The Exoplanet Population Observation Simulator.
I. The Inner Edges of Planetary Systems

Item Type	Article
Authors	Mulders, Gijs D.; Pascucci, Ilaria; Apai, Dániel; Ciesla, Fred J.
Citation	Gijs D. Mulders et al 2018 AJ 156 24
DOI	10.3847/1538-3881/aac5ea
Publisher	IOP PUBLISHING LTD
Journal	ASTRONOMICAL JOURNAL
Rights	© 2018. The American Astronomical Society. All rights reserved.
Download date	27/08/2022 11:35:58
Item License	http://rightsstatements.org/vocab/InC/1.0/
Version	Final published version
Link to Item	http://hdl.handle.net/10150/628623

The Exoplanet Population Observation Simulator. I. The Inner Edges of Planetary Systems

Gijs D. Mulders[1,2] , Ilaria Pascucci[1,2,3] , Dániel Apai[1,2,3,4] , and Fred J. Ciesla[2,5]
[1] Lunar and Planetary Laboratory, The University of Arizona, Tucson, AZ 85721, USA; mulders@lpl.arizona.edu
[2] Earths in Other Solar Systems Team, NASA Nexus for Exoplanet System Science, USA
[3] Max Planck Institute for Astronomy, Königstuhl 17, Heidelberg, D-69117, Germany
[4] Department of Astronomy, The University of Arizona, Tucson, AZ 85721, USA
[5] Department of the Geophysical Sciences, The University of Chicago, 5734 South Ellis Avenue, Chicago, IL 60637, USA

Abstract

The *Kepler* survey provides a statistical census of planetary systems out to the habitable zone. Because most planets are non-transiting, orbital architectures are best estimated using simulated observations of ensemble populations. Here, we introduce EPOS, the Exoplanet Population Observation Simulator, to estimate the prevalence and orbital architectures of multi-planet systems based on the latest *Kepler* data release, DR25. We estimate that at least 42% of Sun-like stars have nearly coplanar planetary systems with seven or more exoplanets. The fraction of stars with at least one planet within 1 au could be as high as 100% depending on assumptions about the distribution of single transiting planets. We estimate an occurrence rate of planets in the habitable zone around Sun-like stars of $\eta_\oplus = 36 \pm 14\%$. The innermost planets in multi-planet systems are clustered around an orbital period of 10 days (0.1 au), reminiscent of the protoplanetary disk inner edge, or which could be explained by a planet trap at that location. Only a small fraction of planetary systems have the innermost planet at long orbital periods, with fewer than $\approx 8\%$ and $\approx 3\%$ having no planet interior to the orbit of Mercury and Venus, respectively. These results reinforce the view that the solar system is not a typical planetary system, but an outlier among the distribution of known exoplanetary systems. We predict that at least half of the habitable zone exoplanets are accompanied by (non-transiting) planets at shorter orbital periods, hence knowledge of a close-in exoplanet could be used as a way to optimize the search for Earth-size planets in the Habitable Zone with future direct imaging missions.

Key words: methods: statistical – planetary systems – planets and satellites: formation – protoplanetary disks – surveys

1. Introduction

The *Kepler* exoplanet survey has revolutionized our understanding of planetary systems around other stars (e.g., Borucki 2017). During the first four years of the mission, hereafter *Kepler*, thousands of exoplanets and exoplanet candidates were discovered, the majority of which are smaller than Neptune and orbit close to their hosts stars. Because transit surveys can detect only a small fraction of the exoplanets, a completeness correction is necessary to understand the true population of exoplanets (e.g., Batalha 2014).

The planet occurrence rate, defined as the average number of planets per star, is typically found to be of order unity for planets in the orbital period and planet radius range detectable with *Kepler* (Youdin 2011; Fressin et al. 2013). Much progress has been made in recent years in understanding the survey detection efficiency of the *Kepler* mission (e.g., Christiansen et al. 2015, 2016). While Earth-sized planets in the habitable zone of Sun-like stars are just below the detection limits of *Kepler* (Thompson et al. 2018), occurrence rates of Earth-sized planets in the habitable zone, $\eta_\oplus$, can be estimated by considering smaller M and K dwarf stars (Dressing & Charbonneau 2013, 2015) or by extrapolating from shorter orbital periods and/or larger planet radii (Petigura et al. 2013a; Foreman-Mackey et al. 2014; Burke et al. 2015).

The exoplanet population discovered by the *Kepler* mission has been characterized in great detail. Key findings are that planet occurrence rates increase with decreasing planet size (Howard et al. 2012) and remain roughly constant for planets smaller than 3 $R_\oplus$ (Petigura et al. 2013b; Morton & Swift 2014; Mulders et al. 2015b). Recently, a gap in the planet radius distribution has been identified around 1.5–2.0 $R_\oplus$ (Fulton et al. 2017; Van Eylen et al. 2017). The occurrence of sub-Neptunes increases with distance from the host star up to an orbital period of ∼10 days, after which it remains roughly constant (Howard et al. 2012; Mulders et al. 2015a), in contrast to planets larger than Neptune whose occurrence increases with the orbital period (Dong & Zhu 2013).

The presence of multiple transiting planets around the same star provides important additional constraints on planetary orbital architectures. The large number of observed systems with multiple transiting planets indicate that planets are preferentially located in systems with small mutual inclinations (Lissauer et al. 2011, 2012; Fang & Margot 2012). Their orbital spacings follow a peaked distribution without a clear preference for being in orbital resonances (Fabrycky et al. 2014). However, the majority of planets in multi-planet systems are not transiting, and modifying planetary occurrence rates from those based on the observed planetary architectures to account for non-transiting or otherwise undetected planets is not straightforward. The true multiplicity of planetary systems can only be constrained from the observed multiplicity and by making assumptions about their mutual inclination distribution (Tremaine & Dong 2012, see also Brakensiek & Ragozzine 2016). Ensemble populations of systems with six or more planets with mutual inclinations of a few degrees provide a good match to the observed population of multi-planet systems
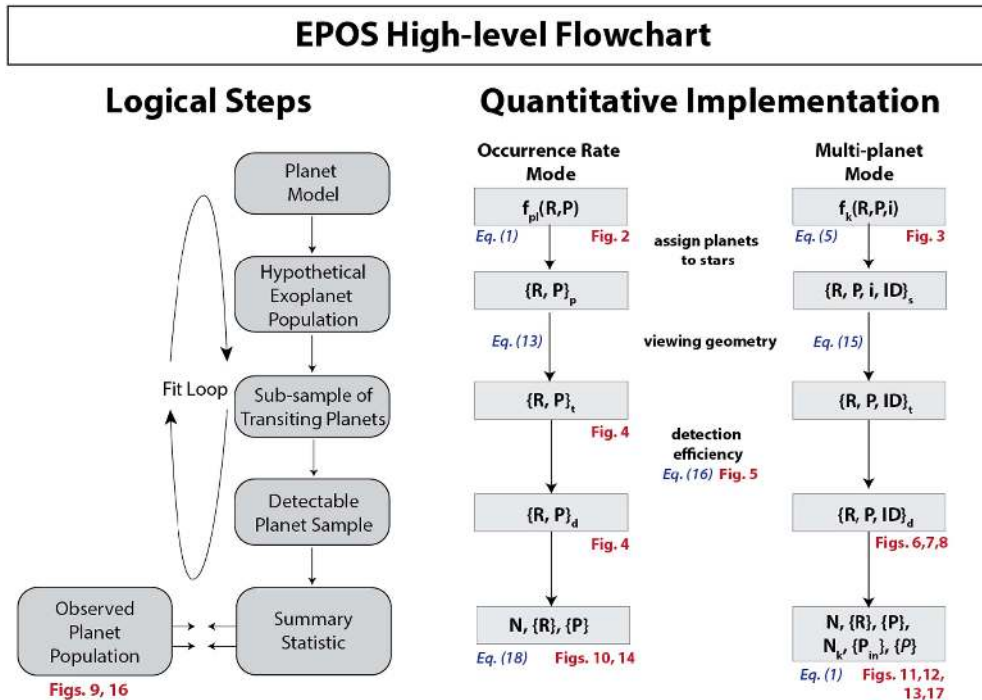
Figure 1. Flowchart of the Exoplanet Population Observation Simulator. A description of all the mathematical symbols can be found in Table 3.

(Lissauer et al. 2011; Fang & Margot 2012; Fabrycky et al. 2014; Ballard & Johnson 2016). However, these simulations underpredict the number of single transiting systems by a factor of 2, indicating that additional populations of planets must be present that are either intrinsically single or have high mutual inclinations.

The occurrence rates and planetary architectures place strong constraints on planet formation models (e.g., Hansen & Murray 2013; Dawson et al. 2016). However, the quantitative comparison of planet formation and orbital evolution models and the observed exoplanet population has been greatly complicated by observational biases. Therefore, published comparisons resorted to trying to explain the presence of specific planet sub-populations by identifying processes that can give rise to such planets. While these studies are essential, such qualitative comparisons can not make full use of the wealth of information represented by the overall exoplanet population statistics. To address this limitation we introduce EPOS,[6] the Exoplanet Population Observation Simulator, that provides a Python interface comparing planet population synthesis models to exoplanet survey data, taking into account the detection biases that skew the observed distributions of planet properties. EPOS uses a forward modeling approach to simulate observable exoplanet populations and constrain their properties via Markov Chain Monte Carlo simulation using emcee (Foreman-Mackey et al. 2013), and has already been employed in two different studies (Kopparapu et al. 2018; Pascucci et al. 2018).

In this paper, we verify this approach using parametric models of planet populations, that we compare to the final data release of the *Kepler* mission, DR25. From this, we are able to make a statistical evaluation of the properties of exoplanetary systems. Among other results, we report on the location of the innermost planet in planetary systems to place our Solar

System in the context of exoplanet systems. We examine how the innermost planets in most of the *Kepler* systems are located much closer in than Mercury and Venus. In an upcoming paper, we will use EPOS to make a direct comparison between planet formation models and exoplanet populations.

2. Code Description

The Exoplanet Population Observation Simulator, EPOS, employs a forward modeling approach to constrain exoplanet populations through the following iterative procedure:

Step 1: Define a distribution of planetary systems from analytic forms for planet occurrence rates and planetary architectures or from a planet population synthesis model.

Step 2: From this, derive a transiting planet population by assigning random orientations to each system and evaluating which planets transit their host stars.

Step 3: Determine which of the transiting planets would be detected by *Kepler*, accounting for detection efficiency.

Step 4: Compare the detectable planet population with exoplanet survey data using a summary statistic.

Step 5: Repeat steps 1–4 until the simulated detectable planet population matches the observed planet population to constrain the intrinsic distribution of planetary systems.

In this section we will describe each step in greater detail. Steps 1–4 take less than a second to run on a single-core CPU, allowing for an efficient sampling of the parameter space using MCMC methods. Figure 1 summarizes these steps and their quantitative implementation in a flowchart. A description of all the mathematical symbols used in this paper can be found in the appendix, Table 3.

2.1. Step 1: Planet Distributions

Planet populations are generated using a Monte Carlo simulation by random draws from a multi-dimensional

probability distribution function, f, which represents the intrinsic distribution of planets and planetary systems. We simulate a planet survey equal in size to the *Kepler* survey, roughly 160,000 stars. The parametric descriptions of f are based on studies of planet occurrence rate and planet multiplicity from *Kepler*. Distributions based on planet formation models will be described in an upcoming paper.

Here we describe two parametric descriptions for the planet population f that correspond to two different modes in EPOS:

Occurrence rate mode: Simulates only the distribution of planet radius, R, and orbital period, P, as $f = f_{pl}(R, P)$. This mode is similar to occurrence rate calculations that estimate the average number of planets per star, η, in a certain period and radius range. This mode can also be used to estimate the planet mass distribution for comparison with microlensing data, see Pascucci et al. (2018) for details.

Multi-planet mode: Simulates multiple planets per system, taking into account their relative spacing and mutual inclination, Δi, to estimate orbital architectures. The properties of each planet in the system are drawn from a distribution $f = f_k(R, P, i)$, where i is the planet's orbital inclination and k is an index for each planet in the system.

2.1.1. Occurrence Rate Mode

The central assumption we use in this paper is that the planet occurrence rate distribution is a separable function in the planet radius and orbital period,

$$f_{pl}(P, R) \propto f_P(P)f_R(R). \tag{1}$$

The distribution is normalized such that the integral over the simulated period and planet radius range equals the number of planets per star, η:

$$\eta = \int_P \int_R f_{pl}(P, R)\, d\log P\, d\log R \tag{2}$$

The planet orbital period distribution is described by a broken power law

$$f_P(P) = \begin{cases} (P/P_{break})^{a_P}, & \text{if } P < P_{break} \\ (P/P_{break})^{b_P}, & \text{otherwise} \end{cases} \tag{3}$$

where the break reflects the observed flattening of sub-Neptune occurrence rates around an orbital period of ten days (e.g., Howard et al. 2012; Mulders et al. 2015a). In this example we use $a_P = 1.5$ for the power law index of the increase interior to $P_{break} = 10$ days, and a flat occurrence rate $b_P = 0$, at longer periods but we will refine these values later.

The planet radius distribution function is similarly described by a broken power law

$$f_R(R) = \begin{cases} (R/R_{break})^{a_R}, & \text{if } R < R_{break} \\ (R/R_{break})^{b_R}, & \text{otherwise} \end{cases} \tag{4}$$

reflecting the observed increase of small planets with decreasing radius ($b_R = -4$), and the "plateau" ($a_R = 0$) of constant occurrence rates for planets smaller than $R_{break} = 3\,R_\oplus$ (e.g., Howard et al. 2012; Petigura et al. 2013b).

The probability distribution function of planet occurrence thus has seven free parameters (η, P_{break}, a_P, b_P, R_{break}, a_R, b_R). An example for $\eta = 2$ is shown in Figure 2.
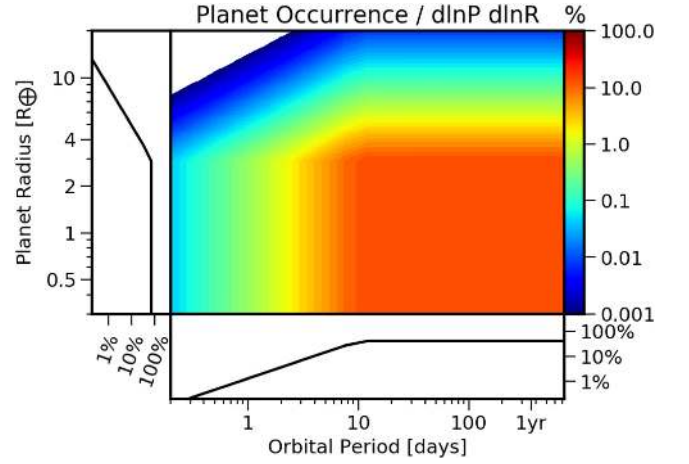
Figure 2. Example planet probability distribution function, $f_{pl}(P, R)$. The color indicates the planet occurrence rate in percent per unit area of $d \ln P\, d \ln R$. The side panels show the marginalized distributions, $f_P(P)$ in units of $d \ln P$ and $f_R(R)$ in units of $d \ln R$.

2.1.2. Multi-planet Mode

In its multi-planet mode EPOS simulates a planetary system for each star. Each planet in the system, denoted by subscript k, is characterized by its radius, orbital period, and orbital inclination with respect to the observer, i, and there are m planets per system. Because the properties of planets in multi-planet systems tend to be correlated, we draw a typical set of planet properties for each system (R_s, P_s, i_s) on which the properties of each planet in the system (R_k, P_k, i_k) are dependent:

$$\begin{aligned} f_{pl,k}(R, P, i\,|\, R_s, P_s, i_s) &\propto f_{R,k}(R|\,R_s) \\ &\quad f_{P,k}(P|\,P_s) \\ &\quad f_{i,k}(i|\,i_s), \end{aligned} \tag{5}$$

where we make the assumption that distributions of planet size, period, and inclination are not interdependent. Previous studies have shown that planets in multi-planet systems tend to be more similar in size and more regularly spaced than random pairings from the overall distribution (Millholland et al. 2017; Weiss et al. 2018) and have low mutual inclinations of a few degrees (e.g., Fabrycky et al. 2014). This motivates our choice to not draw the properties of each planet independently from a distribution $f_{pl}(R, P, i)$.

The system properties (R_s, P_s, i_s) are drawn from a distribution

$$g_{pl}(R_s, P_s, i_s) \propto g_R(R_s)g_P(P_s)g_i(i_s) \tag{6}$$

where once again we have made the assumption that the distributions of planet size, period, and inclination are separable functions of each variable. The distribution of g is normalized such that the integral over the simulated parameter range is equal to the fraction of stars with planetary systems, η_s:

$$\eta_s = \int_P \int_R \int_i g_{pl}(P, R, i)\, d\log P\, d\log R\, di \tag{7}$$

The parameterization of the system and planet properties are described below. The system architecture is illustrated in Figure 3.
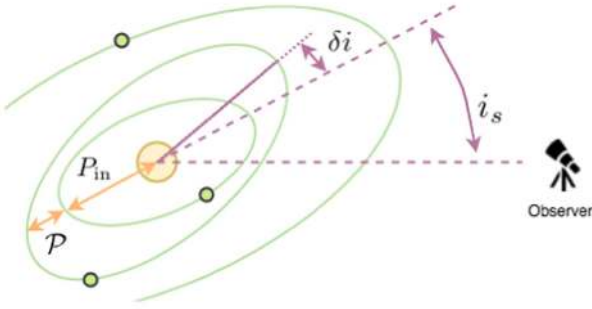
Figure 3. Illustration of the planetary system architecture. P_{in} denotes the orbital period of the innermost planet, while $\mathcal{P}$ denotes the period ratio between adjacent planets. The system is inclined with respect to the observer by an angle i_s, and each planet has a mutual inclination δi with respect to this system inclination. Note that each planet has a different mutual inclination and therefore a different inclination with respect to the line of sight. All orbits are assumed to be circular and the ellipticity of the projected orbits are due to their respective inclinations. The probability distributions for i_s, δi, $\mathcal{P}$, and P_{in} are described in the text.

Orbital Inclination—A planet's orbital inclination with the respect to the line of sight, i_p, is dependent on the inclination of the system, i_s, the mutual inclination of the planet with respect to that system, δi, and the longitude of the ascending node, $\delta\Omega$.

We assume that the planet's mutual inclinations, δi, follow a Rayleigh distribution:

$$f_{\delta i,k}(\delta i) = \delta i \Big/ \Delta i e^{-\delta i^2/(2\Delta i^2)}, \tag{8}$$

where Δi is the mode of the mutual inclination distribution of planetary orbits, which is typically $1°$–$3°$ (Fabrycky et al. 2014). The distribution of system inclinations with respect to the observer, i_s, is proportionate to $\cos(i)$.

$$g_i(i_s) \propto \cos(i_s) \tag{9}$$

where $i_s = 0$ is an edge-on orbit. The distribution of planet inclinations with respect to the observer is then

$$f_{i,k}(i|i_s) = |i_s + f_{\delta i,k}(\delta i)\cos(\pi\delta\Omega)| \tag{10}$$

where $\delta\Omega$ is the longitude of the ascending node with respect to the observer.

Orbital Period—We assume that the planets are regularly spaced in the logarithm of the orbital period, with the location of the first planet ($k = 0$) defining the location of additional planets ($k = 1...m$) in the system. The period ratio of adjacent planets is denoted by $\mathcal{P}_k \equiv P_{k+1}/P_k$, where P_{k+1} is always the planet with the larger orbital period.

The location of the first planet in the system, ($P_0 \equiv P_s$), is parameterized by the broken power law distribution described by Equation (3), but with a steeper decline in planet occurrence with orbital period ($b_P \lesssim 0$) that we will constrain in the fitting process.

Observed orbital spacings follow a broad range in period ratios, and we follow Malhotra (2015) in parameterizing the distribution of dimensionless spacings, $D_k = 2\frac{\mathcal{P}_k^{2/3}-1}{\mathcal{P}_k^{2/3}+1}$, as a log-normal distribution. The orbital period distribution of the kth planet in the system is then given by

$$f_{P,k}(P|P_{k-1}) = \frac{1}{\sqrt{2\pi}\,\sigma}e^{\frac{(\log D_k - D)^2}{2\sigma^2}} \tag{11}$$

with respect to the orbital period of the previous planet, P_{k-1}. D and σ are free parameters that characterize the median and
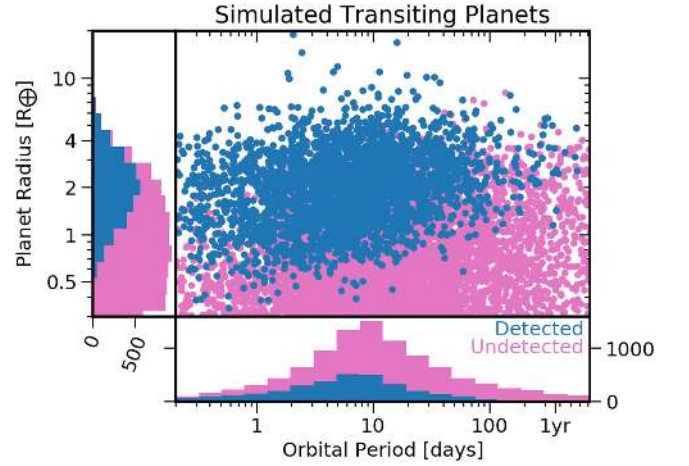
width of the distribution, with typical values of $D \approx -0.4$ and $\sigma \approx 0.2$.

Planet Radius—We assume all planets in the system are of equal size
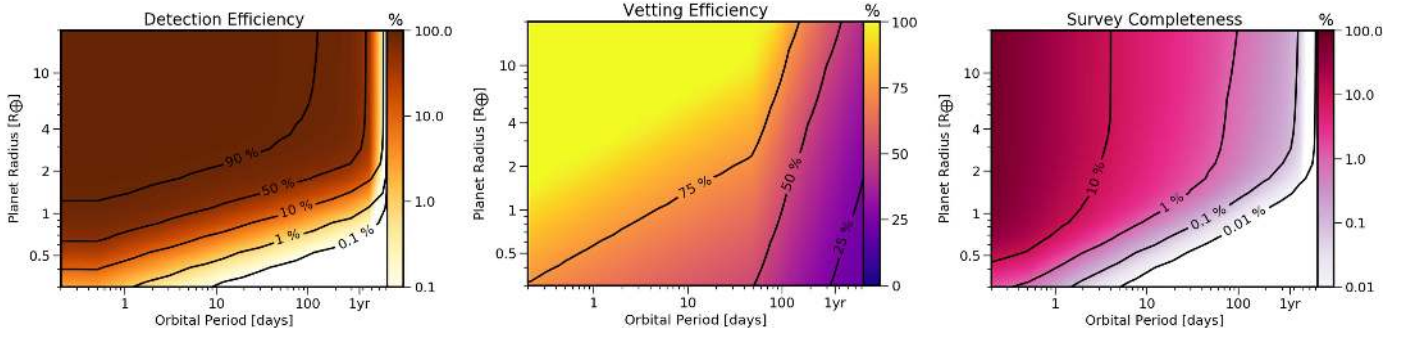
$$f_{R,k}(R|R_s) = g_R(R_s) \tag{12}$$

where we assume the radius distribution follows Equation (4). We choose equal-sized planets over randomly assigned sizes because planets in multi-planet systems tend to be of similar size (Lissauer et al. 2011; Millholland et al. 2017; Weiss et al. 2018). While there are observed trends of increasing planet size with the orbital period, these trends are strongest at short orbital periods (e.g., Carrera et al. 2018) and for large planet sizes (Ciardi et al. 2013; Huang et al. 2016), which we will exclude from our observational comparison, see Section 2.4.

Thus, the number of free parameters describing the population of planetary systems is 11 (η, m, Δi, P_{break}, a_P, b_P, D, σ, R_{break}, a_R, b_R).

2.2. Step 2: Transiting Planet Populations

The synthetic planet population is simulated using a Monte Carlo approach by sampling the distribution function of planet properties (Equation (1) or (5)) n_p times. We simulate planetary systems in the period range [0.5, 730] days and [0.3, 20]$R_\oplus$ (Figure 4). For each planet, 2 random numbers are drawn to determine its properties. The first random number is used to determine the planet orbital period from the cumulative distribution function of Equation (3). The second random number is used to determine the planet radius from the cumulative distribution function of Equation (4).

The total number of draws in the synthetic survey is $n_p = \eta n_\star$, where $n_\star$ is the number of stars in the survey ($n_\star \approx 160,000$), multiplied by the average number of planets per star, η. In occurrence rate mode, with $\eta \approx 2$, the simulated sample $\{R, P\}_p$ consists of $n_s \approx 240,000$ planets. In multi-planet mode, with $\eta_s \approx 0.35$ and $m \approx 6$, the simulated sample $\{R, P, i\}_s$ consists of the same number of planets distributed across $n_s \approx 40,000$ systems. Each star in the survey is assigned a unique identifier, ID, to keep track of the observable planet multiplicity.

Figure 4. Simulated sample of transiting (pink) and detectable (blue) planets generated in occurrence rate mode, see Figure 2.

Figure 5. Planet detection efficiency for dwarf stars from the *Kepler* mission for the DR25 data release as a function of the orbital period and planet radius. Left Panel: The detection efficiency, $f_{S/N}$, the probability that a planet candidate is detected based on the signal-to-noise of the transit. Middle Panel: The vetting efficiency, f_{vet}, the probability that a planet is classified as a reliable planet candidate (a *Robovetter* disposition score of $\geqslant 0.9$). Right Panel: The survey completeness, f_{det}, which includes both the detection and vetting efficiency as well as the the geometric transit probability, $f_{det} = f_{geo}\, f_{S/N}\, f_{vet}$.

In occurrence rate mode, the subset of transiting planets can be simulated by considering the geometric transit probability

$$f_{geo} = R_\star / a. \tag{13}$$

The semimajor axis a is calculated using Kepler's third law. The stellar mass and radius are the average values of the surveyed stars (see next section). A planet is transiting if

$$\chi < f_{geo} \tag{14}$$

where χ is a continuous random variable between 0 and 1. The transiting planet sample, $\{R, P\}_t$, typically contains 10,000 planets. An example is shown in Figure 4.

2.2.1. Multi-planet Transit Probability

The Monte Carlo simulation of transiting planetary systems goes as follows. First, the system inclination, i_s, is drawn according to Equation (9). The draw from a distribution is performed by generating a random number between 0 and 1 and then we interpolates the parameter value from the normalized cumulative distribution. Next, the mutual inclination of the orbit of each planet in the system, $\delta i_{p,k}$, is drawn from Equation (8). Then the longitude of the ascending node with respect to the observer, $\delta\Omega$, is randomly drawn for each planet from a uniform distribution, $\delta\Omega = \chi$. The inclination, i_k, of each planet's orbit with respect to the line of sight can then be calculated from Equation (10). The transiting planet population, $\{R, P\}_t$, is defined by planets that traverse the stellar limb, given by

$$i_k < \arcsin(f_{geo}), \tag{15}$$

where f_{geo} is the geometric transit probability from Equation (13). We note that for a single planet ($\delta i_{p,k} = 0$), this expression is equivalent to the geometric transit probability of Equation (13). This expression is only valid for small mutual inclinations.

To account for the *Kepler dichotomy*, the apparent excess of single transiting systems (e.g., Ballard & Johnson 2016), we assume that a fraction f_{iso} of planetary systems has an isotropic distribution of orbits described by Equation (13) instead of Equation (15). Typically, $f_{iso} \approx 0.5$, but we will treat it as a free parameter to be constrained from the data.

2.3. Step 3: Survey Detection Efficiency

We simulate a detectable planet sample, $\{R, P\}_d$, from the simulated sample of transiting planets, $\{R, P\}_t$, by taking into account the survey detection efficiency, $f_{S/N}(R, P)$, and the vetting completeness, $f_{vet}(R, P)$, described below and displayed in Figure 5. A planet is detectable if:

$$\chi < f_{S/N}\, f_{vet} \tag{16}$$

where χ is a continuous random variable between 0 and 1. $f_{S/N}(R, P)$ is the detection efficiency based on the combined signal-to-noise ratio of all planet transits. $f_{vet}(R, P)$ is the detection efficiency of the *Kepler Robovetter* for a planet candidate sample with high reliability.

We calculate detection efficiency contours for each individual star observed by *Kepler* using KeplerPORTs[7] (Burke & Catanzarite 2017). We included all stars that were fully searched by the *Kepler* pipeline, for which stellar properties were available in the Mathur et al. (2017) catalog, and for which all detection metrics were available on the exoplanet archive.[8]

The detection efficiencies were evaluated on a grid with 20 logarithmically spaced orbital period bins between 0.2 and 730 days assuming planets on circular orbits, and 21 logarithmically spaced planet radius bins between 0.2 and 20 $R_\oplus$. After removing giant and sub-giant stars according to the effective temperature dependent surface gravity criterion in Huber et al. (2016), the sample consists of $n_\star = 159{,}238$ stars, with a median mass of $M_\star = 0.95\, M_\odot$ and median radius of $R_\star = 0.94\, R_\odot$. We calculate the survey detection efficiency, $f_{S/N}(R, P)$, by averaging the individual contributions of each star, displayed in the upper panel of Figure 5.

Not all transiting planets that are detectable based on the detection efficiency are vetted as planet candidates by the *Kepler Robovetter* (Coughlin et al. 2016). Including only reliable planet candidates, here defined with a disposition score larger than 0.9, further reduces the vetting completeness. The vetting efficiencies were calculated following Thompson et al. (2018) based on the *Kepler* simulated data products.[9] The vetting efficiency was evaluated on the same radius and period grid as the detection efficiency and using the same criterion to

---

[7] https://github.com/nasa/KeplerPORTs
[8] https://exoplanetarchive.ipac.caltech.edu/docs/Kepler_completeness_reliability.html
[9] https://exoplanetarchive.ipac.caltech.edu/docs/KeplerSimulated.html
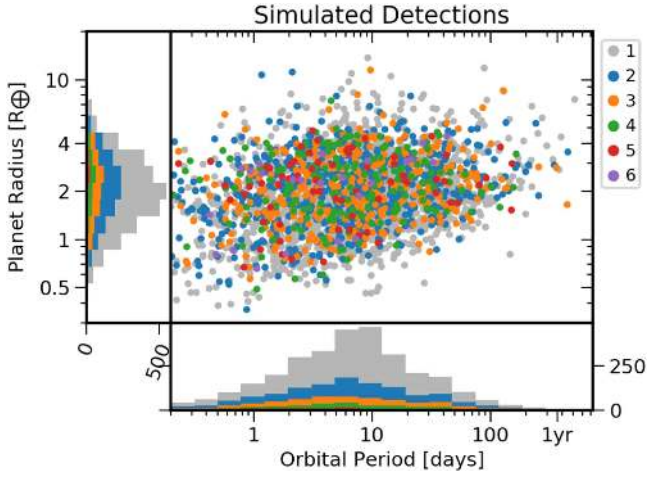
Figure 6. Simulated sample of detectable planetary systems. Planets with no additional planets detected in the system are color coded in gray. Colors indicate the number of observed planets per system.

Figure 7. Period ratio distribution of adjacent planets in multi-planet systems, $\mathcal{P} = P_{k+1}/P_k$. The intrinsic distribution is shown with the red line while the solid blue histograms show the distribution of detections in the simulated survey. The observable distribution is skewed toward shorter orbital period ratios by detection biases. The hatched region indicates the observable planet pairs with at least one non-detected planet between them.

Figure 8. Orbital period distribution of the innermost planet in each system. The intrinsic distribution is shown with the red line while the solid blue histograms show the distribution of detections in the simulated survey. The observable distribution is skewed toward shorter orbital periods by detection biases.

select main-sequence stars. We use a power law in the planet radius and a broken power law in the orbital period to obtain a smooth function for the vetting completeness $f_{\rm vet}$

$$f_{\rm vet} = c \, R_R^a \begin{cases} (P/P_{\rm break})^{a_P}, & \text{if } P < P_{\rm break} \\ (P/P_{\rm break})^{b_P}, & \text{otherwise} \end{cases} \quad (17)$$

with $c = 0.63$, $a_R = 0.19$, $P_{\rm break} = 53$, $a_P = -0.07$, $b_P = -0.39$ (see Appendix B for details). The best-fit vetting efficiency is shown in Figure 5. The vetting efficiency varies between close to 100% for Neptune-sized planets at short orbital periods to 25% near the habitable zone. The break in the power law is needed to match the reduced vetting efficiency for planets at orbital periods larger than $\sim$100 days as described in Thompson et al. (2018).
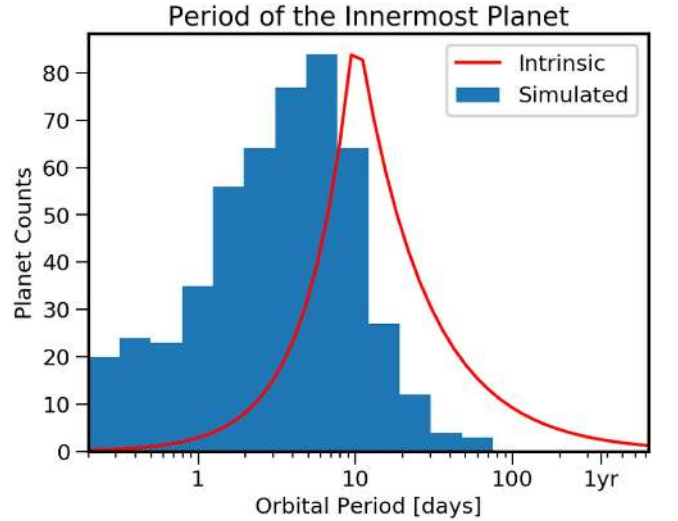
The detectable planet sample, $\{R, P\}_d$, typically contains about 4000 planets (Figure 4). In multi-planet mode, EPOS also keeps track of which planets are observable as part of multi-planet systems (Figure 6). From the observed multi-planet systems we generate a set of summary statistics: N_k, the number of stars with k detectable planets; $\mathcal{P}$, the orbital period ratio between adjacent planets (Figure 7); $P_{\rm in}$, the orbital period of the innermost planet in the system (Figure 8).

It is worth noting that the properties of the observable multi-planet systems are significantly different from the distributions from which they are generated. When the intrinsic population contains only planetary systems with $m \geqslant 7$ planets, the majority of systems have $k = 2$–6 transiting detectable planets. In addition, the observable period and period ratio distributions are skewed by detection biases.

The geometric transit probability favors the detection of planets at shorter orbital periods. The distribution of the location of the innermost *observed* planet in the systems peaks at $P_{\rm in} \sim 6$ days, compared to $P_{\rm in} = 10$ days in the intrinsic distribution (Figure 8). We also find that planetary systems at very short orbital periods ($P_{\rm in} \sim 1$ day) are overrepresented in the detectable distribution by an order of magnitude, while systems with orbital periods similar to the terrestrial planets ($P_{\rm in} \sim 100$ days) are underrepresented by an order of magnitude, due to the same detection biases (Figure 8).

Pairs of planets with smaller orbital period ratios are more likely to be both transiting, shifting the peak of the observable orbital period ratio distribution to smaller period ratios (Figure 7). The hatched area in the bottom panel of Figure 7 shows simulated planet pairs that are observable as adjacent but have a non-transiting planet between them. These planet pairs dominate the period ratio distribution at large orbital period ratios ($\mathcal{P} \gtrsim 4$).

2.4. Step 4: Observational Comparison

In this step we compare the simulated planet populations to the observed exoplanet properties to evaluate how well the simulated planet population reproduces the collection of observed planetary systems. We generate a set of summary statistics for both the simulated data and the *Kepler* survey. We
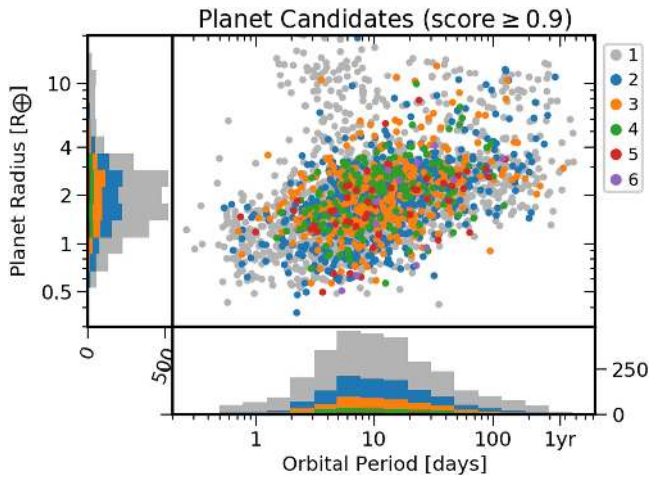
Figure 9. Observed sample of planetary systems. Planets with no additional planets detected in the system are color coded in gray. Colors indicate the number of observed planets per system.

Figure 10. Comparison of simulated planets for the example model (blue) with detected planets (orange). The comparison region (black box) excludes hot Neptunes ($P < 2$ days) and giant planets ($R > 6\,R_\oplus$).

evaluate the survey as a whole, and do not consider dependencies on stellar properties such as stellar mass (Mulders et al. 2015a, 2015b) or metallicity (Mulders et al. 2016; Petigura et al. 2018).

In occurrence rate mode, the summary statistics are the planet radius distribution, $\{R\}_d$, the orbital period distribution, $\{P\}_d$, and the total number of planets, N. In multi-planet mode, additional summary statistics are calculated for the number of stars with k planets, N_k, the period ratio distribution between adjacent planets, $\mathcal{P}$, and the period of the innermost planet in the system, $P_{\rm in}$. We evaluate this summary statistic in the range of $R = [0.5,\ 6]R_\oplus$ and $P = [2,\ 400]$ days. These ranges exclude two regions where the assumptions of separability of parameters clearly breaks down. The maximum planet size of $6\,R_\oplus$ is chosen to exclude giant planets, which have a different distribution of orbital periods than sub-Neptunes (Dong & Zhu 2013; Santerne et al. 2016), are less often part of multi-planet systems (Steffen et al. 2012), or have dissimilar sizes than other planets in the system (Huang et al. 2016). The minimum orbital period of two days is chosen to exclude the photo-evaporation desert (e.g., Lundkvist et al. 2016) where the planet radius distribution deviates significantly from that at larger orbital periods. The other bounds are chosen because there are very few planet detections outside this range.

We then generate a summary statistic from the *Kepler* DR25 catalog. The planet candidate list is taken from Thompson et al. (2018). We include only the main-sequence planet hosts by removing giant and sub-giant stars according to the effective temperature dependent surface gravity criterion (in Huber et al. 2016). We also use a disposition score cut of 0.9 to select a more reliable sample of planet candidates (see Thompson et al. 2018 for details). We account for the lower completeness of this high-reliability planet sample by explicitly taking into account the vetting completeness in the calculation of the survey detection efficiency.

The final list containing 3041 planet candidates is shown in Figure 9. The list contains 1840 observed single systems and 324 double, 113 triple, 38 quadruple, 10 quintuples, and two sextuple systems within the region where the summary statistic is evaluated.
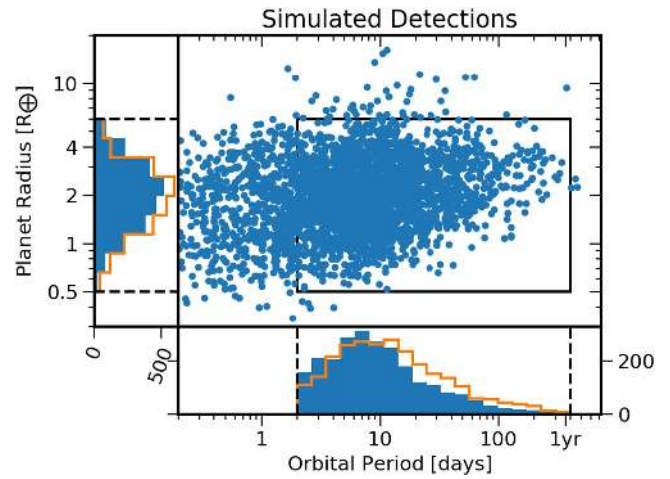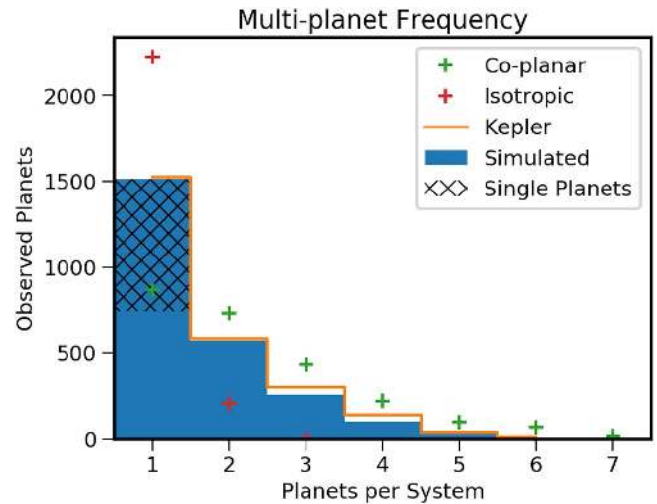

Figure 11. Simulated vs. observed frequency of multi-planet systems. The blue histogram shows the example model with an average mutual inclination of $\Delta i = 2$. Multi-planet statistics from *Kepler* derived in the same radius and period range are shown in orange. The hatched region indicates the excess of single transiting planets, here 40% of systems ($f_{\rm iso} = 0.4$). Crosses indicate a population of planetary systems on coplanar orbits (green, $\Delta i = 0$) and on isotropic orbits (red, $f_{\rm iso} = 1.0$).

2.4.1. Occurrence Rate Mode

Figure 10 shows how the summary statistics of planet radius and orbital period are generated from the detectable planet population (blue) and from the *Kepler* exoplanet population (orange). We compare the planet radius distributions and orbital period distributions separately. While this approach ignores any covariances between planet radius and orbital period that are present in the *Kepler* data, it is consistent with the assumption made in Equation (1) that these functions are separable. We quantify the distance between the two distributions using the two-sample Kolmogorov–Smirnoff (KS) test and calculated the associated probabilities, p_P and p_R, that the observed and simulated distributions are drawn from the same data. We will minimize the differences between these distributions in the fitting step.

Period Ratio of Adjacent Planets
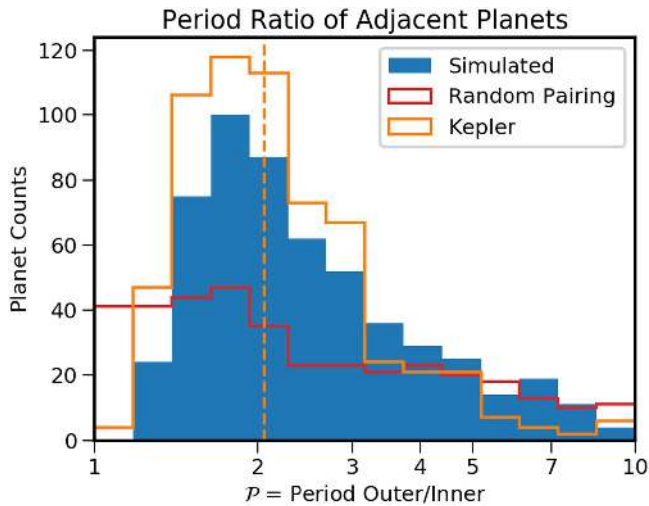
Figure 12. Simulated (blue) vs. observed (orange) period ratio between adjacent planets in multi-planet systems. The red line shows, for comparison, a simulation where planet orbits are randomly drawn from the period-radius distribution (Figure 2) instead of regularly spaced.

Period of the Innermost Planet

Figure 13. Simulated (blue) vs. observed (orange) location of the innermost planet in each system.

2.4.2. Multi-planet Statistics

We calculate three additional summary statistics for multi-planet systems: the frequency of multi-planet systems (N_k, Figure 11); the period ratio of adjacent planet pairs ($\mathcal{P}$, Figure 12); and the distribution of the locations of the innermost planet in each system ($P_{\rm in}$, Figure 13), all evaluated within $R = [0.5, 6]R_{\oplus}$ and $P = [2, 400]$. As discussed in the previous section, the observed planet populations are subject to detection biases, and a proper comparison requires a comparison of the observed distribution to the simulated distribution, which is what EPOS does. All summary statistics are calculated in the same way for observed planets (orange lines) and simulated planets (blue histograms) from the set of orbital periods and host star IDs. We use 2-sample KS tests to calculate the probabilities that the distribution of inner planet orbital periods ($p_{\rm in}$) and orbital period ratios ($p_{\mathcal{P}}$) are drawn from the same distribution as the observations. We use a Pearson χ^2 test to calculate the probability, p_N, that the multi-planet frequencies of the simulated sample are drawn from the same distribution as the observations.

The frequency distribution of planets in multi-planet systems is shown in Figure 11 for a model with $m = 7$ planets per system, a mode for the mutual inclination of $\Delta i = 2°$, and $f_{\rm iso} = 0.4$. The hatched region indicates simulated planets that are on isotropic orbits to match the excess of single transiting systems. We also show a model with only coplanar orbits ($i = 0°$ and $f_{\rm iso} = 0$, green) that overpredicts the frequency of multi-planet systems, particularly at high numbers of planets per system. A model with planets on isotropic orbits ($f_{\rm iso} = 1$, red) underpredicts the frequency of all multi-planet systems.

The period ratio of adjacent planet pairs is shown in Figure 12. The blue histogram shows the observable period ratios of the example model where planets are regularly spaced according to the period ratio distribution of Equation (11). The shape of the period ratio distribution qualitatively reproduces the observed distribution (orange), with a peak near a period ratio of $\mathcal{P} \approx 2$ and a tail toward large orbital period ratios. For comparison, the red line shows a simulation where planetary systems are not regularly spaced, which is constructed by randomly drawing $m = 7$ planets from the period-radius
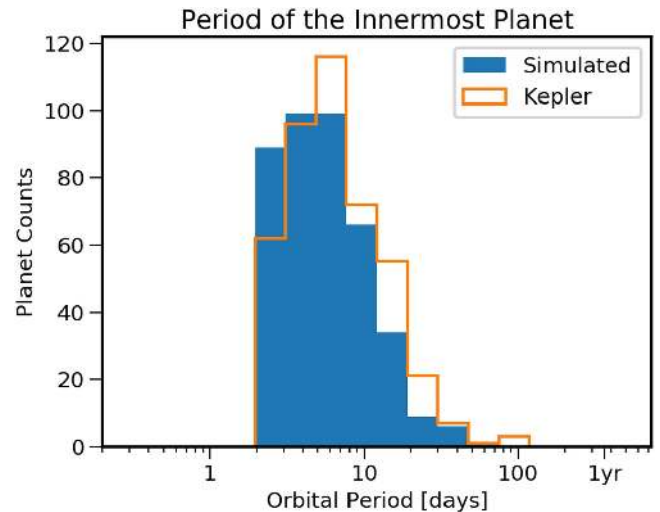
distribution of Equation (1) (Figure 2). These simulated systems have an observable period ratio distribution that is much wider than observed, indicating that planetary systems are regularly spaced (See also Weiss et al. 2018).

The distribution of the location of the innermost detected planet in each of the multi-planet systems is shown in Figure 13. We only focus on the innermost planet of a detected multi-planet system, as for observed single-planet systems it can not be derived from the observable if they are intrinsically single or part of a multi-planet system with non-transiting planets. The distribution of innermost detected planets constrains the fraction of planetary systems without planets at short ($\lesssim$20 days) orbital periods.

2.5. Step 5: Fitting Procedure

With the framework to compare the parameterized distributions of exoplanets to observables we proceed to constrain the parameters to provide the best match to the observed planetary systems. The runtime of steps 1–4 is less than a tenth of a second in occurrence rate mode and less than a second in multi-planet mode, allowing for a comprehensive sampling of parameter space. We use emcee (Foreman-Mackey et al. 2013), an open-source Python implementation of the algorithm by Goodman & Weare (2010) to sample the parameter space and estimate the posterior distribution of parameters.

2.5.1. Occurrence Rate Mode

In occurrence rate mode, we explore the seven-dimensional parameter space consisting of the number of planets per star, η, the orbital period distribution, $P_{\rm break}$, a_P, b_P, and the planet radius distribution, $R_{\rm break}$, a_R, b_R. The summary statistics to evaluate the model given the observations are the number of detected planets, N; the planet orbital period distribution, $\{P\}$; and the planet radius distribution $\{R\}$. We use Fisher's method (Fisher 1925) to combine the probabilities from the summary statistic into a single parameter:

$$L_{\rm occ} = -2(\ln(p_N) + \ln(p_P) + \ln(p_R)), \qquad (18)$$

8

Parameter	Example	Best-fit
η	2.0	$4.9^{+1.3}_{-1.2}$
$P_{\text{break}}(\text{days})$	10	12^{+5}_{-3}
a_P	1.5	$1.5^{+0.5}_{-0.3}$
b_P	0.0	$0.3^{+0.1}_{-0.2}$
$R_{\text{break}}(R_{\oplus})$	3.0	$3.3^{+0.3}_{-0.4}$
a_R	0.0	$-0.5^{+0.2}_{-0.2}$
b_R	-4	-6^{+2}_{-3}

which we use as the likelihood of the model given the data in `emcee`, e.g., $\mathcal{L} \propto \mathcal{L}(N, \{P\}, \{R\}|\eta, P_{\text{break}}, a_P, b_P, R_{\text{break}}, a_R, b_R)$.

We run `emcee` with 200 walkers for 5000 iterations, allowing for a 1000-step burn-in to reach convergence. For the initial positions of the walkers we use the parameters of the example model, see Table 1. The MCMC chain and parameter covariances are shown in the appendix (Figures 22 and 23). The best-fit values and their 1σ confidence intervals are calculated as the 50%, and 16% and 84% percentiles, respectively. The simulated model for the best-fit parameters and for 30 samples of the posterior is shown in Figure 14.

The posterior distribution of planet radius and orbital period are shown in Figure 15. As a sanity check, we calculate occurrence rates as a function of planet radius and orbital period using the inverse detection efficiency method. The occurrence per bin is calculated as

$$\eta_{\text{bin}} = \frac{1}{n_{\star}}\Sigma^{n_p}_j \frac{1}{\text{comp}_j} \qquad (19)$$

where comp_j is the survey completeness evaluated at the radius and orbital period of each planet in the bin (Figure 16). The posterior distributions provide a decent match to the binned occurrence rates, with three notable deviations. At $P \gtrsim 50$ days and $R \lesssim 1.4\,R_{\oplus}$ occurrence rates are lower than the posterior. We attribute this to the bins including regions where *Kepler* has not detected planet candidates (see Figure 16) and occurrence rates are therefore underestimated. The broken power law in planet radius does not describe the population of giant planets at $\sim 10\,R_{\oplus}$, and therefore, in the following, we restrict our observational comparison to $<6\,R_{\oplus}$.

2.6. Multi-planet Systems

In multi-planet mode, we explore the nine-dimensional parameter space consisting of: the fraction of stars with planetary systems, η_s; the mode of the mutual inclination distribution, Δi; the orbital period distribution of the inner planet, $P_{\text{break}}, a_P, b_P$; the period ratio distribution, D and σ; and the fraction of isotropic systems, f_{iso}. The summary statistics to evaluate the model given the observations are the frequency of multi-planet systems, $\{N_k\}$; the orbital period distribution, $\{P\}$; the planet orbital period ratio distribution, $\{\mathcal{P}\}$; the distribution of the innermost planet in the system, $\{P_{\text{in}}\}$. We use Fisher's method to combine the probabilities from the summary statistic
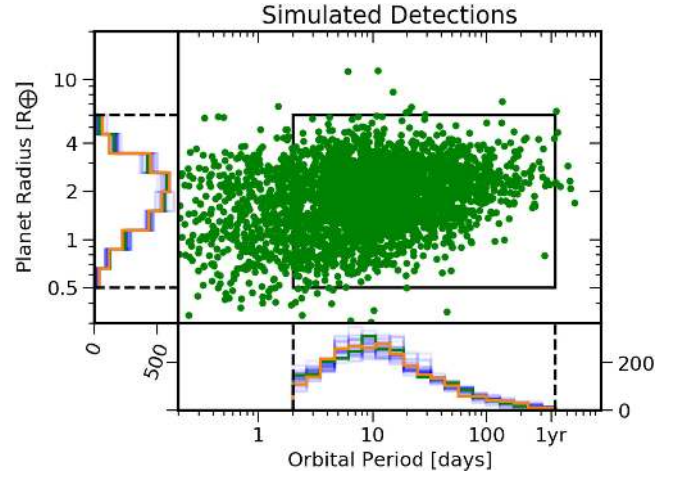
Figure 14. Period–radius distribution of the detected planets in occurrence rate mode. The green population shows the population generated from the best-fit parameters. The black box and dashed lines indicate the range of orbital periods and planet sizes that is included in the observational comparison. The side histograms show the marginalized simulated distributions compared with the observed populations in orange. The blue lines show 30 samples from the posterior.

into a single parameter:

$$\begin{aligned} L_{\text{multi}} = -2(\ln(p_N) + \ln(p_{N,k}) \\ + \ln(p_P) + \ln(p_P) + \ln(p_{\text{in}})), \end{aligned} \qquad (20)$$
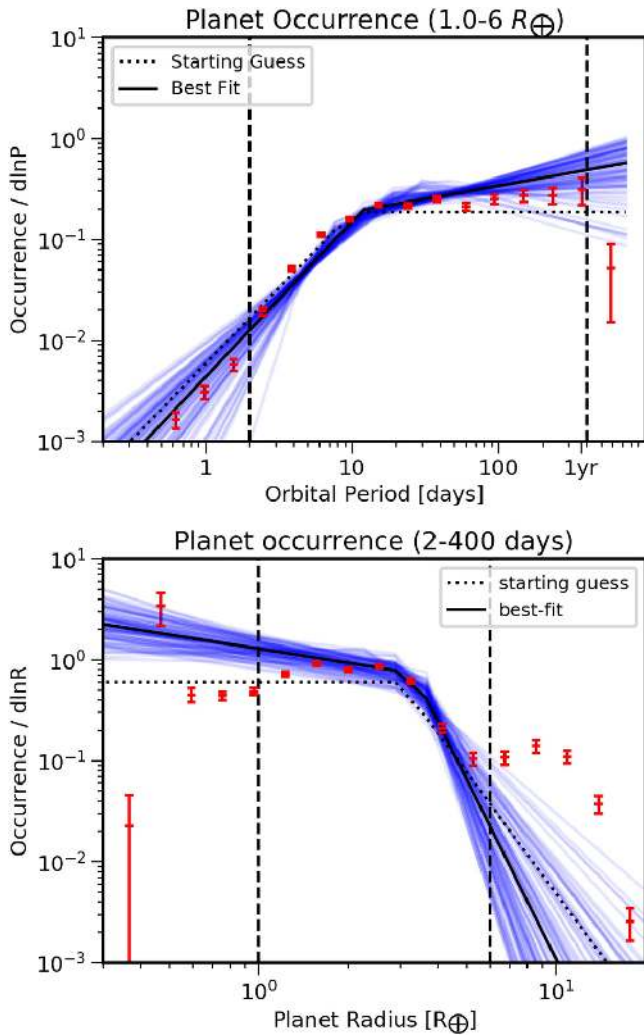
which we use as the likelihood of the model given the data in `emcee`, e.g., $\mathcal{L} \propto \mathcal{L}(N_k, \{P\}, \{\mathcal{P}\}, \{P_{\text{in}}\}|\eta, P_{\text{break}}, a_P, b_P, D, \sigma, \Delta i, f_{\text{iso}})$.

The planet radius distribution is fixed to minimize the amount of free parameters, and we use the best-fit values constrained in the previous section and listed in Table 2. The number of planets per system, m, is fixed to 10 because it is not well constrained in the fitting: Systems with fewer than six planets do not reproduce the observed multiplicity distribution as there are two detected sextuple systems ($N_6 = 2$). Systems with more than seven planets have the same observation signature as the 8th, 9th, etc., planet in the system and will typically remain undetected, as the combined likelihood of detecting all the planets in such systems is too small to allow detection given the size of the *Kepler* sample.

We sample the parameter space with `emcee` using 100 walkers for 2000 iterations, allowing for a 500-step burn-in. The MCMC chain and parameter covariances are shown in the appendix (Figure 24). The simulated observables of the best-fit models and of 30 samples from the posterior are shown in Figure 17. We will discuss the results in the next section.

3. Results

By using `EPOS` in parametric mode we find that the planet occurrence rates of the *Kepler* mission are well described by a broken power law in the orbital period and planet radius in the region $2 < P < 400$ days and $0.5 < R_{\oplus} < 6$. We estimate the planet occurrence rate, the average number of planets per star, to be $\eta = 2.4^{+0.5}_{-0.5}$ in this regime and $\eta = 4.9^{+1.3}_{-1.2}$ in the simulated range ($0.4 < P < 730$ days and $0.3 < R_{\oplus} < 20$). This number is higher than the occurrence rate calculated from inverse detection efficiencies ($\eta_{\text{inv}} = 1.40 \pm 0.03$), which underestimate the occurrence rates for small planets at long orbital periods. These results are largely consistent with
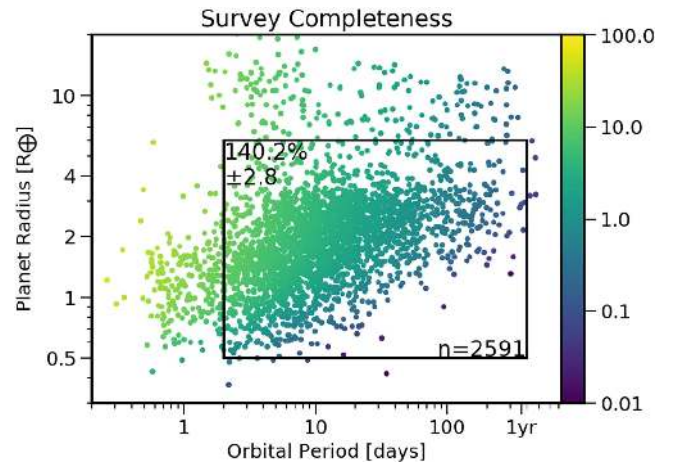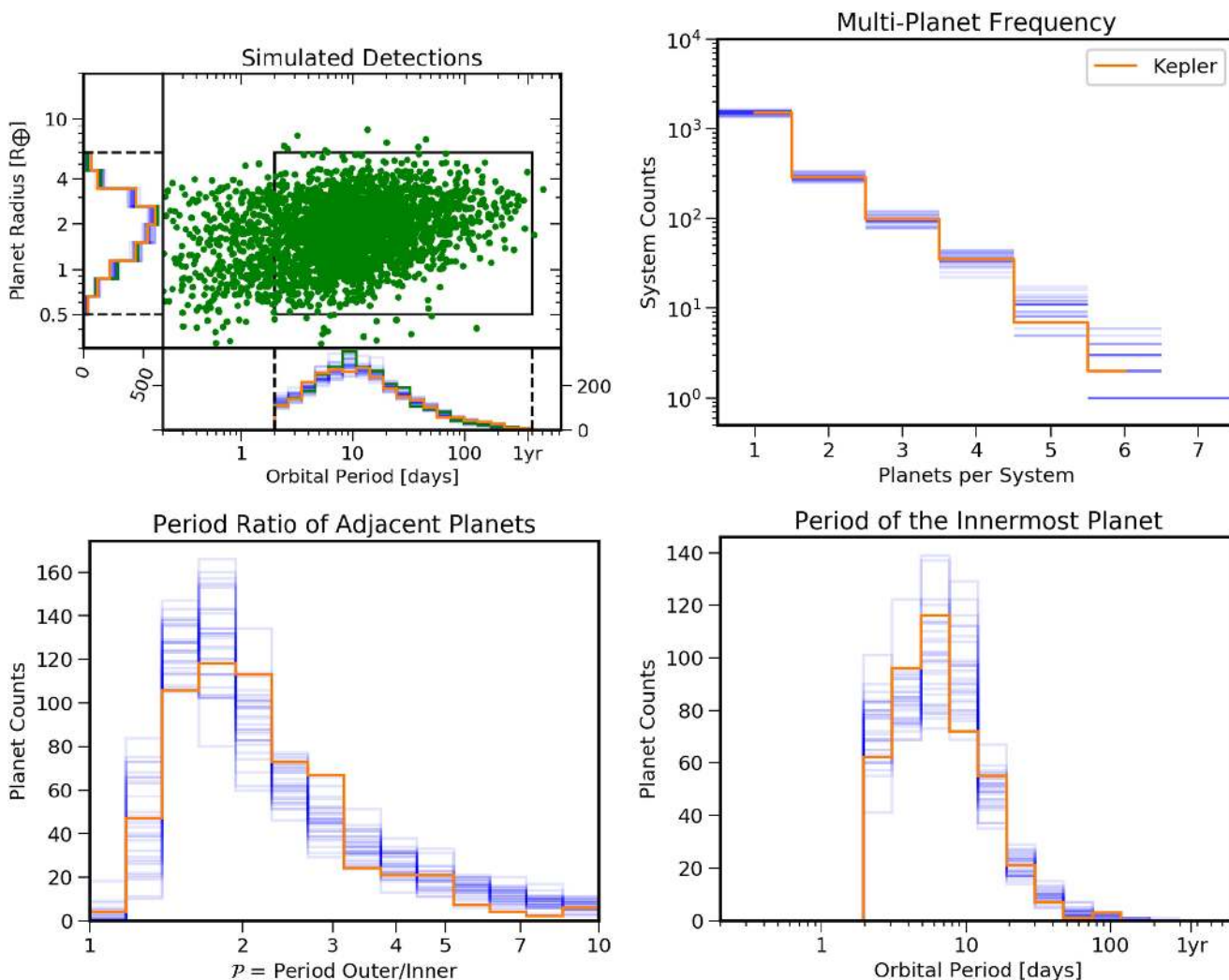
Figure 16. *Kepler* DR25 candidate list, color coded by survey completeness. The sample includes only dwarf stars (log $g < 4.2$) and planet candidates with a disposition score larger than 0.9. The planet occurrence rate estimated from Equation (19) is $\eta_{\rm obs} = 1.40 \pm 0.03$.

Table 2
Fit Parameters for the Example Model in Multi-planet Mode and the Best-fit Solutions with 1σ Confidence Intervals

Parameter	Example	Best-fit
η_s	0.4	$0.67^{+0.17}_{-0.12}$
$P_{\rm in}$ (days)	10	12^{+3}_{-2}
a_P	1.5	$1.6^{+0.4}_{-0.2}$
b_P	-1	$-0.9^{+0.4}_{-0.5}$
$R_{\rm break}$ ($R_\oplus$)	3.0	3.3
a_R	0.0	-0.5
b_R	-4	-6
$\log D$	0.4	$-0.39^{+0.07}_{-0.05}$
σ	0.2	$0.18^{+0.05}_{-0.04}$
m	10	10
Δi (°)	2	$2.1^{+1.0}_{-0.8}$
$f_{\rm iso}$	0.4	$0.38^{+0.08}_{-0.08}$

Note. Parameters $R_{\rm break}$, a_R, and b_R were fixed to their best-fit solutions from the occurrence rate mode.

Figure 15. Posterior orbital period distribution (top) and planet radius distribution (bottom). The red bars show the occurrence rates estimated using the inverse detection efficiencies for comparison. Note that the occurrence rates underestimate the true distribution in bins that include regions where *Kepler* has not detected any planet candidates, in particular $R < 1.5\,R_\oplus$ and $P > 50$ days (see Figure 16).

previous occurrence rate studies collected in the SAG13[10] literature study on planet occurrence rates (see Kopparapu et al. 2018 for details).

The observed population of exoplanets is well described by a population of regularly spaced planetary systems with seven or more planets that orbit $\eta_s = 67^{+17}_{-12}\%$ of stars. We find that the mutual inclinations are consistent with the *Kepler* dichotomy: Half the planet population is in planetary systems that have nearly coplanar orbits, here described by a Rayleigh distribution with $\Delta i = 2.1^{+1}_{-0.8}{}^\circ$, consistent with previous estimates that find ranges between 1° and 3° (Lissauer et al. 2011; Fang & Margot 2012; Fabrycky et al. 2014; Ballard & Johnson 2016). The other half of planets appear as single-planet transiting systems, which we model as multi-planet systems with isotropically distributed orbital inclinations. The typical orbital period ratio between adjacent planets is $\mathcal{P} = 1.8$ but with a wide distribution, consistent with previous analysis of spacings between planets (Fang & Margot 2013).

We discover that the inner edge of planetary systems are clustered around an orbital period of 10 days (Figure 18), reminiscent of the protoplanetary disk inner edge which is located at $\sim$0.1 au for pre-main-sequence Sun-like stars (Millan-Gabet et al. 2007). While the break in planet occurrence rate around 10 days has been previously connected to the disk inner edge (Mulders et al. 2015a; Lee & Chiang 2017), this is the first time a peak in the occurrence rate distribution of sub-Neptunes has been identified. The posterior distribution of the innermost planet peaks at an orbital period of 12 ± 2 days and decays toward shorter and longer orbital periods with a power law index of $a_P = 1.6^{+0.4}_{-0.2}$ and $b_P = -0.9^{+0.4}_{-0.5}$, respectively. The decay toward long orbital periods is surprising, since planet occurrence rates are constant or slightly increasing in this range (green line). Planets exterior to the break are therefore mostly the 2nd, 3rd, etc., planet in the system. We also show that systems with inner planets at orbital periods of 100 days are intrinsically rare, and we discuss the

---
[10] https://exoplanets.nasa.gov/system/internal_resources/details/original/680_SAG13_closeout_8.3.17.pdf

Figure 17. Posterior predictive plots of simulated planetary systems. The 30 samples from the posterior are shown in blue and the observed systems are shown in orange. The top left panel shows the best-fit period–radius distribution in green (see Figure 14). The black box and dashed lines indicate the range of orbital periods and planet sizes that is included in the observational comparison. The top right panel shows the frequency distribution of multi-planet systems with k planets. The bottom left panel shows the period ratio distribution of the adjacent planets. The bottom right panel shows the orbital period distribution of the innermost planet in each multi-planet system.
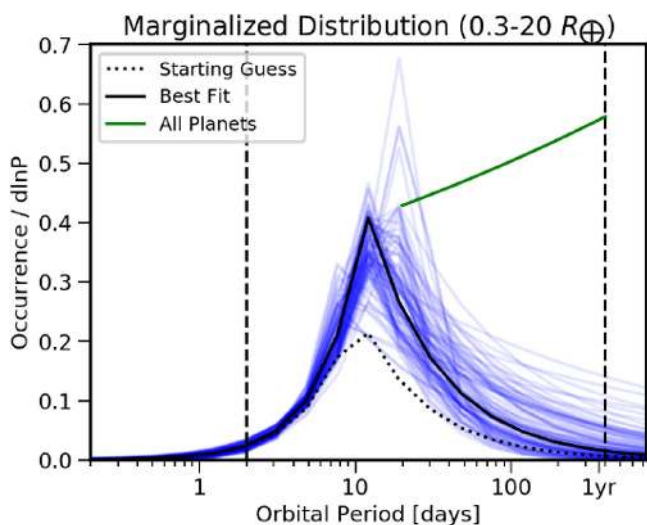
Figure 18. Marginalized orbital period distribution of the innermost planet in the system. The green line indicates the distribution of all the planets in the system.

implications for planet formation theories and the origins of the solar system below.

4. Discussion

4.1. How Rare is the Solar System?

Our modeling analysis indicates that multi-planet systems without planets interior to $P_{in} \sim 100$ days are intrinsically rare. We estimate, based on parametric distributions of planet parameters, that $8^{+10}_{-5}\%$ of planetary systems have no planet interior to the orbit of Mercury and $3^{+5}_{-2}\%$ have no planet interior to Venus. This implies that the solar system may simply be in the tail of the distribution of exoplanet systems.

On the other hand, this comparison with the solar system is made under the assumption that planet orbital architectures (inclinations, spacings, inner planet location) are independent of radius and orbital period. This assumption will need to be verified with additional data. The orbital architectures of the *Kepler* planetary systems are most constrained by
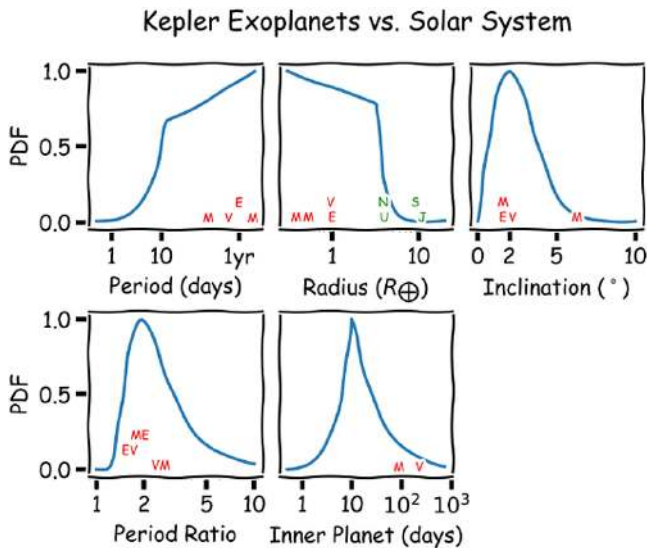
Figure 19. Best-fit distribution of planetary system properties from *Kepler* (blue) compared to the solar system terrestrial planets (red letters). The inclinations of the terrestrial planets are with respect to the invariable plane. The orbital periods, radii, inclinations, and period ratios of the terrestrial planets lie near the peak of the distribution of *Kepler* systems. The only notable exception is the orbital period of the innermost planet, where Mercury (93th percentile) and Venus (97th percentile) lie in the tail of the distribution.

planets that are larger and closer in, typically $P \sim 10$ days and $R \sim 2\,R_\oplus$ (see Figure 14). A solar system analog, if detected, would most likely appear as single transiting systems in the *Kepler* data due to the low probability that multiple terrestrial planets transit (e.g., Brakensiek & Ragozzine 2016). Hence, we can not rule out that the solar system may be part of a population of planetary systems with different orbital characteristics than those detected by the *Kepler* mission.

We make a direct estimate of how many planetary systems without planets interior to Mercury or Venus could be present in the *Kepler* data, without extrapolating the distribution of orbital periods from closer-in planets. We generate a population of planetary systems with similar orbital properties as the solar system where we place the innermost planet at the orbital period of Mercury (or Venus) but otherwise keep the same parameters as in the best-fit model. Such a simulated planet population with $\eta_s = 67\%$ and $f_{\mathrm{iso}} = 0$ matches planet occurrence rates exterior to $P = 50$ days for $P_{\mathrm{in}} = 88$ days. The simulated observations predict 30 multi-planet systems with an innermost planet exterior to $P = 50$ days, while only 4 are observed. This indicates that if a population of planetary systems without planets interior to Mercury existed it would be detectable in the *Kepler* data. We do note that most detectable planets in this range are larger than one Earth radius, so the data do not directly constrain true solar system analogs. Using a mixture of coplanar systems and highly inclined systems (which appear as intrinsically single), we can rule out that more than 10% of systems ($f_{\mathrm{iso}} = 0.9$) have $P_{\mathrm{in}} = 88$ days and $\Delta i = 2$.

We repeat this exercise for systems where the innermost planet shares the orbital period of Venus ($P = 225$ days). None of these simulated systems would be detectable as multi-planet systems, and no planetary systems with an orbital period larger than 150 days are detected with *Kepler*. Hence, we can not rule out that the *Kepler* exoplanet population contains a significant

population of multi-planet systems without planets interior to Venus, though we do not find evidence that such a population exists.

Overall, the solar system seems to be a typical planetary system in most diagnostics of orbital architectures (planetary radii, periods, period ratios, inclinations, and planets per system, Figure 19). The only clear difference is in the period of the innermost planet where the solar system is an outlier. While the lack of super-Earths/mini-Neptunes in the solar system is also notable (Martin & Livio 2015), Earth-sized planets are not intrinsically rare, and the size of the terrestrial planets does not make the solar system an outlier in the exoplanet distribution.

4.2. Systems with Habitable Zone Planets

We also estimate $\eta_\oplus$, the number of Earth-sized planets with Earth-like orbital periods ("habitable zone" planets), here defined as $0.9\,P_\oplus < P < 2.2\,P_\oplus$ and $0.7\,R_\oplus < R < 1.5\,R_\oplus$ based on the Kopparapu et al. (2013) conservative habitable zone for a Sun-like star that is representative of the most common star in the *Kepler* sample. By integrating the posterior distribution over the radius and orbital period range we find $\eta_\oplus = 36^{+14}_{-14}\%$ or $\Gamma_\oplus = \frac{\eta_\oplus}{d\ln P \; d\ln R} = 53^{+20}_{-21}\%$. These results are consistent with the estimate of $\Gamma_\oplus = 60\%$ for GK dwarfs by Burke et al. (2015), though we note that there is a large dispersion in the literature[11] on habitable zone planet occurrence rates based on the adopted completeness correction, the method of extrapolation into the habitable zone, and the planet sample selection. The confidence intervals on $\eta_\oplus$ include counting statistics and systematic uncertainties in extrapolating the planet radius and orbital period distribution out to the habitable zone. They do not include systematic uncertainties on the adopted stellar parameters, in particular the stellar radii which directly impacts the planet radii. For example, large uncertainties in the stellar radius of unresolved stellar binaries lead to over-estimating the occurrence of small planets (Ciardi et al. 2015; Silburt et al. 2015). Better observational constraints on the stellar properties are expected based on parallax measurements from the ESO *Gaia* mission. As a consistency check, we have repeated all of the calculations in this paper with the improved stellar radii and giant star classification from (Berger et al. 2018). The best-fit parameteric distributions are consistent with those in Table 1 within errors, and the habitable zone planet occurrence rate of $\eta_\oplus = 40^{+14}_{-14}\%$ is not significantly different.

The habitable zone planet occurrence rate we estimate is consistent with that of Burke et al. (2015) based on the Q1–Q16 catalog, but about a factor four higher than that of the earliest estimate based on *Kepler* data from Petigura et al. (2013a), which we attribute to an improved understanding of detection efficiency of *Kepler*. This rate is comparable to that of M dwarfs estimated with inverse detection efficiency methods (Dressing & Charbonneau 2015), though we leave a comparison between M dwarfs and FGK stars with consistent methodology for a future paper.

The majority of simulated planets at long orbital periods ($P \sim 100$ days) are not intrinsically single, but are part of multi-planet systems where the innermost planets are typically at an orbital period of $P \sim 10$ days. While the

---

[11] https://exoplanets.nasa.gov/system/internal_resources/details/original/680_SAG13_closeout_8.3.17.pdf

Mulders et al.
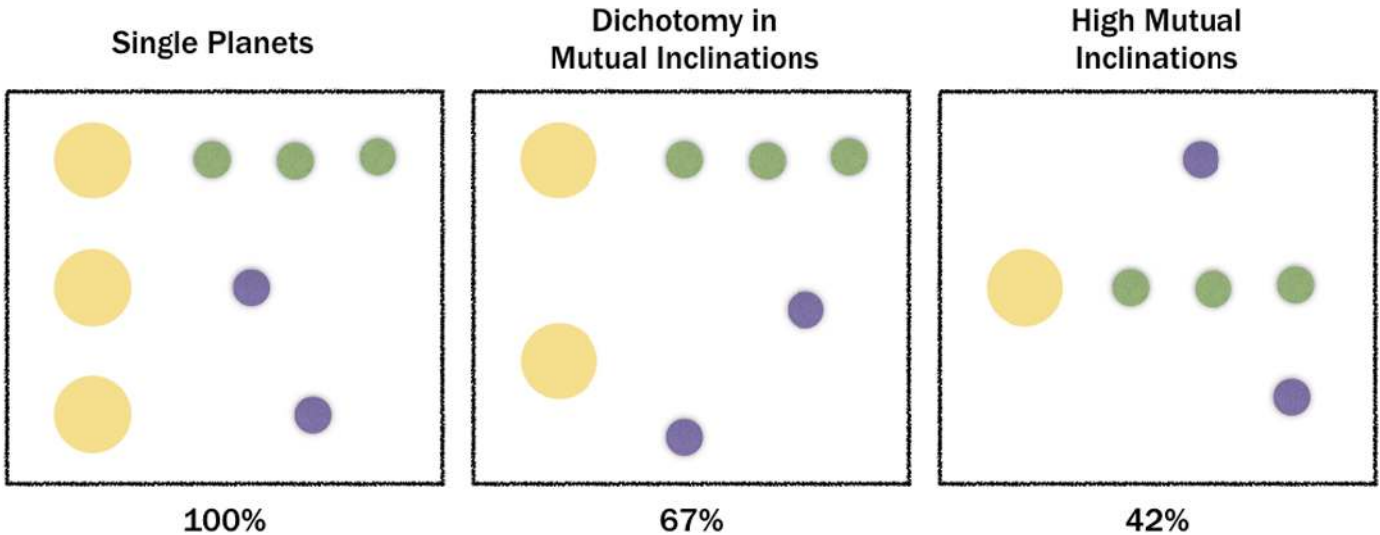
Figure 20. Illustration of how the estimated fraction of stars with planetary systems depends on the distribution of planets across stars. Nearly all coplanar systems orbit 31% of stars (green) while the distribution of observed single transiting planets (purple) is less well constrained.

multi-planet statistics do not directly constrain planetary systems of Earth-sized planets in the habitable zone of Sun-like stars, the presence of planets on orbital periods of 10 days does provide a way to identify targets for future direct imaging missions, such as the HabEx and LUVOIR mission concepts.

Kipping & Lam (2017) have shown that the presence of transiting planets at short orbital periods increases the chance of finding transiting planets at larger orbital periods. Our analysis indicates that, because most planets in multi-planet systems are not transiting, the probability of finding any planet (transiting or non-transiting) is even higher. For example, the detection of an Earth-mass planet at an orbital period of ten days with radial velocity measurements indicates a higher probability of finding a planet in the habitable zone. Depending on whether single transiting planets are intrinsically single or highly inclined multiples, 50%–100% of systems with close-in planets may also have planets in the habitable zone.

4.3. Protoplanetary Disk Inner Edges

The peak in the location of the innermost planet in the system points to a preferred location of planet formation/ migration in protoplanetary disks. This location may reflect either the inner edge of the region where planets form (Chiang & Laughlin 2013; Mulders et al. 2015a; Lee & Chiang 2017) or that of a planet trap where planet migration stalls (Terquem & Papaloizou 2007). While the break in planet occurrence at ∼10 days has been attributed to the disk inner edge before (Mulders et al. 2015a; Lee & Chiang 2017), the peak in the location of the innermost planet presents solid evidence that there is a preferred location for planet formation at 0.1 au, rather than a continuum of location between 0.1 and 1 au. In particular, this result is inconsistent with wide regions between 0.1 and 1 au acting as a trap for individual planets (e.g., Dittkrist et al. 2014), and is more reminiscent of the inward migration of multiple planets where the first planet is trapped at the inner disk edge and

halts the migration of other planets (Cossou et al. 2014; Coleman & Nelson 2016; Izidoro et al. 2017).

4.4. Do Most Stars Have Planets?

Our analysis confirms that the *average* number of planets per (Sun-like) star is larger than one (Section 2.1.1). However, this does not imply that most stars have planets, since planets can be unevenly distributed among stars. In Section 2.1.1 we show that $\eta_s = 67^{+17}_{-12}\%$ of stars can host multi-planet systems. An important assumption we made is that the apparent excess of single transits is because $f_{\rm iso} = 38\%$ of multi-planet systems have high mutual inclinations. The inclination distribution of exoplanets can not be uniquely constrained from transit survey data (Tremaine & Dong 2012). Different assumptions on the mutual inclination distribution lead to different estimates of the fraction of stars with planets, illustrated in Figure 20.

On the conservative end, the highly inclined planets in otherwise coplanar multi-planet systems could contribute to the observed number of single transits (Figure 20, right panel). In this case, the dichotomy is an artifact of the assumption that mutual inclinations follow a Rayleigh distribution. A distribution with a larger tail toward higher inclinations could possibly fit the multi-planet statistics with a single population. In this case, the fraction of stars with planets drops to ∼41%. Alternatively, a population of intrinsically single planets could contribute to the observed single transiting systems (Figure 20, right panel). In this case, the estimated fraction of stars with planets increases to unity.

A way of discriminating between both scenarios may be the use of stellar obliquity. Transiting planets with high mutual inclinations or isotropic distributions have large obliquities with respect to the line of sight. There are indications that the stellar obliquity is larger for single-transit systems (Morton & Winn 2014), indicating that they may have highly inclined orbits. Several mechanisms have been proposed to disrupt initially coplanar systems, reducing their multiplicity and/or increasing their mutual inclinations, and giving rise to the population of single transiting systems. These mechanisms

include dynamical instabilities within systems (Pu & Wu 2015; Volk & Gladman 2015), external perturbations from giant planets or stars (Johansen et al. 2012; Hansen 2017; Izidoro et al. 2017; Lai & Pu 2017; Cai et al. 2018) and host star misalignment (Spalding & Batygin 2016). Alternatively, planets may form as intrinsically single systems in half the cases (Moriarty & Ballard 2016). Additional constraints on the single planet population are needed to constrain these scenarios. However, we stress that roughly half of the observed single transiting systems are part of multi-planet systems and have non-transiting or non-detected planets in the system, and hence the sample of "singles" is diluted with multi-planets, weakening any potential trends. Transit timing variations can also be used to detect additional non-transiting planets in the system (Nesvorný et al. 2012).

The *Kepler* mission only detects planets out to ∼1 au and down to ∼0.5 $R_\oplus$. Hence, any estimate of the fraction of stars with planets is likely to be a lower limit. We do not see a clear indication that planet occurrence rate decreases toward the detection limits of *Kepler*. While planet occurrence rates calculated using the inverse detection efficiency seem to decrease for Earth-size and smaller planets (Figure 15), this is also the region where the *Kepler* exoplanet list becomes incomplete. The posterior planet size distribution does not show this trend and is almost flat in log R_p, indicating that planets below the detection efficiency of *Kepler* could be extremely common. Since planetary systems tend to have planets of similar size (Millholland et al. 2017; Weiss et al. 2018), planetary systems with only planets smaller than the detection limit would be missed. Extending the planet size distribution down to the size of Ceres (0.07 $R_\oplus$) would increase the fraction of stars with planets to 100%.

We find that the orbital period distribution of sub-Neptunes is nearly flat in the logarithm of orbital period. However, planets at long orbital periods are mostly members of systems with planets also at shorter orbital periods: there is no evidence for a large population of planetary systems with only long-period planets. Extrapolating the distribution of planetary systems to orbital periods larger than a year would add planets to existing systems, increasing the number of planets per star but not the fraction of stars with planets. Of course, such an extrapolation does not take into account the presence of additional populations of planets, which could be important, for example, if planets form more frequently at the snow line (exterior to 1 au). Giant planets orbit 10%–20% of Sun-like stars (Cumming et al. 2008), most of them beyond 1 au. Microlensing surveys hint at the existence of a population of Neptune-mass planets around a significant fraction of M dwarfs (Cassan et al. 2012; Suzuki et al. 2016). Since these populations remain mostly undetected in the *Kepler* survey, it is not clear if they belong to the same stars, though there are some indications that they do (Huang et al. 2016).

5. Summary

We present the Exoplanet Population Observation Simulator, EPOS, a Python code to constrain the properties of exoplanet populations from biased survey data. We showcase how planet occurrence rates and orbital architectures can be constrained from the latest *Kepler* data release, DR25. We find that:

1. The *Kepler* exoplanet population between orbital periods of 2–400 days and planet radii of 0.5–6 $R_\oplus$ are well described by broken power-laws. The planet occurrence rate in this regime is $\eta = 2.4 \pm 0.5$ consistent with previous works. The estimated planet occurrence rate in the habitable zone is $\eta_\oplus = 36 \pm 14\%$.

2. The observed multi-planet frequencies are consistent with ensemble populations that are a mix of nearly coplanar systems and a population of planets whose orbital architectures are unconstrained, consistent with the previously reported *Kepler* dichotomy. $62 \pm 8\%$ of exoplanets are in systems with six or more planets, orbiting 42% of Sun-like stars. The remaining $38 \pm 8\%$ of exoplanets could be intrinsically single planets or be part of multi-planet systems with high mutual inclinations, raising the fraction of stars with planets to somewhere between 45% and 100%.

3. The mutual inclinations of planetary orbits can be described by a Rayleigh distribution with a mode of $i = 2° \pm 1°$. The spacing between adjacent planets follows a wide distribution with a peak at an orbital period ratio $\mathcal{P} = 1.8$, though detection biases shift the peak of the observed distribution to shorter orbital period ratios.

4. The distribution of the innermost planet in the system peaks at an orbital period of $P \approx 10$ days. Planetary systems without planets interior to Mercury's and Venus' orbit are rare at 8% and 3%, respectively.

The EPOS code presented in this paper provides a first step in a larger effort to constrain planetary orbital architectures from exoplanet populations. The next step will be to use more complex models of planetary system architectures based on planet formation models. Future directions include incorporating additional exoplanet survey data from radial velocity, microlensing, and direct imaging, as well as from transit surveys such as K2 and TESS.

Appendix A
Table of Mathematical Symbols

All mathematical symbols used in this paper are summarized in Table 3.

Table 3
List of Mathematical Symbols used in this Paper

Parameter	Unit	Description
R	$R_\oplus$	Planet radius
P	day	Planet orbital period
$\mathcal{P}$		Orbital period ratio of adjacent planets
i	°	Orbital inclination
δi	°	Planet mutual inclination
Δi	°	Mode of mutual inclination distribution
Ω	rad	Longitude of ascending node
f		Planet probability density function
g		Planetary system probability density function
χ		Continuous random variable
f_{det}		Survey completeness
f_{geo}		Geometric transit probability
$f_{\mathrm{S/N}}$		Detection efficiency
f_{vet}		Vetting efficiency
f_{iso}		Fraction of isotropic systems
η		Average number of planets per star
η_s		Fraction of stars with planets
m		Planets per system (multiplicity)
ID		Host star identifier
$\{\}$		Ensemble of parameters
n		Sample size
N		Number of detected planets
s		Subscript for planetary systems
p		Subscript for planets
t		Subscript for transiting planets
d		Subscript for detectable planets
in		Subscript for innermost planet
k		Index for kth planet in system
a, b		Power law indices
c		Normalization factor
D, σ		Dimensionless spacing parameters

Appendix B
Vetting Completeness

The *Kepler* Robovetter (Coughlin et al. 2016; Coughlin 2017) was designed to discriminate between real transiting events and instrumental or astrophysical signals that mimic that of a transiting planet, a process called "vetting." For planets with few transits and a low signal-to-noise ratio, vetting is a trade-off between detecting planets (completeness) or removing false positives (reliability). A planet candidate sample with high completeness has a low reliability, and vice versa. The Robovetter was calibrated by evaluating the ability to identify the signals of a number of simulated transiting events to better quantify the efficiency and confidence with which planets can be detected.

Each planet candidate is assigned a disposition score between 0 and 1, which can be used to create a more reliable sample (fewer false positives classified as planet candidates) at the the cost of having a less complete sample (more planets classified as false positives), see Thompson et al. (2018) for details. For EPOS we opt to use a planet sample with high reliability by using a disposition score cut of 0.9, reducing the sample completeness which, however, can be quantified and corrected for as we discuss below. The score cut of 0.9 was chosen to eliminate the peak of false positives in the

habitable zone that coincides with the orbital period of the spacecraft.

An empirical procedure for estimating the vetting completeness is described in Thompson et al. (2018), based on the injection of simulated transits into the *Kepler* data (Christiansen 2017) and analyzed by the Robovetter (Coughlin 2017). We downloaded the Robovetter results for simulated planet injections (group 1) from the *Kepler* simulated data page.[12] We removed all injections not on main-sequence stars according to the prescription in Huber et al. (2016). We only considered injections inside the orbital period and radius range where we use EPOS ($P = [0.5, 730]$ days and $R = [0.3, 20]R_\oplus$). There are 44981 injections, of which 77% (34558) are classified as planet candidates and 43% (19247) have a disposition score >0.9.

The top left panel of Figure 21 shows the vetting completeness, the fraction of injections classified as planet candidates with a score larger than 0.9, binned to the same grid as the survey detection efficiency. Planet injections are not equally distributed across this parameter range, with the majority of injections around sub-Neptunes at long orbital periods, as shown in the top right panel. The vetting completeness is $\sim$75% at orbital periods less than roughly 50 days and decreases below 25% at the longest orbital periods. The completeness also drops below 25% for giant planets, though this is most likely an artifact of those planets only being injected around the most noisy stars, see Thompson et al. (2018), and does not reflect a real decrease in the detectability of giant planets in the *Kepler* pipeline.

We parametrize the vetting completeness to obtain a smooth distribution of radius and orbital period that covers all of the regions we are simulating with EPOS. We use a functional form for the vetting completenss, f_{vet}, that is a double broken power law:

$$f_{\mathrm{vet}} = c\, f_{\mathrm{vet},P}(P)\, f_{\mathrm{vet},R}(R) \qquad (21)$$

with

$$f_{\mathrm{vet},P}(P) = \begin{cases} (P/P_{\mathrm{break}})^{a_P}, & \text{if } P < P_{\mathrm{break}} \\ (P/P_{\mathrm{break}})^{b_P}, & \text{otherwise} \end{cases} \qquad (22)$$

and

$$f_{\mathrm{vet},P}(R) = \begin{cases} (R/R_{\mathrm{break}})^{a_R}, & \text{if } R < R_{\mathrm{break}} \\ (R/R_{\mathrm{break}})^{b_R}, & \text{otherwise.} \end{cases} \qquad (23)$$

We find the best-fit parameters using scipy.optimize. curve_fit, where we minimize the difference between the measured and parametrized vetting completeness over all grid cells. The best-fit solution is shown in the bottom left panel of Figure 21 and the bottom right panel shows the residuals per grid cell. The best-fit parameters are $c = 0.88 \pm 0.08$, $P_{\mathrm{break}} = 53 \pm 8$ days, $a_P = -0.07 \pm 0.03$, $b_P = -0.39 \pm 0.02$, $R_{\mathrm{break}} = 5.7 \pm 2.1\,R_\oplus$, $a_R = 0.19 \pm 0.03$, and $b_R = -2.7 \pm 6.2$. The break in the power law with orbital period corresponds to the decreased detection efficiency at long orbital periods as documented in Thompson et al. (2018), Coughlin (2017). The break in the power law of planet radius corresponds to a decreased detection efficiency for giant planets. However, as mentioned above, this is likely an artifact of the way the injection tests were defined, and there is no reason to assume that the low

---

[12] https://exoplanetarchive.ipac.caltech.edu/docs/KeplerSimulated.html
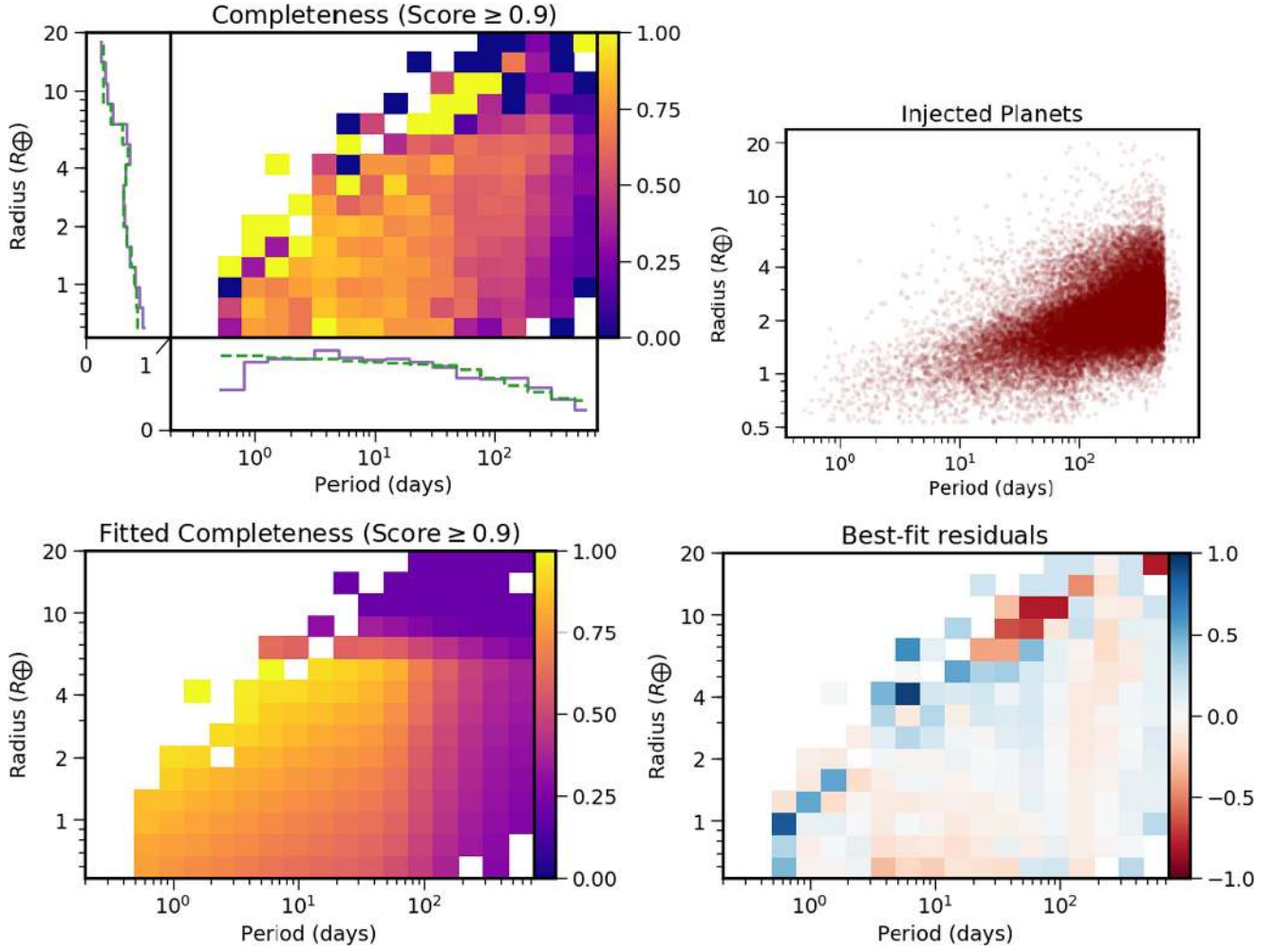
Figure 21. The top left panel shows the vetting efficiency per grid cell for a Robovetter score larger than 0.9. The side panels show marginalized detection efficiency (solid purple) compared to the best-fit double broken power law model (green dashed). The top right panel shows the locations of the planet injections from Christiansen (2017). The bottom left panel shows the best-fit detection efficiency evaluated per grid cell. The bottom right panel shows the difference between the measured and fitted vetting completeness per grid cell.

vetting efficiency for giant planets is a real feature (Thompson et al. 2018). Therefore, we do not include the power law index radius break in EPOS. Instead, we assume the planet radius dependence of the detection is described by the power law in planet radius that was fitted to the small planets. This yields a vetting efficiency of

$$f_{\text{vet}}(R, P) = 0.63\, R^{0.19} \begin{cases} (P/P_{\text{break}})^{-0.07}, & \text{if } P < P_{\text{break}} \\ (P/P_{\text{break}})^{-0.39}, & \text{otherwise} \end{cases} \quad (24)$$

with $P_{\text{break}} = 53$, which is shown in Figure 5.

Appendix C
Posterior Distributions

The MCMC chain and corner plots for the occurrence rate mode are shown in Figures 22 and 23. The corner plots for the multi-planet mode are shown in Figure 24.
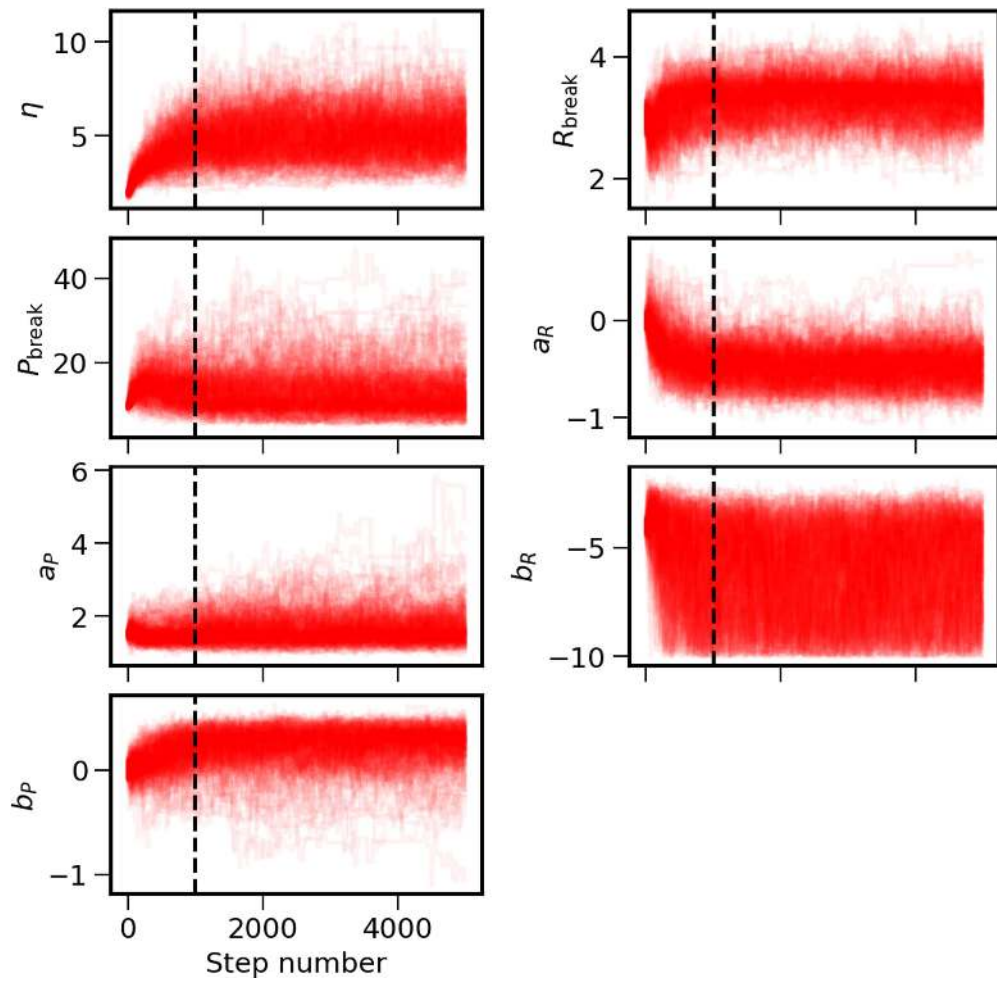
Figure 22. Position in seven-dimensional parameter space of 200 walkers in the MCMC chain. The dashed line indicates the end of the burn-in phase.
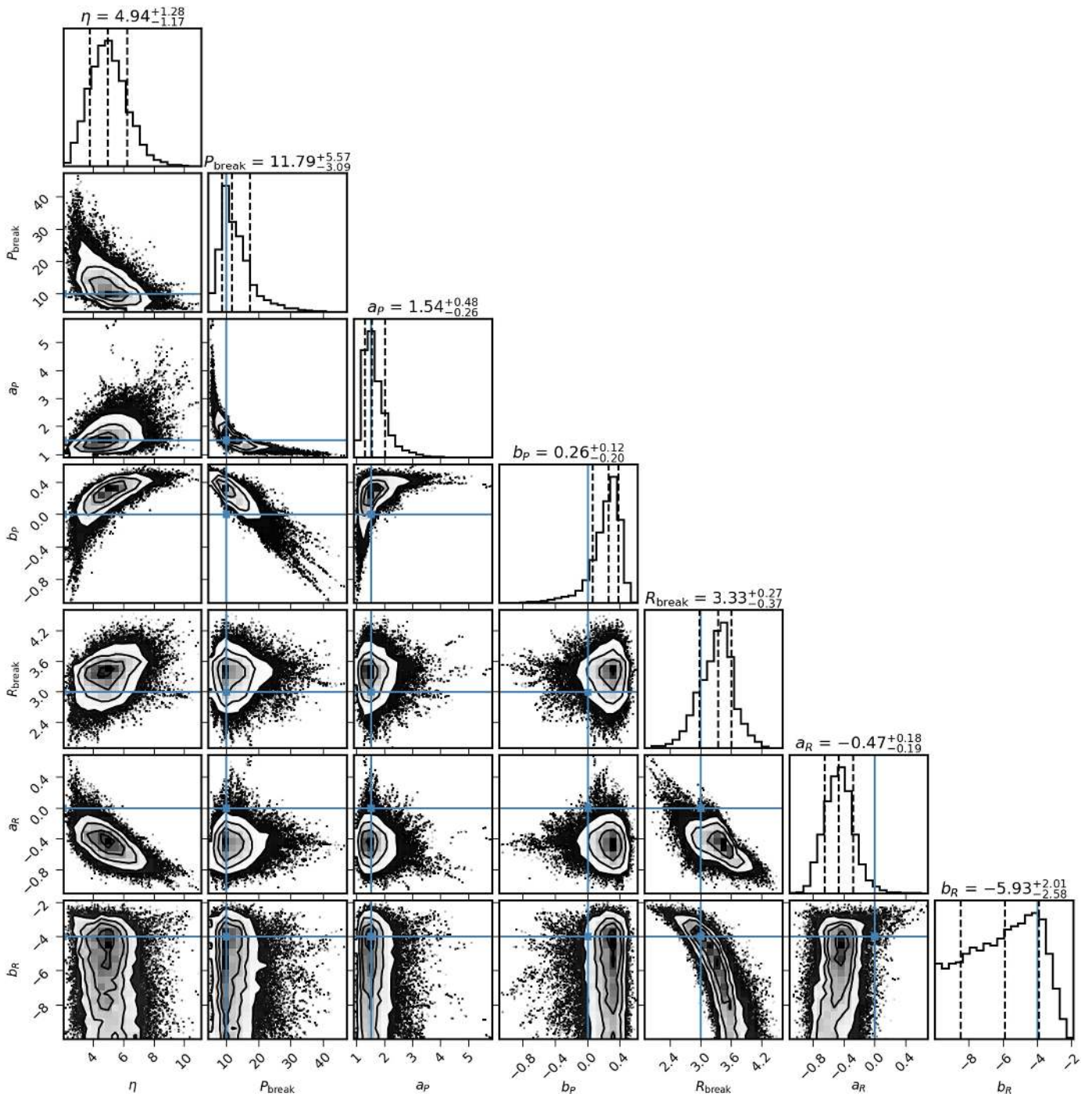
Figure 23. Corner plots for the parametric distribution, generated using an open-source `Python` package `corner`. Blue lines indicate the initial guess that was based on previous studies.
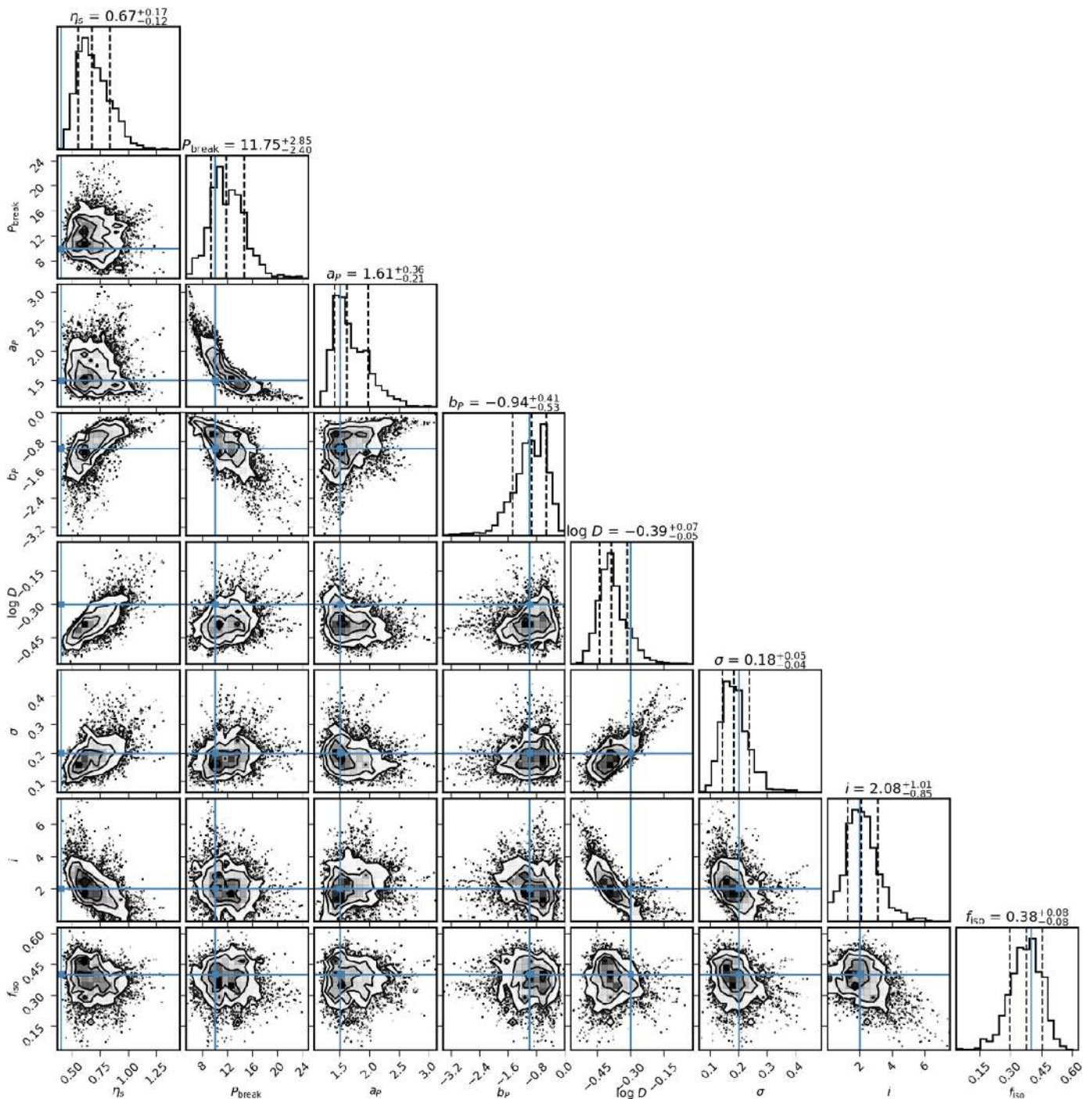
Figure 24. Corner plots for the multi-planet parametric distribution, generated using an open-source `Python` package `corner`. Blue lines indicate the initial guess.

ORCID iDs

Gijs D. Mulders ⬤ https://orcid.org/0000-0002-1078-9493
Ilaria Pascucci ⬤ https://orcid.org/0000-0001-7962-1683
Dániel Apai ⬤ https://orcid.org/0000-0003-3714-5855
Fred J. Ciesla ⬤ https://orcid.org/0000-0002-0093-065X

References

Ballard, S., & Johnson, J. A. 2016, ApJ, 816, 66
Batalha, N. 2014, PNAS, 111, 12647
Berger, T. A., Huber, D., Gaidos, E., & van Saders, J. L. 2018, arXiv:1805.00231
Borucki, W. J. 2017, PAPhS, 161, 38
Brakensiek, J., & Ragozzine, D. 2016, ApJ, 821, 47
Burke, C. J., & Catanzarite, J. 2017, Planet Detection Metrics: Per-Target Detection Contours for Data Release 25 (KSCI-19111-002)
Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, ApJ, 809, 8
Cai, M. X., Portegies Zwart, S., & van Elteren, A. 2018, MNRAS, 474, 5114
Carrera, D., Ford, E. B., Izidoro, A., et al. 2018, arXiv:1804.05069
Cassan, A., Kubas, D., Beaulieu, J. P., et al. 2012, Natur, 481, 167
Chiang, E., & Laughlin, G. 2013, MNRAS, 431, 3444
Christiansen, J. L. 2017, Planet Detection Metrics: Pixel-Level Transit Injection Tests of Pipeline Detection Efficiency for Data Release 25(KSCI-19110-001)
Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2015, ApJ, 810, 95
Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2016, ApJ, 828, 99
Ciardi, D. R., Beichman, C. A., Horch, E. P., & Howell, S. B. 2015, ApJ, 805, 16

Ciardi, D. R., Fabrycky, D. C., Ford, E. B., et al. 2013, ApJ, 763, 41
Coleman, G. A. L., & Nelson, R. P. 2016, MNRAS, 457, 2480
Cossou, C., Raymond, S. N., Hersant, F., & Pierens, A. 2014, A&A, 569, A56
Coughlin, J. L. 2017, Planet Detection Metrics: Robovetter Completeness andEffectiveness for Data Release 25 (KSCI-19114-001)
Coughlin, J. L., Rowe, J. F., Caldwell, D. A., et al. 2016, ApJS, 224, 12
Cumming, A., Butler, R. P., Marcy, G. W., et al. 2008, PASP, 120, 531
Dawson, R. I., Lee, E. J., & Chiang, E. 2016, ApJ, 822, 54
Dittkrist, K. M., Mordasini, C., Klahr, H., Alibert, Y., & Henning, T. 2014, A&A, 567, A121
Dong, S., & Zhu, Z. 2013, ApJ, 778, 53
Dressing, C., & Charbonneau, D. 2013, ApJ, 767, 95
Dressing, C., & Charbonneau, D. 2015, ApJ, 807, 45
Fabrycky, D. C., Lissauer, J. J., Ragozzine, D., et al. 2014, ApJ, 790, 146
Fang, J., & Margot, J.-L. 2012, ApJ, 761, 92
Fang, J., & Margot, J.-L. 2013, ApJ, 767, 115
Fisher, R. 1925, Statistical Methods for Research Workers (Edinburgh: Oliver & Boyd)
Foreman-Mackey, D. 2016, JOSS, 24
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, ApJ, 795, 64
Fressin, F., Torres, G., Charbonneau, D., et al. 2013, ApJ, 766, 81
Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, AJ, 154, 109
Goodman, J., & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65
Hansen, B. M. S. 2017, MNRAS, 467, 1531
Hansen, B. M. S., & Murray, N. 2013, ApJ, 775, 53
Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, ApJS, 201, 15
Huang, C., Wu, Y., & Triaud, A. H. M. J. 2016, ApJ, 825, 98
Huber, D., Bryson, S. T., Haas, M. R., et al. 2016, ApJS, 224, 2
Hunter, J. D. 2007, CSE, 9, 90
Izidoro, A., Ogihara, M., Raymond, S. N., et al. 2017, MNRAS, 470, 1750
Johansen, A., Davies, M. B., Church, R. P., & Holmelin, V. 2012, ApJ, 758, 39
Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python, https://www.scipy.org/
Kipping, D. M., & Lam, C. 2017, MNRAS, 465, 3495
Kopparapu, R. K., Hebrard, E., Belikov, R., et al. 2018, arXiv:1802.09602
Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, ApJ, 765, 131
Lai, D., & Pu, B. 2017, AJ, 153, 42
Lee, E. J., & Chiang, E. 2017, ApJ, 842, 40

Lissauer, J. J., Marcy, G. W., Rowe, J. F., et al. 2012, ApJ, 750, 112
Lissauer, J. J., Ragozzine, D., Fabrycky, D. C., et al. 2011, ApJS, 197, 8
Lundkvist, M. S., Kjeldsen, H., Albrecht, S., et al. 2016, NatCo, 7, 11201
Malhotra, R. 2015, ApJ, 808, 71
Martin, R. G., & Livio, M. 2015, ApJ, 810, 105
Mathur, S., Huber, D., Batalha, N., et al. 2017, ApJS, 229, 30
Millan-Gabet, R., Malbet, F., Akeson, R., et al. 2007, in Protostars and Planets V, ed. B. Reipurth, D. Jewitt, & K. Keils (Tucson, AZ: Univ. Arizona Press), 539
Millholland, S., Wang, S., & Laughlin, G. 2017, ApJL, 849, L33
Moriarty, J., & Ballard, S. 2016, ApJ, 832, 34
Morton, T. D., & Swift, J. 2014, ApJ, 791, 10
Morton, T. D., & Winn, J. N. 2014, ApJ, 796, 47
Mulders, G. D. 2018, EPOS: The Exoplanet Population Observation Simulator, Zenodo, doi:10.5281/zenodo.1247569
Mulders, G. D., Pascucci, I., & Apai, D. 2015a, ApJ, 798, 112
Mulders, G. D., Pascucci, I., & Apai, D. 2015b, ApJ, 814, 130
Mulders, G. D., Pascucci, I., Apai, D., Frasca, A., & Molenda-Zakowicz, J. 2016, AJ, 152, 187
Nesvorný, D., Kipping, D. M., Buchhave, L. A., et al. 2012, Sci, 336, 1133
Pascucci, I., Mulders, G. D., Gould, A., & Fernandes, R. 2018, arXiv:1803.00777
Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013a, PNAS, 110, 19273
Petigura, E. A., Marcy, G. W., & Howard, A. W. 2013b, ApJ, 770, 69
Petigura, E. A., Marcy, G. W., Winn, J. N., et al. 2018, AJ, 155, 89
Pu, B., & Wu, Y. 2015, ApJ, 807, 44
Santerne, A., Moutou, C., Tsantaki, M., et al. 2016, A&A, 587, A64
Silburt, A., Gaidos, E., & Wu, Y. 2015, ApJ, 799, 180
Spalding, C., & Batygin, K. 2016, ApJ, 830, 5
Steffen, J. H., Ragozzine, D., Fabrycky, D. C., et al. 2012, PNAS, 109, 7982
Suzuki, D., Bennett, D. P., Sumi, T., et al. 2016, ApJ, 833, 145
Terquem, C., & Papaloizou, J. 2007, ApJ, 654, 1110
The Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, arXiv:1801.02634
Thompson, S. E., Coughlin, J. L., Hoffman, K., et al. 2018, ApJS, 235, 38
Tremaine, S., & Dong, S. 2012, AJ, 143, 94
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
Van Eylen, V., Agentoft, C., Lundkvist, M. S., et al. 2017, arXiv:1710.05398
Volk, K., & Gladman, B. 2015, ApJL, 806, L26
Weiss, L. M., Marcy, G. W., Petigura, E. A., et al. 2018, AJ, 155, 48
Youdin, A. N. 2011, ApJ, 742, 38