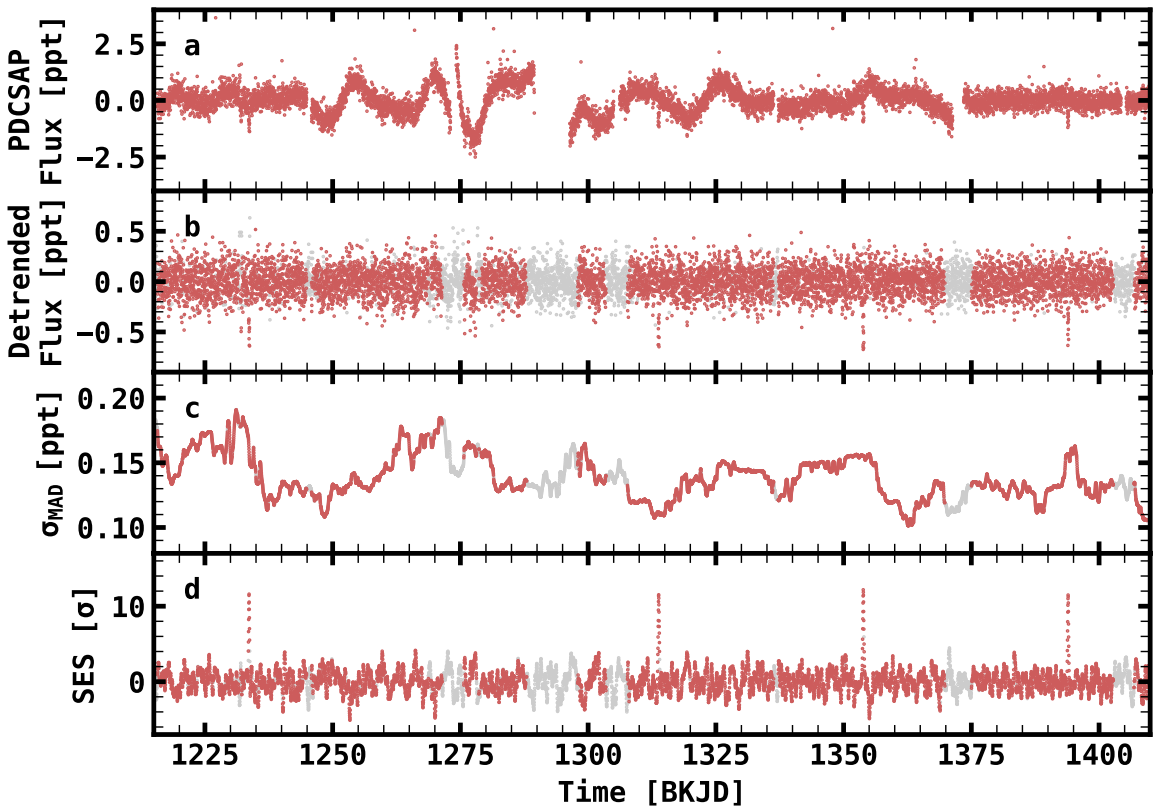


Fig. 4.11.— The processing stages before applying `TraSH-DUMP` to a *Kepler* light curve. Here we show a light curve of the confirmed planet-hosting star Kepler-874 (Morton et al. 2016) as a demonstration of the detection process. (a) The PDCSAP light curve of Kepler-874, centered at zero, and with flux in units of parts per thousand (ppt). (b) The detrended, gap-filled light curve of Kepler-874. Note the removal of positive flux outliers, and the artificial data that are added via our gap-filling algorithm (gray points). (c) The variance estimate at each time, taken as 1.4826× the running MAD. In this case the window used to calculate the running MAD is 9× the estimated transit duration of 0.25 days. (d) The single event statistic (SES) calculated at each cadence. Note the peaks at $t \approx 1234$, 1314, 1354, and 1394 BKJD, corresponding to transits from Kepler-874b, a confirmed Sub-Neptune ($R_p = 3.6 R_\oplus$) with an orbital period of $P \approx 40$ days.

outliers, and there are no other points within $\pm 3\sigma$ of the putative negative outlier.

After removing outliers, a gap-filling algorithm is applied to the detrended time-series. Long data gaps (>1 day) are filled using a linear combination of an autoregressive model both before and after the gap in data. This approach preserves the photometric noise profile on each side of the gap which helps to prevent False Alarms due to abrupt changes in photometric noise (caused by, e.g., transitioning from one CCD to another). Shorter

data gaps are filled by reflecting the flux before and after the gap (see Figure 4.11b).

The last adjustment made to the time-series before applying the matched filter is to pad the time-series with an equal number of points at the beginning and end of the light curve. The light curve is padded for two reasons. First, by padding the light curve so that the resulting number of data points is a multiple of two, we can improve our computational efficiency by applying the convolution in Equation 4.4 via a fast Fourier transform. Second, the padding is required to have an adequate estimate for the whitening coefficients. The whitening coefficients are estimated via a running MAD filter, and the padded ends of the time-series are used to prevent losses in efficiency (i.e., accuracy) in estimating the central mean of the window used. Without the padding, the MAD estimate suffers from few data points near the edges of the time-series and as a result can be poorly constrained. The padding is implemented via an autoregressive model similar to the model used for long data gaps. After the time series is detrended, the padded, gap-filled light curve is ready for the application of the matched filter.

### Calculation of Transit Templates and Application of the Matched Filter

After processing the *Kepler* light curve, a series of transit templates are calculated with which to perform the correlation. In this processing stage, we choose the grid of transit durations applied from the properties of the star being searched and the temporal baseline of the time-series. For our purposes, we require three transits (though the algorithm can be applied to search for systems with two-transits as well), and so the maximum period considered is half the temporal baseline, as three transits could theoretically occur at exactly the beginning, middle, and end of the time series. We then choose the maximum transit duration, $t_{\mathrm{dur,\,max}}$, searched via the expected duration for a circular transit with

an impact parameter, $b = 0$:

$$t_{\mathrm{dur,\,max}} = 0.54 \text{ days } \left(\frac{P_{\mathrm{max}}}{365 \text{ days}}\right)^{1/3} \left(\frac{\rho}{\rho_{\odot}}\right)^{-1/3} . \tag{4.7}$$

The minimum transit duration searched is the larger of either 2.5× the exposure time, or a duration of 1/8 the period corresponding to a circular orbit with a semi-major axis the size of the stellar radius. The grid constructed between these two extremes is such that $t_{\mathrm{dur,\,n}} = (1.1)^n t_{\mathrm{dur,\,min}}$ for $n = 1, 2, ..., N$ for all $t_{\mathrm{dur,\,n}} < t_{\mathrm{dur,\,max}}$. We employ a non-linear limb-darkening law to calculate the signal template. The limb-darkening coefficients for the transit template are chosen from the closest match from Claret (2018) based on the $T_{\mathrm{eff}}$ and $\log g$ of the star being searched, and a transit template for each duration is calculated via the python package `batman` (Kreidberg 2015). The template is calculated for a transit with a depth of 100 ppm, the typical scatter in a *Kepler* light curve, so as to optimize our search for low signal to noise transits

Finally, we calculate the overcomplete discrete wavelet transform for the padded, gap-filled flux array and for each of the transit templates. The overcomplete wavelet transform is implemented via the `Pywavelets` software package (Lee et al. 2019). From each band in the overcomplete wavelet transform, we calculate whitening coefficients via a running MAD filter (see Figure 4.11c), and the transit template is correlated with the data array to calculate the SES via Equation 4.4 (see Figure 4.11d).

### Period Search and Identification of Threshold-Crossing Events

After calculating the SES at each cadence, we replace the artificial points that fill gaps in the time series and pad the array with $\mathbb{N} = 0$. $\mathbb{N}(n)$ and $\mathbb{D}(n)$ are binned at widths of $t_{\mathrm{dur}}/3$, and the period search is conducted using an implementation of the Fast Folding Algorithm often used to search for pulsars (Staelin 1969). The Fast Folding Algorithm ensures run times that scale as $\mathcal{O}(n \log n)$ which results in a computationally tractable

search without sacrificing period resolution. The range of periods searched for each transit duration are determined by inverting Equation 4.7 to estimate the minimum and maximum period for an implied durations of $t_{\rm dur}/3$ to $2\,t_{\rm dur}$. The maximum MES is determined at each combination of period and transit duration, and, if above some user-defined threshold, the period and time of mid transit are catalogued.

Each combination of period, transit duration, and time of mid transit above a user defined threshold in MES represents a potential Threshold Crossing Event (TCE). After completing the search over all transit durations and periods, each potential TCE is refined by calculating the MES with the unbinned SES arrays. At this stage we also filter out TCEs that are within 1% of a harmonic or subharmonic of a stronger period, as well as TCEs with an average of <3 cadences at points within one transit duration of the expected transit, and TCEs that are dominated by the SES from a single event in the light curve.

Beginning from the potential TCE with the largest MES, we then mask out each TCE and check whether the remaining events remain significant. In this way we test for aliasing before promoting each potential TCE. If the MES of a potential TCE is not significantly reduced after another TCE is masked and the time of the TCE remains the same, we catalog the additional TCE in the light curve, and repeat this process. In this way, we allow for multiple detections on each pass before running the search again. `TraSH-DUMP` goes through up to three iterations total, masking all current TCEs and recalculating SES(n), $\mathbb{N}(n)$, and $\mathbb{D}(n)$ between each iteration.

### 4.3.3 Comparison to TLS with Real Data

To test the sensitivity of `TraSH-DUMP`, we can compare it to other pipelines commonly used by the scientific community. To this end, we compare `TraSH-DUMP` to the Transit Least Squares algorithm (TLS; Hippke et al. 2019). TLS is a reformulation of Box Least Squares (BLS; Kovács et al. 2002), which has been shown to increase detection

sensitivity significantly compared to BLS by application of a realistic transit template instead of a box-shaped transit. This is the same principle upon which `TraSH-DUMP` is built; however, one difference between the two algorithms is that `TraSH-DUMP` takes the effects of smearing caused by finite exposure times into account, while TLS does not.

Both the TLS and BLS algorithms use the Signal Detection Efficiency (SDE) as their detection statistic. This parameter is analogous to the MES, if the MES were normalized further at each period using the mean and variance of the MES distributions at differing transit epochs and durations. Because our formulation makes it so that this is already the case (at least for white noise), the SDE and MES should be somewhat comparable statistics. However, one difference between our formulation of the MES and the TLS formulation of the SDE is that TLS adds an additional step to constrain their detection statistic better. What is classically referred to as the SDE, these authors call the "Raw SDE", and define their SDE as,

$$\mathrm{SDE}(P) = \frac{\mathrm{SDE}_{\mathrm{raw}}(P) - \mathrm{SDE}_{\mathrm{trend}}(P)}{\mathrm{SDE}_{\mathrm{scatter}}(P)} \quad , \tag{4.8}$$

where the trend ($\mathrm{SDE}_{\mathrm{trend}}$) is estimated via a median filtered periodogram and the scatter ($\mathrm{SDE}_{\mathrm{scatter}}$) is estimated via the point to point differences in the periodogram (as in Ofir 2014), respectively. This step typically boosts the signal to noise of a putative transiting signal by reducing the noise ramp in the background TLS spectrum. While `TraSH-DUMP` is not equipped to measure the significance of signals in the same way, such an exercise may prove fruitful in estimating a modified version of the MES. However, in `TraSH-DUMP`'s current iteration, the $\mathrm{SDE}_{\mathrm{raw}}$ is a more analogous statistic with which to compare against TLS.

In comparing with TLS and BLS, we provide both the corrected and the raw SDE, as the raw SDE is a more analogous comparison to the MES applied in `TraSH-DUMP`. While these detection statistics are analogous, they do note confer a direct one to one mapping.

Our intent in comparing these different pipelines to `TraSH-DUMP` is to show that we are able to recover planets with similarly high efficiency, though a more thorough comparison between each pipeline would be needed to determine in which areas of parameter space one may prove more efficient than another.

To assess the performance of `TraSH-DUMP` on real data, we utilize *K2* light curves, which should have similar detection biases as the *Kepler* mission since they are in the same bandpass and therefore transits have equivalent limb-darkening parameters. However searching planets on *K2* light curves is more computationally efficient given the shorter temporal baseline. For each of the systems analyzed in this section, light curves were downloaded from the `lightkurve` python package (Lightkurve Collaboration et al. 2018), and systematics were removed via the Self Flat-Fielding method (SFF; Vanderburg & Johnson 2014).

### K2-3 (EPIC 201367065)

K2-3 is an early M dwarf ($T_{\rm eff} = 3835$ K, $\log g = 4.7$, $\rho_\star = 3.5\rho_\odot$) with three confirmed small transiting planets (Crossfield et al. 2016), K2-3b ($R_p = 2R_\oplus, P = 10$ days), K2-3c ($R_p = 1.7R_\oplus, P = 24.6$ days), and K2-3d ($R_p = 1.6R_\oplus, P = 45.5$ days). This star was chosen for a comparison because it has well-defined and known transits, and because the shape of the transits deviate significantly from that of a transiting planet around a solar-type star.

The results of both `TraSH-DUMP` and TLS analyses of the processed K2-3 light curve are shown in Figure 4.12. We uncover planets K2-3b and -c, though `TraSH-DUMP` confuses the combination of one transit from K2-3d and systematic noise as a third candidate. Such confusion stems from masking errors. Because `TraSH-DUMP` was run in its binning mode, the transit epoch may have only centered to within $t_{\rm dur}/3$ on its initial pass. To recover K2-3d more effectively will require a more precise centroiding algorithm to accurately measure the transit epochs of K2-3b and -c to more effectively mask their

transits, as well as improved systematics removal and identification in the processing steps between masking transits and recalculating $\mathbb{N}(n)$ and $\mathbb{D}(n)$.

Running TLS twice, the second time with the largest periodic signal from the first run (K2-3b) masked, TLS recovers both K2-3b and -c. For K2-3b, `TraSH-DUMP` recovers the signal with MES = 36.8, and TLS recovers the signal with SDE = 31.4 ($\mathrm{SDE_{raw}}$ = 28.0). After masking the transit signal from K2-3b, `TraSH-DUMP` recovers the signal from K2-3c with MES = 15.7, and TLS recovers the transiting signal with SDE = 24.1 ($\mathrm{SDE_{raw}}$ = 16.2). Thus, comparing TLS against the `TraSH-DUMP`'s sensitivity, the two detection algorithms tend to recover signals with similar significance.

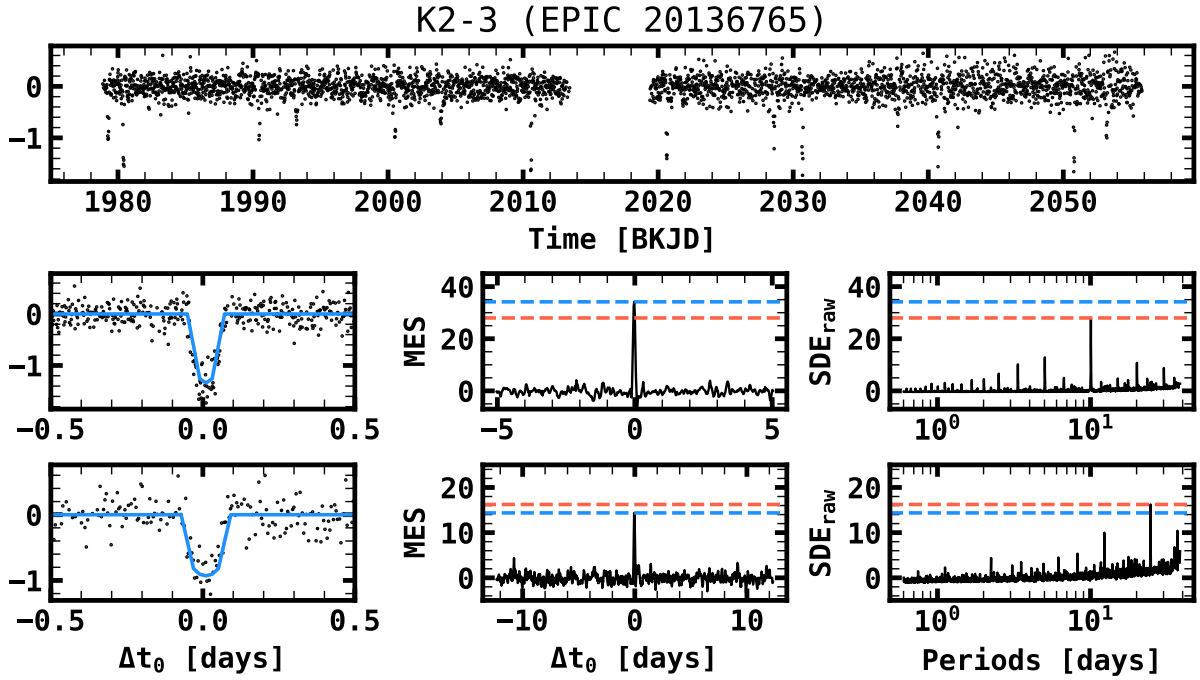


Fig. 4.12.— The results of the `TraSH-DUMP` search on K2-3, a star with three known, small transiting planets. *Top Panel:* The detrended light curve of K2-3 with systematics removed. The flux is in units of parts per thousand. *Middle Row:* From left to right, the light curve phase-folded at the period of K2-3b, with the transit template that was used for the search overplotted, the MES distribution at the planet period, and the SDE distribution in period from the TLS search. The dashed blue and red lines in the middle and right panels denote the detection metric from `TraSH-DUMP` and TLS, respectively. The transit template in the left-most figure is not a fit to the data, it is merely the search template adopted by `TraSH-DUMP` for K2-3, and scaled to a similar transit depth. *Bottom Row:* The same as the middle row, but for K2-3c.

### K2-132 (EPIC 201367065)

K2-132 is a bright ($Kp = 11.7$) evolved star at the base of the red giant branch ($T_{\rm eff}$ = 4840 K, $\log g = 3.3$), with a transiting hot Jupiter ($P = 9.18$ days, $R_p = 14.6\,R_\oplus$) and precise stellar constraints via asteroseismology (Grunblatt et al. 2017). As a result, K2-132 is a particularly interesting target for testing our algorithms because it lies in a region of parameter space where hot Jupiters may be severely affected by tides, and its light curve shows high, correlated variability at shorter timescales that would not be removed by our detrending process. The results of the `TraSH-DUMP` and TLS transit search are shown in Figure 4.13.

The hot Jupiter, K2-132b, is recovered by both pipelines, and again with similar sensitivity. Because there is only one known transiting planet, we do not iterate over more periods for either TLS or `TraSH-DUMP`, though `TraSH-DUMP` did recover a False Alarm due to the feature at ∼2784.5 BKJD which was missed by the SFF systematics removal routine. `TraSH-DUMP` recovered the transiting planet at a significance of MES = 12.0, and TLS recovered the transiting planet at a significance of SDE = 15.1 ($\rm SDE_{raw}$ = 9.9). Therefore, we find once again that `TraSH-DUMP` is able to recover the transiting planet with similar significance to TLS, at least when comparing the raw periodogram, though `TraSH-DUMP` still recovers the signal at a lower significance than the smoothed, systematics-corrected periodogram calculated by the TLS algorithm. Applying similar post-processing analyses to a `TraSH-DUMP`-derived periodogram is likely to provide similar improvements.

The post-processing step of smoothing the periodogram that is applied by TLS has one disadvantage over `TraSH-DUMP`, which is that the SDE derived is global in nature. In other words, an accurate assessment of the significance of a signal can only be made if a wide range of periods are searched. While this is not an issue if searching for a single planet, it can become a computational barrier when searching for multiple planets, as the SDE would have to be recomputed for every period before testing the significance of

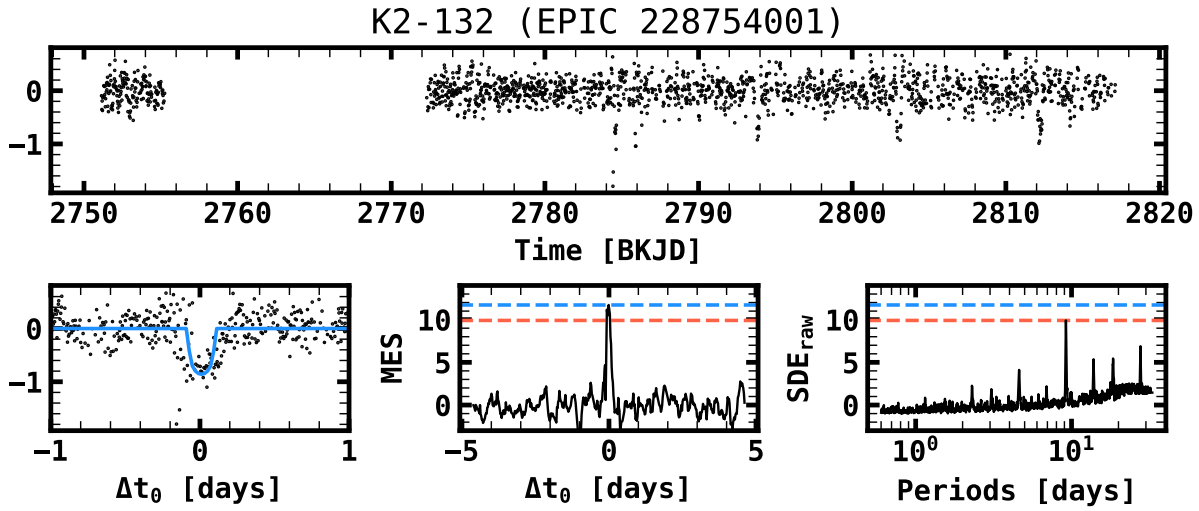


Fig. 4.13.— The results of the TraSH-DUMP search on K2-132, a star at the base of the RGB with a transiting hot Jupiter. *Top Panel:* The detrended light curve of K2-132 with systematics removed. The flux is in units of parts per thousand. There are regions of the light curve where the systematics removal is slightly overfit, which has caused a deformation in the transit shape. *Bottom Row:* From left to right, the light curve phase-folded at the period of K2-b, with the transit template used by the TraSH-DUMP algorithm overlaid in blue, the MES distribution at the planet period, and the TLS periodogram. The dashed blue and red lines in the middle and right panels denote the detection metric from TraSH-DUMP and TLS, respectively. The transit template in the left-most figure is not a fit to the data, it is merely the search template adopted by TraSH-DUMP for K2-132, and scaled to a similar transit depth.

a second putative signal after removing the most significant signal.

## 4.4 Discussion

### 4.4.1 Expected Gain in Detection Efficiency from CaTShIP

While deriving fundamental stellar parameters for stars in the *Kepler* field is useful for a plethora of science cases, the question remains, what is the expected increase in detection efficiency gained from incorporating CaTShIP parameters into a TraSH-DUMP planet search? This is a somewhat complicated question, because while template mismatch does decrease the sensitivity of the *Kepler* pipeline for stars dissimilar to the Sun, it doesn't seem to impact light curves with little correlated noise, as evidenced by running

*Kepler*'s TPS on light curves with simulated transits injected (Christiansen et al. 2020). For light curves dominated by red noise, the sensitivity of the *Kepler* transiting planet search module drops by $\gtrsim$10%. Subdividing the red noise-dominated stars into their respective temperatures shows an additional drop of ∼5-10% in the sensitivity for stars that differ from the sun by $\gtrsim$1000 K (Christiansen et al. 2020). This additional drop may be explained by differing transit shapes. In the case of white noise, the adaptive power spectrum calculated by *Kepler*'s Transiting Planet Search module is only contaminated by noise at frequencies near the Nyquist sampling rate. However, for stars with strong correlated noise, the power spectrum is contaminated at lower frequencies, so having a more precise estimate of the transit shape (and therefore the power spectrum) seems to increase the transit detection efficiency, all other variables being equal (Christiansen et al. 2020).

At face value, these results seem to suggest that adopting the limb-darkening parameters from the `CaTShIP` may allow one to avoid similar drops in detection efficiency, but this is likely only the case in particular areas of parameter space. For instance, due to the finite exposure time of the *Kepler* photometer (∼30 minutes), the expected transit shape will be "smeared", erasing high frequency structure in the transit shape which effectively eliminates any unique structure at transit ingress and egress dictated by limb-darkening parameters. However, the lower-frequency structure of the transit shape dictated by limb-darkening, such as the "U-shape" formed between ingress and egress, may still be preserved at finite exposure times. Therefore, what we expect to find is that adopting limb-darkening parameters from the `CaTShIP` should significantly increase the sensitivity of `TraSH-DUMP` (and TLS, for the same reasons) compared to adopting a box-shaped template, but it is only likely to significantly increase sensitivity for long duration ($t_{\rm dur} \gg t_{\rm exp}$) transits compared to adopting a global transit template (i.e., using the same limb-darkening parameters for each star, as with the *Kepler* mission).

Assuming similar properties to the *Kepler*'s TPS, incorporating stellar parameters

from the `CaTShIP` into the transit templates used to search for stars should provide as much as a ∼10% increase in sensitivity overall, with the largest boost in the face of strong correlated noise and for long transit durations. While an enhanced sensitivity of ∼10% may not represent a paradigm-shifting difference, for the *Kepler* field it may translate to ∼100+ new planet candidates.

### 4.4.2 Additional Challenges for Transit-Search Surveys of Evolved Hosts

While we've discussed the effects of limb-darkening parameters and red noise on the detectability of transiting planets, these topics have all been discussed in the context of light curves with systematic trends removed. The removal of such instrumental artifacts can be non-trivial, and is usually performed by identifying common patterns within groups of nearby stars and removing those systematics (Smith et al. 2012). Instrumental systematics such as these typically act on timescales of a few days, which is a similar timescale to that of transit durations one might expect from a planet orbiting an evolved host. Because of this, transiting signals with long durations are susceptible to being corrupted by common modes for removing systematic trends. One must ensure that timescales over which corrections are applied are longer than any expected transit duration. It's not clear to what extent *Kepler*'s systematics corrections have affected long duration transit signals in the data, but such a consideration is important.

Another complication that will naturally arise in the search for long-duration transit signals is a likely increase in the False Alarm rate. Because stellar activity often acts on timescales of a few days, stellar activity is likely to contaminate any planet catalog searching for long duration transits. Mitigating such contaminants may prove a difficult task. One possible strategy that may be useful for an occurrence rate measurement is to jointly model the distribution of transiting planets with the false alarm rate. While any one transiting candidate may not be a reliable detection, information can still be gleaned

from a large enough sample of unreliable candidates.

Employing this strategy requires a precise understanding of the false alarm rate. Such measurements have been made by the *Kepler* team by running *Kepler*'s TPS on inverted and scrambled light curves (Thompson et al. 2018). Inverted light curves are used to invert any potential transit signals in the data, so that no such signals should be detected, and scrambled light curves are used to dephase any periodic signals so that any false alarm detections must be due to spurious events, rather than astrophysical or instrumental signals.

In conclusion, the search for transiting planets around evolved stars offers several new challenges, but deriving independent constraints on the occurrence of TraSH is key to understanding the end stages of a planet's lifetime.

## 4.5 Conclusions

Stellar evolution is expected to play a dramatic role in disrupting and re-sculpting planetary architectures, through effects such as increased tides and increased stellar irradiation. However, the precise understanding needed to measure the onset of these effects is lacking, with current literature suggesting that changes in planetary architecture driven by stellar evolution occur as early as the MS (Hamer & Schlaufman 2019), while other studies suggest that such effects may not play a role until well after the host star begins its ascent up the RGB (Grunblatt et al. 2018). To understand the onset of planetary disruption, detailed planet demographics are needed across a wide range of stellar host masses, evolutionary states, and planet orbital separations. Particularly key to this understanding is measuring the demographics of planets with subgiant host stars because of the precise age information available via isochrone modeling.

Because the primary goal of the *Kepler* mission was to understand the demographics of Earth-like planets around Sun-like stars, detailed understanding of the completeness

and reliability of planet candidates orbiting subgiant and RGB stars requires a targeted transiting planet search across all phases of stellar evolution. In this work we lay the foundation for performing such a study for the stars in the *Kepler* field. In particular we make two important advances in regard to this goal; we (1) homogeneously infer ages, masses, radii, and other fundamental parameters for ∼163,000 stars in the *Kepler* field, improving upon the literature by incorporating uniform photometric and spectroscopic metallicities for each star in the sample, and (2) introduce a new detection algorithm, `TraSH-DUMP` (TRAnsiting planets with Subgiant Hosts – Detection with an Unbiased Matched-filter Pipeline), designed to improve sensitivity and avoid second-order detection biases driven by transit template mismatch in correlated noise.

Our catalog of stellar parameters, the `CaTShIP`, includes precisely measured radii, ($\sigma_{R_\star} \approx 2.7\%$), masses ($\sigma_{M_\star} \approx 6\%$), and ages across the main sequence accurate to within a factor of ∼2, and in agreement with asteroseismic ages to within ∼30%. In particular, we demonstrate that we are able to distinguish the evolutionary state of high-mass stars ($\gtrsim 1.2 M_\odot$) on the Zero-Age, Intermediate-Age, and Terminal-Age main sequence as a function of mass. This precision results in age uncertainties as low as ∼10-100 Myr and provides the precision in evolutionary state needed to infer the interior structure of such stars – a vital component for predictions of orbital decay via evolving equilibrium and dynamical tidal interactions.

In detailing the methodology for `TraSH-DUMP`, we compare its sensitivity against TLS, a state-of-the-art open-source detection algorithm, demonstrating similar sensitivity while providing a computational load that scales well for stars with long temporal baselines typical of the *Kepler* field, and with minimal loss in period resolution. In particular, we expect our pipeline to provide an improvement in the face of correlated red noise, such as that caused by stellar variability, and for long period, long duration signals, where *Kepler*'s TPS module displayed decreased sensitivity in a parameter space occupied by putative planets of interest for the goals of this work.

The union of the `CaTShIP` and `TraSH-DUMP` provide the necessary tools to time the onset of stellar evolution driven planetary engulfment, migration, and disruption.

# Chapter 5

# Summary, Future Directions, and Unanswered Questions

The *Kepler* mission has provided the means necessary for the ensemble study of hundreds to thousands of exoplanets, allowing the first real census estimate of planets ranging from small, terrestrial planets about the size of the Earth to large planets up to twice the size of Jupiter. Despite the incredible diversity of planets found, from the surprising abundance of Super-Earths, a planet type not found in our own Solar system, to the very existence of hot Jupiters, which disrupted astronomers' understanding of planet formation, one constant has been the role of stars in shaping their own planetary systems. The properties of such systems cannot be decoupled from the very stars they orbit, and as such a proper understanding of planetary systems requires a similarly rigid understanding of the star around which they form and are found.

In this thesis we've explored the connection inherent between planetary systems and their host stars via the joint ensemble study of hundreds of exoplanets discovered by the *Kepler* mission with thoroughly characterized host stars via the APOGEE survey. Throughout this work we've stressed the importance of well understood stellar knowledge, and shown how measuring stellar properties leads to a better, more robust understanding

of exoplanet populations.

## 5.1 The Role of Stellar Chemistry in Planet Demographics

More specifically, in our quest to understand the planet populations, we've highlighted stellar chemistry and composition as a probe of planet-forming environments, finding clear evidence that stellar chemistry facilitates not only the growth of large planets, but the formation, evolution, and migration of planets at all sizes and orbital separations probed by the *Kepler* survey. More specifically:

- In Chapter 2, we find a connection between the number density of iron atoms in the atmospheres of stars and the distance (i.e. orbital period) with which planets orbit them. This connection is seemingly present for planets of all sizes. We characterized this correlation, inferring a transition region at a period of $P \sim 8$ days, which corresponds to orbital separations of 0.07 au, ∼5-6 times closer than the average distance of Mercury to the Sun, beyond which the properties of planet host stars are significantly different.

- We hypothesized that the transition region at $P \sim 8$ days (∼0.07 au) represents a threshold with either planet formation or migration, where planets are not able to form closer in, or are stopped from migrating closer than this distance. However, metal-rich systems seem to control this stopping radius, possibly through the depletion/enhancement of planet-forming material or through the presence of planetesimals or asteroids that may drive planetary migration.

- This critical orbital separation also seems to play a role in the size of planets, as planets interior to this critical distance are significantly smaller despite having metal-rich hosts, and planets at further orbital distances are larger in radius, despite

having metal-poor hosts. This is an indication that something other than core accretion is shaping the distribution of small planets, because in the core-accretion paradigm close-in planets should be larger due to the enhanced host star metallicity and as a result, implied accretion rate. We interpret this apparent distinction as a signature of photoevaporation, where close-in planets with metal-rich hosts have been stripped of their atmospheres and planets with larger orbital separations are more likely to retain their atmospheres. Thus, because stellar metallicity seems to play an important role in shaping the orbital distance of planets of all sizes, it also shapes the distribution of planet radii via not only enhanced accretion, but by enhanced photoevaporation.

- In addition to our interpretation that photoevaporation has shaped the distribution of close-in planets, we also take one step further in comparing our results to theory, and interpret that all the planets that are close-in have rocky cores. Were that not the case, and the cores of such planets were drawn from multiple different compositions, we would find a gradient of planet radii, which doesn't seem to be the case.

- In Chapter 3, we apply a more rigorous statistical analysis to understand not only the metallicity of planet hosting stars in the APOGEE survey, but the overall enhancement of ten additional chemical elements. By carefully controlling for biases inherent in the *Kepler* detection pipeline and the APOGEE follow-up survey we are able to make the first ever inference about the population of small planets and their dependence on detailed stellar chemistry, free of detection bias.

- We find that the occurrence of small planets is highly correlated with the enhancement of any element at short orbital distances. For instance, a change in the stellar composition of any one element as small as 0.1 dex correlates with an increase in the number of planets by anywhere from $\sim$40-100%.

Based on the conglomeration of all of these results, we've compiled convincing evidence that stellar chemistry is a strong factor in shaping the distribution of close-in ($\lesssim$1 au) exoplanets. And, these correlations are an important indicator to consider when making any kind of inference about the distribution of planet types, sizes, and orbital properties.

## 5.2 Stellar Age-Dating and Long-term Planetary Evolution

One particularly exciting opportunity is to couple studies of planet demographics with inferences of stellar age. Such inferences have only recently become available thanks to the large scale astrometric survey, *Gaia*. In this thesis, we build the foundation to understand the connection between stellar evolution and the long-term evolution of planetary systems. In particular:

- In Chapter 3, we make a concerted effort to understand the biases that are likely to be present in any statistical study of the planet population combined with stellar ages. In particular, we show that small differences in metallicity, even as small as 0.05 dex, can lead to significant biases even for small planets, due to the strong correlation between planet occurrence and the enhancement of metals. This motivates the need for large-scale spectroscopic surveys to understand planet demographics more fully.

- Based on the biases present in stellar ages, we call into question the causation of correlations between planet properties with metallicity, and provide an alternative hypothesize that some such correlations may actually be driven by age, rather than the enhancement of metals and planet forming material which is the classical interpretation of such trends.

- We also make an argument for measuring planet occurrence rates for stars in open clusters as a way to make inferences about changes in the planet population with age, and provide estimates of the planet occurrence rates in open clusters as a means of facilitating such studies.

- Based on the discussion and results in Chapter 3, in Chapter 4 we motivate the need for an independent planet search in the *Kepler* field to better understand long-term planetary evolution, particularly with regard to tidal forces. We provide the means for performing such a search, by measuring fundamental stellar properties, including age, for ∼163,000 stars in the *Kepler* field and developing an independent transit-search algorithm, the TRAnsiting planets with Subgiant Hosts – Detection with an Unbiased Matched filter Pipeline (`TraSH-DUMP`), to detect planets in the light curves of well-characterized stars.

- In setting the foundation for a large-scale search for TRAnsiting planets with Subgiant Hosts (TraSH), we create a catalog of ∼163,000 well characterized stars in the *Kepler* field. We demonstrate that the precision and accuracy of the stellar properties in this catalog, the CAtolog of Transit Search Input Paramters (`CaTShIP`), are consistent with other methodologies, and our improvements in precision derive particularly from our use of a single scale for measuring spectroscopic and photometric metallicities, once again emphasizing the need for demographic studies to consider such properties.

- Importantly, we demonstrate the precision in stellar parameters needed to resolve multiple stages of evolution on and beyond the main sequence for higher mass stars ($\gtrsim 1.2 M_{\odot}$), a critical component for inferring the interior structure of such stars and understanding predictions from tidal decay theory.

- Finally, we demonstrate the sensitivity of `TraSH-DUMP` by comparing it to the current state of the art transit detection algorithms and show that it is competitive

with such algorithms and able to detect transiting planets with similar or greater significance, while still providing a computationally feasible search for light curves with long temporal baselines. These properties of `TraSH-DUMP` are considered in addition to searching a more robust area of parameter space than traditional planet search algorithms such as that used by the *Kepler* team.

From these foundations, we set the stage for a new initiative aimed at obtaining a complete understanding of the evolution of planetary systems, in a statistical sense. While it is difficult to measure any changes in the demographics of planets due to the timescales of such evolution, the era of precision astrophysics via *Gaia* is setting the stage for inferring stellar ages from which we can compare the evolving demographics of exoplanets with age.

## 5.3 Future Work

### 5.3.1 Targeted Searches for New Planet Populations in Different Parts of the Galaxy

One of the primary takeaways from Chapter 3 was the difficulty in understanding the role of specific chemical elements in shaping the planetary distribution function due to degeneracies in the *Kepler* field. These degeneracies arise from the fact that the stars in *Kepler* are dominated by members of the thin disk. As a result, they all share a similar star formation history, and so the enhancement of any one element is well correlated with the enhancement of all other elements. As a result it is difficult to isolate the effects of specific elements. The solution for breaking these degeneracies is searching for planets in different parts of the Galaxy. It has been shown that the location of stars in the Milky Way vary compositionally due to effects such as radial migration and differences in star formation history (see Figure 5.1). Therefore, in order to probe the populations of

planets that have host stars with differing abundance patterns, a targeted search among such stars is needed.

The primary challenges in constructing a sample of such planet-search stars is the degree of difficulty in detecting planets orbiting such stars. Due to the photometric precision needed for transit surveys, or the high-resolution spectra needed for RV surveys, planet searches are typically limited to bright stars. *Kepler* for instance, was limited to stars brighter than $Kp \sim 16$ (Batalha et al. 2010), which for a Sun-like star corresponds to a distance of $\sim$1 kpc.

However, to observe a large sample of stars with fundamentally different chemical signatures than those in the Solar neighborhood, one would need to perform a targeted planet search for a sample of stars $\gtrsim$1-2 kpc from the Solar neighborhood (see Figure 5.1), which would result in a dim sample. As a result, current transit surveys such as *TESS* (Ricker et al. 2014), *K2*, and *Kepler* are unlikely to uncover a significant sample of such planets. While such a planet search would be difficult for a transit survey, a dedicated RV survey should be possible with the next generation of precision Doppler spectrographs on the next generation of large telescopes (e.g., G-CLEF – the GMT Chicago Large Earth Finder; Szentgyorgyi et al. 2012).

### 5.3.2 The Search for TraSH in *Kepler* and *TESS*

In Chapter 4 we laid the foundation for a targeted search and planet demographics study of TRAnsiting planets with Subgiant Hosts (TraSH) in the *Kepler* field. While the *Kepler* field offers an opportunity to understand TraSH at all orbital separations, a more statistically significant sample may be gleaned from *TESS*. To understand the role of tides and the destruction of close-in planets with stellar evolution, a large, statistically significant population of stars with precisely measured ages and masses, across several evolutionary phases, and with a large number of detected planets, is needed. *TESS* may be the best opportunity to make such a measurement, as it is observing $\gtrsim$80 million stars

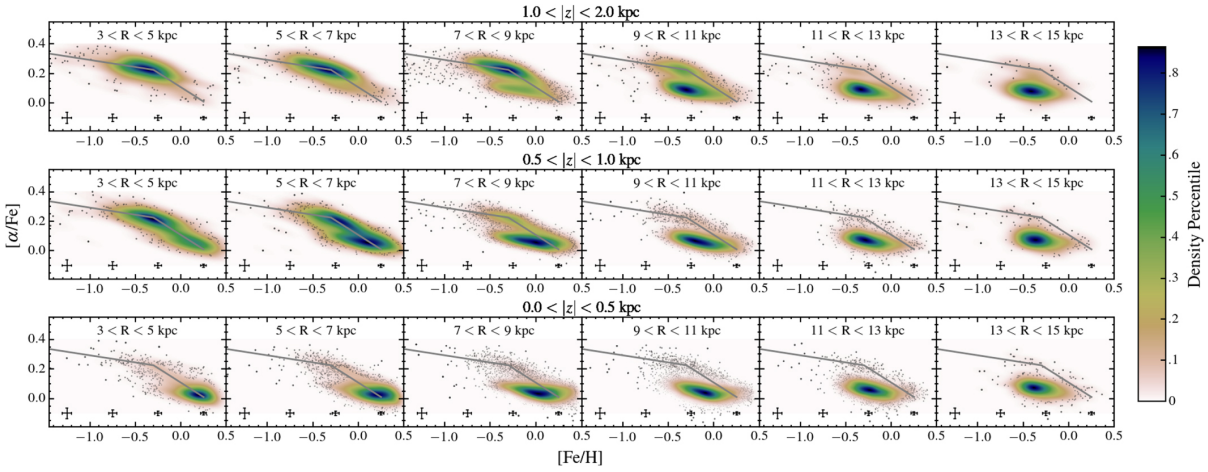


Fig. 5.1.— The joint metallicity and alpha abundance distribution functions as a function of location in the Galaxy as derived with APOGEE data from Hayden et al. (2015). The chemical signatures of stars in the Solar neighborhood ($7 < R < 9$ kpc and $0 < |z| < 1$ kpc) are predominantly from the thin disk, meaning that all elements more or less correlate with one another. To gain a population of stars with differing chemistry, one must observe stars in the thick disk, which becomes the dominant population at $|z| \gtrsim$ 1-2 kpc.

bright enough (*TESS* mag $\lesssim$ 15) to detect large planets, an increase of over two orders of magnitude compared to the number of stars observed in *Kepler* or *K2*.

An additional advantage with *TESS* target stars is the higher cadence (∼2-10 minutes, compared to ∼30 minutes with *Kepler*) with which observations are taken. The high cadence allows for the asteroseismic characterization of a much larger sample of stars than in *Kepler*. Combining seismic parameters with *Gaia* parallaxes and precision photometry, we should be able to infer ages for subgiant stars with precisions of ∼10% (see Figure 5.2), even without metallicity information. While *Kepler* offers the opportunity to measure precise ages (as evidenced in Chapter 4), the asteroseismic sample from *Kepler* may be more biased and as a result may not offer quite the same uniformity in characterization as the *TESS* sample. However, while *TESS* offers the opportunity to study a vast number of stars, the detailed understanding from the *Kepler* mission is still needed to probe a larger range of orbital separations, as *TESS* is unlikely to discover as many planets at longer periods.

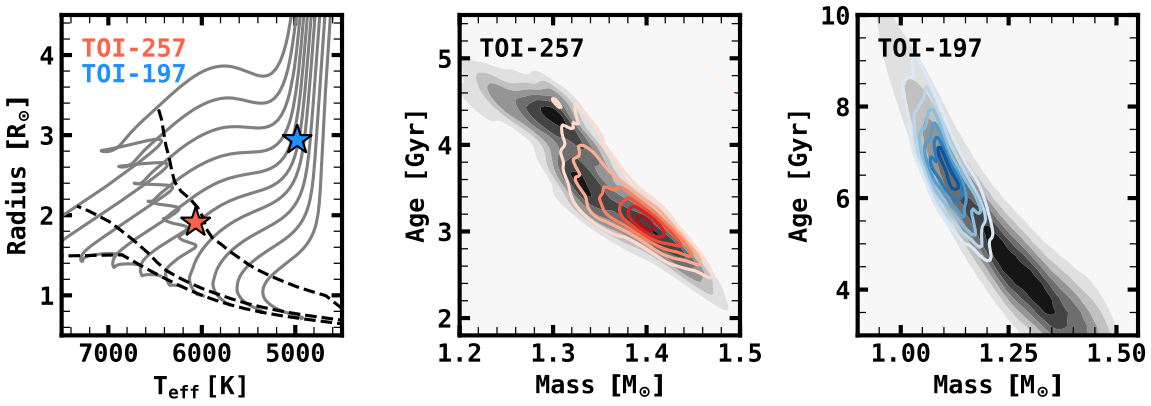


Fig. 5.2.— Stellar parameters derived from `isofit` for confirmed planet-hosting stars TOI-257 and TOI-197, assuming a Milky Way thin disk metallicity prior. With asteroseismic constraints, ages were inferred with precisions of ∼10% and ∼15% for TOI-257 and TOI-197, respectively. This is a substantial increase in precision from ∼50% as inferred without asteroseismic constraints. *Left:* The effective temperatures and radii of TOI-257 and TOI-197. The solid lines show stellar evolutionary tracks (Choi et al. 2016) ranging from 0.8-1.7 $M_{\odot}$. The dashed lines show the Zero-Age, Intermediate-Age, and Terminal-Age Main Sequence. *Center:* The mass and age of TOI-257, a star near the terminal age main sequence (Addison et al. 2020). The contours show the posterior distribution in age and mass with asteroseismic constraints (grayscale), and without asteroseismic constraints (red). *Right:* Same as (center), but for TOI-197, a star near the base of the Red Giant Branch (Huber et al. 2019).

## 5.4 Key Questions Still Unanswered

While our understanding of planet formation and the characteristics of exoplanets in our Galaxy has continued to grow substantially over the last decade, there are still key questions driving this field forward. Among the most relevant questions pertaining to this work, can hot Jupiters survive the main sequence? Due to the uncertainty in inferring stellar ages, this is by no means a settled debate. The difficulty in understanding the complexity of tidal forces and their effects on planets derives not only from observations, but also in the complex nature of the physics itself.

Does the chemical composition of planetesimals dictate planet formation efficiency? And can the chemical composition of planetesimals truly be estimated via host star

abundances? While our work in Chapter 3 would (naively) suggest that the enhancement of any element contributes to the occurrence of planets equally, it is likely that our sample was too biased to truly probe such a question in the first place.

What is the minimum metal content needed to form a planet? We've probed planet formation for stars as low as [Fe/H] $\approx -0.5$, which still form terrestrial planets very efficiently. However, there are stars with metallicities as low as [Fe/H] $\approx -7$ (Nordlander et al. 2017), though we don't even know of planet host stars with metallicities below [Fe/H] $\approx -0.5$. Can more extreme low-metallicity stars host or form planets at all? At what point are there no longer enough metals to form planets? These questions are still driving the field forward, and motivate the search for exoplanets in wildly varying regions of the Galaxy.

# References

Abolfathi, B., Aguado, D. S., Aguilar, G., Prieto, C. A., Almeida, A., Ananna, T. T., Anders, F., Anderson, S. F., Andrews, B. H., Anguiano, B., Aragón-Salamanca, A., Argudo-Fernández, M., Armengaud, E., Ata, M., Aubourg, E., Avila-Reese, V., Badenes, C., Bailey, S., Balland, C., Barger, K. A., Barrera-Ballesteros, J., Bartosz, C., Bastien, F., Bates, D., Baumgarten, F., Bautista, J., Beaton, R., Beers, T. C., Belfiore, F., Bender, C. F., Bernardi, M., Bershady, M. A., Beutler, F., Bird, J. C., Bizyaev, D., Blanc, G. A., Blanton, M. R., Blomqvist, M., Bolton, A. S., Boquien, M., Borissova, J., Bovy, J., Diaz, C. A. B., Brandt, W. N., Brinkmann, J., Brownstein, J. R., Bundy, K., Burgasser, A. J., Burtin, E., Busca, N. G., Cañas, C. I., Cano-Díaz, M., Cappellari, M., Carrera, R., Casey, A. R., Sodi, B. C., Chen, Y., Cherinka, B., Chiappini, C., Choi, P. D., Chojnowski, D., Chuang, C.-H., Chung, H., Clerc, N., Cohen, R. E., Comerford, J. M., Comparat, J., do Nascimento, J. C., da Costa, L., Cousinou, M.-C., Covey, K., Crane, J. D., Cruz-Gonzalez, I., Cunha, K., da Silva Ilha, G., Damke, G. J., Darling, J., Jr., J. W. D., Dawson, K., de Icaza Lizaola, M. A. C., de la Macorra, A., de la Torre, S., Lee, N. D., de Sainte Agathe, V., Machado, A. D., Dell'Agli, F., Delubac, T., Diamond-Stanic, A. M., Donor, J., Downes, J. J., Drory, N., du Mas des Bourboux, H., Duckworth, C. J., Dwelly, T., Dyer, J., Ebelke, G., Eigenbrot, A. D., Eisenstein, D. J., Elsworth, Y. P., Emsellem, E., Eracleous, M., Erfanianfar, G., Escoffier, S., Fan, X., Alvar, E. F., Fernandez-Trincado, J. G., Cirolini,

R. F., Feuillet, D., Finoguenov, A., Fleming, S. W., Font-Ribera, A., Freischlad, G., Frinchaboy, P., Fu, H., Chew, Y. G. M., Galbany, L., Pérez, A. E. G., Garcia-Dias, R., García-Hernández, D. A., Oehmichen, L. A. G., Gaulme, P., Gelfand, J., Gil-Marín, H., Gillespie, B. A., Goddard, D., Hernández, J. I. G., Gonzalez-Perez, V., Grabowski, K., Green, P. J., Grier, C. J., Gueguen, A., Guo, H., Guy, J., Hagen, A., Hall, P., Harding, P., Hasselquist, S., Hawley, S., Hayes, C. R., Hearty, F., Hekker, S., Hernandez, J., Toledo, H. H., Hogg, D. W., Holley-Bockelmann, K., Holtzman, J. A., Hou, J., Hsieh, B.-C., Hunt, J. A. S., Hutchinson, T. A., Hwang, H. S., Angel, C. E. J., Johnson, J. A., Jones, A., Jönsson, H., Jullo, E., Khan, F. S., Kinemuchi, K., Kirkby, D., IV, C. C. K., Kitaura, F.-S., Knapp, G. R., Kneib, J.-P., Kollmeier, J. A., Lacerna, I., Lane, R. R., Lang, D., Law, D. R., Goff, J.-M. L., Lee, Y.-B., Li, H., Li, C., Lian, J., Liang, Y., Lima, M., Lin, L., Long, D., Lucatello, S., Lundgren, B., Mackereth, J. T., MacLeod, C. L., Mahadevan, S., Maia, M. A. G., Majewski, S., Manchado, A., Maraston, C., Mariappan, V., Marques-Chaves, R., Masseron, T., Masters, K. L., McDermid, R. M., McGreer, I. D., Melendez, M., Meneses-Goytia, S., Merloni, A., Merrifield, M. R., Meszaros, S., Meza, A., Minchev, I., Minniti, D., Mueller, E.-M., Muller-Sanchez, F., Muna, D., Muñoz, R. R., Myers, A. D., Nair, P., Nandra, K., Ness, M., Newman, J. A., Nichol, R. C., Nidever, D. L., Nitschelm, C., Noterdaeme, P., O'Connell, J., Oelkers, R. J., Oravetz, A., Oravetz, D., Ortíz, E. A., Osorio, Y., Pace, Z., Padilla, N., Palanque-Delabrouille, N., Palicio, P. A., Pan, H.-A., Pan, K., Parikh, T., Pâris, I., Park, C., Peirani, S., Pellejero-Ibanez, M., Penny, S., Percival, W. J., Perez-Fournon, I., Petitjean, P., Pieri, M. M., Pinsonneault, M., Pisani, A., Prada, F., Prakash, A., de Andrade Queiroz, A. B., Raddick, M. J., Raichoor, A., Rembold, S. B., Richstein, H., Riffel, R. A., Riffel, R., Rix, H.-W., Robin, A. C., Torres, S. R., Román-Zúñiga, C., Ross, A. J., Rossi, G., Ruan, J., Ruggeri, R., Ruiz, J., Salvato, M., Sánchez, A. G., Sánchez, S. F., Almeida, J. S., Sánchez-Gallego, J. R., Rojas, F. A. S., Santiago, B. X., Schiavon, R. P., Schimoia, J. S., Schlafly, E., Schlegel,

D., Schneider, D. P., Schuster, W. J., Schwope, A., Seo, H.-J., Serenelli, A., Shen, S., Shen, Y., Shetrone, M., Shull, M., Aguirre, V. S., Simon, J. D., Skrutskie, M., Slosar, A., Smethurst, R., Smith, V., Sobeck, J., Somers, G., Souter, B. J., Souto, D., Spindler, A., Stark, D. V., Stassun, K., Steinmetz, M., Stello, D., Storchi-Bergmann, T., Streblyanska, A., Stringfellow, G. S., Suárez, G., Sun, J., Szigeti, L., Taghizadeh-Popp, M., Talbot, M. S., Tang, B., Tao, C., Tayar, J., Tembe, M., Teske, J., Thakar, A. R., Thomas, D., Tissera, P., Tojeiro, R., Tremonti, C., Troup, N. W., Urry, M., Valenzuela, O., van den Bosch, R., Vargas-González, J., Vargas-Magaña, M., Vazquez, J. A., Villanova, S., Vogt, N., Wake, D., Wang, Y., Weaver, B. A., Weijmans, A.-M., Weinberg, D. H., Westfall, K. B., Whelan, D. G., Wilcots, E., Wild, V., Williams, R. A., Wilson, J., Wood-Vasey, W. M., Wylezalek, D., Xiao, T., Yan, R., Yang, M., Ybarra, J. E., Yèche, C., Zakamska, N., Zamora, O., Zarrouk, P., Zasowski, G., Zhang, K., Zhao, C., Zhao, G.-B., Zheng, Z., Zheng, Z., Zhou, Z.-M., Zhu, G., Zinn, J. C., & Zou, H. 2018, The Astrophysical Journal Supplement Series, 235, 42

Addison, B. C., Wright, D. J., Nicholson, B. A., Cale, B., Mocnik, T., Huber, D., Plavchan, P., Wittenmyer, R. A., Vanderburg, A., Chaplin, W. J., Chontos, A., Clark, J. T., Eastman, J. D., Ziegler, C., Brahm, R., Carter, B. D., Clerte, M., Espinoza, N., Horner, J., Bentley, J., Kane, S. R., Kielkopf, J. F., Laychock, E., Mengel, M. W., Okumura, J., Stassun, K. G., Bedding, T. R., Bowler, B. P., Burnelis, A., Collins, M., Crossfield, I., Davis, A. B., Evensberget, D., Heitzmann, A., Howell, S. B., Law, N., Mann, A. W., Marsden, S., O'Connor, J., Shporer, A., Stevens, C., Tinney, C. G., Tylor, C., Wang, S., Zhang, H., Henning, T., Kossakowski, D., Ricker, G., Sarkis, P., Vanderspek, R., Latham, D. W., Seager, S., Winn, J. N., Jenkins, J. M., Mireles, I., Rowden, P., Pepper, J., Daylan, T., Schlieder, J. E., Collins, K. A., Collins, K. I., Tan, T.-G., Ball, W. H., Basu, S., Buzasi, D. L., Campante, T. L., Corsaro, E., Gonzalez-Cuesta, L., Davies, G. R., A. Garcia, R., Guo, Z., Handberg, R., Hekker, S., Hey, D. R., Kallinger, T., Kawaler, S. D., Kayhan, C., Kuszlewicz, J. S., Lund, M. N.,

Lyttle, A., Mathur, S., Miglio, A., Mosser, B., Nielsen, M. B., Serenelli, A. M., Silva Aguirre, V., & Themessl, N. 2020, arXiv e-prints, arXiv:2001.07345

Adibekyan, V. Z., Delgado Mena, E., Sousa, S. G., Santos, N. C., Israelian, G., González Hernández, J. I., Mayor, M., & Hakobyan, A. A. 2012a, A&A, 547, A36

Adibekyan, V. Z., Figueira, P., Santos, N. C., Mortier, A., Mordasini, C., Delgado Mena, E., Sousa, S. G., Correia, A. C. M., Israelian, G., & Oshagh, M. 2013, A&A, 560, A51

Adibekyan, V. Z., Sousa, S. G., Santos, N. C., Delgado Mena, E., González Hernández, J. I., Israelian, G., Mayor, M., & Khachatryan, G. 2012b, A&A, 545, A32

Ahumada, R., Prieto, C. A., Almeida, A., Anders, F., Anderson, S. F., Andrews, B. H., Anguiano, B., Arcodia, R., Armengaud, E., Aubert, M., Avila, S., Avila-Reese, V., Badenes, C., Balland, C., Barger, K., Barrera-Ballesteros, J. K., Basu, S., Bautista, J., Beaton, R. L., Beers, T. C., Benavides, B. I. T., Bender, C. F., Bernardi, M., Bershady, M., Beutler, F., Bidin, C. M., Bird, J., Bizyaev, D., Blanc, G. A., Blanton, M. R., Boquien, M., Borissova, J., Bovy, J., Brandt, W. N., Brinkmann, J., Brownstein, J. R., Bundy, K., Bureau, M., Burgasser, A., Burtin, E., Cano-Díaz, M., Capasso, R., Cappellari, M., Carrera, R., Chabanier, S., Chaplin, W., Chapman, M., Cherinka, B., Chiappini, C., Doohyun Choi, P., Chojnowski, S. D., Chung, H., Clerc, N., Coffey, D., Comerford, J. M., Comparat, J., da Costa, L., Cousinou, M.-C., Covey, K., Crane, J. D., Cunha, K., Ilha, G. d. S., Dai, Y. S., Damsted, S. B., Darling, J., Davidson, James W., J., Davies, R., Dawson, K., De, N., de la Macorra, A., De Lee, N., Queiroz, A. B. d. A., Deconto Machado, A., de la Torre, S., Dell'Agli, F., du Mas des Bourboux, H., Diamond-Stanic, A. M., Dillon, S., Donor, J., Drory, N., Duckworth, C., Dwelly, T., Ebelke, G., Eftekharzadeh, S., Davis Eigenbrot, A., Elsworth, Y. P., Eracleous, M., Erfanianfar, G., Escoffier, S., Fan, X., Farr, E., Fernández-Trincado, J. G., Feuillet, D., Finoguenov, A., Fofie, P., Fraser-McKelvie, A., Frinchaboy, P. M., Fromenteau, S., Fu, H., Galbany, L., Garcia, R. A., García-Hernández, D. A., Oehmichen, L. A. G., Ge, J.,

Maia, M. A. G., Geisler, D., Gelfand, J., Goddy, J., Gonzalez-Perez, V., Grabowski, K., Green, P., Grier, C. J., Guo, H., Guy, J., Harding, P., Hasselquist, S., Hawken, A. J., Hayes, C. R., Hearty, F., Hekker, S., Hogg, D. W., Holtzman, J. A., Horta, D., Hou, J., Hsieh, B.-C., Huber, D., Hunt, J. A. S., Chitham, J. I., Imig, J., Jaber, M., Angel, C. E. J., Johnson, J. A., Jones, A. M., Jönsson, H., Jullo, E., Kim, Y., Kinemuchi, K., Kirkpatrick, Charles C., I., Kite, G. W., Klaene, M., Kneib, J.-P., Kollmeier, J. A., Kong, H., Kounkel, M., Krishnarao, D., Lacerna, I., Lan, T.-W., Lane, R. R., Law, D. R., Le Goff, J.-M., Leung, H. W., Lewis, H., Li, C., Lian, J., Lin, L., Long, D., Longa-Peña, P., Lundgren, B., Lyke, B. W., Ted Mackereth, J., MacLeod, C. L., Majewski, S. R., Manchado, A., Maraston, C., Martini, P., Masseron, T., Masters, K. L., Mathur, S., McDermid, R. M., Merloni, A., Merrifield, M., Mészáros, S., Miglio, A., Minniti, D., Minsley, R., Miyaji, T., Mohammad, F. G., Mosser, B., Mueller, E.-M., Muna, D., Muñoz-Gutiérrez, A., Myers, A. D., Nadathur, S., Nair, P., Nandra, K., do Nascimento, J. C., Nevin, R. J., Newman, J. A., Nidever, D. L., Nitschelm, C., Noterdaeme, P., O'Connell, J. E., Olmstead, M. D., Oravetz, D., Oravetz, A., Osorio, Y., Pace, Z. J., Padilla, N., Palanque-Delabrouille, N., Palicio, P. A., Pan, H.-A., Pan, K., Parker, J., Paviot, R., Peirani, S., Ramŕez, K. P., Penny, S., Percival, W. J., Perez-Fournon, I., Pérez-Ràfols, I., Petitjean, P., Pieri, M. M., Pinsonneault, M., Poovelil, V. J., Povick, J. T., Prakash, A., Price-Whelan, A. M., Raddick, M. J., Raichoor, A., Ray, A., Rembold, S. B., Rezaie, M., Riffel, R. A., Riffel, R., Rix, H.-W., Robin, A. C., Roman-Lopes, A., Román-Zúñiga, C., Rose, B., Ross, A. J., Rossi, G., Rowlands, K., Rubin, K. H. R., Salvato, M., Sánchez, A. G., Sánchez-Menguiano, L., Sánchez-Gallego, J. R., Sayres, C., Schaefer, A., Schiavon, R. P., Schimoia, J. S., Schlafly, E., Schlegel, D., Schneider, D. P., Schultheis, M., Schwope, A., Seo, H.-J., Serenelli, A., Shafieloo, A., Shamsi, S. J., Shao, Z., Shen, S., Shetrone, M., Shirley, R., Aguirre, V. S., Simon, J. D., Skrutskie, M. F., Slosar, A., Smethurst, R., Sobeck, J., Sodi, B. C., Souto, D., Stark, D. V., Stassun, K. G., Steinmetz, M., Stello, D., Stermer,

J., Storchi-Bergmann, T., Streblyanska, A., Stringfellow, G. S., Stutz, A., Suárez, G., Sun, J., Taghizadeh-Popp, M., Talbot, M. S., Tayar, J., Thakar, A. R., Theriault, R., Thomas, D., Thomas, Z. C., Tinker, J., Tojeiro, R., Toledo, H. H., Tremonti, C. A., Troup, N. W., Tuttle, S., Unda-Sanzana, E., Valentini, M., Vargas-González, J., Vargas-Magaña, M., Vázquez-Mata, J. A., Vivek, M., Wake, D., Wang, Y., Weaver, B. A., Weijmans, A.-M., Wild, V., Wilson, J. C., Wilson, R. F., Wolthuis, N., Wood-Vasey, W. M., Yan, R., Yang, M., Yèche, C., Zamora, O., Zarrouk, P., Zasowski, G., Zhang, K., Zhao, C., Zhao, G., Zheng, Z., Zheng, Z., Zhu, G., & Zou, H. 2020, ApJS, 249, 3

Akeson, R. L., Christiansen, J., Ciardi, D. R., Ramirez, S., Schlieder, J., Van Eyken, J. C., & NASA Exoplanet Archive Team. 2017, in American Astronomical Society Meeting Abstracts, Vol. 229, American Astronomical Society Meeting Abstracts, 146.16

Alibert, Y., Mordasini, C., & Benz, W. 2011, A&A, 526, A63

Allende Prieto, C., Beers, T. C., Wilhelm, R., Newberg, H. J., Rockosi, C. M., Yanny, B., & Lee, Y. S. 2006, ApJ, 636, 804

Anguiano, B., Majewski, S. R., Freeman, K. C., Mitschang, A. W., & Smith, M. C. 2018, MNRAS, 474, 854

Armitage, P. J. A Brief Overview of Planet Formation, ed. H. J. Deeg & J. A. Belmonte, 135

Bailer-Jones, C. A. L. 2015, PASP, 127, 994

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., & Andrae, R. 2018, AJ, 156, 58

Batalha, N. M. 2014, Proceedings of the National Academy of Science, 111, 12647

Batalha, N. M., Borucki, W. J., Koch, D. G., Brown, T. M., Caldwell, D. A., & Latham, D. W. 2010, Highlights of Astronomy, 15, 712

Beaugé, C. & Nesvorný, D. 2013, ApJ, 763, 12

Bedell, M., Bean, J. L., Meléndez, J., Mills, S. M., Fabrycky, D. C., Freitas, F. C., Ramírez, I., Asplund, M., Liu, F., & Yong, D. 2017, ApJ, 839, 94

Berger, T. A., Huber, D., Gaidos, E., & van Saders, J. L. 2018, ArXiv e-prints

Berger, T. A., Huber, D., Gaidos, E., van Saders, J. L., & Weiss, L. M. 2020a, AJ, 160, 108

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

Bertran de Lis, S., Allende Prieto, C., Majewski, S. R., Schiavon, R. P., Holtzman, J. A., Shetrone, M., Carrera, R., García Pérez, A. E., Mészáros, S., Frinchaboy, P. M., Hearty, F. R., Nidever, D. L., Zasowski, G., & Ge, J. 2016, A&A, 590, A74

Blanton, M. R., Bershady, M. A., Abolfathi, B., Albareti, F. D., Prieto, C. A., Almeida, A., Alonso-García, J., Anders, F., Anderson, S. F., Andrews, B., Aquino-Ortíz, E., Aragón-Salamanca, A., Argudo-Fernández, M., Armengaud, E., Aubourg, E., Avila-Reese, V., Badenes, C., Bailey, S., Barger, K. A., Barrera-Ballesteros, J., Bartosz, C., Bates, D., Baumgarten, F., Bautista, J., Beaton, R., Beers, T. C., Belfiore, F., Bender, C. F., Berlind, A. A., Bernardi, M., Beutler, F., Bird, J. C., Bizyaev, D., Blanc, G. A., Blomqvist, M., Bolton, A. S., Boquien, M., Borissova, J., van den Bosch, R., Bovy, J., Brandt, W. N., Brinkmann, J., Brownstein, J. R., Bundy, K., Burgasser, A. J., Burtin, E., Busca, N. G., Cappellari, M., Carigi, M. L. D., Carlberg, J. K., Rosell, A. C., Carrera, R., Chanover, N. J., Cherinka, B., Cheung, E., Chew, Y. G. M., Chiappini, C., Choi, P. D., Chojnowski, D., Chuang, C.-H., Chung, H., Cirolini, R. F.,

Clerc, N., Cohen, R. E., Comparat, J., da Costa, L., Cousinou, M.-C., Covey, K., Crane, J. D., Croft, R. A. C., Cruz-Gonzalez, I., Cuadra, D. G., Cunha, K., Damke, G. J., Darling, J., Davies, R., Dawson, K., de la Macorra, A., Dell'Agli, F., Lee, N. D., Delubac, T., Mille, F. D., Diamond-Stanic, A., Cano-Díaz, M., Donor, J., Downes, J. J., Drory, N., du Mas des Bourboux, H., Duckworth, C. J., Dwelly, T., Dyer, J., Ebelke, G., Eigenbrot, A. D., Eisenstein, D. J., Emsellem, E., Eracleous, M., Escoffier, S., Evans, M. L., Fan, X., Fernández-Alvar, E., Fernandez-Trincado, J. G., Feuillet, D. K., Finoguenov, A., Fleming, S. W., Font-Ribera, A., Fredrickson, A., Freischlad, G., Frinchaboy, P. M., Fuentes, C. E., Galbany, L., Garcia-Dias, R., García-Hernández, D. A., Gaulme, P., Geisler, D., Gelfand, J. D., Gil-Marín, H., Gillespie, B. A., Goddard, D., Gonzalez-Perez, V., Grabowski, K., Green, P. J., Grier, C. J., Gunn, J. E., Guo, H., Guy, J., Hagen, A., Hahn, C., Hall, M., Harding, P., Hasselquist, S., Hawley, S. L., Hearty, F., Hernández, J. I. G., Ho, S., Hogg, D. W., Holley-Bockelmann, K., Holtzman, J. A., Holzer, P. H., Huehnerhoff, J., Hutchinson, T. A., Hwang, H. S., Ibarra-Medel, H. J., da Silva Ilha, G., Ivans, I. I., Ivory, K., Jackson, K., Jensen, T. W., Johnson, J. A., Jones, A., Jönsson, H., Jullo, E., Kamble, V., Kinemuchi, K., Kirkby, D., Kitaura, F.-S., Klaene, M., Knapp, G. R., Kneib, J.-P., Kollmeier, J. A., Lacerna, I., Lane, R. R., Lang, D., Law, D. R., Lazarz, D., Lee, Y., Goff, J.-M. L., Liang, F.-H., Li, C., Li, H., Lian, J., Lima, M., Lin, L., Lin, Y.-T., de Lis, S. B., Liu, C., de Icaza Lizaola, M. A. C., Long, D., Lucatello, S., Lundgren, B., MacDonald, N. K., Machado, A. D., MacLeod, C. L., Mahadevan, S., Maia, M. A. G., Maiolino, R., Majewski, S. R., Malanushenko, E., Malanushenko, V., Manchado, A., Mao, S., Maraston, C., Marques-Chaves, R., Masseron, T., Masters, K. L., McBride, C. K., McDermid, R. M., McGrath, B., McGreer, I. D., Peña, N. M., Melendez, M., Merloni, A., Merrifield, M. R., Meszaros, S., Meza, A., Minchev, I., Minniti, D., Miyaji, T., More, S., Mulchaey, J., Müller-Sánchez, F., Muna, D., Munoz, R. R., Myers, A. D., Nair, P., Nandra, K., do Nascimento, J. C., Negrete, A., Ness, M., Newman, J. A., Nichol, R. C.,

Nidever, D. L., Nitschelm, C., Ntelis, P., O'Connell, J. E., Oelkers, R. J., Oravetz, A., Oravetz, D., Pace, Z., Padilla, N., Palanque-Delabrouille, N., Palicio, P. A., Pan, K., Parejko, J. K., Parikh, T., Pâris, I., Park, C., Patten, A. Y., Peirani, S., Pellejero-Ibanez, M., Penny, S., Percival, W. J., Perez-Fournon, I., Petitjean, P., Pieri, M. M., Pinsonneault, M., Pisani, A., Poleski, R., Prada, F., Prakash, A., de Andrade Queiroz, A. B., Raddick, M. J., Raichoor, A., Rembold, S. B., Richstein, H., Riffel, R. A., Riffel, R., Rix, H.-W., Robin, A. C., Rockosi, C. M., Rodríguez-Torres, S., Roman-Lopes, A., Román-Zúñiga, C., Rosado, M., Ross, A. J., Rossi, G., Ruan, J., Ruggeri, R., Rykoff, E. S., Salazar-Albornoz, S., Salvato, M., Sánchez, A. G., Aguado, D. S., Sánchez-Gallego, J. R., Santana, F. A., Santiago, B. X., Sayres, C., Schiavon, R. P., da Silva Schimoia, J., Schlafly, E. F., Schlegel, D. J., Schneider, D. P., Schultheis, M., Schuster, W. J., Schwope, A., Seo, H.-J., Shao, Z., Shen, S., Shetrone, M., Shull, M., Simon, J. D., Skinner, D., Skrutskie, M. F., Slosar, A., Smith, V. V., Sobeck, J. S., Sobreira, F., Somers, G., Souto, D., Stark, D. V., Stassun, K., Stauffer, F., Steinmetz, M., Storchi-Bergmann, T., Streblyanska, A., Stringfellow, G. S., Suárez, G., Sun, J., Suzuki, N., Szigeti, L., Taghizadeh-Popp, M., Tang, B., Tao, C., Tayar, J., Tembe, M., Teske, J., Thakar, A. R., Thomas, D., Thompson, B. A., Tinker, J. L., Tissera, P., Tojeiro, R., Toledo, H. H., de la Torre, S., Tremonti, C., Troup, N. W., Valenzuela, O., Valpuesta, I. M., Vargas-González, J., Vargas-Magaña, M., Vazquez, J. A., Villanova, S., Vivek, M., Vogt, N., Wake, D., Walterbos, R., Wang, Y., Weaver, B. A., Weijmans, A.-M., Weinberg, D. H., Westfall, K. B., Whelan, D. G., Wild, V., Wilson, J., Wood-Vasey, W. M., Wylezalek, D., Xiao, T., Yan, R., Yang, M., Ybarra, J. E., Yèche, C., Zakamska, N., Zamora, O., Zarrouk, P., Zasowski, G., Zhang, K., Zhao, G.-B., Zheng, Z., Zheng, Z., Zhou, X., Zhou, Z.-M., Zhu, G. B., Zoccali, M., & Zou, H. 2017, The Astronomical Journal, 154, 28

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

Borucki, W. J., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., Caldwell, J., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Dupree, A. K., Gautier, T. N., Geary, J. C., Gilliland, R., Gould, A., Howell, S. B., Jenkins, J. M., Kondo, Y., Latham, D. W., Marcy, G. W., Meibom, S., Kjeldsen, H., Lissauer, J. J., Monet, D. G., Morrison, D., Sasselov, D., Tarter, J., Boss, A., Brownlee, D., Owen, T., Buzasi, D., Charbonneau, D., Doyle, L., Fortney, J., Ford, E. B., Holman, M. J., Seager, S., Steffen, J. H., Welsh, W. F., Rowe, J., Anderson, H., Buchhave, L., Ciardi, D., Walkowicz, L., Sherry, W., Horch, E., Isaacson, H., Everett, M. E., Fischer, D., Torres, G., Johnson, J. A., Endl, M., MacQueen, P., Bryson, S. T., Dotson, J., Haas, M., Kolodziejczak, J., Van Cleve, J., Chandrasekaran, H., Twicken, J. D., Quintana, E. V., Clarke, B. D., Allen, C., Li, J., Wu, H., Tenenbaum, P., Verner, E., Bruhweiler, F., Barnes, J., & Prsa, A. 2010, Science, 327, 977

Bossini, D., Vallenari, A., Bragaglia, A., Cantat-Gaudin, T., Sordo, R., Balaguer-Núñez, L., Jordi, C., Moitinho, A., Soubiran, C., Casamiquela, L., Carrera, R., & Heiter, U. 2019, A&A, 623, A108

Bouvier, J. 2013, in EAS Publications Series, Vol. 62, EAS Publications Series, ed. P. Hennebelle & C. Charbonnel, 143–168

Bowler, B. P., Shkolnik, E. L., Liu, M. C., Schlieder, J. E., Mann, A. W., Dupuy, T. J., Hinkley, S., Crepp, J. R., Johnson, J. A., Howard, A. W., Flagg, L., Weinberger, A. J., Aller, K. M., Allers, K. N., Best, W. M. J., Kotson, M. C., Montet, B. T., Herczeg, G. J., Baranec, C., Riddle, R., Law, N. M., Nielsen, E. L., Wahhaj, Z., Biller, B. A., & Hayward, T. L. 2015, ApJ, 806, 62

Boyajian, T. S., von Braun, K., van Belle, G., Farrington, C., Schaefer, G., Jones, J., White, R., McAlister, H. A., ten Brummelaar, T. A., Ridgway, S., Gies, D., Sturmann, L., Sturmann, J., Turner, N. H., Goldfinger, P. J., & Vargas, N. 2013, The Astrophysical Journal, 771, 40

Brewer, J. M., Fischer, D. A., Valenti, J. A., & Piskunov, N. 2016, ApJS, 225, 32

Brown, T. M. 1991, ApJ, 371, 396

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

Brugamyer, E., Dodson-Robinson, S. E., Cochran, W. D., & Sneden, C. 2011, ApJ, 738, 97

Bruntt, H., Basu, S., Smalley, B., Chaplin, W. J., Verner, G. A., Bedding, T. R., Catala, C., Gazzano, J.-C., Molenda-Żakowicz, J., Thygesen, A. O., Uytterhoeven, K., Hekker, S., Huber, D., Karoff, C., Mathur, S., Mosser, B., Appourchaux, T., Campante, T. L., Elsworth, Y., García, R. A., Handberg, R., Metcalfe, T. S., Quirion, P.-O., Régulo, C., Roxburgh, I. W., Stello, D., Christensen-Dalsgaard, J., Kawaler, S. D., Kjeldsen, H., Morris, R. L., Quintana, E. V., & Sanderfer, D. T. 2012, MNRAS, 423, 122

Bruntt, H., Bedding, T. R., Quirion, P.-O., Lo Curto, G., Carrier, F., Smalley, B., Dall, T. H., Arentoft, T., Bazot, M., & Butler, R. P. 2010, MNRAS, 405, 1907

Bryson, S., Coughlin, J. L., Kunimoto, M., & Mullally, S. E. 2020, AJ, 160, 200

Buchhave, L. A., Bizzarro, M., Latham, D. W., Sasselov, D., Cochran, W. D., Endl, M., Isaacson, H., Juncher, D., & Marcy, G. W. 2014, Nature, 509, 593

Buchhave, L. A. & Latham, D. W. 2015, ApJ, 808, 187

Buchhave, L. A., Latham, D. W., Johansen, A., Bizzarro, M., Torres, G., Rowe, J. F., Batalha, N. M., Borucki, W. J., Brugamyer, E., Caldwell, C., Bryson, S. T., Ciardi, D. R., Cochran, W. D., Endl, M., Esquerdo, G. A., Ford, E. B., Geary, J. C., Gilliland, R. L., Hansen, T., Isaacson, H., Laird, J. B., Lucas, P. W., Marcy, G. W., Morse, J. A., Robertson, P., Shporer, A., Stefanik, R. P., Still, M., & Quinn, S. N. 2012, Nature, 486, 375

Burke, C. J. & Catanzarite, J. 2017a, Planet Detection Metrics: Per-Target Flux-Level Transit Injection Tests of TPS for Data Release 25, Kepler Science Document KSCI-19109-002

—. 2017b, Planet Detection Metrics: Per-Target Flux-Level Transit Injection Tests of TPS for Data Release 25, Kepler Science Document

Burke, C. J., Christiansen, J. L., Mullally, F., Seader, S., Huber, D., Rowe, J. F., Coughlin, J. L., Thompson, S. E., Catanzarite, J., Clarke, B. D., Morton, T. D., Caldwell, D. A., Bryson, S. T., Haas, M. R., Batalha, N. M., Jenkins, J. M., Tenenbaum, P., Twicken, J. D., Li, J., Quintana, E., Barclay, T., Henze, C. E., Borucki, W. J., Howell, S. B., & Still, M. 2015, ApJ, 809, 8

Cantat-Gaudin, T., Jordi, C., Vallenari, A., Bragaglia, A., Balaguer-Núñez, L., Soubiran, C., Bossini, D., Moitinho, A., Castro-Ginard, A., Krone-Martins, A., Casamiquela, L., Sordo, R., & Carrera, R. 2018, A&A, 618, A93

Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, ApJ, 529, L45

Choi, J., Dotter, A., Conroy, C., Cantiello, M., Paxton, B., & Johnson, B. D. 2016, ApJ, 823, 102

Choi, P. I. & Herbst, W. 1996, AJ, 111, 283

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

Christiansen, J. L., Clarke, B. D., Burke, C. J., Jenkins, J. M., Bryson, S. T., Coughlin, J. L., Mullally, F., Thompson, S. E., Twicken, J. D., Batalha, N. M., Haas, M. R., Catanzarite, J., Campbell, J. R., Kamal Uddin, A., Zamudio, K., Smith, J. C., & Henze, C. E. 2016, ApJ, 828, 99

Christiansen, J. L., Clarke, B. D., Burke, C. J., Jenkins, J. M., Bryson, S. T., Coughlin, J. L., Mullally, S. E., Twicken, J. D., Batalha, N. M., Catanzarite, J., Uddin, A. K., Zamudio, K., Smith, J. C., Henze, C. E., & Campbell, J. 2020, AJ, 160, 159

Christiansen, J. L., Clarke, B. D., Burke, C. J., Seader, S., Jenkins, J. M., Twicken, J. D., Catanzarite, J. D., Smith, J. C., Batalha, N. M., Haas, M. R., Thompson, S. E., Campbell, J. R., Sabale, A., & Kamal Uddin, A. 2015, ApJ, 810, 95

Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., Burke, C. J., Tenenbaum, P., Seader, S., Thompson, S. E., Barclay, T. S., Clarke, B. D., Li, J., Smith, J. C., Stumpe, M. C., Twicken, J. D., & Van Cleve, J. 2012, PASP, 124, 1279

Claret, A. 2018, A&A, 618, A20

Coughlin, J. 2015, IAU General Assembly, 22, 2257510

Crossfield, I. J. M., Ciardi, D. R., Petigura, E. A., Sinukoff, E., Schlieder, J. E., Howard, A. W., Beichman, C. A., Isaacson, H., Dressing, C. D., Christiansen, J. L., Fulton, B. J., Lépine, S., Weiss, L., Hirsch, L., Livingston, J., Baranec, C., Law, N. M., Riddle, R., Ziegler, C., Howell, S. B., Horch, E., Everett, M., Teske, J., Martinez, A. O., Obermeier, C., Benneke, B., Scott, N., Deacon, N., Aller, K. M., Hansen, B. M. S., Mancini, L., Ciceri, S., Brahm, R., Jordán, A., Knutson, H. A., Henning, T., Bonnefoy, M., Liu, M. C., Crepp, J. R., Lothringer, J., Hinz, P., Bailey, V., Skemer, A., & Defrere, D. 2016, ApJS, 226, 7

Crouzet, N., Guillot, T., Agabi, A., Rivet, J. P., Bondoux, E., Challita, Z., Fanteï-Caujolle, Y., Fressin, F., Mékarnia, D., Schmider, F. X., Valbousquet, F., Blazit, A., Bonhomme, S., Abe, L., Daban, J. B., Gouvret, C., Fruth, T., Rauer, H., Erikson, A., Barbieri, M., Aigrain, S., & Pont, F. 2010, A&A, 511, A36

Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., Li, G.-P., Li, Q., Zhang, L.-P., Su, H.-J., Yao, Z.-Q., Wang, Y.-N., Xing, X.-Z., Li, X.-N., Zhu, Y.-T., Wang, G., Gu, B.-Z., Luo, A.-L., Xu,

X.-Q., Zhang, Z.-C., Liu, G.-R., Zhang, H.-T., Yang, D.-H., Cao, S.-Y., Chen, H.-Y., Chen, J.-J., Chen, K.-X., Chen, Y., Chu, J.-R., Feng, L., Gong, X.-F., Hou, Y.-H., Hu, H.-Z., Hu, N.-S., Hu, Z.-W., Jia, L., Jiang, F.-H., Jiang, X., Jiang, Z.-B., Jin, G., Li, A.-H., Li, Y., Li, Y.-P., Liu, G.-Q., Liu, Z.-G., Lu, W.-Z., Mao, Y.-D., Men, L., Qi, Y.-J., Qi, Z.-X., Shi, H.-M., Tang, Z.-H., Tao, Q.-S., Wang, D.-Q., Wang, D., Wang, G.-M., Wang, H., Wang, J.-N., Wang, J., Wang, J.-L., Wang, J.-P., Wang, L., Wang, S.-Q., Wang, Y., Wang, Y.-F., Xu, L.-Z., Xu, Y., Yang, S.-H., Yu, Y., Yuan, H., Yuan, X.-Y., Zhai, C., Zhang, J., Zhang, Y.-X., Zhang, Y., Zhao, M., Zhou, F., Zhou, G.-H., Zhu, J., & Zou, S.-C. 2012, Research in Astronomy and Astrophysics, 12, 1197

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

Cunha, K., Smith, V. V., Johnson, J. A., Bergemann, M., Mészáros, S., Shetrone, M. D., Souto, D., Allende Prieto, C., Schiavon, R. P., Frinchaboy, P., Zasowski, G., Bizyaev, D., Holtzman, J., García Pérez, A. E., Majewski, S. R., Nidever, D., Beers, T., Carrera, R., Geisler, D., Gunn, J., Hearty, F., Ivans, I., Martell, S., Pinsonneault, M., Schneider, D. P., Sobeck, J., Stello, D., Stassun, K. G., Skrutskie, M., & Wilson, J. C. 2015, ApJ, 798, L41

David, T. J., Contardo, G., Sandoval, A., Angus, R., Lu, Y. L., Bedell, M., Curtis, J. L., Foreman-Mackey, D., Fulton, B. J., Grunblatt, S. K., & Petigura, E. A. 2021, AJ, 161, 265

Dawson, R. I., Chiang, E., & Lee, E. J. 2015, MNRAS, 453, 1471

Dawson, R. I. & Murray-Clay, R. A. 2013, ApJ, 767, L24

De Cat, P., Fu, J. N., Ren, A. B., Yang, X. H., Shi, J. R., Luo, A. L., Yang, M., Wang, J. L., Zhang, H. T., Shi, H. M., Zhang, W., Dong, S., Catanzaro, G., Corbally, C. J., Frasca, A., Gray, R. O., Molenda-Żakowicz, J., Uytterhoeven, K., Briquet, M., Bruntt,

H., Frandsen, S., Kiss, L., Kurtz, D. W., Marconi, M., Niemczura, E., Østensen, R. H., Ripepi, V., Smalley, B., Southworth, J., Szabó, R., Telting, J. H., Karoff, C., Silva Aguirre, V., Wu, Y., Hou, Y. H., Jin, G., & Zhou, X. L. 2015, ApJS, 220, 19

Dong, S., Xie, J.-W., Zhou, J.-L., Zheng, Z., & Luo, A. 2017, ArXiv e-prints

Dong, S., Zheng, Z., Zhu, Z., De Cat, P., Fu, J. N., Yang, X. H., Zhang, H., Jin, G., & Zhang, Y. 2014, ApJ, 789, L3

Donor, J., Frinchaboy, P. M., Cunha, K., O'Connell, J. E., Allende Prieto, C., Almeida, A., Anders, F., Beaton, R., Bizyaev, D., Brownstein, J. R., Carrera, R., Chiappini, C., Cohen, R., García-Hernández, D. A., Geisler, D., Hasselquist, S., Jönsson, H., Lane, R. R., Majewski, S. R., Minniti, D., Bidin, C. M., Pan, K., Roman-Lopes, A., Sobeck, J. S., & Zasowski, G. 2020, AJ, 159, 199

Dorn, C., Hinkel, N. R., & Venturini, J. 2017a, A&A, 597, A38

Dorn, C., Venturini, J., Khan, A., Heng, K., Alibert, Y., Helled, R., Rivoldini, A., & Benz, W. 2017b, A&A, 597, A37

Dotter, A. 2016, ApJS, 222, 8

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

—. 2015, ApJ, 807, 45

Eisenstein, D. J., Weinberg, D. H., Agol, E., Aihara, H., Allende Prieto, C., Anderson, S. F., Arns, J. A., Aubourg, É., Bailey, S., Balbinot, E., & et al. 2011, AJ, 142, 72

Eisenstein, D. J., Weinberg, D. H., Agol, E., Aihara, H., Prieto, C. A., Anderson, S. F., Arns, J. A., Éric Aubourg, Bailey, S., Balbinot, E., Barkhouser, R., Beers, T. C., Berlind, A. A., Bickerton, S. J., Bizyaev, D., Blanton, M. R., Bochanski, J. J., Bolton, A. S., Bosman, C. T., Bovy, J., Brandt, W. N., Breslauer, B., Brewington, H. J.,

Brinkmann, J., Brown, P. J., Brownstein, J. R., Burger, D., Busca, N. G., Campbell, H., Cargile, P. A., Carithers, W. C., Carlberg, J. K., Carr, M. A., Chang, L., Chen, Y., Chiappini, C., Comparat, J., Connolly, N., Cortes, M., Croft, R. A. C., Cunha, K., da Costa, L. N., Davenport, J. R. A., Dawson, K., Lee, N. D., de Mello, G. F. P., de Simoni, F., Dean, J., Dhital, S., Ealet, A., Ebelke, G. L., Edmondson, E. M., Eiting, J. M., Escoffier, S., Esposito, M., Evans, M. L., Fan, X., Castellá, B. F., Ferreira, L. D., Fitzgerald, G., Fleming, S. W., Font-Ribera, A., Ford, E. B., Frinchaboy, P. M., Pérez, A. E. G., Gaudi, B. S., Ge, J., Ghezzi, L., Gillespie, B. A., Gilmore, G., Girardi, L., Gott, J. R., Gould, A., Grebel, E. K., Gunn, J. E., Hamilton, J.-C., Harding, P., Harris, D. W., Hawley, S. L., Hearty, F. R., Hennawi, J. F., Hernández, J. I. G., Ho, S., Hogg, D. W., Holtzman, J. A., Honscheid, K., Inada, N., Ivans, I. I., Jiang, L., Jiang, P., Johnson, J. A., Jordan, C., Jordan, W. P., Kauffmann, G., Kazin, E., Kirkby, D., Klaene, M. A., Knapp, G. R., Kneib, J.-P., Kochanek, C. S., Koesterke, L., Kollmeier, J. A., Kron, R. G., Lampeitl, H., Lang, D., Lawler, J. E., Goff, J.-M. L., Lee, B. L., Lee, Y. S., Leisenring, J. M., Lin, Y.-T., Liu, J., Long, D. C., Loomis, C. P., Lucatello, S., Lundgren, B., Lupton, R. H., Ma, B., Ma, Z., MacDonald, N., Mack, C., Mahadevan, S., Maia, M. A. G., Majewski, S. R., Makler, M., Malanushenko, E., Malanushenko, V., Mandelbaum, R., Maraston, C., Margala, D., Maseman, P., Masters, K. L., McBride, C. K., McDonald, P., McGreer, I. D., McMahon, R. G., Requejo, O. M., Ménard, B., Miralda-Escudé, J., Morrison, H. L., Mullally, F., Muna, D., Murayama, H., Myers, A. D., Naugle, T., Neto, A. F., Nguyen, D. C., Nichol, R. C., Nidever, D. L., O'Connell, R. W., Ogando, R. L. C., Olmstead, M. D., Oravetz, D. J., Padmanabhan, N., Paegert, M., Palanque-Delabrouille, N., Pan, K., Pandey, P., Parejko, J. K., Pâris, I., Pellegrini, P., Pepper, J., Percival, W. J., Petitjean, P., Pfaffenberger, R., Pforr, J., Phleps, S., Pichon, C., Pieri, M. M., Prada, F., Price-Whelan, A. M., Raddick, M. J., Ramos, B. H. F., Reid, I. N., Reyle, C., Rich, J., Richards, G. T., Rieke, G. H., Rieke, M. J., Rix, H.-W., Robin, A. C., Rocha-Pinto,

H. J., Rockosi, C. M., Roe, N. A., Rollinde, E., Ross, A. J., Ross, N. P., Rossetto, B., Sánchez, A. G., Santiago, B., Sayres, C., Schiavon, R., Schlegel, D. J., Schlesinger, K. J., Schmidt, S. J., Schneider, D. P., Sellgren, K., Shelden, A., Sheldon, E., Shetrone, M., Shu, Y., Silverman, J. D., Simmerer, J., Simmons, A. E., Sivarani, T., Skrutskie, M. F., Slosar, A., Smee, S., Smith, V. V., Snedden, S. A., Stassun, K. G., Steele, O., Steinmetz, M., Stockett, M. H., Stollberg, T., Strauss, M. A., Szalay, A. S., Tanaka, M., Thakar, A. R., Thomas, D., Tinker, J. L., Tofflemire, B. M., Tojeiro, R., Tremonti, C. A., Magaña, M. V., Verde, L., Vogt, N. P., Wake, D. A., Wan, X., Wang, J., Weaver, B. A., White, M., White, S. D. M., Wilson, J. C., Wisniewski, J. P., Wood-Vasey, W. M., Yanny, B., Yasuda, N., Yèche, C., York, D. G., Young, E., Zasowski, G., Zehavi, I., & Zhao, B. 2011, The Astronomical Journal, 142, 72

Eisner, J. A., Hillenbrand, L. A., White, R. J., Akeson, R. L., & Sargent, A. I. 2005, ApJ, 623, 952

Eistrup, C., Walsh, C., & van Dishoeck, E. F. 2018, A&A, 613, A14

Ercolano, B. & Clarke, C. J. 2010, MNRAS, 402, 2735

Everett, M. E., Howell, S. B., & Kinemuchi, K. 2012, PASP, 124, 316

Everett, M. E., Howell, S. B., Silva, D. R., & Szkody, P. 2013, ApJ, 771, 107

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

Fleming, S. W., Mahadevan, S., Deshpande, R., Bender, C. F., Terrien, R. C., Marchwinski, R. C., Wang, J., Roy, A., Stassun, K. G., Allende Prieto, C., Cunha, K., Smith, V. V., Agol, E., Ak, H., Bastien, F. A., Bizyaev, D., Crepp, J. R., Ford, E. B., Frinchaboy, P. M., García-Hernández, D. A., García Pérez, A. E., Gaudi, B. S., Ge, J., Hearty, F., Ma, B., Majewski, S. R., Mészáros, S., Nidever, D. L., Pan, K., Pepper, J., Pinsonneault, M. H., Schiavon, R. P., Schneider, D. P., Wilson, J. C., Zamora, O., & Zasowski, G. 2015, AJ, 149, 143

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

Fressin, F., Torres, G., Charbonneau, D., Bryson, S. T., Christiansen, J., Dressing, C. D., Jenkins, J. M., Walkowicz, L. M., & Batalha, N. M. 2013, ApJ, 766, 81

Frinchaboy, P. M., Thompson, B., Jackson, K. M., O'Connell, J., Meyer, B., Zasowski, G., Majewski, S. R., Chojnowksi, S. D., Johnson, J. A., Allende Prieto, C., Beers, T. C., Bizyaev, D., Brewington, H., Cunha, K., Ebelke, G., García Pérez, A. E., Hearty, F. R., Holtzman, J., Kinemuchi, K., Malanushenko, E., Malanushenko, V., Marchante, M., Mészáros, S., Muna, D., Nidever, D. L., Oravetz, D., Pan, K., Schiavon, R. P., Schneider, D. P., Shetrone, M., Simmons, A., Snedden, S., Smith, V. V., & Wilson, J. C. 2013, ApJ, 777, L1

Fulton, B. J. & Petigura, E. A. 2018, ArXiv e-prints

Fulton, B. J., Petigura, E. A., Howard, A. W., Isaacson, H., Marcy, G. W., Cargile, P. A., Hebb, L., Weiss, L. M., Johnson, J. A., Morton, T. D., Sinukoff, E., Crossfield, I. J. M., & Hirsch, L. A. 2017, ArXiv e-prints

Furlan, E., Ciardi, D. R., Cochran, W. D., Everett, M. E., Latham, D. W., Marcy, G. W., Buchhave, L. A., Endl, M., Isaacson, H., Petigura, E. A., Gautier, T. N., I., Huber, D., Bieryla, A., Borucki, W. J., Brugamyer, E., Caldwell, C., Cochran, A., Howard, A. W., Howell, S. B., Johnson, M. C., MacQueen, P. J., Quinn, S. N., Robertson, P., Mathur, S., & Batalha, N. M. 2018, ApJ, 861, 149

Gaia Collaboration, Babusiaux, C., van Leeuwen, F., Barstow, M. A., Jordi, C., Vallenari, A., Bossini, D., Bressan, A., Cantat-Gaudin, T., van Leeuwen, M., Brown, A. G. A., Prusti, T., de Bruijne, J. H. J., Bailer-Jones, C. A. L., Biermann, M., Evans, D. W., Eyer, L., Jansen, F., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Panem, C., Pourbaix, D., Randich, S., Sartoretti, P., Siddiqui, H. I., Soubiran, C., Walton, N. A., Arenou, F., Bastian, U., Cropper, M., Drimmel, R., Katz, D.,

Lattanzi, M. G., Bakker, J., Cacciari, C., Castañeda, J., Chaoul, L., Cheek, N., De Angeli, F., Fabricius, C., Guerra, R., Holl, B., Masana, E., Messineo, R., Mowlavi, N., Nienartowicz, K., Panuzzo, P., Portell, J., Riello, M., Seabroke, G. M., Tanga, P., Thévenin, F., Gracia-Abril, G., Comoretto, G., Garcia-Reinaldos, M., Teyssier, D., Altmann, M., Andrae, R., Audard, M., Bellas-Velidis, I., Benson, K., Berthier, J., Blomme, R., Burgess, P., Busso, G., Carry, B., Cellino, A., Clementini, G., Clotet, M., Creevey, O., Davidson, M., De Ridder, J., Delchambre, L., Dell'Oro, A., Ducourant, C., Fernández-Hernández, J., Fouesneau, M., Frémat, Y., Galluccio, L., García-Torres, M., González-Núñez, J., González-Vidal, J. J., Gosset, E., Guy, L. P., Halbwachs, J. L., Hambly, N. C., Harrison, D. L., Hernández, J., Hestroffer, D., Hodgkin, S. T., Hutton, A., Jasniewicz, G., Jean-Antoine-Piccolo, A., Jordan, S., Korn, A. J., Krone-Martins, A., Lanzafame, A. C., Lebzelter, T., Löffler, W., Manteiga, M., Marrese, P. M., Martín-Fleitas, J. M., Moitinho, A., Mora, A., Muinonen, K., Osinde, J., Pancino, E., Pauwels, T., Petit, J. M., Recio-Blanco, A., Richards, P. J., Rimoldini, L., Robin, A. C., Sarro, L. M., Siopis, C., Smith, M., Sozzetti, A., Süveges, M., Torra, J., van Reeven, W., Abbas, U., Abreu Aramburu, A., Accart, S., Aerts, C., Altavilla, G., Álvarez, M. A., Alvarez, R., Alves, J., Anderson, R. I., Andrei, A. H., Anglada Varela, E., Antiche, E., Antoja, T., Arcay, B., Astraatmadja, T. L., Bach, N., Baker, S. G., Balaguer-Núñez, L., Balm, P., Barache, C., Barata, C., Barbato, D., Barblan, F., Barklem, P. S., Barrado, D., Barros, M., Bartholomé Muñoz, L., Bassilana, J. L., Becciani, U., Bellazzini, M., Berihuete, A., Bertone, S., Bianchi, L., Bienaymé, O., Blanco-Cuaresma, S., Boch, T., Boeche, C., Bombrun, A., Borrachero, R., Bouquillon, S., Bourda, G., Bragaglia, A., Bramante, L., Breddels, M. A., Brouillet, N., Brüsemeister, T., Brugaletta, E., Bucciarelli, B., Burlacu, A., Busonero, D., Butkevich, A. G., Buzzi, R., Caffau, E., Cancelliere, R., Cannizzaro, G., Carballo, R., Carlucci, T., Carrasco, J. M., Casamiquela, L., Castellani, M., Castro-Ginard, A., Charlot, P., Chemin, L., Chiavassa, A., Cocozza, G., Costigan, G., Cowell, S., Crifo, F., Crosta, M., Crowley,

C., Cuypers, J., Dafonte, C., Damerdji, Y., Dapergolas, A., David, P., David, M., de Laverny, P., De Luise, F., De March, R., de Martino, D., de Souza, R., de Torres, A., Debosscher, J., del Pozo, E., Delbo, M., Delgado, A., Delgado, H. E., Diakite, S., Diener, C., Distefano, E., Dolding, C., Drazinos, P., Durán, J., Edvardsson, B., Enke, H., Eriksson, K., Esquej, P., Eynard Bontemps, G., Fabre, C., Fabrizio, M., Faigler, S., Falcão, A. J., Farràs Casas, M., Federici, L., Fedorets, G., Fernique, P., Figueras, F., Filippi, F., Findeisen, K., Fonti, A., Fraile, E., Fraser, M., Frézouls, B., Gai, M., Galleti, S., Garabato, D., García-Sedano, F., Garofalo, A., Garralda, N., Gavel, A., Gavras, P., Gerssen, J., Geyer, R., Giacobbe, P., Gilmore, G., Girona, S., Giuffrida, G., Glass, F., Gomes, M., Granvik, M., Gueguen, A., Guerrier, A., Guiraud, J., Gutié, R., Haigron, R., Hatzidimitriou, D., Hauser, M., Haywood, M., Heiter, U., Helmi, A., Heu, J., Hilger, T., Hobbs, D., Hofmann, W., Holland, G., Huckle, H. E., Hypki, A., Icardi, V., Janßen, K., Jevardat de Fombelle, G., Jonker, P. G., Juhász, Á. L., Julbe, F., Karampelas, A., Kewley, A., Klar, J., Kochoska, A., Kohley, R., Kolenberg, K., Kontizas, M., Kontizas, E., Koposov, S. E., Kordopatis, G., Kostrzewa-Rutkowska, Z., Koubsky, P., Lambert, S., Lanza, A. F., Lasne, Y., Lavigne, J. B., Le Fustec, Y., Le Poncin-Lafitte, C., Lebreton, Y., Leccia, S., Leclerc, N., Lecoeur-Taibi, I., Lenhardt, H., Leroux, F., Liao, S., Licata, E., Lindstrøm, H. E. P., Lister, T. A., Livanou, E., Lobel, A., López, M., Managau, S., Mann, R. G., Mantelet, G., Marchal, O., Marchant, J. M., Marconi, M., Marinoni, S., Marschalkó, G., Marshall, D. J., Martino, M., Marton, G., Mary, N., Massari, D., Matijevič, G., Mazeh, T., McMillan, P. J., Messina, S., Michalik, D., Millar, N. R., Molina, D., Molinaro, R., Molnár, L., Montegriffo, P., Mor, R., Morbidelli, R., Morel, T., Morris, D., Mulone, A. F., Muraveva, T., Musella, I., Nelemans, G., Nicastro, L., Noval, L., O'Mullane, W., Ordénovic, C., Ordóñez-Blanco, D., Osborne, P., Pagani, C., Pagano, I., Pailler, F., Palacin, H., Palaversa, L., Panahi, A., Pawlak, M., Piersimoni, A. M., Pineau, F. X., Plachy, E., Plum, G., Poggio, E., Poujoulet, E., Prša, A., Pulone, L., Racero, E.,

Ragaini, S., Rambaux, N., Ramos-Lerate, M., Regibo, S., Reylé, C., Riclet, F., Ripepi, V., Riva, A., Rivard, A., Rixon, G., Roegiers, T., Roelens, M., Romero-Gómez, M., Rowell, N., Royer, F., Ruiz-Dern, L., Sadowski, G., Sagristà Sellés, T., Sahlmann, J., Salgado, J., Salguero, E., Sanna, N., Santana-Ros, T., Sarasso, M., Savietto, H., Schultheis, M., Sciacca, E., Segol, M., Segovia, J. C., Ségransan, D., Shih, I. C., Siltala, L., Silva, A. F., Smart, R. L., Smith, K. W., Solano, E., Solitro, F., Sordo, R., Soria Nieto, S., Souchay, J., Spagna, A., Spoto, F., Stampa, U., Steele, I. A., Steidelmüller, H., Stephenson, C. A., Stoev, H., Suess, F. F., Surdej, J., Szabados, L., Szegedi-Elek, E., Tapiador, D., Taris, F., Tauran, G., Taylor, M. B., Teixeira, R., Terrett, D., Teyssandier, P., Thuillot, W., Titarenko, A., Torra Clotet, F., Turon, C., Ulla, A., Utrilla, E., Uzzi, S., Vaillant, M., Valentini, G., Valette, V., van Elteren, A., Van Hemelryck, E., Vaschetto, M., Vecchiato, A., Veljanoski, J., Viala, Y., Vicente, D., Vogt, S., von Essen, C., Voss, H., Votruba, V., Voutsinas, S., Walmsley, G., Weiler, M., Wertz, O., Wevers, T., Wyrzykowski, Ł., Yoldas, A., Žerjal, M., Ziaeepour, H., Zorec, J., Zschocke, S., Zucker, S., Zurbach, C., & Zwitter, T. 2018a, A&A, 616, A10

Gaia Collaboration, Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., Babusiaux, C., Biermann, M., Creevey, O. L., Evans, D. W., Eyer, L., Hutton, A., Jansen, F., Jordi, C., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Panem, C., Pourbaix, D., Randich, S., Sartoretti, P., Soubiran, C., Walton, N. A., Arenou, F., Bailer-Jones, C. A. L., Bastian, U., Cropper, M., Drimmel, R., Katz, D., Lattanzi, M. G., van Leeuwen, F., Bakker, J., Cacciari, C., Castañeda, J., De Angeli, F., Ducourant, C., Fabricius, C., Fouesneau, M., Frémat, Y., Guerra, R., Guerrier, A., Guiraud, J., Jean-Antoine Piccolo, A., Masana, E., Messineo, R., Mowlavi, N., Nicolas, C., Nienartowicz, K., Pailler, F., Panuzzo, P., Riclet, F., Roux, W., Seabroke, G. M., Sordo, R., Tanga, P., Thévenin, F., Gracia-Abril, G., Portell, J., Teyssier, D., Altmann, M., Andrae, R., Bellas-Velidis, I., Benson, K., Berthier, J., Blomme, R., Brugaletta, E., Burgess, P. W., Busso, G., Carry, B., Cellino, A., Cheek, N., Clementini, G.,

Damerdji, Y., Davidson, M., Delchambre, L., Dell'Oro, A., Fernández-Hernández, J., Galluccio, L., García-Lario, P., Garcia-Reinaldos, M., González-Núñez, J., Gosset, E., Haigron, R., Halbwachs, J. L., Hambly, N. C., Harrison, D. L., Hatzidimitriou, D., Heiter, U., Hernández, J., Hestroffer, D., Hodgkin, S. T., Holl, B., Janßen, K., Jevardat de Fombelle, G., Jordan, S., Krone-Martins, A., Lanzafame, A. C., Löffler, W., Lorca, A., Manteiga, M., Marchal, O., Marrese, P. M., Moitinho, A., Mora, A., Muinonen, K., Osborne, P., Pancino, E., Pauwels, T., Petit, J. M., Recio-Blanco, A., Richards, P. J., Riello, M., Rimoldini, L., Robin, A. C., Roegiers, T., Rybizki, J., Sarro, L. M., Siopis, C., Smith, M., Sozzetti, A., Ulla, A., Utrilla, E., van Leeuwen, M., van Reeven, W., Abbas, U., Abreu Aramburu, A., Accart, S., Aerts, C., Aguado, J. J., Ajaj, M., Altavilla, G., Álvarez, M. A., Álvarez Cid-Fuentes, J., Alves, J., Anderson, R. I., Anglada Varela, E., Antoja, T., Audard, M., Baines, D., Baker, S. G., Balaguer-Núñez, L., Balbinot, E., Balog, Z., Barache, C., Barbato, D., Barros, M., Barstow, M. A., Bartolomé, S., Bassilana, J. L., Bauchet, N., Baudesson-Stella, A., Becciani, U., Bellazzini, M., Bernet, M., Bertone, S., Bianchi, L., Blanco-Cuaresma, S., Boch, T., Bombrun, A., Bossini, D., Bouquillon, S., Bragaglia, A., Bramante, L., Breedt, E., Bressan, A., Brouillet, N., Bucciarelli, B., Burlacu, A., Busonero, D., Butkevich, A. G., Buzzi, R., Caffau, E., Cancelliere, R., Cánovas, H., Cantat-Gaudin, T., Carballo, R., Carlucci, T., Carnerero, M. I., Carrasco, J. M., Casamiquela, L., Castellani, M., Castro-Ginard, A., Castro Sampol, P., Chaoul, L., Charlot, P., Chemin, L., Chiavassa, A., Cioni, M. R. L., Comoretto, G., Cooper, W. J., Cornez, T., Cowell, S., Crifo, F., Crosta, M., Crowley, C., Dafonte, C., Dapergolas, A., David, M., David, P., de Laverny, P., De Luise, F., De March, R., De Ridder, J., de Souza, R., de Teodoro, P., de Torres, A., del Peloso, E. F., del Pozo, E., Delbo, M., Delgado, A., Delgado, H. E., Delisle, J. B., Di Matteo, P., Diakite, S., Diener, C., Distefano, E., Dolding, C., Eappachen, D., Edvardsson, B., Enke, H., Esquej, P., Fabre, C., Fabrizio, M., Faigler, S., Fedorets, G., Fernique, P., Fienga, A., Figueras, F., Fouron, C., Fragkoudi, F.,

Fraile, E., Franke, F., Gai, M., Garabato, D., Garcia-Gutierrez, A., García-Torres, M., Garofalo, A., Gavras, P., Gerlach, E., Geyer, R., Giacobbe, P., Gilmore, G., Girona, S., Giuffrida, G., Gomel, R., Gomez, A., Gonzalez-Santamaria, I., González-Vidal, J. J., Granvik, M., Gutiérrez-Sánchez, R., Guy, L. P., Hauser, M., Haywood, M., Helmi, A., Hidalgo, S. L., Hilger, T., Hładczuk, N., Hobbs, D., Holland, G., Huckle, H. E., Jasniewicz, G., Jonker, P. G., Juaristi Campillo, J., Julbe, F., Karbevska, L., Kervella, P., Khanna, S., Kochoska, A., Kontizas, M., Kordopatis, G., Korn, A. J., Kostrzewa-Rutkowska, Z., Kruszyńska, K., Lambert, S., Lanza, A. F., Lasne, Y., Le Campion, J. F., Le Fustec, Y., Lebreton, Y., Lebzelter, T., Leccia, S., Leclerc, N., Lecoeur-Taibi, I., Liao, S., Licata, E., Lindstrøm, E. P., Lister, T. A., Livanou, E., Lobel, A., Madrero Pardo, P., Managau, S., Mann, R. G., Marchant, J. M., Marconi, M., Marcos Santos, M. M. S., Marinoni, S., Marocco, F., Marshall, D. J., Martin Polo, L., Martín-Fleitas, J. M., Masip, A., Massari, D., Mastrobuono-Battisti, A., Mazeh, T., McMillan, P. J., Messina, S., Michalik, D., Millar, N. R., Mints, A., Molina, D., Molinaro, R., Molnár, L., Montegriffo, P., Mor, R., Morbidelli, R., Morel, T., Morris, D., Mulone, A. F., Munoz, D., Muraveva, T., Murphy, C. P., Musella, I., Noval, L., Ordénovic, C., Orrù, G., Osinde, J., Pagani, C., Pagano, I., Palaversa, L., Palicio, P. A., Panahi, A., Pawlak, M., Peñalosa Esteller, X., Penttilä, A., Piersimoni, A. M., Pineau, F. X., Plachy, E., Plum, G., Poggio, E., Poretti, E., Poujoulet, E., Prša, A., Pulone, L., Racero, E., Ragaini, S., Rainer, M., Raiteri, C. M., Rambaux, N., Ramos, P., Ramos-Lerate, M., Re Fiorentin, P., Regibo, S., Reylé, C., Ripepi, V., Riva, A., Rixon, G., Robichon, N., Robin, C., Roelens, M., Rohrbasser, L., Romero-Gómez, M., Rowell, N., Royer, F., Rybicki, K. A., Sadowski, G., Sagristà Sellés, A., Sahlmann, J., Salgado, J., Salguero, E., Samaras, N., Sanchez Gimenez, V., Sanna, N., Santoveña, R., Sarasso, M., Schultheis, M., Sciacca, E., Segol, M., Segovia, J. C., Ségransan, D., Semeux, D., Shahaf, S., Siddiqui, H. I., Siebert, A., Siltala, L., Slezak, E., Smart, R. L., Solano, E., Solitro, F., Souami, D., Souchay, J., Spagna, A., Spoto, F., Steele,

I. A., Steidelmüller, H., Stephenson, C. A., Süveges, M., Szabados, L., Szegedi-Elek, E., Taris, F., Tauran, G., Taylor, M. B., Teixeira, R., Thuillot, W., Tonello, N., Torra, F., Torra, J., Turon, C., Unger, N., Vaillant, M., van Dillen, E., Vanel, O., Vecchiato, A., Viala, Y., Vicente, D., Voutsinas, S., Weiler, M., Wevers, T., Wyrzykowski, Ł., Yoldas, A., Yvard, P., Zhao, H., Zorec, J., Zucker, S., Zurbach, C., & Zwitter, T. 2021, A&A, 649, A1

Gaia Collaboration, Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., Babusiaux, C., Bailer-Jones, C. A. L., Biermann, M., Evans, D. W., Eyer, L., Jansen, F., Jordi, C., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Panem, C., Pourbaix, D., Randich, S., Sartoretti, P., Siddiqui, H. I., Soubiran, C., van Leeuwen, F., Walton, N. A., Arenou, F., Bastian, U., Cropper, M., Drimmel, R., Katz, D., Lattanzi, M. G., Bakker, J., Cacciari, C., Castañeda, J., Chaoul, L., Cheek, N., De Angeli, F., Fabricius, C., Guerra, R., Holl, B., Masana, E., Messineo, R., Mowlavi, N., Nienartowicz, K., Panuzzo, P., Portell, J., Riello, M., Seabroke, G. M., Tanga, P., Thévenin, F., Gracia-Abril, G., Comoretto, G., Garcia-Reinaldos, M., Teyssier, D., Altmann, M., Andrae, R., Audard, M., Bellas-Velidis, I., Benson, K., Berthier, J., Blomme, R., Burgess, P., Busso, G., Carry, B., Cellino, A., Clementini, G., Clotet, M., Creevey, O., Davidson, M., De Ridder, J., Delchambre, L., Dell´Oro, A., Ducourant, C., Fernández-Hernández, J., Fouesneau, M., Frémat, Y., Galluccio, L., García-Torres, M., González-Núñez, J., González-Vidal, J. J., Gosset, E., Guy, L. P., Halbwachs, J.-L., Hambly, N. C., Harrison, D. L., Hernández, J., Hestroffer, D., Hodgkin, S. T., Hutton, A., Jasniewicz, G., Jean-Antoine-Piccolo, A., Jordan, S., Korn, A. J., Krone-Martins, A., Lanzafame, A. C., Lebzelter, T., Löffler, W., Manteiga, M., Marrese, P. M., Martín-Fleitas, J. M., Moitinho, A., Mora, A., Muinonen, K., Osinde, J., Pancino, E., Pauwels, T., Petit, J.-M., Recio-Blanco, A., Richards, P. J., Rimoldini, L., Robin, A. C., Sarro, L. M., Siopis, C., Smith, M., Sozzetti, A., Süveges, M., Torra, J., van Reeven, W., Abbas, U., Abreu Aramburu, A., Accart, S., Aerts, C., Altavilla, G., Álvarez, M. A.,

Alvarez, R., Alves, J., Anderson, R. I., Andrei, A. H., Anglada Varela, E., Antiche, E., Antoja, T., Arcay, B., Astraatmadja, T. L., Bach, N., Baker, S. G., Balaguer-Núñez, L., Balm, P., Barache, C., Barata, C., Barbato, D., Barblan, F., Barklem, P. S., Barrado, D., Barros, M., Barstow, M. A., Bartholomé Muñoz, S., Bassilana, J.-L., Becciani, U., Bellazzini, M., Berihuete, A., Bertone, S., Bianchi, L., Bienaymé, O., Blanco-Cuaresma, S., Boch, T., Boeche, C., Bombrun, A., Borrachero, R., Bossini, D., Bouquillon, S., Bourda, G., Bragaglia, A., Bramante, L., Breddels, M. A., Bressan, A., Brouillet, N., Brüsemeister, T., Brugaletta, E., Bucciarelli, B., Burlacu, A., Busonero, D., Butkevich, A. G., Buzzi, R., Caffau, E., Cancelliere, R., Cannizzaro, G., Cantat-Gaudin, T., Carballo, R., Carlucci, T., Carrasco, J. M., Casamiquela, L., Castellani, M., Castro-Ginard, A., Charlot, P., Chemin, L., Chiavassa, A., Cocozza, G., Costigan, G., Cowell, S., Crifo, F., Crosta, M., Crowley, C., Cuypers+, J., Dafonte, C., Damerdji, Y., Dapergolas, A., David, P., David, M., de Laverny, P., De Luise, F., De March, R., de Martino, D., de Souza, R., de Torres, A., Debosscher, J., del Pozo, E., Delbo, M., Delgado, A., Delgado, H. E., Di Matteo, P., Diakite, S., Diener, C., Distefano, E., Dolding, C., Drazinos, P., Durán, J., Edvardsson, B., Enke, H., Eriksson, K., Esquej, P., Eynard Bontemps, G., Fabre, C., Fabrizio, M., Faigler, S., Falcão, A. J., Farràs Casas, M., Federici, L., Fedorets, G., Fernique, P., Figueras, F., Filippi, F., Findeisen, K., Fonti, A., Fraile, E., Fraser, M., Frézouls, B., Gai, M., Galleti, S., Garabato, D., García-Sedano, F., Garofalo, A., Garralda, N., Gavel, A., Gavras, P., Gerssen, J., Geyer, R., Giacobbe, P., Gilmore, G., Girona, S., Giuffrida, G., Glass, F., Gomes, M., Granvik, M., Gueguen, A., Guerrier, A., Guiraud, J., Gutiérrez-Sánchez, R., Haigron, R., Hatzidimitriou, D., Hauser, M., Haywood, M., Heiter, U., Helmi, A., Heu, J., Hilger, T., Hobbs, D., Hofmann, W., Holland, G., Huckle, H. E., Hypki, A., Icardi, V., Janßen, K., Jevardat de Fombelle, G., Jonker, P. G., Juhász, Á. L., Julbe, F., Karampelas, A., Kewley, A., Klar, J., Kochoska, A., Kohley, R., Kolenberg, K., Kontizas, M., Kontizas, E., Koposov, S. E., Kordopatis, G., Kostrzewa-Rutkowska, Z.,

Koubsky, P., Lambert, S., Lanza, A. F., Lasne, Y., Lavigne, J.-B., Le Fustec, Y., Le Poncin-Lafitte, C., Lebreton, Y., Leccia, S., Leclerc, N., Lecoeur-Taibi, I., Lenhardt, H., Leroux, F., Liao, S., Licata, E., Lindstrøm, H. E. P., Lister, T. A., Livanou, E., Lobel, A., López, M., Managau, S., Mann, R. G., Mantelet, G., Marchal, O., Marchant, J. M., Marconi, M., Marinoni, S., Marschalkó, G., Marshall, D. J., Martino, M., Marton, G., Mary, N., Massari, D., Matijevic, G., Mazeh, T., McMillan, P. J., Messina, S., Michalik, D., Millar, N. R., Molina, D., Molinaro, R., Molnár, L., Montegriffo, P., Mor, R., Morbidelli, R., Morel, T., Morris, D., Mulone, A. F., Muraveva, T., Musella, I., Nelemans, G., Nicastro, L., Noval, L., O´Mullane, W., Ordénovic, C., Ordóñez-Blanco, D., Osborne, P., Pagani, C., Pagano, I., Pailler, F., Palacin, H., Palaversa, L., Panahi, A., Pawlak, M., Piersimoni, A. M., Pineau, F.-X., Plachy, E., Plum, G., Poggio, E., Poujoulet, E., Prsa, A., Pulone, L., Racero, E., Ragaini, S., Rambaux, N., Ramos-Lerate, M., Regibo, S., Reylé, C., Riclet, F., Ripepi, V., Riva, A., Rivard, A., Rixon, G., Roegiers, T., Roelens, M., Romero-Gómez, M., Rowell, N., Royer, F., Ruiz-Dern, L., Sadowski, G., Sagristà Sellés, T., Sahlmann, J., Salgado, J., Salguero, E., Sanna, N., Santana-Ros, T., Sarasso, M., Savietto, H., Schultheis, M., Sciacca, E., Segol, M., Segovia, J. C., Ségransan, D., Shih, I-C., Siltala, L., Silva, A. F., Smart, R. L., Smith, K. W., Solano, E., Solitro, F., Sordo, R., Soria Nieto, S., Souchay, J., Spagna, A., Spoto, F., Stampa, U., Steele, I. A., Steidelmüller, H., Stephenson, C. A., Stoev, H., Suess, F. F., Surdej, J., Szabados, L., Szegedi-Elek, E., Tapiador, D., Taris, F., Tauran, G., Taylor, M. B., Teixeira, R., Terrett, D., Teyssandier, P., Thuillot, W., Titarenko, A., Torra Clotet, F., Turon, C., Ulla, A., Utrilla, E., Uzzi, S., Vaillant, M., Valentini, G., Valette, V., van Elteren, A., Van Hemelryck, E., van Leeuwen, M., Vaschetto, M., Vecchiato, A., Veljanoski, J., Viala, Y., Vicente, D., Vogt, S., von Essen, C., Voss, H., Votruba, V., Voutsinas, S., Walmsley, G., Weiler, M., Wertz, O., Wevers, T., Wyrzykowski, L., Yoldas, A., Zerjal, M., Ziaeepour, H., Zorec, J., Zschocke, S., Zucker, S., Zurbach, C., & Zwitter, T. 2018b, A&A, 616, A1

Gaidos, E. & Mann, A. W. 2013, ApJ, 762, 41

Gaidos, E., Mann, A. W., Rizzuto, A., Nofi, L., Mace, G., Vanderburg, A., Feiden, G., Narita, N., Takeda, Y., Esposito, T. M., De Rosa, R. J., Ansdell, M., Hirano, T., Graham, J. R., Kraus, A., & Jaffe, D. 2017, MNRAS, 464, 850

García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., Shetrone, M., Mészáros, S., Bizyaev, D., Carrera, R., Cunha, K., García-Hernández, D. A., Johnson, J. A., Majewski, S. R., Nidever, D. L., Schiavon, R. P., Shane, N., Smith, V. V., Sobeck, J., Troup, N., Zamora, O., Weinberg, D. H., Bovy, J., Eisenstein, D. J., Feuillet, D., Frinchaboy, P. M., Hayden, M. R., Hearty, F. R., Nguyen, D. C., O'Connell, R. W., Pinsonneault, M. H., Wilson, J. C., & Zasowski, G. 2016, AJ, 151, 144

Gautier, T. N., Borucki, W. J., Caldwell, D. A., & Koch, D. G. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 366, Transiting Extrapolar Planets Workshop, ed. C. Afonso, D. Weldrake, & T. Henning, 219

Ghezzi, L., Cunha, K., Smith, V. V., de Araújo, F. X., Schuler, S. C., & de la Reza, R. 2010, ApJ, 720, 1290

Ghezzi, L., Montet, B. T., & Johnson, J. A. 2018, ApJ, 860, 109

Gonzalez, G. 1997, MNRAS, 285, 403

—. 1998, A&A, 334, 221

González Hernández, J. I. & Bonifacio, P. 2009, A&A, 497, 497

Green, G. M. 2018, The Journal of Open Source Software, 3, 695

Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, ApJ, 887, 93

Grunblatt, S. K., Huber, D., Gaidos, E., Hon, M., Zinn, J. C., & Stello, D. 2019, AJ, 158, 227

Grunblatt, S. K., Huber, D., Gaidos, E., Lopez, E. D., Barclay, T., Chontos, A., Sinukoff, E., Van Eylen, V., Howard, A. W., & Isaacson, H. T. 2018, ApJ, 861, L5

Grunblatt, S. K., Huber, D., Gaidos, E., Lopez, E. D., Howard, A. W., Isaacson, H. T., Sinukoff, E., Vanderburg, A., Nofi, L., Yu, J., North, T. S. H., Chaplin, W., Foreman-Mackey, D., Petigura, E., Ansdell, M., Weiss, L., Fulton, B., & Lin, D. N. C. 2017, AJ, 154, 254

Gunn, J. E., Siegmund, W. A., Mannery, E. J., Owen, R. E., Hull, C. L., Leger, R. F., Carey, L. N., Knapp, G. R., York, D. G., Boroski, W. N., Kent, S. M., Lupton, R. H., Rockosi, C. M., Evans, M. L., Waddell, P., Anderson, J. E., Annis, J., Barentine, J. C., Bartoszek, L. M., Bastian, S., Bracker, S. B., Brewington, H. J., Briegel, C. I., Brinkmann, J., Brown, Y. J., Carr, M. A., Czarapata, P. C., Drennan, C. C., Dombeck, T., Federwitz, G. R., Gillespie, B. A., Gonzales, C., Hansen, S. U., Harvanek, M., Hayes, J., Jordan, W., Kinney, E., Klaene, M., Kleinman, S. J., Kron, R. G., Kresinski, J., Lee, G., Limmongkol, S., Lindenmeyer, C. W., Long, D. C., Loomis, C. L., McGehee, P. M., Mantsch, P. M., Neilsen, Jr., E. H., Neswold, R. M., Newman, P. R., Nitta, A., Peoples, Jr., J., Pier, J. R., Prieto, P. S., Prosapio, A., Rivetta, C., Schneider, D. P., Snedden, S., & Wang, S.-i. 2006, AJ, 131, 2332

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

—. 2020, MNRAS, 493, 792

Gustafsson, B., Edvardsson, B., Eriksson, K., Jørgensen, U. G., Nordlund, Å., & Plez, B. 2008, A&A, 486, 951

Hamer, J. H. & Schlaufman, K. C. 2019, AJ, 158, 190

Hartman, J. D., Bakos, G., Stanek, K. Z., & Noyes, R. W. 2004, AJ, 128, 1761

Hayden, M. R., Bovy, J., Holtzman, J. A., Nidever, D. L., Bird, J. C., Weinberg, D. H., Andrews, B. H., Majewski, S. R., Allende Prieto, C., Anders, F., Beers, T. C., Bizyaev, D., Chiappini, C., Cunha, K., Frinchaboy, P., García-Herńandez, D. A., García Pérez, A. E., Girardi, L., Harding, P., Hearty, F. R., Johnson, J. A., Mészáros, S., Minchev, I., O'Connell, R., Pan, K., Robin, A. C., Schiavon, R. P., Schneider, D. P., Schultheis, M., Shetrone, M., Skrutskie, M., Steinmetz, M., Smith, V., Wilson, J. C., Zamora, O., & Zasowski, G. 2015, ApJ, 808, 132

Hayden, M. R., Holtzman, J. A., Bovy, J., Majewski, S. R., Johnson, J. A., Allende Prieto, C., Beers, T. C., Cunha, K., Frinchaboy, P. M., García Pérez, A. E., Girardi, L., Hearty, F. R., Lee, Y. S., Nidever, D., Schiavon, R. P., Schlesinger, K. J., Schneider, D. P., Schultheis, M., Shetrone, M., Smith, V. V., Zasowski, G., Bizyaev, D., Feuillet, D., Hasselquist, S., Kinemuchi, K., Malanushenko, E., Malanushenko, V., O'Connell, R., Pan, K., & Stassun, K. 2014, AJ, 147, 116

Heiter, U. & Luck, R. E. 2003a, AJ, 126, 2015

—. 2003b, AJ, 126, 2015

Hinkel, N. R. & Unterborn, C. T. 2018, ApJ, 853, 83

Hippke, M., David, T. J., Mulders, G. D., & Heller, R. 2019, AJ, 158, 143

Holtzman, J. A., Hasselquist, S., Shetrone, M., Cunha, K., Allende Prieto, C., Anguiano, B., Bizyaev, D., Bovy, J., Casey, A., Edvardsson, B., Johnson, J. A., Jönsson, H., Meszaros, S., Smith, V. V., Sobeck, J., Zamora, O., Chojnowski, S. D., Fernandez-Trincado, J., Garcia-Hernandez, D. A., Majewski, S. R., Pinsonneault, M., Souto, D., Stringfellow, G. S., Tayar, J., Troup, N., & Zasowski, G. 2018, AJ, 156, 125

Holtzman, J. A., Shetrone, M., Johnson, J. A., Allende Prieto, C., Anders, F., Andrews, B., Beers, T. C., Bizyaev, D., Blanton, M. R., Bovy, J., Carrera, R., Chojnowski, S. D., Cunha, K., Eisenstein, D. J., Feuillet, D., Frinchaboy, P. M., Galbraith-Frew, J., García Pérez, A. E., García-Hernández, D. A., Hasselquist, S., Hayden, M. R., Hearty, F. R., Ivans, I., Majewski, S. R., Martell, S., Meszaros, S., Muna, D., Nidever, D., Nguyen, D. C., O'Connell, R. W., Pan, K., Pinsonneault, M., Robin, A. C., Schiavon, R. P., Shane, N., Sobeck, J., Smith, V. V., Troup, N., Weinberg, D. H., Wilson, J. C., Wood-Vasey, W. M., Zamora, O., & Zasowski, G. 2015, AJ, 150, 148

Howard, A. W., Marcy, G. W., Bryson, S. T., Jenkins, J. M., Rowe, J. F., Batalha, N. M., Borucki, W. J., Koch, D. G., Dunham, E. W., Gautier, III, T. N., Van Cleve, J., Cochran, W. D., Latham, D. W., Lissauer, J. J., Torres, G., Brown, T. M., Gilliland, R. L., Buchhave, L. A., Caldwell, D. A., Christensen-Dalsgaard, J., Ciardi, D., Fressin, F., Haas, M. R., Howell, S. B., Kjeldsen, H., Seager, S., Rogers, L., Sasselov, D. D., Steffen, J. H., Basri, G. S., Charbonneau, D., Christiansen, J., Clarke, B., Dupree, A., Fabrycky, D. C., Fischer, D. A., Ford, E. B., Fortney, J. J., Tarter, J., Girouard, F. R., Holman, M. J., Johnson, J. A., Klaus, T. C., Machalek, P., Moorhead, A. V., Morehead, R. C., Ragozzine, D., Tenenbaum, P., Twicken, J. D., Quinn, S. N., Isaacson, H., Shporer, A., Lucas, P. W., Walkowicz, L. M., Welsh, W. F., Boss, A., Devore, E., Gould, A., Smith, J. C., Morris, R. L., Prsa, A., Morton, T. D., Still, M., Thompson, S. E., Mullally, F., Endl, M., & MacQueen, P. J. 2012, ApJS, 201, 15

Howell, S. B., Sobeck, C., Haas, M., Still, M., Barclay, T., Mullally, F., Troeltzsch, J., Aigrain, S., Bryson, S. T., Caldwell, D., Chaplin, W. J., Cochran, W. D., Huber, D., Marcy, G. W., Miglio, A., Najita, J. R., Smith, M., Twicken, J. D., & Fortney, J. J. 2014, PASP, 126, 398

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

Huang, Y., Liu, X. W., Yuan, H. B., Xiang, M. S., Chen, B. Q., & Zhang, H. W. 2015, MNRAS, 454, 2863

Huber, D., Chaplin, W. J., Chontos, A., Kjeldsen, H., Christensen-Dalsgaard, J., Bedding, T. R., Ball, W., Brahm, R., Espinoza, N., Henning, T., Jordán, A., Sarkis, P., Knudstrup, E., Albrecht, S., Grundahl, F., Fredslund Andersen, M., Pallé, P. L., Crossfield, I., Fulton, B., Howard, A. W., Isaacson, H. T., Weiss, L. M., Handberg, R., Lund, M. N., Serenelli, A. M., Rørsted Mosumgaard, J., Stokholm, A., Bieryla, A., Buchhave, L. A., Latham, D. W., Quinn, S. N., Gaidos, E., Hirano, T., Ricker, G. R., Vanderspek, R. K., Seager, S., Jenkins, J. M., Winn, J. N., Antia, H. M., Appourchaux, T., Basu, S., Bell, K. J., Benomar, O., Bonanno, A., Buzasi, D. L., Campante, T. L., Çelik Orhan, Z., Corsaro, E., Cunha, M. S., Davies, G. R., Deheuvels, S., Grunblatt, S. K., Hasanzadeh, A., Di Mauro, M. P., García, R. A., Gaulme, P., Girardi, L., Guzik, J. A., Hon, M., Jiang, C., Kallinger, T., Kawaler, S. D., Kuszlewicz, J. S., Lebreton, Y., Li, T., Lucas, M., Lundkvist, M. S., Mann, A. W., Mathis, S., Mathur, S., Mazumdar, A., Metcalfe, T. S., Miglio, A., Monteiro, M. J. P. F. G., Mosser, B., Noll, A., Nsamba, B., Ong, J. M. J., Örtel, S., Pereira, F., Ranadive, P., Régulo, C., Rodrigues, T. S., Roxburgh, I. W., Silva Aguirre, V., Smalley, B., Schofield, M., Sousa, S. G., Stassun, K. G., Stello, D., Tayar, J., White, T. R., Verma, K., Vrard, M., Yıldız, M., Baker, D., Bazot, M., Beichmann, C., Bergmann, C., Bugnet, L., Cale, B., Carlino, R., Cartwright, S. M., Christiansen, J. L., Ciardi, D. R., Creevey, O., Dittmann, J. A., Do Nascimento, Jose-Dias, J., Van Eylen, V., Fürész, G., Gagné, J., Gao, P., Gazeas, K., Giddens, F., Hall, O. J., Hekker, S., Ireland , M. J., Latouf, N., LeBrun, D., Levine, A. M., Matzko, W., Natinsky, E., Page, E., Plavchan, P., Mansouri-Samani, M., McCauliff, S., Mullally, S. E., Orenstein, B., Garcia Soto, A., Paegert, M., van Saders, J. L., Schnaible, C., Soderblom, D. R., Szabó, R., Tanner, A., Tinney, C. G., Teske, J., Thomas, A., Trampedach, R., Wright, D., Yuan, T. T., & Zohrabi, F. 2019, AJ, 157, 245

Huber, D., Chaplin, W. J., Christensen-Dalsgaard, J., Gilliland, R. L., Kjeldsen, H., Buchhave, L. A., Fischer, D. A., Lissauer, J. J., Rowe, J. F., Sanchis-Ojeda, R., Basu, S., Handberg, R., Hekker, S., Howard, A. W., Isaacson, H., Karoff, C., Latham, D. W., Lund, M. N., Lundkvist, M., Marcy, G. W., Miglio, A., Silva Aguirre, V., Stello, D., Arentoft, T., Barclay, T., Bedding, T. R., Burke, C. J., Christiansen, J. L., Elsworth, Y. P., Haas, M. R., Kawaler, S. D., Metcalfe, T. S., Mullally, F., & Thompson, S. E. 2013, ApJ, 767, 127

Huber, D., Zinn, J., Bojsen-Hansen, M., Pinsonneault, M., Sahlholdt, C., Serenelli, A., Silva Aguirre, V., Stassun, K., Stello, D., Tayar, J., Bastien, F., Bedding, T. R., Buchhave, L. A., Chaplin, W. J., Davies, G. R., García, R. A., Latham, D. W., Mathur, S., Mosser, B., & Sharma, S. 2017, ApJ, 844, 102

Iben, I. 1968, Nature, 220, 143

Ida, S. & Lin, D. N. C. 2004, ApJ, 616, 567

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

Jin, S. & Mordasini, C. 2017, ArXiv e-prints

Johnson, J. A., Petigura, E. A., Fulton, B. J., Marcy, G. W., Howard, A. W., Isaacson, H., Hebb, L., Cargile, P. A., Morton, T. D., Weiss, L. M., Winn, J. N., Rogers, L. A., Sinukoff, E., & Hirsch, L. A. 2017, AJ, 154, 108

Jönsson, H., Holtzman, J. A., Allende Prieto, C., Cunha, K., García-Hernández, D. A., Hasselquist, S., Masseron, T., Osorio, Y., Shetrone, M., Smith, V., Stringfellow, G. S., Bizyaev, D., Edvardsson, B., Majewski, S. R., Mészáros, S., Souto, D., Zamora, O., Beaton, R. L., Bovy, J., Donor, J., Pinsonneault, M. H., Poovelil, V. J., & Sobeck, J. 2020, AJ, 160, 120

Kaufer, A., Stahl, O., Tubbesing, S., Nørregaard, P., Avila, G., Francois, P., Pasquini, L., & Pizzella, A. 1999, The Messenger, 95, 8

Kendall, M. G. 1938, Biometrika, 30, 81

Kennedy, G. M. & Kenyon, S. J. 2008, ApJ, 673, 502

Kjeldsen, H. & Bedding, T. R. 1995, A&A, 293, 87

Koch, D. G., Borucki, W. J., Basri, G., Batalha, N. M., Brown, T. M., Caldwell, D., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Gautier, III, T. N., Geary, J. C., Gilliland, R. L., Gould, A., Jenkins, J., Kondo, Y., Latham, D. W., Lissauer, J. J., Marcy, G., Monet, D., Sasselov, D., Boss, A., Brownlee, D., Caldwell, J., Dupree, A. K., Howell, S. B., Kjeldsen, H., Meibom, S., Morrison, D., Owen, T., Reitsema, H., Tarter, J., Bryson, S. T., Dotson, J. L., Gazis, P., Haas, M. R., Kolodziejczak, J., Rowe, J. F., Van Cleve, J. E., Allen, C., Chandrasekaran, H., Clarke, B. D., Li, J., Quintana, E. V., Tenenbaum, P., Twicken, J. D., & Wu, H. 2010, ApJ, 713, L79

Kovács, G., Zucker, S., & Mazeh, T. 2002, A&A, 391, 369

Kreidberg, L. 2015, PASP, 127, 1161

Kuchner, M. J. & Lecar, M. 2002, ApJ, 574, L87

Kunitomo, M., Ikoma, M., Sato, B., Katsuta, Y., & Ida, S. 2011, in AAS/Division for Extreme Solar Systems Abstracts, Vol. 2, AAS/Division for Extreme Solar Systems Abstracts, 30.05

Kurucz, R. L. 1993, in Astronomical Society of the Pacific Conference Series, Vol. 44, IAU Colloq. 138: Peculiar versus Normal Phenomena in A-type and Related Stars, ed. M. M. Dworetsky, F. Castelli, & R. Faraggiana, 87

Lambrechts, M. & Lega, E. 2017, A&A, 606, A146

Lee, G. R., Gommers, R., Waselewski, F., Wohlfahrt, K., & O’Leary, A. 2019, Journal of Open Source Software, 4, 1237

Lightkurve Collaboration, Cardoso, J. V. d. M., Hedges, C., Gully-Santiago, M., Saunders, N., Cody, A. M., Barclay, T., Hall, O., Sagear, S., Turtelboom, E., Zhang, J., Tzanidakis, A., Mighell, K., Coughlin, J., Bell, K., Berta-Thompson, Z., Williams, P., Dotson, J., & Barentsen, G. 2018, Lightkurve: Kepler and TESS time series analysis in Python, Astrophysics Source Code Library

Linden, S. T., Pryal, M., Hayes, C. R., Troup, N. W., Majewski, S. R., Andrews, B. H., Beers, T. C., Carrera, R., Cunha, K., Fernández-Trincado, J. G., Frinchaboy, P., Geisler, D., Lane, R. R., Nitschelm, C., Pan, K., Allende Prieto, C., Roman-Lopes, A., Smith, V. V., Sobeck, J., Tang, B., Villanova, S., & Zasowski, G. 2017, ApJ, 842, 49

Liu, F., Asplund, M., Ramírez, I., Yong, D., & Meléndez, J. 2014, MNRAS, 442, L51

Liu, F., Asplund, M., Yong, D., Meléndez, J., Ramírez, I., Karakas, A. I., Carlos, M., & Marino, A. F. 2016, MNRAS, 463, 696

Lo Curto, G., Mayor, M., Benz, W., Bouchy, F., Lovis, C., Moutou, C., Naef, D., Pepe, F., Queloz, D., Santos, N. C., Segransan, D., & Udry, S. 2010, A&A, 512, A48

Lodders, K. 2003, ApJ, 591, 1220

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

Mack, III, C. E., Schuler, S. C., Stassun, K. G., & Norris, J. 2014, ApJ, 787, 98

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., Allende Prieto, C., Barkhouser, R., Bizyaev, D., Blank, B., Brunner, S., Burton, A., Carrera, R., Chojnowski, S. D., Cunha, K., Epstein, C., Fitzgerald, G., Garcia Perez, A. E., Hearty, F. R., Henderson,

C., Holtzman, J. A., Johnson, J. A., Lam, C. R., Lawler, J. E., Maseman, P., Meszaros, S., Nelson, M., Coung Nguyen, D., Nidever, D. L., Pinsonneault, M., Shetrone, M., Smee, S., Smith, V. V., Stolberg, T., Skrutskie, M. F., Walker, E., Wilson, J. C., Zasowski, G., Anders, F., Basu, S., Beland, S., Blanton, M. R., Bovy, J., Brownstein, J. R., Carlberg, J., Chaplin, W., Chiappini, C., Eisenstein, D. J., Elsworth, Y., Feuillet, D., Fleming, S. W., Galbraith-Frew, J., Garcia, R. A., Anibal Garcia-Hernandez, D., Gillespie, B. A., Girardi, L., Gunn, J. E., Hasselquist, S., Hayden, M. R., Hekker, S., Ivans, I., Kinemuchi, K., Klaene, M., Mahadevan, S., Mathur, S., Mosser, B., Muna, D., Munn, J. A., Nichol, R. C., O'Connell, R. W., Robin, A. C., Rocha-Pinto, H., Schultheis, M., Serenelli, A. M., Shane, N., Silva Aguirre, V., Sobeck, J. S., Thompson, B., Troup, N. W., Weinberg, D. H., & Zamora, O. 2015, ArXiv e-prints

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., Prieto, C. A., Barkhouser, R., Bizyaev, D., Blank, B., Brunner, S., Burton, A., Carrera, R., Chojnowski, S. D., Cunha, K., Epstein, C., Fitzgerald, G., Perez, A. E. G., Hearty, F. R., Henderson, C., Holtzman, J. A., Johnson, J. A., Lam, C. R., Lawler, J. E., Maseman, P., Meszaros, S., Nelson, M., Nguyen, D. C., Nidever, D. L., Pinsonneault, M., Shetrone, M., Smee, S., Smith, V. V., Stolberg, T., Skrutskie, M. F., Walker, E., Wilson, J. C., Zasowski, G., Anders, F., Basu, S., Beland, S., Blanton, M. R., Bovy, J., Brownstein, J. R., Carlberg, J., Chaplin, W., Chiappini, C., Eisenstein, D. J., Elsworth, Y., Feuillet, D., Fleming, S. W., Galbraith-Frew, J., Garcia, R. A., Garcia-Hernandez, D. A., Gillespie, B. A., Girardi, L., Gunn, J. E., Hasselquist, S., Hayden, M. R., Hekker, S., Ivans, I., Kinemuchi, K., Klaene, M., Mahadevan, S., Mathur, S., Mosser, B., Muna, D., Munn, J. A., Nichol, R. C., O'Connell, R. W., Parejko, J. K., Robin, A. C., Rocha-Pinto, H., Schultheis, M., Serenelli, A. M., Shane, N., Aguirre, V. S., Sobeck, J. S., Thompson, B., Troup, N. W., Weinberg, D. H., & Zamora, O. 2017, AJ, 154, 94

Maldonado, J., Villaver, E., Eiroa, C., & Micela, G. 2019, A&A, 624, A94

Mann, A. W., Dupuy, T., Kraus, A. L., Gaidos, E., Ansdell, M., Ireland, M., Rizzuto, A. C., Hung, C.-L., Dittmann, J., Factor, S., Feiden, G., Martinez, R. A., Ruíz-Rodríguez, D., & Thao, P. C. 2019, ApJ, 871, 63

Mann, A. W., Feiden, G. A., Gaidos, E., Boyajian, T., & von Braun, K. 2015, ApJ, 804, 64

Mann, A. W., Gaidos, E., Mace, G. N., Johnson, M. C., Bowler, B. P., LaCourse, D., Jacobs, T. L., Vanderburg, A., Kraus, A. L., Kaplan, K. F., & Jaffe, D. T. 2016, ApJ, 818, 46

Mann, A. W., Vanderburg, A., Rizzuto, A. C., Kraus, A. L., Berlind, P., Bieryla, A., Calkins, M. L., Esquerdo, G. A., Latham, D. W., Mace, G. N., Morris, N. R., Quinn, S. N., Sokal, K. R., & Stefanik, R. P. 2018, AJ, 155, 4

Mann, H. B. & Whitney, D. R. 1947, Ann. Math. Statist., 18, 50

Marboeuf, U., Thiabaud, A., Alibert, Y., Cabral, N., & Benz, W. 2014, A&A, 570, A36

Martin, R. G. & Livio, M. 2012, MNRAS, 425, L6

Martinez, C. F., Cunha, K., Ghezzi, L., & Smith, V. V. 2019, ApJ, 875, 29

Mathur, S., Huber, D., Batalha, N. M., Ciardi, D. R., Bastien, F. A., Bieryla, A., Buchhave, L. A., Cochran, W. D., Endl, M., Esquerdo, G. A., Furlan, E., Howard, A., Howell, S. B., Isaacson, H., Latham, D. W., MacQueen, P. J., & Silva, D. R. 2017, ApJS, 229, 30

Mayor, M., Marmier, M., Lovis, C., Udry, S., Ségransan, D., Pepe, F., Benz, W., Bertaux, J. L., Bouchy, F., Dumusque, X., Lo Curto, G., Mordasini, C., Queloz, D., & Santos, N. C. 2011, arXiv e-prints, arXiv:1109.2497

Mayor, M., Pepe, F., Queloz, D., Bouchy, F., Rupprecht, G., Lo Curto, G., Avila, G., Benz, W., Bertaux, J.-L., Bonfils, X., Dall, T., Dekker, H., Delabre, B., Eckert,

W., Fleury, M., Gilliotte, A., Gojak, D., Guzman, J. C., Kohler, D., Lizon, J.-L., Longinotti, A., Lovis, C., Megevand, D., Pasquini, L., Reyes, J., Sivan, J.-P., Sosnowska, D., Soto, R., Udry, S., van Kesteren, A., Weber, L., & Weilenmann, U. 2003, The Messenger, 114, 20

Mayor, M. & Queloz, D. 1995, Nature, 378, 355

McQuillan, A., Mazeh, T., & Aigrain, S. 2013, ApJ, 775, L11

McWilliam, A. 1997, ARA&A, 35, 503

Meléndez, J., Asplund, M., Gustafsson, B., & Yong, D. 2009, ApJ, 704, L66

Min, M., Dullemond, C. P., Kama, M., & Dominik, C. 2011, Icarus, 212, 416

Molenda-Żakowicz, J., Sousa, S. G., Frasca, A., Uytterhoeven, K., Briquet, M., Van Winckel, H., Drobek, D., Niemczura, E., Lampens, P., Lykke, J., Bloemen, S., Gameiro, J. F., Jean, C., Volpi, D., Gorlova, N., Mortier, A., Tsantaki, M., & Raskin, G. 2013, MNRAS, 434, 1422

Mordasini, C., Alibert, Y., Benz, W., Klahr, H., & Henning, T. 2012a, A&A, 541, A97

Mordasini, C., Alibert, Y., Georgy, C., Dittkrist, K.-M., Klahr, H., & Henning, T. 2012b, A&A, 547, A112

Morton, T. D. 2015, isochrones: Stellar model grid package

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

Mulders, G. D., Ciesla, F. J., Min, M., & Pascucci, I. 2015a, ApJ, 807, 9

Mulders, G. D., Pascucci, I., & Apai, D. 2015b, ApJ, 798, 112

—. 2015c, ApJ, 798, 112

Mulders, G. D., Pascucci, I., Apai, D., & Ciesla, F. J. 2018, AJ, 156, 24

Mulders, G. D., Pascucci, I., Apai, D., Frasca, A., & Molenda-Żakowicz, J. 2016, AJ, 152, 187

Mullally, F., Coughlin, J. L., Thompson, S. E., Christiansen, J., Burke, C., Clarke, B. D., & Haas, M. R. 2016, PASP, 128, 074502

Mullally, F., Coughlin, J. L., Thompson, S. E., Rowe, J., Burke, C., Latham, D. W., Batalha, N. M., Bryson, S. T., Christiansen, J., Henze, C. E., Ofir, A., Quarles, B., Shporer, A., Van Eylen, V., Van Laerhoven, C., Shah, Y., Wolfgang, A., Chaplin, W. J., Xie, J.-W., Akeson, R., Argabright, V., Bachtell, E., Barclay, T., Borucki, W. J., Caldwell, D. A., Campbell, J. R., Catanzarite, J. H., Cochran, W. D., Duren, R. M., Fleming, S. W., Fraquelli, D., Girouard, F. R., Haas, M. R., Hełminiak, K. G., Howell, S. B., Huber, D., Larson, K., Gautier, III, T. N., Jenkins, J. M., Li, J., Lissauer, J. J., McArthur, S., Miller, C., Morris, R. L., Patil-Sabale, A., Plavchan, P., Putnam, D., Quintana, E. V., Ramirez, S., Silva Aguirre, V., Seader, S., Smith, J. C., Steffen, J. H., Stewart, C., Stober, J., Still, M., Tenenbaum, P., Troeltzsch, J., Twicken, J. D., & Zamudio, K. A. 2015, ApJS, 217, 31

Muzerolle, J., Calvet, N., Hartmann, L., & D'Alessio, P. 2003, ApJ, 597, L149

Nadaraya, E. A. 1964, Theory of Probability and its Applications, 9, 141

Najita, J. R., Carr, J. S., Glassgold, A. E., & Valenti, J. A. 2007, Protostars and Planets V, 507

Nakatani, R., Hosokawa, T., Yoshida, N., Nomura, H., & Kuiper, R. 2017, ArXiv e-prints

Narang, M., Manoj, P., Furlan, E., Mordasini, C., Henning, T., Mathew, B., Banyal, R. K., & Sivarani, T. 2018, AJ, 156, 221

Netopil, M., Paunzen, E., Heiter, U., & Soubiran, C. 2016, A&A, 585, A150

Nidever, D. L., Holtzman, J. A., Allende Prieto, C., Beland, S., Bender, C., Bizyaev, D., Burton, A., Desphande, R., Fleming, S. W., García Pérez, A. E., Hearty, F. R., Majewski, S. R., Mészáros, S., Muna, D., Nguyen, D., Schiavon, R. P., Shetrone, M., Skrutskie, M. F., Sobeck, J. S., & Wilson, J. C. 2015, AJ, 150, 173

Nissen, P. E., Chen, Y. Q., Carigi, L., Schuster, W. J., & Zhao, G. 2014, A&A, 568, A25

Nordlander, T., Amarsi, A. M., Lind, K., Asplund, M., Barklem, P. S., Casey, A. R., Collet, R., & Leenaarts, J. 2017, A&A, 597, A6

Öberg, K. I., Murray-Clay, R., & Bergin, E. A. 2011, ApJ, 743, L16

Ofir, A. 2014, A&A, 561, A138

Owen, J. E. & Wu, Y. 2013, ApJ, 775, 105

—. 2017, ArXiv e-prints

Owen, R. E., Siegmund, W. A., Limmongkol, S., & Hull, C. L. 1994, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 2198, Instrumentation in Astronomy VIII, ed. D. L. Crawford & E. R. Craine, 110–114

Patra, K. C., Winn, J. N., Holman, M. J., Gillon, M., Burdanov, A., Jehin, E., Delrez, L., Pozuelos, F. J., Barkaoui, K., Benkhaldoun, Z., Narita, N., Fukui, A., Kusakabe, N., Kawauchi, K., Terada, Y., Bouma, L. G., Weinberg, N. N., & Broome, M. 2020, AJ, 159, 150

Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, Proceedings of the National Academy of Science, 110, 19273

Petigura, E. A., Howard, A. W., Marcy, G. W., Johnson, J. A., Isaacson, H., Cargile, P. A., Hebb, L., Fulton, B. J., Weiss, L. M., Morton, T. D., Winn, J. N., Rogers, L. A., Sinukoff, E., Hirsch, L. A., & Crossfield, I. J. M. 2017a, ArXiv e-prints

Petigura, E. A., Marcy, G. W., Winn, J. N., Weiss, L. M., Fulton, B. J., Howard, A. W., Sinukoff, E., Isaacson, H., Morton, T. D., & Johnson, J. A. 2018, AJ, 155, 89

Petigura, E. A., Sinukoff, E., Lopez, E. D., Crossfield, I. J. M., Howard, A. W., Brewer, J. M., Fulton, B. J., Isaacson, H. T., Ciardi, D. R., Howell, S. B., Everett, M. E., Horch, E. P., Hirsch, L. A., Weiss, L. M., & Schlieder, J. E. 2017b, AJ, 153, 142

Pinsonneault, M. H., Elsworth, Y. P., Tayar, J., Serenelli, A., Stello, D., Zinn, J., Mathur, S., García, R. A., Johnson, J. A., Hekker, S., Huber, D., Kallinger, T., Mészáros, S., Mosser, B., Stassun, K., Girardi, L., Rodrigues, T. S., Silva Aguirre, V., An, D., Basu, S., Chaplin, W. J., Corsaro, E., Cunha, K., García-Hernández, D. A., Holtzman, J., Jönsson, H., Shetrone, M., Smith, V. V., Sobeck, J. S., Stringfellow, G. S., Zamora, O., Beers, T. C., Fernández-Trincado, J. G., Frinchaboy, P. M., Hearty, F. R., & Nitschelm, C. 2018, ApJS, 239, 32

Pinte, C., Ménard, F., Berger, J. P., Benisty, M., & Malbet, F. 2008, ApJ, 673, L63

Plavchan, P. & Bilinski, C. 2013, ApJ, 769, 86

Pollacco, D. L., Skillen, I., Collier Cameron, A., Christian, D. J., Hellier, C., Irwin, J., Lister, T. A., Street, R. A., West, R. G., Anderson, D. R., Clarkson, W. I., Deeg, H., Enoch, B., Evans, A., Fitzsimmons, A., Haswell, C. A., Hodgkin, S., Horne, K., Kane, S. R., Keenan, F. P., Maxted, P. F. L., Norton, A. J., Osborne, J., Parley, N. R., Ryans, R. S. I., Smalley, B., Wheatley, P. J., & Wilson, D. M. 2006, PASP, 118, 1407

Price-Whelan, A. M., Hogg, D. W., Rix, H.-W., Beaton, R. L., Lewis, H. M., Nidever, D. L., Almeida, A., Badenes, C., Barba, R., Beers, T. C., Carlberg, J. K., De Lee, N., Fernández-Trincado, J. G., Frinchaboy, P. M., García-Hernández, D. A., Green, P. J., Hasselquist, S., Longa-Peña, P., Majewski, S. R., Nitschelm, C., Sobeck, J., Stassun, K. G., Stringfellow, G. S., & Troup, N. W. 2020, ApJ, 895, 2

Ramírez, I., Asplund, M., Baumann, P., Meléndez, J., & Bensby, T. 2010, A&A, 521, A33

Ramírez, I., Meléndez, J., & Asplund, M. 2014, A&A, 561, A7

Rein, H. 2012, MNRAS, 427, L21

Rice, W. K. M. & Armitage, P. J. 2003, ApJ, 598, L55

Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. Á., Bean, J. L., Berta-Thompson, Z. K., Brown, T. M., Buchhave, L., Butler, N. R., Butler, R. P., Chaplin, W. J., Charbonneau, D., Christensen-Dalsgaard, J., Clampin, M., Deming, D., Doty, J., De Lee, N., Dressing, C., Dunham, E. W., Endl, M., Fressin, F., Ge, J., Henning, T., Holman, M. J., Howard, A. W., Ida, S., Jenkins, J. M., Jernigan, G., Johnson, J. A., Kaltenegger, L., Kawai, N., Kjeldsen, H., Laughlin, G., Levine, A. M., Lin, D., Lissauer, J. J., MacQueen, P., Marcy, G., McCullough, P. R., Morton, T. D., Narita, N., Paegert, M., Palle, E., Pepe, F., Pepper, J., Quirrenbach, A., Rinehart, S. A., Sasselov, D., Sato, B., Seager, S., Sozzetti, A., Stassun, K. G., Sullivan, P., Szentgyorgyi, A., Torres, G., Udry, S., & Villasenor, J. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. Á., Bean, J. L., Berta-Thompson, Z. K., Brown, T. M., Buchhave, L., Butler, N. R., Butler, R. P., Chaplin, W. J., Charbonneau, D., Christensen-Dalsgaard, J., Clampin, M., Deming, D., Doty, J., De Lee, N., Dressing, C., Dunham, E. W., Endl, M., Fressin, F., Ge, J., Henning, T., Holman, M. J., Howard, A. W., Ida, S., Jenkins, J., Jernigan, G., Johnson, J. A., Kaltenegger, L., Kawai, N., Kjeldsen, H., Laughlin, G., Levine, A. M., Lin, D., Lissauer, J. J., MacQueen, P., Marcy, G., McCullough, P. R., Morton, T. D., Narita, N., Paegert, M., Palle, E., Pepe, F., Pepper, J., Quirrenbach, A., Rinehart, S. A., Sasselov, D., Sato, B., Seager, S., Sozzetti, A., Stassun, K. G., Sullivan, P., Szent-

gyorgyi, A., Torres, G., Udry, S., & Villasenor, J. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9143, Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave, ed. J. Oschmann, Jacobus M., M. Clampin, G. G. Fazio, & H. A. MacEwen, 914320

Rizzuto, A. C., Mann, A. W., Vanderburg, A., Kraus, A. L., & Covey, K. R. 2017, AJ, 154, 224

Rizzuto, A. C., Vanderburg, A., Mann, A. W., Kraus, A. L., Dressing, C. D., Agüeros, M. A., Douglas, S. T., & Krolikowski, D. M. 2018, AJ, 156, 195

Sahu, K. C., Casertano, S., Bond, H. E., Valenti, J., Ed Smith, T., Minniti, D., Zoccali, M., Livio, M., Panagia, N., Piskunov, N., Brown, T. M., Brown, T., Renzini, A., Rich, R. M., Clarkson, W., & Lubow, S. 2006, Nature, 443, 534

Sandquist, E. L., Jessen-Hansen, J., Shetrone, M. D., Brogaard, K., Meibom, S., Leitner, M., Stello, D., Bruntt, H., Antoci, V., Orosz, J. A., Grundahl, F., & Frandsen, S. 2016, ApJ, 831, 11

Santos, N. C., Israelian, G., & Mayor, M. 2004, A&A, 415, 1153

Santos, N. C., Mayor, M., Bonfils, X., Dumusque, X., Bouchy, F., Figueira, P., Lovis, C., Melo, C., Pepe, F., Queloz, D., Ségransan, D., Sousa, S. G., & Udry, S. 2011, A&A, 526, A112

Schlaufman, K. C. 2015, ApJ, 799, L26

Schlaufman, K. C. & Laughlin, G. 2011, ApJ, 738, 177

Schlaufman, K. C. & Winn, J. N. 2013, ApJ, 772, 143

Schuler, S. C., Vaz, Z. A., Katime Santrich, O. J., Cunha, K., Smith, V. V., King, J. R., Teske, J. K., Ghezzi, L., Howell, S. B., & Isaacson, H. 2015, ApJ, 815, 5

Scott, D. W. 2010, WIREs Comput. Stat., 2, 497–502

Sellwood, J. A. & Binney, J. J. 2002, MNRAS, 336, 785

Serenelli, A., Johnson, J., Huber, D., Pinsonneault, M., Ball, W. H., Tayar, J., Silva Aguirre, V., Basu, S., Troup, N., Hekker, S., Kallinger, T., Stello, D., Davies, G. R., Lund, M. N., Mathur, S., Mosser, B., Stassun, K. G., Chaplin, W. J., Elsworth, Y., García, R. A., Handberg, R., Holtzman, J., Hearty, F., García-Hernández, D. A., Gaulme, P., & Zamora, O. 2017, ApJS, 233, 23

Serenelli, A. M., Bergemann, M., Ruchti, G., & Casagrande, L. 2013, MNRAS, 429, 3645

Shetrone, M., Bizyaev, D., Lawler, J. E., Allende Prieto, C., Johnson, J. A., Smith, V. V., Cunha, K., Holtzman, J., García Pérez, A. E., Mészáros, S., Sobeck, J., Zamora, O., García-Hernández, D. A., Souto, D., Chojnowski, D., Koesterke, L., Majewski, S., & Zasowski, G. 2015, ApJS, 221, 24

Silburt, A., Gaidos, E., & Wu, Y. 2015, ApJ, 799, 180

Skrutskie, M. F., Cutri, R. M., Stiening, R., Weinberg, M. D., Schneider, S., Carpenter, J. M., Beichman, C., Capps, R., Chester, T., Elias, J., Huchra, J., Liebert, J., Lonsdale, C., Monet, D. G., Price, S., Seitzer, P., Jarrett, T., Kirkpatrick, J. D., Gizis, J. E., Howard, E., Evans, T., Fowler, J., Fullmer, L., Hurt, R., Light, R., Kopan, E. L., Marsh, K. A., McCallon, H. L., Tam, R., Van Dyk, S., & Wheelock, S. 2006, AJ, 131, 1163

Smith, J. C., Stumpe, M. C., Van Cleve, J. E., Jenkins, J. M., Barclay, T. S., Fanelli, M. N., Girouard, F. R., Kolodziejczak, J. J., McCauliff, S. D., Morris, R. L., & Twicken, J. D. 2012, PASP, 124, 1000

Smith, V. V., Bizyaev, D., Cunha, K., Shetrone, M. D., Souto, D., Allende Prieto, C., Masseron, T., Mészáros, S., Jönsson, H., Hasselquist, S., Osorio, Y., García-Hernández,

D. A., Plez, B., Beaton, R. L., Holtzman, J., Majewski, S. R., Stringfellow, G. S., & Sobeck, J. 2021, AJ, 161, 254

Sneden, C. 1973, ApJ, 184, 839

Sousa, S. G., Santos, N. C., Israelian, G., Mayor, M., & Monteiro, M. J. P. F. G. 2007, A&A, 469, 783

Sousa, S. G., Santos, N. C., Mayor, M., Udry, S., Casagrande, L., Israelian, G., Pepe, F., Queloz, D., & Monteiro, M. J. P. F. G. 2008, A&A, 487, 373

Southworth, J., Hinse, T. C., Jørgensen, U. G., Dominik, M., Ricci, D., Burgdorf, M. J., Hornstrup, A., Wheatley, P. J., Anguita, T., Bozza, V., Novati, S. C., Harpsøe, K., Kjærgaard, P., Liebig, C., Mancini, L., Masi, G., Mathiasen, M., Rahvar, S., Scarpetta, G., Snodgrass, C., Surdej, J., Thöne, C. C., & Zub, M. 2009, MNRAS, 396, 1023

Souto, D., Allende Prieto, C., Cunha, K., Pinsonneault, M., Smith, V. V., Garcia-Dias, R., Bovy, J., García-Hernández, D. A., Holtzman, J., Johnson, J. A., Jönsson, H., Majewski, S. R., Shetrone, M., Sobeck, J., Zamora, O., Pan, K., & Nitschelm, C. 2019, ApJ, 874, 97

Souto, D., Cunha, K., García-Hernández, D. A., Zamora, O., Allende Prieto, C., Smith, V. V., Mahadevan, S., Blake, C., Johnson, J. A., Jönsson, H., Pinsonneault, M., Holtzman, J., Majewski, S. R., Shetrone, M., Teske, J., Nidever, D., Schiavon, R., Sobeck, J., García Pérez, A. E., Gómez Maqueo Chew, Y., & Stassun, K. 2017, ApJ, 835, 239

Souto, D., Cunha, K., Smith, V. V., Allende Prieto, C., García-Hernández, D. A., Pinsonneault, M., Holzer, P., Frinchaboy, P., Holtzman, J., Johnson, J. A., Jönsson, H., Majewski, S. R., Shetrone, M., Sobeck, J., Stringfellow, G., Teske, J., Zamora, O., Zasowski, G., Carrera, R., Stassun, K., Fernandez-Trincado, J. G., Villanova, S., Minniti, D., & Santana, F. 2018, ApJ, 857, 14

Spearman, C. 1904, The American Journal of Psychology, 15, 72

Staelin, D. H. 1969, IEEE Proceedings, 57, 724

Stassun, K. G., Mathieu, R. D., Mazeh, T., & Vrba, F. J. 1999, AJ, 117, 2941

Stefansson, G., Mahadevan, S., Hebb, L., Wisniewski, J., Huehnerhoff, J., Morris, B., Halverson, S., Zhao, M., Wright, J., O'rourke, J., Knutson, H., Hawley, S., Kanodia, S., Li, Y., Hagen, L. M. Z., Liu, L. J., Beatty, T., Bender, C., Robertson, P., Dembicky, J., Gray, C., Ketzeback, W., McMillan, R., & Rudyk, T. 2017, ApJ, 848, 9

Sun, M., Arras, P., Weinberg, N. N., Troup, N. W., & Majewski, S. R. 2018, MNRAS, 481, 4077

Szentgyorgyi, A., Frebel, A., Furesz, G., Hertz, E., Norton, T., Bean, J., Bergner, H., Crane, J., Evans, J., Evans, I., Gauron, T., Jordán, A., Park, S., Uomoto, A., Barnes, S., Davis, W., Eisenhower, M., Epps, H., Guzman, D., McCracken, K., Ordway, M., Plummer, D., Podgorski, W., & Weaver, D. 2012, in Ground-based and Airborne Instrumentation for Astronomy IV, ed. I. S. McLean, S. K. Ramsay, & H. Takami, Vol. 8446, International Society for Optics and Photonics (SPIE), 496 – 510

Teske, J. K., Ghezzi, L., Cunha, K., Smith, V. V., Schuler, S. C., & Bergemann, M. 2015, ApJ, 801, L10

Teske, J. K., Khanal, S., & Ramírez, I. 2016a, ApJ, 819, 19

Teske, J. K., Shectman, S. A., Vogt, S. S., Díaz, M., Butler, R. P., Crane, J. D., Thompson, I. B., & Arriagada, P. 2016b, AJ, 152, 167

Thomas, H. C. 1967, ZAp, 67, 420

Thompson, S. E., Coughlin, J. L., Hoffman, K., Mullally, F., Christiansen, J. L., Burke, C. J., Bryson, S., Batalha, N., Haas, M. R., Catanzarite, J., Rowe, J. F., Barentsen,

G., Caldwell, D. A., Clarke, B. D., Jenkins, J. M., Li, J., Latham, D. W., Lissauer, J. J., Mathur, S., Morris, R. L., Seader, S. E., Smith, J. C., Klaus, T. C., Twicken, J. D., Van Cleve, J. E., Wohler, B., Akeson, R., Ciardi, D. R., Cochran, W. D., Henze, C. E., Howell, S. B., Huber, D., Prša, A., Ramírez, S. V., Morton, T. D., Barclay, T., Campbell, J. R., Chaplin, W. J., Charbonneau, D., Christensen-Dalsgaard, J., Dotson, J. L., Doyle, L., Dunham, E. W., Dupree, A. K., Ford, E. B., Geary, J. C., Girouard, F. R., Isaacson, H., Kjeldsen, H., Quintana, E. V., Ragozzine, D., Shabram, M., Shporer, A., Silva Aguirre, V., Steffen, J. H., Still, M., Tenenbaum, P., Welsh, W. F., Wolfgang, A., Zamudio, K. A., Koch, D. G., & Borucki, W. J. 2018, ApJS, 235, 38

Thorngren, D. P., Fortney, J. J., Murray-Clay, R. A., & Lopez, E. D. 2016, ApJ, 831, 64

Troup, N. W., Nidever, D. L., De Lee, N., Carlberg, J., Majewski, S. R., Fernandez, M., Covey, K., Chojnowski, S. D., Pepper, J., Nguyen, D. T., Stassun, K., Nguyen, D. C., Wisniewski, J. P., Fleming, S. W., Bizyaev, D., Frinchaboy, P. M., García-Hernández, D. A., Ge, J., Hearty, F., Meszaros, S., Pan, K., Allende Prieto, C., Schneider, D. P., Shetrone, M. D., Skrutskie, M. F., Wilson, J., & Zamora, O. 2016, AJ, 151, 85

Turner, J. D., Ridden-Harper, A., & Jayawardhana, R. 2021, AJ, 161, 72

Twicken, J. D., Jenkins, J. M., Seader, S. E., Tenenbaum, P., Smith, J. C., Brownston, L. S., Burke, C. J., Catanzarite, J. H., Clarke, B. D., Cote, M. T., Girouard, F. R., Klaus, T. C., Li, J., McCauliff, S. D., Morris, R. L., Wohler, B., Campbell, J. R., Kamal Uddin, A., Zamudio, K. A., Sabale, A., Bryson, S. T., Caldwell, D. A., Christiansen, J. L., Coughlin, J. L., Haas, M. R., Henze, C. E., Sand erfer, D. T., & Thompson, S. E. 2016, AJ, 152, 158

Ulrich, R. K. 1986, ApJ, 306, L37

Valenti, J. A. & Fischer, D. A. 2005, ApJS, 159, 141

Valenti, J. A. & Piskunov, N. 1996, A&AS, 118, 595

Vanderburg, A. & Johnson, J. A. 2014, PASP, 126, 948

Vanderburg, A., Mann, A. W., Rizzuto, A., Bieryla, A., Kraus, A. L., Berlind, P., Calkins, M. L., Curtis, J. L., Douglas, S. T., Esquerdo, G. A., Everett, M. E., Horch, E. P., Howell, S. B., Latham, D. W., Mayo, A. W., Quinn, S. N., Scott, N. J., & Stefanik, R. P. 2018, AJ, 156, 46

Veras, D. 2016, Royal Society Open Science, 3, 150571

Villaver, E. & Livio, M. 2009, ApJ, 705, L81

Villaver, E., Livio, M., Mustill, A. J., & Siess, L. 2014, ApJ, 794, 3

Vogt, S. S., Allen, S. L., Bigelow, B. C., Bresee, L., Brown, B., Cantrall, T., Conrad, A., Couture, M., Delaney, C., Epps, H. W., Hilyard, D., Hilyard, D. F., Horn, E., Jern, N., Kanto, D., Keane, M. J., Kibrick, R. I., Lewis, J. W., Osborne, J., Pardeilhan, G. H., Pfister, T., Ricketts, T., Robinson, L. B., Stover, R. J., Tucker, D., Ward, J., & Wei, M. Z. 1994, in Proc. SPIE, Vol. 2198, Instrumentation in Astronomy VIII, ed. D. L. Crawford & E. R. Craine, 362

von Braun, K. & Boyajian, T. 2017, arXiv e-prints, arXiv:1707.07405

Walkowicz, L. M. & Basri, G. S. 2013, MNRAS, 436, 1883

Wang, J. & Fischer, D. A. 2015, AJ, 149, 14

Wang, S. & Chen, X. 2019, ApJ, 877, 116

Watson, G. S. 1964, Sankhyā Ser., 26, 359

Weinberg, N. N., Sun, M., Arras, P., & Essick, R. 2017, ApJ, 849, L11

Wilson, J. C., Hearty, F., Skrutskie, M. F., Majewski, S., Schiavon, R., Eisenstein, D., Gunn, J., Blank, B., Henderson, C., Smee, S., Barkhouser, R., Harding, A., Fitzgerald, G., Stolberg, T., Arns, J., Nelson, M., Brunner, S., Burton, A., Walker, E., Lam, C., Maseman, P., Barr, J., Leger, F., Carey, L., MacDonald, N., Horne, T., Young, E., Rieke, G., Rieke, M., O'Brien, T., Hope, S., Krakula, J., Crane, J., Zhao, B., Carr, M., Harrison, C., Stoll, R., Vernieri, M. A., Holtzman, J., Shetrone, M., Allende-Prieto, C., Johnson, J., Frinchaboy, P., Zasowski, G., Bizyaev, D., Gillespie, B., & Weinberg, D. 2010, in Proc. SPIE, Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III, 77351C

Wilson, J. C., Hearty, F., Skrutskie, M. F., Majewski, S. R., Schiavon, R., Eisenstein, D., Gunn, J., Gillespie, B., Weinberg, D., Blank, B., Henderson, C., Smee, S., Barkhouser, R., Harding, A., Hope, S., Fitzgerald, G., Stolberg, T., Arns, J., Nelson, M., Brunner, S., Burton, A., Walker, E., Lam, C., Maseman, P., Barr, J., Leger, F., Carey, L., MacDonald, N., Ebelke, G., Beland, S., Horne, T., Young, E., Rieke, G., Rieke, M., O'Brien, T., Crane, J., Carr, M., Harrison, C., Stoll, R., Vernieri, M., Holtzman, J., Nidever, D., Shetrone, M., Allende-Prieto, C., Johnson, J., Frinchaboy, P., Zasowski, G., Garcia Perez, A., Bizyaev, D., & Zhao, B. 2012, in American Astronomical Society Meeting Abstracts, Vol. 219, American Astronomical Society Meeting Abstracts #219, 428.02

Wilson, J. C., Hearty, F. R., Skrutskie, M. F., Majewski, S. R., Holtzman, J. A., Eisenstein, D., Gunn, J., Blank, B., Henderson, C., Smee, S., Nelson, M., Nidever, D., Arns, J., Barkhouser, R., Barr, J., Beland, S., Bershady, M. A., Blanton, M. R., Brunner, S., Burton, A., Carey, L., Carr, M., Colque, J. P., Crane, J., Damke, G. J., Davidson, J. W., J., Dean, J., Di Mille, F., Don, K. W., Ebelke, G., Evans, M., Fitzgerald, G., Gillespie, B., Hall, M., Harding, A., Harding, P., Hammond, R., Hancock, D., Harrison, C., Hope, S., Horne, T., Karakla, J., Lam, C., Leger, F., MacDonald, N., Maseman, P.,

Matsunari, J., Melton, S., Mitcheltree, T., O'Brien, T., O'Connell, R. W., Patten, A., Richardson, W., Rieke, G., Rieke, M., Roman-Lopes, A., Schiavon, R. P., Sobeck, J. S., Stolberg, T., Stoll, R., Tembe, M., Trujillo, J. D., Uomoto, A., Vernieri, M., Walker, E., Weinberg, D. H., Young, E., Anthony-Brumfield, B., Bizyaev, D., Breslauer, B., De Lee, N., Downey, J., Halverson, S., Huehnerhoff, J., Klaene, M., Leon, E., Long, D., Mahadevan, S., Malanushenko, E., Nguyen, D. C., Owen, R., Sánchez-Gallego, J. R., Sayres, C., Shane, N., Shectman, S. A., Shetrone, M., Skinner, D., Stauffer, F., & Zhao, B. 2019, PASP, 131, 055001

Wilson, R. F., Teske, J., Majewski, S. R., Cunha, K., Smith, V., Souto, D., Bender, C., Mahadevan, S., Troup, N., Allende Prieto, C., Stassun, K. G., Skrutskie, M. F., Almeida, A., García-Hernández, D. A., Zamora, O., & Brinkmann, J. 2018, AJ, 155, 68

Winn, J. N. Planet Occurrence: Doppler and Transit Surveys, ed. H. J. Deeg & J. A. Belmonte, 195

Winn, J. N. & Fabrycky, D. C. 2015, ARA&A, 53, 409

Winn, J. N., Sanchis-Ojeda, R., Rogers, L., Petigura, E. A., Howard, A. W., Isaacson, H., Marcy, G. W., Schlaufman, K., Cargile, P., & Hebb, L. 2017, ArXiv e-prints

Wolszczan, A. & Frail, D. A. 1992, Nature, 355, 145

Wyse, R. F. G. 1995, in IAU Symposium, Vol. 164, Stellar Populations, ed. P. C. van der Kruit & G. Gilmore, 133

Xiao, L., Niu, R., & Zhang, H. 2017, MNRAS, 467, 2869

Yadav, R. K. S., Bedin, L. R., Piotto, G., Anderson, J., Cassisi, S., Villanova, S., Platais, I., Pasquini, L., Momany, Y., & Sagar, R. 2008, A&A, 484, 609

Yasui, C., Kobayashi, N., Tokunaga, A. T., Saito, M., & Tokoku, C. 2009, ApJ, 705, 54

—. 2010, ApJ, 723, L113

Youdin, A. N. 2011, ApJ, 742, 38

Youdin, A. N. & Goodman, J. 2005, ApJ, 620, 459

Zamora, O., García-Hernández, D. A., Allende Prieto, C., Carrera, R., Koesterke, L., Edvardsson, B., Castelli, F., Plez, B., Bizyaev, D., Cunha, K., García Pérez, A. E., Gustafsson, B., Holtzman, J. A., Lawler, J. E., Majewski, S. R., Manchado, A., Mészáros, S., Shane, N., Shetrone, M., Smith, V. V., & Zasowski, G. 2015, AJ, 149, 181

Zasowski, G., Cohen, R. E., Chojnowski, S. D., Santana, F., Oelkers, R. J., Andrews, B., Beaton, R. L., Bender, C., Bird, J. C., Bovy, J., Carlberg, J. K., Covey, K., Cunha, K., Dell'Agli, F., Fleming, S. W., Frinchaboy, P. M., García-Hernández, D. A., Harding, P., Holtzman, J., Johnson, J. A., Kollmeier, J. A., Majewski, S. R., Mészáros, S., Munn, J., Muñoz, R. R., Ness, M. K., Nidever, D. L., Poleski, R., Román-Zúñiga, C., Shetrone, M., Simon, J. D., Smith, V. V., Sobeck, J. S., Stringfellow, G. S., Szigetiáros, L., Tayar, J., & Troup, N. 2017, AJ, 154, 198

Zasowski, G., Johnson, J. A., Frinchaboy, P. M., Majewski, S. R., Nidever, D. L., Rocha Pinto, H. J., Girardi, L., Andrews, B., Chojnowski, S. D., Cudworth, K. M., Jackson, K., Munn, J., Skrutskie, M. F., Beaton, R. L., Blake, C. H., Covey, K., Deshpande, R., Epstein, C., Fabbian, D., Fleming, S. W., Garcia Hernandez, D. A., Herrero, A., Mahadevan, S., Mészáros, S., Schultheis, M., Sellgren, K., Terrien, R., van Saders, J., Allende Prieto, C., Bizyaev, D., Burton, A., Cunha, K., da Costa, L. N., Hasselquist, S., Hearty, F., Holtzman, J., García Pérez, A. E., Maia, M. A. G., O'Connell, R. W., O'Donnell, C., Pinsonneault, M., Santiago, B. X., Schiavon, R. P., Shetrone, M., Smith, V., & Wilson, J. C. 2013, AJ, 146, 81

Zhou, G., Bakos, G. Á., Hartman, J. D., Latham, D. W., Torres, G., Bhatti, W., Penev,

K., Buchhave, L., Kovács, G., Bieryla, A., Quinn, S., Isaacson, H., Fulton, B. J., Falco, E., Csubry, Z., Everett, M., Szklenar, T., Esquerdo, G., Berlind, P., Calkins, M. L., Béky, B., Knox, R. P., Hinz, P., Horch, E. P., Hirsch, L., Howell, S. B., Noyes, R. W., Marcy, G., de Val-Borro, M., Lázár, J., Papp, I., & Sári, P. 2017, AJ, 153, 211

Zhu, W. 2019, ApJ, 873, 8

Ziegler, C., Law, N. M., Baranec, C., Riddle, R., Duev, D. A., Howard, W., Jensen-Clem, R., Kulkarni, S. R., Morton, T., & Salama, M. 2018, The Astronomical Journal, 155, 161

Zinn, J. C., Pinsonneault, M. H., Huber, D., & Stello, D. 2019a, ApJ, 878, 136

Zinn, J. C., Pinsonneault, M. H., Huber, D., Stello, D., Stassun, K., & Serenelli, A. 2019b, ApJ, 885, 166

Zong, W., Fu, J.-N., De Cat, P., Wang, J., Shi, J., Luo, A., Zhang, H., Frasca, A., Molenda-Żakowicz, J., Gray, R. O., Corbally, C. J., Catanzaro, G., Cang, T., Wang, J., Chen, J., Hou, Y., Liu, J., Niu, H., Pan, Y., Tian, H., Yan, H., Zhang, Y., & Zuo, H. 2020, ApJS, 251, 15