



The Occurrence of Rocky Habitable-zone Planets around Solar-like Stars from Kepler Data

Steve Bryson¹, Michelle Kunimoto², Ravi K. Kopparapu³, Jeffrey L. Coughlin^{1,4}, William J. Borucki¹, David Koch^{1,46}, Victor Silva Aguirre⁵, Christopher Allen⁶, Geert Barentsen⁷, Natalie M. Batalha⁸, Travis Berger⁹, Alan Boss¹⁰, Lars A. Buchhave¹¹, Christopher J. Burke¹², Douglas A. Caldwell^{1,4}, Jennifer R. Campbell¹, Joseph Catanzarite¹³, Hema Chandrasekaran¹⁴, William J. Chaplin^{5,15}, Jessie L. Christiansen¹⁶, Jørgen Christensen-Dalsgaard⁵, David R. Ciardi¹⁶, Bruce D. Clarke^{1,4}, William D. Cochran¹⁷, Jessie L. Dotson¹, Laurance R. Doyle⁴, Eduardo Seperuelo Duarte^{1,18}, Edward W. Dunham¹⁹, Andrea K. Dupree²⁰, Michael Endl¹⁷, James L. Fanson²¹, Eric B. Ford^{22,23,24,25}, Maura Fujieh¹, Thomas N. Gautier III²¹, John C. Geary^{26,47}, Ronald L Gilliland²⁷, Forrest R. Girouard^{1,28,46}, Alan Gould²⁹, Michael R. Haas^{1,47}, Christopher E. Henze¹, Matthew J. Holman²⁰, Andrew W. Howard³⁰, Steve B. Howell¹, Daniel Huber⁹, Roger C. Hunter¹, Jon M. Jenkins¹, Hans Kjeldsen⁵, Jeffery Kolodziejczak³¹, Kipp Larson³², David W. Latham²⁰, Jie Li^{1,4}, Savita Mathur^{33,34}, Søren Meibom²⁰, Chris Middour³⁵, Robert L. Morris^{1,4}, Timothy D. Morton³⁶, Fergal Mullally^{1,4}, Susan E. Mullally²⁷, David Pletcher¹, Andrej Prsa³⁷, Samuel N. Quinn²⁰, Elisa V. Quintana³, Darin Ragozzine³⁸, Solange V. Ramirez³⁹, Dwight T. Sanderfer^{1,40}, Dimitar Sasselov²⁰, Shawn E. Seader⁴¹, Megan Shabram¹, Avi Shporer¹², Jeffrey C. Smith^{1,4}, Jason H. Steffen⁴², Martin Still⁴³, Guillermo Torres²⁰, John Troeltzsch³², Joseph D. Twicken^{1,4}, Akm Kamal Uddin³⁵, Jeffrey E. Van Cleve³², Janice Voss^{1,46}, Lauren M. Weiss⁹, William F. Welsh⁴⁴, Bill Wohler^{1,4}, and Khadeejah A Zamudio^{1,45}

¹ NASA Ames Research Center, Moffett Field, CA 94035, USA; steve.bryson@nasa.gov

² Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada

³ NASA Goddard Space Flight Center, Greenbelt, MD, USA

⁴ SETI Institute, 189 Bernardo Ave, Suite 200, Mountain View, CA 94043, USA

⁵ Stellar Astrophysics Centre (SAC), Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark

⁶ Orbital Sciences Corporation, NASA Ames Research Center, Moffett Field, CA 94035, USA

⁷ Bay Area Environmental Research Institute, 625 2nd Street, Suite 209, Petaluma, CA 94952, USA

⁸ University of California Santa Cruz, Santa Cruz, CA, USA

⁹ Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA

¹⁰ Earth & Planets Laboratory, Carnegie Institution for Science, USA

¹¹ DTU Space, National Space Institute, Technical University of Denmark, Elektrovej 328, DK-2800 Kgs. Lyngby, Denmark

¹² Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

¹³ Cross-Entropy Consulting, USA

¹⁴ Computational Engineering Division, Lawrence Livermore National Laboratory, L-261, Livermore, CA-94550, USA

¹⁵ School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK

¹⁶ Caltech/IPAC NASA Exoplanet Science Institute, Pasadena, CA 91125, USA

¹⁷ McDonald Observatory and Center for Planetary Systems Habitability, The University of Texas at Austin, Austin TX 78712, USA

¹⁸ Instituto Federal de Educação Ciência e Tecnologia do Rio de Janeiro, Nilópolis, RJ, Brasil

¹⁹ Lowell Observatory, 1400 West Mars Hill Road, Flagstaff, AZ 86001, USA

²⁰ Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

²¹ Jet Propulsion Laboratory, California Institute of Technology, USA

²² Department of Astronomy & Astrophysics, 525 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802, USA

²³ Center for Exoplanets & Habitable Worlds, 525 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802, USA

²⁴ Center for Astrostatistics, 525 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802, USA

²⁵ Institute for Computational & Data Sciences, The Pennsylvania State University, University Park, PA 16802, USA

²⁶ Smithsonian Astrophysical Observatory, USA

²⁷ Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD, 21218, USA

²⁸ Orbital Sciences Corporation, 2401 East El Segundo Boulevard, Suite 200, El Segundo, CA 90245, USA

²⁹ Lawrence Hall of Science, University of California Berkeley, USA

³⁰ Department of Astronomy, California Institute of Technology, Pasadena, CA 91125, USA

³¹ NASA/Marshall Space Flight Center, USA

³² Ball Aerospace and Technologies Corporation, Boulder, CO 80301, USA

³³ Instituto de Astrofísica de Canarias, E-38200, La Laguna, Tenerife, Spain

³⁴ Universidad de La Laguna, Dpto. de Astrofísica, E-38205, La Laguna, Tenerife, Spain

³⁵ Millennium Engineering & Integration Services, Moffett Field, CA 94035, USA

³⁶ Department of Physics and Astronomy, University of Southern California, Los Angeles, CA 90089-0484, USA

³⁷ Villanova University, Department of Astrophysics and Planetary Science, 800 Lancaster Avenue, Villanova PA 19085, USA

³⁸ Brigham Young University, Department of Physics and Astronomy, N283 ESC, Provo, UT 84602, USA

³⁹ Carnegie Observatories, USA

⁴⁰ Jacobs Engineering, USA

⁴¹ Rincon Research Corporation, 101 North Wilmot Road, Tucson, AZ 85711, USA

⁴² University of Nevada, Las Vegas, 4505 South Maryland Parkway, Las Vegas, NV 89154, USA

⁴³ Division of Astronomical Sciences, National Science Foundation, 2415 Eisenhower Avenue, Alexandria, VA 22314, USA⁴⁴ Department of Astronomy, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-1221 USA⁴⁵ KBRwyle, Houston, TX, USA

Received 2020 September 1; revised 2020 October 19; accepted 2020 October 20; published 2020 December 22

Abstract

We present the occurrence rates for rocky planets in the habitable zones (HZs) of main-sequence dwarf stars based on the Kepler DR25 planet candidate catalog and Gaia-based stellar properties. We provide the first analysis in terms of star-dependent instellation flux, which allows us to track HZ planets. We define η_{\oplus} as the HZ occurrence of planets with radii between 0.5 and $1.5 R_{\oplus}$ orbiting stars with effective temperatures between 4800 and 6300 K. We find that η_{\oplus} for the conservative HZ is between $0.37^{+0.48}_{-0.21}$ (errors reflect 68% credible intervals) and $0.60^{+0.90}_{-0.36}$ planets per star, while the optimistic HZ occurrence is between $0.58^{+0.73}_{-0.33}$ and $0.88^{+1.28}_{-0.51}$ planets per star. These bounds reflect two extreme assumptions about the extrapolation of completeness beyond orbital periods where DR25 completeness data are available. The large uncertainties are due to the small number of detected small HZ planets. We find similar occurrence rates between using Poisson likelihood Bayesian analysis and using Approximate Bayesian Computation. Our results are corrected for catalog completeness and reliability. Both completeness and the planet occurrence rate are dependent on stellar effective temperature. We also present occurrence rates for various stellar populations and planet size ranges. We estimate with 95% confidence that, on average, the nearest HZ planet around G and K dwarfs is ~ 6 pc away and there are ~ 4 HZ rocky planets around G and K dwarfs within 10 pc of the Sun.

Unified Astronomy Thesaurus concepts: Exoplanets (498)

Supporting material: machine-readable table

1. Introduction

One of the primary goals of the Kepler mission (Borucki et al. 2010; Koch et al. 2010; Borucki 2016) was to determine the frequency of occurrence of habitable-zone (HZ) rocky planets around Sun-like stars, also known as “ η_{\oplus} .“ HZ rocky planets are broadly construed as any rocky planet in its star’s HZ, roughly defined as being at the right distance from the star so that its surface temperature would permit liquid water (see Section 2). Measuring η_{\oplus} informs theories of planet formation, helping us to understand why we are here, and is an important input to mission design for instruments designed to detect and characterize HZ planets such as LUVOIR (The LUVOIR Team 2019) and HabEX (Gaudi et al. 2020).

Kepler’s strategy to measure η_{\oplus} was to continuously observe $>150,000$ solar-like main-sequence dwarf stars (primarily F, G, and K) with a highly sensitive photometer in solar orbit, identifying planets through the detection of transits. In the process, Kepler revolutionized our perspective of exoplanets in the Galaxy. The planet catalog in the final Kepler data release 25 (DR25) contains 4034 planet candidates (Thompson et al. 2018), leading to the confirmation or statistical validation of over 2300 exoplanets⁴⁸—more than half of all exoplanets known today.

Identifying HZ rocky planets proved to be a greater challenge than anticipated. Based on the sensitivity of the Kepler photometer and the expectation that solar variability was typical of quiet main-sequence dwarfs, it was believed that 4 years of observation would detect a sufficient number of rocky HZ planets to constrain their frequency of occurrence. However, Kepler observations showed that stellar variability is, on average, $\sim 50\%$ higher than

solar variability (Gilliland et al. 2011), which suppressed the number of HZ rocky planets that could be detected in 4 years. In response, Kepler’s observational time was extended to 8 years, but the failure of reaction wheels, required to maintain precise pointing, prevented the continuation of high-precision observations in the original Kepler field after 4 years (Howell et al. 2014). Furthermore, by definition, Kepler planet candidates must have at least three observed transits. The longest orbital period with three transits that can be observed in the 4 years of Kepler data is 710 days (assuming fortuitous timing when the transits occurred). Given that the HZs of many F and late G stars require orbital periods longer than 710 days, Kepler was not capable of detecting all HZ planets around these stars.

The result is Kepler data in which transiting rocky HZ planets are often near or beyond Kepler’s detection limit. Of the thousands of planets in the DR25 catalog, relatively few are unambiguously rocky and near their HZs: there are 56 such planet candidates with a radius of $\leq 2.5 R_{\oplus}$ and 9 planet candidates with a radius of $\leq 1.5 R_{\oplus}$ (using the planet radii from Berger et al. 2020a). As described in Section 2, we expect many planets near the HZ larger than $1.5 R_{\oplus}$ to be non-rocky. These small numbers present challenges in the measurement of the frequency of occurrence of HZ planets.

Converting a planet catalog into an underlying occurrence rate is also challenging due to the existence of selection effects and biases, with issues only exacerbated in the η_{\oplus} regime. Planet candidate catalogs generally suffer from three types of error:

1. The catalog is incomplete, missing real planets.
2. The catalog is unreliable, being polluted with false positives.
3. The catalog is inaccurate, with observational error leading to inaccurate planet properties.

Near the detection limit, both completeness and reliability can be low, requiring careful correction for the computation of occurrence rates. The DR25 planet candidate catalog includes several products that facilitate the characterization of and correction for completeness

⁴⁶ Deceased.⁴⁷ Retired.⁴⁸ <https://exoplanetarchive.ipac.caltech.edu>

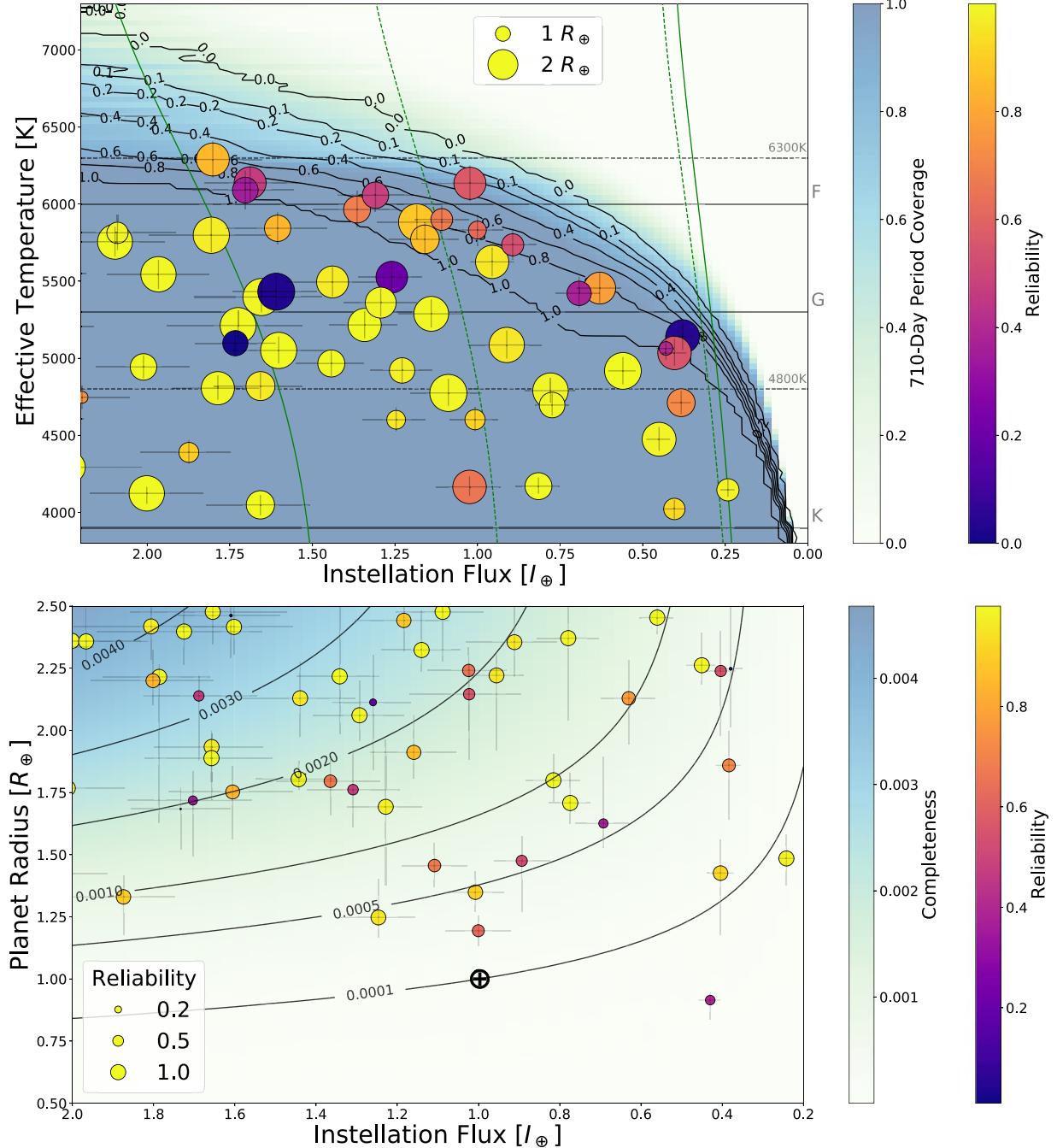


Figure 1. Two views of the DR25 planet candidate population with radii smaller than $2.5 R_{\oplus}$ and instellation flux near the host star’s HZ around main-sequence dwarf stars. Top: Instellation flux versus stellar effective temperature, showing the HZ and Kepler observational coverage. The background color map gives, at each point, the fraction of stars at that effective temperature and instellation flux that may have planets with orbital periods of 710 days or less, so it is possible to observe three transits. The contours show the fraction of planets with periods of 500 days or less, indicating available completeness measurements. The solid green lines are the boundaries of the optimistic HZ, while the dashed green lines are the boundaries of the conservative HZ (see Section 2). The planets are sized by their radius and colored by their reliability. Bottom: Instellation flux versus planet radius. The color map and contours show the average completeness for the stellar population (Section 3.3.1). The planets are sized and colored by reliability (Section 3.3.3), with radius and instellation flux error bars. In the lower panel the \oplus symbol shows the Earth.

and reliability (Thompson et al. 2018). Data supporting completeness characterization, however, are only supplied for orbital periods of 500 days or less, requiring extrapolation of completeness for planets beyond these orbital periods.

These issues are summarized in Figure 1, which shows the DR25 planet candidate population and its observational coverage, observational error, completeness, and reliability. The details of these populations are given in Appendix C.

Our calculation of HZ occurrence will be in terms of planet radius and instellation flux, measuring the photon flux incident on the planet from its host star, which allows us to consider each star’s HZ. We will proceed in two steps:

1. Develop a model describing the planet population in the neighborhood of the HZ (Section 3.4). Because this is a statistical study, the model will be based on a large

number of stars, using the observed DR25 planet candidate catalog and Gaia-based stellar properties, and will include corrections for catalog completeness and reliability.

2. Derive the average number of rocky planets per star in each star's HZ from the planet population model (Section 3.5). This will often be done in a subset of the parameter space used to compute the population model.

When computing a quantity over a desired range of parameters such as the radius and instellation flux, it is often the case that using data from a wider range will give better results. For example, it is well known that polynomial fits to data have higher uncertainty near the boundaries of the data. As explained in Section 2, we are primarily interested in rocky HZ planets, with planet radii between $0.5 R_{\oplus}$ and $1.5 R_{\oplus}$ and instellation flux within each star's estimated HZ, for stars with effective temperatures between 4800 and 6300 K. To create our population model, we will use a larger domain with a planet radius range of $0.5 R_{\oplus}$ to $2.5 R_{\oplus}$ and an instellation range 0.2 to 2.2 times Earth's insolation, which encloses the HZs of all the stars we consider. We will focus on using two stellar populations: one exactly matching our desired effective temperature range of 4800–6300 K and one with a larger range of 3900 K to 6300 K to investigate whether the larger range will improve our results. Most of our results will be reported for both stellar populations because it is possible that including stars in the 3900–4800 K range will bias our results. We will have a population model for each choice of stellar population.

Once we have our population models, we will use them to compute our definition of η_{\oplus} , the average number of planets per star with radii between 0.5 and $1.5 R_{\oplus}$, in the star's HZ, averaged over stars with effective temperatures from 4800 to 6300 K. In the end we will find that the two stellar populations predict similar median values for η_{\oplus} , but the model using stars with effective temperatures from 3900 K to 6300 K yields significantly smaller (though still large) uncertainties. While we are focused on our definition of η_{\oplus} , occurrence rates over other ranges of planet radii and stellar temperatures are of interest. We will use a population model based on the 3900–6300 K stellar population to compute the average number of HZ planets per star for various ranges of planet radii and stellar effective temperatures.

1.1. Previous Kepler-based η_{\oplus} Estimates

Attempts to measure η_{\oplus} and general occurrence rates with Kepler have been made since the earliest Kepler catalog releases (Borucki et al. 2011). Youdin (2011) and Howard et al. (2012) were two influential early studies, in which planets found in only the first four months of data (Q0–Q2) were used to constrain Kepler occurrence rates. Youdin (2011) developed a maximum likelihood method to fit an underlying planetary distribution function, which later influenced the Poisson likelihood function method adopted by, e.g., Burke et al. (2015) and Bryson et al. (2020a). Howard et al. (2012) took an alternative approach of estimating occurrence rates in bins of planets, defined over a 2D grid of planet radius and orbital period. In each bin, nondetections were corrected for by weighting each planet by the inverse of its detection efficiency. This inverse detection efficiency method is one of the most popular approaches in the literature.

Catanzarite & Shao (2011) and Traub (2012) (Q0–Q5; Borucki et al. 2011) were among the first papers to focus on the η_{\oplus} question specifically. Later η_{\oplus} papers from Dressing & Charbonneau (2013) (Q1–Q6; Batalha et al. 2013), Kopparapu et al. (2013) (Q1–Q6), Burke et al. (2015) (Q1–Q16; Mullally et al. 2016), and Silburt et al. (2015) (Q1–Q16) were able to take advantage of newer planet catalogs based on increased amounts of data. Other papers have used custom pipelines to search Kepler light curves to estimate η_{\oplus} with independently produced planet catalogs: namely, Petigura et al. (2013) (Q1–Q15), Foreman-Mackey et al. (2014) (Q1–Q15), Dressing & Charbonneau (2015) (Q1–Q16), and Kunimoto & Matthews (2020) (Q1–Q17). Still more have been meta-analyses of results from the exoplanet community based on different Kepler catalogs (Kopparapu 2018; Garrett et al. 2018).

Comparisons between these η_{\oplus} studies are challenging due to the wide variety of catalogs used, some of which are based on only a fraction of the data of the others. Characterization of completeness has also varied between authors, with some assuming a simple analytic model of detection efficiency (e.g., Youdin 2011; Howard et al. 2012), others empirically estimating detection efficiency with transit injection/recovery tests (e.g., Petigura et al. 2013; Burke et al. 2015), and still others simply assuming a catalog is complete beyond some threshold (e.g., Catanzarite & Shao 2011). Borucki et al. (2011) provided a comprehensive analysis of completeness bias, reliability against astrophysical false positives, and reliability against statistical false alarms based on manual vetting and simple noise estimates. Fully automated vetting has been implemented via the Robovetter (Coughlin 2017) for the Kepler DR24 (Coughlin et al. 2016) and DR25 catalogs. The final Kepler data release (DR25), based on the full set of Kepler observations and accompanied by comprehensive data products for characterizing completeness, has been essential for alleviating issues of completeness and reliability. The DR25 catalog is now the standard used by occurrence rate studies (e.g., Mulders et al. 2018; Hsu et al. 2018; Zink et al. 2019; Bryson et al. 2020a).

DR25 was the first catalog to include data products that allowed for the characterization of catalog reliability against false alarms due to noise and systematic instrumental artifacts, which are the most prevalent contaminants in the η_{\oplus} regime. Thus nearly all previous works did not incorporate reliability against false alarms in their estimates. Bryson et al. (2020a) were the first to directly take into account reliability against both noise/systematics and astrophysical false positives and in doing so found that occurrence rates for small planets in long-period orbits drop significantly after reliability correction. Mulders et al. (2018) attempted to mitigate the impact of contamination by using a DR25 Disposition Score cut (see Section 7.3.4 of Thompson et al. 2018) as an alternative to reliability correction. As shown in Bryson et al. (2020b), while this approach does produce a higher-reliability planet catalog, explicit accounting for reliability is still necessary for accurate occurrence rates.

Studies have also varied in stellar property catalogs used, and exoplanet occurrence rates have been shown to be sensitive to such choices. For instance, the discovery of a gap in the radius distribution of small planets, first uncovered in observations by Fulton et al. (2017), was enabled by improvements in stellar radius measurements by the California Kepler Survey (Johnson et al. 2017; Petigura et al. 2017). The use of Gaia DR2 parallaxes,

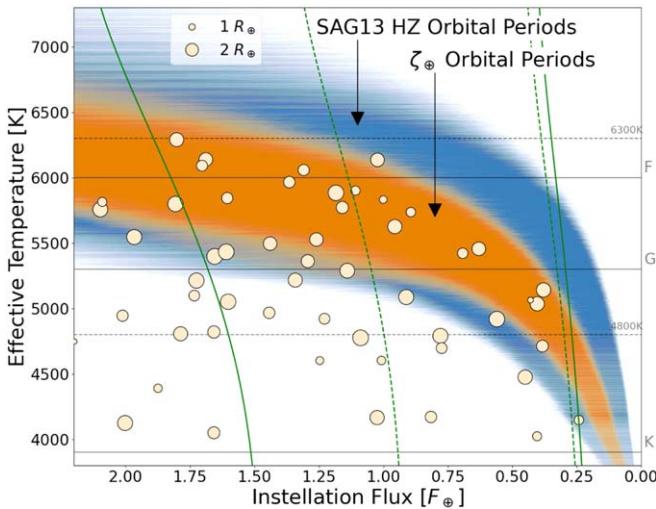


Figure 2. The HZ flux range compared with example orbital periods, previously used to estimate HZ occurrence for G, K, and F stars. For each star in the stellar parent sample, we show the instellation flux range of each orbital period range, with the SAG13 instellation flux range shown as the blue region (comprised of a horizontal blue line for each star showing the flux range for that orbital period range) and ζ_{\oplus} shown as the orange region. The solid green lines are the boundaries of the optimistic HZ, while the dashed green lines are the boundaries of the conservative HZ (see Section 2). The planet population is the same as that in Figure 1, and the planets are sized by their radius.

which have resulted in a reduction in the median stellar radius uncertainty of Kepler stars from $\approx 27\%$ to $\approx 4\%$ (Berger et al. 2020b), has been another significant improvement with important implications for η_{\oplus} . Bryson et al. (2020a) showed that the occurrence rates of planets near Earth’s orbit and size can drop by a factor of 2 if one adopts planet radii based on Gaia stellar properties rather than on pre-Gaia Kepler Input Catalog stellar properties.

1.2. Our Work

Measuring η_{\oplus} requires a definition of what it actually means to be considered a rocky planet in the HZ. Different authors use different definitions, including regarding whether η_{\oplus} refers to the number of rocky HZ planets per star or to the number of stars with rocky HZ planets. In this paper, for reasons detailed in Section 2, we define η_{\oplus} as the average number of planets per star with a planet radius between 0.5 and 1.5 Earth radii in the star’s HZ, where the average is taken over stars with effective temperatures between 4800 and 6300 K. We compute η_{\oplus} for both conservative and optimistic HZs, denoted respectively as η_{\oplus}^C and η_{\oplus}^O .

Most of the existing literature on HZ occurrence rates characterize the HZ using orbital period, where a single period range is adopted to represent the bounds of the HZ for the entire stellar population considered. However, no single period range covers the HZ for a wide variety of stars. Figure 2 shows two example period ranges used for HZ occurrence rate studies relative to the HZ of each star in our stellar parent sample. The SAG13⁴⁹ HZ range of $237 \leq \text{period} \leq 860$ days is shown in blue, and ζ_{\oplus} , defined in Burke et al. (2015) as within 20% of Earth’s orbital period, is shown in orange. While these period ranges cover much of the HZ for G stars, they miss significant portions of the HZs of K and F stars and include regions

outside the HZ even when restricted to G stars. This will be true for any fixed choice of orbital period range for the range of stellar effective temperatures required for good statistical analysis. Such coverage will not lead to accurate occurrence rates of planets in the HZ. Given that the period ranges of many HZ definitions also extend beyond the detection limit of Kepler, computing η_{\oplus} requires extrapolation of a fitted population model to longer orbital periods. Lopez & Rice (2018) and Pascucci et al. (2019) presented evidence and theoretical arguments that inferring the population of small rocky planets at low instellation from a population of larger planets at high instellation can introduce significant overestimates of η_{\oplus} .

For these reasons, we choose to work in terms of the instellation flux, measuring the photon flux incident on the planet from its host star, rather than the orbital period. In Section 3 we describe how we adopt existing occurrence rate methods and completeness characterizations to use instellation flux instead of orbital period. We address concerns with extrapolating completeness to long orbital periods by providing bounds on the impact of the limited coverage of completeness data (Section 3.3). Following Howard et al. (2012), Youdin (2011), and Burke et al. (2015), among others, we compute the number of planets per star f . As in Youdin (2011) and Burke et al. (2015), we first compute a population model in terms of the differential rate $\lambda \equiv d^2f/dr dl$, where r is the planet radius and l is the instellation flux. We consider several possible functional forms for λ and allow λ to depend on the stellar host’s effective temperature. We compute λ over the radius range $0.5 R_{\oplus} \leq r \leq 2.5 R_{\oplus}$ and the instellation flux range $0.2 I_{\oplus} \leq l \leq 2.2 I_{\oplus}$, averaged over the effective temperatures of the stellar population used for the computation (Section 3.4). Occurrence rates will be computed by integrating λ over the desired planet radius and instellation flux range and averaging over the desired effective temperature range to give f , the average number of planets per star (Section 3.5).

By restricting our analysis to planets with $r \leq 2.5 R_{\oplus}$ in regions of instellation flux close to the HZ, we believe we are avoiding the biases pointed out by Lopez & Rice (2018) and Pascucci et al. (2019). As seen in Figure 1, there are more detected planets with $1.5 R_{\oplus} \leq r \leq 2.5 R_{\oplus}$ than planets with $0.5 R_{\oplus} \leq r \leq 1.5 R_{\oplus}$, so our results in Section 4 will be driven by these larger planets, but all planets we consider are at similar low levels of instellation. In Figure 2 of Lopez & Rice (2018) we note that for instellation flux between 10 and 20 there is little change in the predicted relative sizes of the $1.5 R_{\oplus} \leq r \leq 2.5 R_{\oplus}$ and $0.5 R_{\oplus} \leq r \leq 1.5 R_{\oplus}$ planet populations. Naively extrapolating this to instellation of < 2 in the HZ, we infer that the sizes of these larger- and smaller-planet populations in the HZ are similar. Therefore by working at low instellation flux we are likely less vulnerable to overestimating the HZ rocky planet population.

We use both Poisson likelihood-based inference with Markov Chain Monte Carlo (MCMC) and likelihood-free inference with approximate Bayesian computation (ABC). The Poisson likelihood method is one of the most common approaches to calculating exoplanet occurrence rates (e.g., Burke et al. 2015; Zink et al. 2019; Bryson et al. 2020a), while ABC was only applied for the first time in Hsu et al. (2018, 2019). As described in Section 3.4.2, these methods differ in their treatment of reliability and input uncertainty and allow us to assess the possible dependence of our result on the

⁴⁹ <https://exoplanets.nasa.gov/exep/exopag/sag/#sag13>

assumption of a Poisson likelihood function. We present our results in Section 4. Recognizing the importance of reliability correction in the η_{\oplus} regime and confirming the same impact of reliability of Bryson et al. (2020a), we choose to report only our reliability-incorporated results. We also present results both with and without incorporating uncertainties in planet radius, instellation flux, and host star effective temperature in our analysis. Our final recommended population models and implications for η_{\oplus} , as well as how our results relate to previous estimates, are discussed in Section 5.

All results reported in this paper are produced with Python code, mostly in the form of Python Jupyter notebooks, found at the paper’s GitHub site.⁵⁰

1.3. Notation

When summarizing a distribution, we use the notation m_{-e2}^{+e1} to refer to the 68% credible interval $[m - e2, m + e1]$, where m is the median and “ $n\%$ credible interval” means that the central $n\%$ of the values fall in that interval. We caution our distributions are typically not Gaussian, so this interval should not be treated as “ 1σ .” We will also supply 95% and 99% credible intervals for our most important results.

We use the following notation throughout the paper:

1. r : planet radius in units of Earth radii R_{\oplus}
2. I : planet instellation flux in units of Earth instellation I_{\oplus}
3. T_{eff} : stellar effective temperature in kelvins. When referring to a planet, this is the effective temperature of that planet’s host star.
4. f : the number of planets per star, typically a function of r , I , and T_{eff}
5. λ : the differential rate population model $\equiv d^2f/dr dI$, typically a function of r , I , and T_{eff} . λ is defined by several parameters—for example, exponents when λ is a power law.
6. θ : the vector of parameters that define λ , whose contents depend on the particular form of λ

η_{\oplus} without modification refers to the average number of HZ planets per star with $0.5 R_{\oplus} \leq r \leq 1.5 R_{\oplus}$ and a host star effective temperature between 3900 and 6300 K, with the conservative or optimistic HZ specified in that context. η_{\oplus}^C and η_{\oplus}^O , respectively, specifically refer to occurrence in the conservative and optimistic HZs. Additional subscripts on η_{\oplus} refer to different stellar populations. For example $\eta_{\oplus,\text{GK}}^C$ is the occurrence of conservative HZ planets with $0.5 R_{\oplus} \leq r \leq 1.5 R_{\oplus}$ around GK host stars.

2. Habitability

2.1. Characterizing Rocky Planets in the HZ

A key aspect in computing HZ planet occurrence rates is the location and width of the HZ. Classically, it is defined as the region around a star in which a rocky-mass/size planet with an Earth-like atmospheric composition (CO_2 , H_2O , and N_2) can sustain liquid water on its surface. Surface liquid water is important for the development of life as we know it, and the availability of water on the surface assumes that any biological activity on the surface alters the atmospheric composition of

the planet, betraying the presence of life when observed with remote detection techniques.

Various studies estimate the limits and the width of the HZ in the literature (see Kopparapu 2018 and Kopparapu et al. 2019 for a review) and explore the effect of physical processes such as tidal locking, the rotation rate of the planet, the combination of different greenhouse gases, planetary mass, obliquity, and eccentricity on HZ limits. These effects necessitate a more nuanced approach to identifying habitability limits and are particularly useful for exploring those environmental conditions where habitability could be maintained. However, for the purpose of calculating the occurrence rates of planets in the HZ, it is best to use a standard for HZ limits as a first attempt, such as Earth-like conditions. One reason is that it would become computationally expensive to estimate the occurrence rates of HZ planets considering all the various HZ limits arising from these planetary and stellar properties. Furthermore, future flagship mission concept studies like LUVOIR (The LUVOIR Team 2019), HabEX (Gaudi et al. 2020), and OST (Meixner et al. 2019) used the classical HZ limits as their standard case to estimate exoEarth mission yields and identify associated biosignature gases. Therefore, in this study we use the conservative and optimistic HZ estimates from Kopparapu et al. (2014), where the conservative inner and outer edges of the HZ are defined by the “runaway greenhouse” and “maximum greenhouse” limits and the optimistic inner and outer HZ boundaries are the “recent Venus” and “early Mars” limits. By using these HZ limits, we are able to (1) make a consistent comparison with already published occurrence rates of HZ planets in the literature that have also used the same HZ limits, (2) provide a range of values for HZ planet occurrence, and (3) obtain an “average” occurrence rate of planets in the HZ, as the conservative and optimistic HZ limits from Kopparapu et al. (2014) span the range of HZ limits from more complex models and processes.

We consider planets in the $0.5\text{--}1.5 R_{\oplus}$ size range to calculate rocky planet occurrence rates, as studies have suggested that planets that fall within these radius bounds are most likely to be rocky (Rogers 2015; Wolfgang et al. 2016; Chen & Kipping 2017; Fulton et al. 2017). While some studies have indicated that the rocky regime can extend to as large as $2.5 R_{\oplus}$ (Otegi et al. 2020), many of these large-radius regime planets seem to be highly irradiated planets, receiving stellar fluxes much larger than those of planets within the HZ. Nevertheless, we have also calculated the occurrence rates of planets with radii up to $2.5 R_{\oplus}$. We note that Kane et al. (2016) also used Kopparapu et al. (2014) HZ estimates to identify HZ planet candidates using DR24 planet candidate catalog and DR25 stellar properties.

We also limit the host stellar spectral types to stars with $4800 \leq T_{\text{eff}} \leq 6300$ K, covering mid-K to late F. The reason for limiting to $T_{\text{eff}} > 4800$ K is twofold: (1) the inner working angle (the smallest angle on the sky at which a direct-imaging telescope can reach its designed ratio of planet to star flux) for the LUVOIR coronagraph instrument ECLIPS falls off below 48 mas at 1 micron ($3\lambda/D$) for a planet at 10 pc for $T_{\text{eff}} \leq 4800$ K, and (2) planets are likely tidal-locked or synchronously rotating below 4800 K, which could potentially alter the inner HZ limit significantly (Yang et al. 2013, 2014; Godolt et al. 2015; Way et al. 2015; Wolf & Toon 2015; Kopparapu et al. 2016, 2017; Bin et al. 2018). The upper limit of 6300 K is a result of planets in HZs having longer orbital periods around early F stars, where

⁵⁰ <https://github.com/stevepur/DR25-occurrence-public/tree/master/insolation>

Kepler was not capable of detecting these planets, as described in Section 1.

2.2. Effective Temperature Dependence of the Width of the HZ

The width of the HZ for hotter stars is larger than the width for cooler stars, implying that the HZ occurrence rate may be dependent on the host star's effective temperature. In this section we derive an approximate form for this effective temperature dependence, which we refer to as the “geometric effect.”

We compute the instellation flux I on a planet orbiting a particular star as $I = R_*^2 T^4 / a^2$, where R_* is the stellar radius in solar radii, $T = T_{\text{eff}} / T_\odot$ is the effective temperature divided by the solar effective temperature, and a is the semimajor axis of the planet orbit in astronomical units. We assume the orbit is circular. Then the size of the HZ Δa is determined by the instellation flux at the inner and outer HZ boundaries, I_{inner} and I_{outer} , as

$$\begin{aligned} \Delta a &= a_{\text{outer}} - a_{\text{inner}} \\ &= R_* T^2 \left(\frac{1}{\sqrt{I_{\text{outer}}}} - \frac{1}{\sqrt{I_{\text{inner}}}} \right). \end{aligned} \quad (1)$$

The factor $(1/\sqrt{I_{\text{outer}}} - 1/\sqrt{I_{\text{inner}}})$ has a weak T_{eff} dependence, ranging from 1.25 at 3900 K to 0.97 at 6300 K, which we crudely approximate as constant in this paragraph. We also observe that, for the main-sequence dwarf stellar populations we use in our computations (described in Section 3.1), R_* has an approximately linear dependence on T , which we write as $(\tau T + R_0)$ ($\tau \approx 1.8$ and $R_0 \approx -0.74$). Therefore

$$\Delta a \propto (\tau T + R_0) T^2. \quad (2)$$

So even if the differential occurrence rate λ has no dependence on a and therefore no dependence on I , the HZ occurrence rate may depend on T_{eff} simply because hotter stars have larger HZs.

Several studies, such as Burke et al. (2015) and Bryson et al. (2020a), have examined planet occurrence in terms of the orbital period p and have shown that df/dp is well approximated by a power law p^α . In Appendix A we show that this power law, combined with the relationship between instellation flux and period, implies that the instellation flux portion of the differential rate function λ , df/dI , has the form

$$df/dI \approx CI^\nu ((\tau T + R_0) T^4)^\delta \quad (3)$$

where $\nu = -\frac{3}{4}(\alpha - \frac{7}{3})$ and $\delta = -\nu - 1$. This form incorporates the combined effects of the size of the HZ increasing with T_{eff} and dependence on the period power law p^α . The derivation in Appendix A uses several crude approximations, so Equation (3) is qualitative rather than quantitative.

In Section 3.4 we consider forms of the population model λ that separate the geometric effect in Equation (2) from a possible more physical dependence on T_{eff} , and compare them with a direct measurement of the T_{eff} dependence. To separate the geometric effect, we incorporate a geometric factor $g(T_{\text{eff}})$ inspired by Equation (2). Because of the crude approximations used to derive Equations (2) and (3) we use an empirical fit to the HZ width Δa for all stars in our stellar sample. Because we will use this fit in models of the differential population rate df/dI in Section 3.4.1, we perform the fit computing Δa for each star

using a fixed flux interval $\Delta I \in [0.25, 1.8]$. Because $a = R_*^2 T^4 / \sqrt{I}$, Δa is just a scaling of each star's luminance $R_*^2 T^4$ by the factor $1/\sqrt{0.25} - 1/\sqrt{1.8}$. As shown in Figure 3, Δa is well fit, with well-behaved residuals, by the broken power law

$$g(T_{\text{eff}}) = \begin{cases} 10^{-11.84} T_{\text{eff}}^{3.16} & \text{if } T_{\text{eff}} \leq 5117 \text{ K}, \\ 10^{-16.77} T_{\text{eff}}^{4.49} & \text{otherwise.} \end{cases} \quad (4)$$

If the semimajor axes of planets are uniformly distributed in our stellar sample, then we expect that HZ planet occurrence would have a T_{eff} dependence due to Equation (4). In individual planetary systems, however, there is evidence of constant spacing in $\log(a)$ (Weiss et al. 2018), implying spacing proportional to $\Delta a/a$. In this case there would be no impact of larger HZs with increasing T_{eff} : taking a to be the average of the inner and outer semimajor axes, the star-dependent terms cancel each other, so $\Delta a/a$ is the same for all stars, independently of T_{eff} . This would imply that HZ occurrence has no T_{eff} dependence due to the increasing size of the HZ. However, Weiss et al. (2018) also pointed out that the characteristic spacing of systems is at most weakly correlated with stellar type, with stronger correlation of the characteristic spacing with planet radius. A more detailed treatment of planet spacing is necessary to explore how multiplicity influences HZ planet occurrence.

3. Methodology

We base our occurrence rate of f planets per star on a differential population rate model $\lambda(I, r, T_{\text{eff}}) = \frac{d^2 f(I, r, T_{\text{eff}})}{dI dr}$ that describes how f varies as a function of incident stellar flux I and planet radius r . We allow $\lambda(I, r, T_{\text{eff}})$ (and therefore $f(I, r, T_{\text{eff}})$) to depend on the host star's effective temperature T_{eff} . In Section 3.4 we use the DR25 planet candidate catalog to determine λ . We cannot, however, simply take all the planet candidates in the DR25 catalog at face value. We must statistically characterize, and correct for, errors in the catalog.

The DR25 planet candidate catalog contains 4034 planet candidates, identified through a uniform method of separating planet candidates from false positives and false alarms (Thompson et al. 2018). This automated method is known to make mistakes, being both incomplete due to missing true transiting planets and unreliable due to misidentifying various types of astrophysical false positives and instrumental false alarms as transiting planets. Low completeness and low reliability are particularly acute near the Kepler detection limit, which happens to coincide with the HZs of F, G, and K stars. We characterize DR25 completeness and reliability using the synthetic data described in Thompson et al. (2018) with the methods described in Bryson et al. (2020a). We correct for completeness and reliability when determining the population rate λ using the methods of Bryson et al. (2020a) and Kunimoto & Bryson (2020).

The methods used in Bryson et al. (2020a) and Kunimoto & Bryson (2020) computed population models in terms of orbital period and radius. Generalizing these methods to instellation flux, radius, and effective temperature is relatively straightforward, with the treatment of completeness characterization presenting the greatest challenge. In this section we summarize these methods, focusing on the changes required to operate in instellation flux rather than in period and to allow for dependence on T_{eff} .

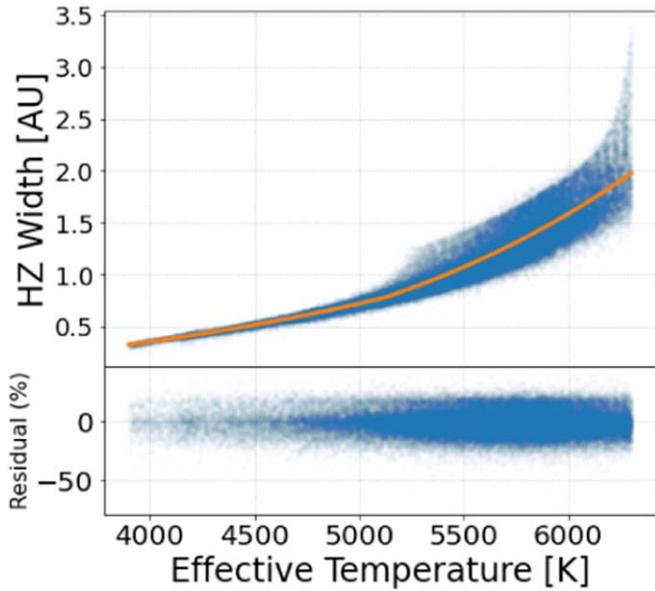


Figure 3. The width of the optimistic HZ (outer HZ boundary minus inner HZ boundary) as a function of effective temperature for all the stars in our parent sample with effective temperatures between 3900 and 6300 K. The line shows the broken power-law fit in Equation (4). The percentage residual from the fit is shown in the lower panel.

3.1. Stellar Populations

As in Bryson et al. (2020a), our stellar catalog uses the Gaia-based stellar properties from Berger et al. (2020b) combined with the DR25 stellar catalog at the NASA Exoplanet Archive (see Footnote 49), with the cuts described in the baseline case of Bryson et al. (2020a). We summarize these cuts here for convenience.

We begin by merging the catalog from Berger et al. (2020b), the DR25 stellar catalog (with its supplement), and the catalog from Berger et al. (2018), keeping only the 177,798 stars that are in all three catalogs. We remove poorly characterized, binary, and evolved stars, as well as stars whose observations were not well suited for long-period transit searches (Burke et al. 2015; Burke & Catanzarite 2017), with the following cuts:

1. Remove stars with a Berger et al. (2020b) goodness of fit (*iso_gof*) of <0.99 and a Gaia renormalized unit weight error (RUWE; Lindegren 2018), as provided by Berger et al. (2020b), of >1.2 , leaving 162,219 stars.
2. Remove stars that, according to Berger et al. (2018), are likely binaries, leaving 160,633 stars.
3. Remove stars that have evolved off the main sequence, recomputing the Evol flag described in Berger et al. (2018) using the Berger et al. (2020b) stellar properties, leaving 105,118 stars.
4. Remove noisy targets identified in the KeplerPorts package,⁵¹ leaving 103,626 stars.
5. Remove stars with NaN limb darkening coefficients, leaving 103,371 stars.
6. Remove stars with a NaN observation duty cycle, leaving 102,909 stars.

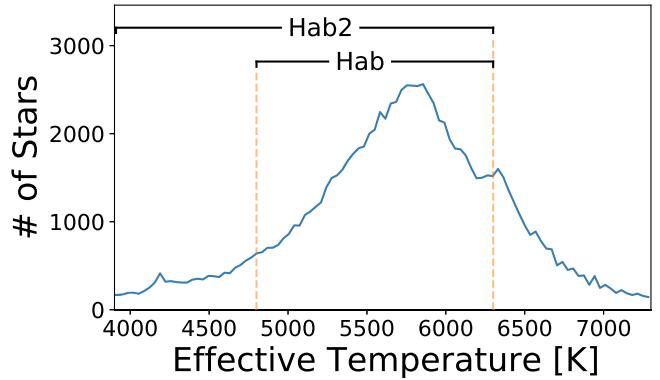


Figure 4. The distribution of stellar effective temperature for the stellar populations used in this paper.

7. Remove stars with a decrease in observation duty cycle of $>30\%$ due to data removal from other transits detected on these stars, leaving 98,672 stars.
8. Remove stars with an observation duty cycle of $<60\%$, leaving 95,335 stars.
9. Remove stars with a data span of <1000 days, leaving 87,765 stars.
10. Remove stars with a DR25 stellar properties table *timeoutsumry* flag $\neq 1$, leaving 82,371 stars.

Selecting FGK stars with effective temperatures between 3900 and 7300 K, which are a superset of the stellar populations we consider in this paper, we have 80,929 stars.

We are primarily interested in HZ occurrence rates for stars with effective temperatures hotter than 4800 K, and Kepler observational coverage is very poor above 6300 K (see Section 2). We fit our population model using two stellar populations and examine the solutions to determine which range of stellar temperatures is best for computing the desired occurrence rate. These stellar populations are as follows:

1. hab: stars with effective temperatures between 4800 and 6300 K (61,913 stars)
2. hab2: stars with effective temperatures between 3900 and 6300 K (68,885 stars)

The effective temperature distribution of the stars in these populations is shown in Figure 4. This distribution has considerably fewer cooler stars than we believe is the actual distribution of stars in the Galaxy. Our analysis is weighted by the number of stars as a function of effective temperature.

There are two stellar population cuts recommended in Burke & Catanzarite (2017) that we do not apply. The first is the requirement that stellar radii be less than $1.35 R_{\odot}$ ($1.25 R_{\odot}$ in Burke & Catanzarite 2017, but Burke now recommends $1.35 R_{\odot}$ (private communication)). We do not impose this stellar radius cut, instead opting for a physically motivated selection based on the Evol flag. After our cuts, 6.8% of the hab2 population are stars that have radii larger than $1.35 R_{\odot}$. The completeness analysis for these stars is not expected to be as accurate as that for smaller stars.

The second recommended cut that we do not apply is the requirement that the longest transit duration be less than 15 hr. This cut is due to the fact planet search in the Kepler pipeline does not examine transit durations longer than 15 hr (Twicken et al. 2016). For the hab2 population, assuming circular orbits, the transit durations of planets at the inner optimistic HZ boundary exceed 15 hr for 2.7% of the stars. The transit durations of planets at the outer optimistic HZ boundary exceed 15 hr for 35% of the stars,

⁵¹ https://github.com/nasa/KeplerPORTs/blob/master/DR25_DEModel_NoisyTargetList.txt

with the duration being less than 25 hr for 98.7% of the stars. While the exact impact on completeness analysis of transit durations longer than 15 hr is unknown, there is evidence that the impact is small. KOI 5236.01, for example, has a transit duration of 14.54 hr, an orbital period of 550.86 days, and a transit signal-to-noise ratio (S/N) of 20.8. KOI 5236.01 is correctly identified in the Kepler pipeline in searches for transit durations of 3.5–15 hr. KOI 7932.01 has a transit duration of 14.84 hr, an orbital period of 502.256 days, and a transit S/N of 8.1, among the lowest transit S/N for planet candidates with periods of >450 days. KOI 7932.01 is correctly identified in searches using transit durations of 9–15 hr. So even for low-S/N transits, the transit can be identified in searches for transit durations 9–15 times the actual duration. If these examples are typical, we can expect that transit durations of up to 25 hr will be detected. While these examples do not show that the impact of long transits on completeness is actually small, the bulk of these long durations occur in orbits beyond 500 days, so they are absorbed by the upper and lower bounds on completeness we describe in Section 3.3.2. We are confident that long transit durations, as well as large stars, cause completeness to decrease, so their impact falls within these upper and lower bounds.

3.2. Planet Input Populations

We use the planet properties from the Kepler DR25 threshold-crossing event (TCE) catalog (Twicken et al. 2016), with the Gaia-based planet radii and instellation fluxes from Berger et al. (2020a).

Three DR25 small-planet candidates that are near their host star’s HZ (planet radius of $\leq 2.5 R_{\oplus}$ and instellation flux between 0.2 and $2.2 I_{\oplus}$) are not included in our planet sample. KOI 854.01 and KOI 4427.01 are orbiting host stars with effective temperatures of ≤ 3900 K, and KOI 7932.01’s host star is cut from our stellar populations because it is marked “evolved” (see Bryson et al. 2020a).

3.3. Completeness and Reliability

3.3.1. Detection and Vetting Completeness

The DR25 completeness products are based on injected data—a ground truth of transiting planets obtained by injecting artificial transit signals with known characteristics on all observed stars at the pixel level (Christiansen et al. 2020). A large number of transits are also injected on a small number of target stars to measure the dependence of completeness on transit parameters and stellar properties (Burke & Catanzarite 2017). The data are then analyzed by the Kepler detection pipeline (Jenkins et al. 2010) to produce a catalog of detections at the injected ephemerides called injected and recovered TCEs, which are then sent through the same Robovetter used to identify planet candidates.

Detection completeness is defined as the fraction of injected transits that are recovered as TCEs by the Kepler detection pipeline, whether or not those TCEs are subsequently identified as planet candidates. We use the detection completeness of Burke & Catanzarite (2017), which was computed for each target star as a function of the period and the simulated multiple-event statistic (MES), based on stellar noise properties measured in that star’s Kepler light curve. MES is a measure of the S/N that is specific to the Kepler pipeline (Jenkins et al. 2010). The result is referred to as completeness detection contours.

Vetting completeness is defined as the fraction of detected injected transits that are identified as planet candidates by the

Robovetter (Coughlin 2017). We compute vetting completeness for a population of stars based on the simulated MES and orbital period of injected transits. We use the method of Bryson et al. (2020a), which models vetting completeness as a binomial problem with a rate given by the product of rotated logistic functions of MES and the orbital period. We assume that vetting completeness and detection completeness are independent of each other, so we can multiply them together to create combined completeness contours.

The product of vetting and detection completeness as a function of period and MES is converted to a function of period and planet radius for each star. This product is further multiplied by the geometric transit probability for each star, which is a function of planet period and radius, given that star’s radius. The final result is a completeness contour for each star that includes detection and vetting completeness and geometric transit probability.

We need to convert the completeness contours from radius and period to radius and instellation flux. For each star, we first set the range of instellation fluxes to $0.2 \leq I \leq 2.2$, which contains the HZ for FGK stars. We then interpolate the completeness contour from period to instellation flux via $I = R_*^2 T^4 / a^2$, where R_* is the stellar radius, $T = T_{\text{eff}} / T_{\odot}$ is the effective temperature relative to the Sun, and a is the semimajor axis of a circular orbit around this star with a given period.

Once the completeness contours are interpolated onto the radius and instellation flux for all stars, they are summed or averaged as required by the inference method used in Section 3.4.2.

3.3.2. Completeness Extrapolation

For most stars in our parent sample, there are regions of the HZ that require orbital periods beyond the 500 day limit of the period–radius completeness contours. Figure 5 shows the distribution of orbital periods at the inner and outer optimistic HZ boundaries for FGK stars in our stellar sample relative to the 500 day limit. We see that a majority of these stars will require some completeness extrapolation to cover their HZs and a small fraction of stars have no completeness information at all. It is unknown precisely how the completeness contours will extrapolate out to longer periods, but we believe that the possible completeness values can be bounded.

We assume that completeness is, on average, a decreasing function of orbital period. Therefore, the completeness beyond 500 days will be less than the completeness at 500 days. While this may not be a correct assumption for a small number of individual stars due to local completeness minima in the period due to the window function (Burke & Catanzarite 2017), we have high confidence that this assumption is true on average. We therefore bound the extrapolated completeness for each star by computing the two extreme extrapolation cases:

1. Constant-completeness extrapolation, where, for each radius bin, completeness for periods greater than 500 days is set to the completeness at 500 days. This extrapolation will have higher completeness than reality, resulting in a smaller completeness correction and lower occurrence rates, which we take to be a lower bound. In the tables below we refer to this lower bound as “low” values. Here “low” refers to the resulting occurrence rates, and some population model parameters in the this case will have higher values.
2. Zero-completeness extrapolation, where, for each radius bin, completeness for periods greater than 500 days is set to zero. Zero completeness will have lower completeness

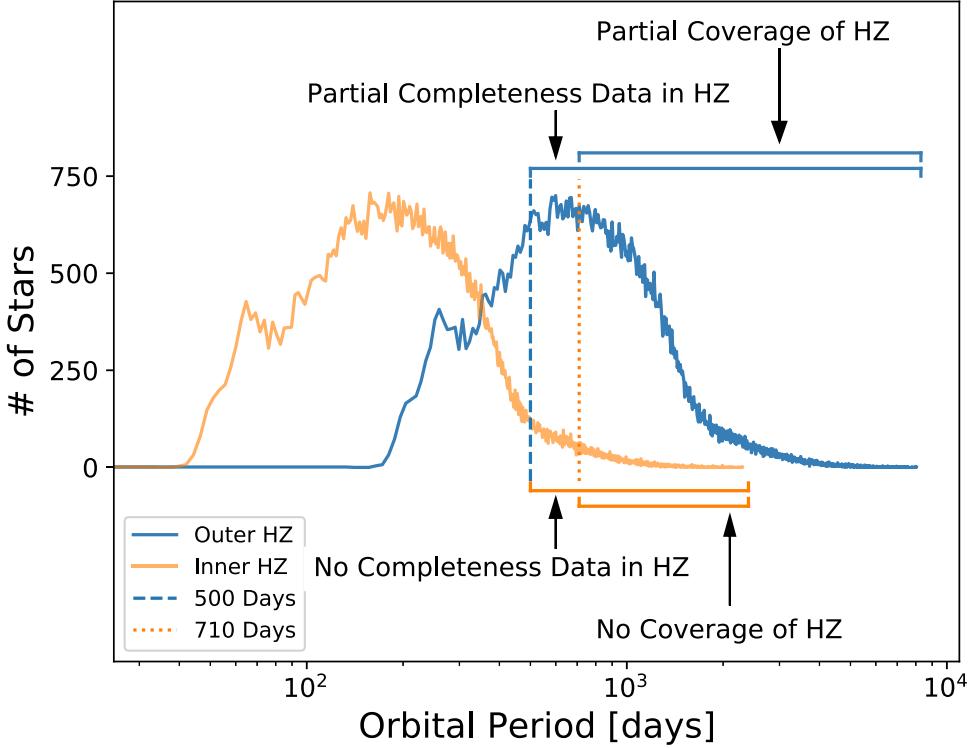


Figure 5. The distribution of orbital periods of the inner and outer optimistic HZ boundaries. We show the orbital period distribution of circular orbits at the outer (blue) and inner (orange) boundaries of the optimistic HZ for our FGK stellar sample. The blue vertical dashed line at 500 days indicates the limit of the completeness contours, beyond which there is no completeness data. The orange vertical dotted line at 710 days shows the limit of Kepler coverage, in the sense that beyond 710 days there is no possibility of three detected transits resulting in a planet detection. Stars whose orbital period for the inner HZ boundary is beyond 500 days have no completeness data in their HZ, while stars whose outer HZ boundary orbital period is beyond 500 days require some completeness extrapolation. Kepler could not detect HZ planets for stars whose inner HZ orbital period is beyond 710 days, while stars whose outer HZ orbital periods are beyond 710 days have only partial coverage, which will decrease completeness.

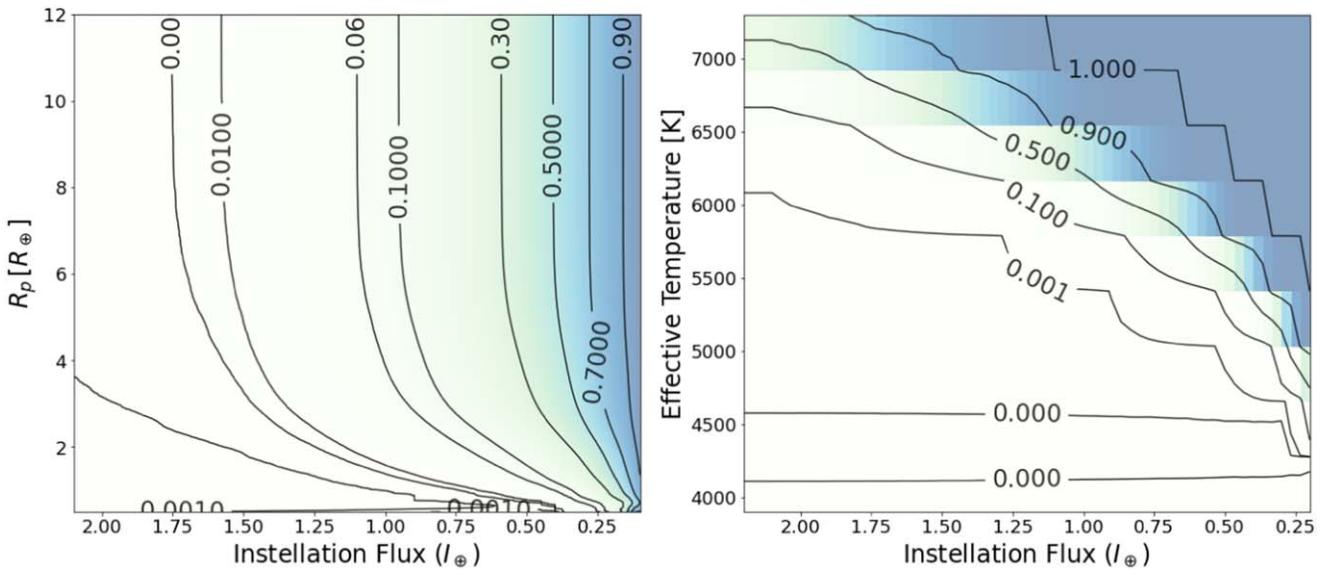


Figure 6. Left: The relative difference (difference divided by value) between the constant extrapolation and zero extrapolation completeness contours, summed over FGK stars, as a function of instellation flux and radius. Right: The relative difference as a function of instellation flux and effective temperature.

than reality, resulting in a larger completeness correction and higher occurrence rates, which we take to be an upper bound. In the tables below we refer to this upper bound as “high” values. Here “high” refers to the resulting occurrence rates, and some population model parameters in this case will have lower values.

We solve for population models and compute occurrence rates for both extrapolation cases. Figure 6 shows the relative difference in the completeness contours summed over all stars. We see that for effective temperatures below ~ 4500 K the difference between constant- and zero-completeness extrapolation is very close to zero, because these cooler stars are well

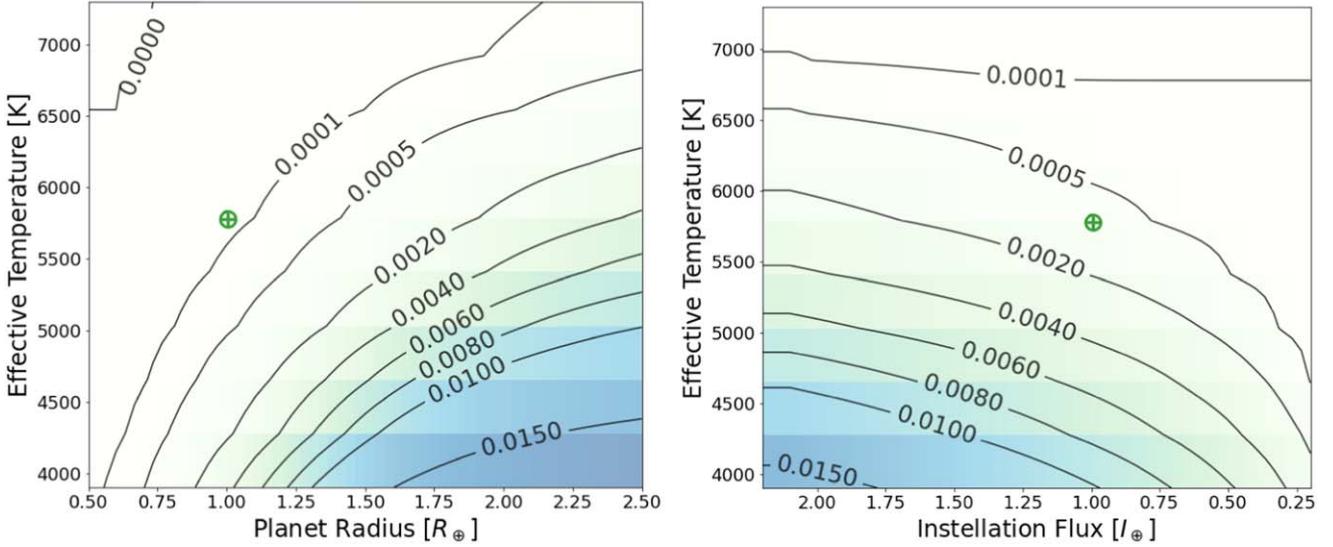


Figure 7. Example dependence of completeness on effective temperature, using the FGK stellar population and constant-completeness extrapolation, which provides an upper completeness bound. Left: Planet radius versus effective temperature. Right: Instellation flux versus effective temperature. The location of the Earth–Sun system is shown with a \oplus symbol.

covered by completeness data (see Figure 1). We therefore expect the upper and lower occurrence rate bounds to converge for these stars.

The Poisson likelihood we use requires the completeness contours to be summed over all stars, while the ABC method requires the completeness to be averaged over all stars. We observe a significant dependence of summed completeness on effective temperature, shown in Figure 7. We address this dependence of completeness on effective temperature by summing (for the Poisson likelihood) or averaging (for ABC) the completeness contours in effective temperature bins, as described in Section 3.4.2.

3.3.3. Reliability

We compute planet reliability as in Bryson et al. (2020a). Because this is done as a function of the MES and period, there is no change from the methods of that paper.

3.4. Computing the Population Model $\lambda(I, r, T)$

As described in Section 1.2, we develop a planet population model using a parameterized differential rate function λ and use Bayesian inference to find the model parameters that best explain the data. To test the robustness of our results, we use both the Poisson likelihood MCMC method of Burke et al. (2015) and the ABC method of Kunimoto & Matthews (2020) to compute our population model. Both methods are modified to account for vetting completeness and reliability, with the Poisson likelihood method described in Bryson et al. (2020a) and the ABC method described in Kunimoto & Bryson (2020) and Bryson et al. (2020b). In contrast to previous work, we also take into account uncertainties in planet radius, instellation flux, and host star effective temperature, described in Section 3.4.2.

3.4.1. Population Model Choices

We consider three population models for the differential rate function $\lambda(I, r, T)$. These models are functions of instellation flux I , planet radius r , and stellar effective temperature T_{eff} .

These models depend on possibly different sets of parameters, which we describe with the parameter vector θ . For each model, we solve for the θ that best describes the planet candidate data:

$$\begin{aligned}\lambda_1(I, r, T, \theta) &= F_0 C_1 r^\alpha I^\beta T^\gamma g(T), \theta = (F_0, \alpha, \beta, \gamma) \\ \lambda_2(I, r, T, \theta) &= F_0 C_2 r^\alpha I^\beta T^\gamma, \theta = (F_0, \alpha, \beta, \gamma) \\ \lambda_3(I, r, T, \theta) &= F_0 C_3 r^\alpha I^\beta g(T), \theta = (F_0, \alpha, \beta)\end{aligned}\quad (5)$$

where $g(T)$ is given by Equation (4). The normalization constants C_i in Equation (5) are chosen so that the integral of λ from r_{\min} to r_{\max} and I_{\min} to I_{\max} , averaged over T_{\min} to T_{\max} , = F_0 , so F_0 is the average number of planets per star in that radius, instellation flux, and effective temperature range.

λ_1 allows for dependence on T_{eff} beyond the geometric dependence described in Section 2.2, breaking possible degeneracy between any intrinsic T_{eff} and the geometric dependence by fixing the geometric dependence as $g(T)$. So, for example, if the planet population rate's dependence is entirely due to the larger HZ for hotter stars, captured in λ_1 by $g(T)$, then there is no additional T_{eff} dependence and $\gamma = 0$. λ_2 does not separate out the geometric T_{eff} dependence. λ_3 assumes that there is no T_{eff} dependence beyond the geometric effect.

All models and inference calculations use uniform uninformative priors: $0 \leq F_0 \leq 50,000$, $-5 \leq \alpha \leq 5$, $-5 \leq \beta \leq 5$, and $-500 \leq \gamma \leq 50$. The computations are initialized to a neighborhood of the maximum likelihood solution obtained with a standard nonlinear solver.

3.4.2. Inference Methods

Both the Poisson likelihood and ABC inference methods use the same stellar and planet populations and the same characterization of completeness and reliability computed using the approach of Bryson et al. (2020a). These steps are as follows:

1. Select a target star population, which will be our parent population of stars that are searched for planets. We apply various cuts intended to select well-behaved and

- well-observed stars. We consider two such populations, defined by the effective temperature range as described in Section 3.1, in order to explore the dependence of our results on the choice of stellar population.
2. Use the injected data to characterize vetting completeness.
 3. Compute the detection completeness using a version of KeplerPorts⁵² modified for vetting completeness and insolation interpolation, incorporating vetting completeness and geometric probability for each star, and sum over the stars in effective temperature bins, as described in Section 3.3.1.
 4. Use observed, inverted, and scrambled data to characterize false-alarm reliability, as described in Section 3.3.3.
 5. Assemble the collection of planet candidates, including by computing the reliability of each candidate from the false-alarm reliability and false-positive probability.
 6. For each model in Equation (5), use the Poisson likelihood or ABC method to infer the model parameters θ that are most consistent with the planet candidate data for the selected stellar population.

Because vetting completeness and reliability depend on the stellar population and the resulting planet catalog, all steps are computed for each choice of stellar population.

A full study of the impact of uncertainties in stellar and planet properties would include the impact of uncertainties on detection contours and is beyond the scope of this paper. However, we study the impact of uncertainties in planet radius, insellation flux, and host star effective temperature, shown Figure 1, on our occurrence rates. For both the Poisson likelihood and ABC methods we perform our inference computation both with and without uncertainties, which allows us to estimate the approximate contribution of input planet property uncertainties to the final occurrence rate uncertainties. In Bryson et al. (2020a) it was shown that uncertainties in reliability characterization have effectively no impact.

For the Poisson likelihood inference of the parameters in Equation (5) without input uncertainty, reliability is implemented by running the MCMC computation 100 times, with the planets removed with a probability given by their reliability. The likelihood used in the Poisson method is Equation (17) in Appendix B. For details see Bryson et al. (2020a).

We treat input uncertainties in a manner similar to how we treat reliability: we run the Poisson MCMC inference 400 times, each time selecting the planet population according to reliability. We then sample the planet insellation flux, radius, and star effective temperature from the two-sided normal distribution with the width given by the respective catalog uncertainties. We perform this sampling prior to restricting to our period and insellation flux ranges of interest so planets whose median property values are outside the range may enter the range due to their uncertainties. The posteriors from the 400 runs are concatenated together to give the posterior distribution of the parameters θ for each model. This approach to uncertainty does not recompute the underlying parent stellar population with resampled effective temperature uncertainties, because that would require recomputation of the completeness contours with each realization, which is beyond our computational resources. Shabram et al. (2020) performed a similar uncertainty study, properly resampling the underlying

parent population, and observed an impact of uncertainty similar to ours (see Section 4.2). Our analysis of uncertainty should be considered an approximation. While the result is not technically a sample from a posterior distribution, in Section 4 we compare the resulting sample to the posterior sample from the model neglecting these uncertainties and find that the population parameter values and resulting occurrence rates change in a predictable way.

The ABC-based inference of the parameters in Equation (5) is computed using the approach of Kunimoto & Bryson (2020), with some modifications to accommodate temperature dependence and uncertainties on planet radius, insellation flux, and temperature.

In the ABC method, the underlying Kepler population is simulated in each completeness effective temperature bin separately. $N_p = F_0 N_s h(T)$ planets are drawn for each bin, where N_s is the number of stars in the bin and $h(T)$ collects the model-dependent temperature terms from Equation (5), averaged over the temperature range of the bin and normalized over the entire temperature range of the sample. Following the procedure of Mulders et al. (2018), we assign each planet an insellation flux between 0.2 and 2.2 I_\oplus from the cumulative distribution function of I^β and a radius between 0.5 and 2.5 R_\oplus from the cumulative distribution function of r^α . The detectable planet sample is then simulated from this underlying population by drawing from a Bernoulli distribution with a detection probability averaged over the bin's stellar population. We compare the detected planets to the observed planet candidate population using a distance function, which quantifies agreement between the flux distributions, radius distributions, and sample sizes of the catalogs. For the distances between the flux and radius distributions, we choose the two-sample Anderson–Darling (AD) statistic, which has been shown to be more powerful than the commonly used Kolmogorov–Smirnov test (Engmann & Cousineau 2011). The third distance is the modified Canberra distance from Hsu et al. (2019),

$$\rho = \sum_{i=1}^N \frac{|n_{s,i} - n_{o,i}|}{\sqrt{n_{s,i} + n_{o,i}}} \quad (6)$$

where $n_{s,i}$ and $n_{o,i}$ are the number of simulated and observed planets within the i th bin's temperature range and the sum is over all N bins. For more details, see Bryson et al. (2020b).

These simulations are repeated within a population Monte Carlo ABC algorithm to infer the parameters that give the closest match between the simulated and observed catalogs. With each iteration of the ABC algorithm, model parameters are accepted when each resulting population's distance from the observed population is less than the 75th quantile of the previous iteration's accepted distances. Following the guidance of Prangle (2017), we confirm that our algorithm converges by observing that the distances between the simulated and observed catalogs approach zero with each iteration, and see that the uncertainties on the model parameters flatten out to a noise floor.

This forward model is appropriate for estimating the average number of planets per star in a given flux, radius, and temperature range, similar to the Poisson likelihood method. However, rather than requiring many inferences on different catalogs to incorporate reliability or input uncertainty, we take a different approach. For reliability, we modify the distance function as described in Bryson et al. (2020b). In summary, we replace the two-sample AD statistic with a generalized AD

⁵² <https://github.com/nasa/KeplerPORTS>

statistic developed in Trusina et al. (2020) that can accept a weight for each data point, and set each observed planet's weight equal to its reliability. We also alter the third distance (Equation (6)) so that a planet's contribution to the total number of planets in its bin is equal to its reliability. As demonstrated in Kunimoto & Bryson (2020) and Bryson et al. (2020b), this weighted distance approach gives results consistent with the Poisson likelihood function method with reliability. Meanwhile, to account for input uncertainty, the observed population is altered for every comparison with a simulated population by randomly assigning each observed planet a new radius, instellation flux, and host star effective temperature from the two-sided normal distribution with the width given by their respective uncertainties.

3.5. Computing Occurrence Rates

Once the population rate model λ has been chosen and its parameters determined as described in Section 3.4, we can compute the number of HZ planets per star. For planets with radii between r_0 and r_1 and instellation fluxes between I_0 and I_1 , for a star with effective temperature T_{eff} the number of planets per star is

$$f(T_{\text{eff}}) = \int_{r_0}^{r_1} \int_{I_0}^{I_1} \lambda(I, r, T_{\text{eff}}, \theta) dI dr. \quad (7)$$

For a collection of stars with effective temperatures ranging from T_0 to T_1 , we compute the average number of planets per star, assuming a uniform distribution of stars in that range, as

$$\bar{f} = \frac{1}{T_1 - T_0} \int_{T_0}^{T_1} f(T) dT. \quad (8)$$

We typically compute Equation (8) for every θ in the posterior of our solution, giving a distribution of occurrence rates.

The HZ is not a rectangular region in the $I-T_{\text{eff}}$ plane (see Figure 1), so to compute occurrence in the HZ for a given T_{eff} , we integrate I from the inner HZ flux $I_{\text{out}}(T_{\text{eff}})$ to the outer flux $I_{\text{in}}(T_{\text{eff}})$:

$$f_{\text{HZ}}(T_{\text{eff}}) = \int_{r_{\min}}^{r_{\max}} \int_{I_{\text{out}}(T_{\text{eff}})}^{I_{\text{in}}(T_{\text{eff}})} \lambda(I, r, T_{\text{eff}}, \theta) dI dr. \quad (9)$$

The functions $I_{\text{out}}(T)$ and $I_{\text{in}}(T)$ are given in Kopparapu et al. (2014) and depend on the choice of conservative versus optimistic HZ. $f_{\text{HZ}}(T_{\text{eff}})$ will be a distribution of occurrence rates if we use a distribution of θ . For a collection of stars with effective temperatures ranging from T_0 to T_1 , we compute $f_{\text{HZ}}(T_{\text{eff}})$ for a sampling of $T \in [T_0, T_1]$ and concatenate these distributions together to make a distribution of HZ occurrence rates f_{HZ} for that radius, flux, and temperature range. When we compute f_{HZ} to determine the occurrence rate for a generic set of stars, we uniformly sample over $[T_0, T_1]$ (in practice we use all integer kelvin values of $T \in [T_0, T_1]$). The resulting distribution is our final result.

Figure 8 shows the impact of uncertainty in stellar effective temperature on the HZ boundaries. For each star we compute the uncertainty in the HZ boundaries with 100 realizations of that star's effective temperature with uncertainty, modeled as a two-sided Gaussian. The gray regions in Figure 8 show the 86% credible intervals of the uncertainty of the HZ boundaries. These intervals are small relative to the size of the HZ and are well centered on the central value. For example, consider the

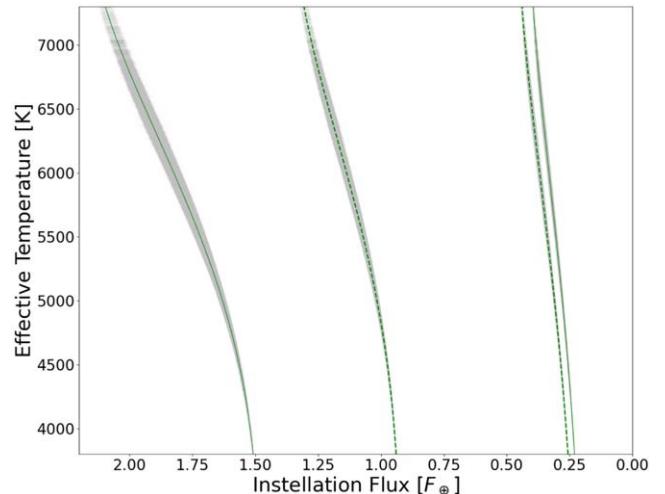


Figure 8. The uncertainty in the HZ boundaries due to uncertainty in stellar effective temperature. For every star, the inner and outer boundaries of the HZ are shown in green, with the 68% credible interval for each boundary shown in gray. The solid green line is the optimistic HZ, and the dashed green line is the conservative HZ.

inner optimistic HZ boundary, which has the widest error distribution in Figure 8. The median of the difference between the median HZ uncertainty and the HZ boundary without uncertainty is less than 0.002%, with a standard deviation less than 0.9%. Therefore, we do not believe that uncertainties in HZ boundaries resulting from stellar effective temperature uncertainties have a significant impact on occurrence rates.

4. Results

4.1. Inferring the Planet Population Model Parameters

For each choice of population differential rate model from Equation (5) and stellar population from Section 3.1, we determine the parameter vector θ with zero and constant extrapolated completeness, giving high and low bounds on the occurrence rates. These solutions are computed over the radius range $0.5 R_{\oplus} \leq r \leq 2.5 R_{\oplus}$ and the instellation flux range $0.2 I_{\oplus} \leq I \leq 2.2 I_{\oplus}$ using the hab and hab2 stellar populations described in Section 3.1. We perform these calculations both without and with input uncertainties in the planet radius, instellation flux, and T_{eff} shown in Figure 1. Examples of the resulting θ posterior distributions are shown in Figure 9 and the corresponding rate functions λ are shown in Figure 10. The solutions for models 1–3 for the hab and hab2 stellar populations computed using the Poisson likelihood method are given in Table 1, and those computed using ABC are given in Table 2. An example of the sampled planet population using input uncertainties is shown in Figure 11.

We compare models 1–3 using the Akaike information criterion (AIC), but the AIC does not indicate that one of the models is significantly more consistent with the data than another. The resulting relative likelihoods from the AIC analysis relative to model 1 are 0.31 for model 2 and 2.07 for model 3. Such low relative likelihood ratios are not considered compelling.

The F_0 parameter, giving the average number of planets per star in the solution domain (see Section 5), indicates that solutions using zero-completeness extrapolation (see Section 3.3.2) yield higher occurrence than those using constant-completeness extrapolation. This is because zero-completeness extrapolation induces larger completeness corrections. The zero-completeness extrapolation

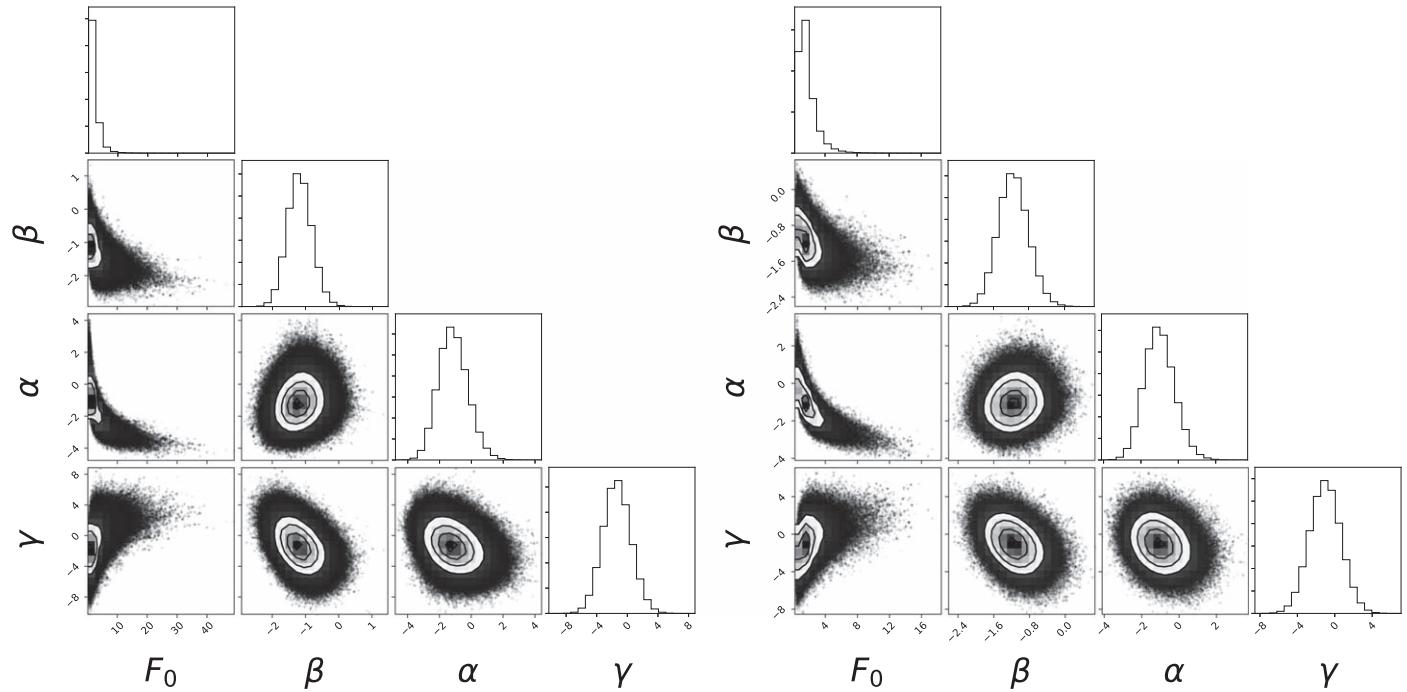


Figure 9. The posterior distributions of model 1 from Equation (5) for the hab2 stellar population and zero extrapolation completeness. Left: With input uncertainty. Right: Without input uncertainty.

solution provides an upper bound on the HZ occurrence rate, and the constant extrapolation solution provides a lower bound. Reality will be somewhere in between and is likely to be closer to the zero extrapolation case for hotter stars, which have lower completeness in their HZs.

The hab stellar population is a subset of the hab2 population, so one may expect that they give similar solutions. But the hab stellar population contains significant regions of extrapolated completeness and low-reliability planet candidates for the flux range considered in our solution (see Figure 1). While the hab2 population contains the same regions, hab2 also contains many cooler stars that host higher-reliability planet candidates. These stars provide a better constraint on the power laws we use to describe the population. The result is that the hab2 solution has significantly smaller uncertainties than the hab solution, as seen in Tables 1 and 2.

4.2. HZ Occurrence Rates

Table 3 gives η_{\oplus} , computed using the Poisson likelihood method, for the optimistic and conservative HZs for the hab and hab2 stellar populations and models 1–3. The low and high values correspond to the solutions using constant- and zero-completeness extrapolation, which bound the actual occurrence rates (see Section 3.3.2). We see the expected behavior of zero-completeness extrapolation leading to higher occurrence due to a larger completeness correction. Table 4 gives the same occurrence rates computed using the ABC method. The distributions of η_{\oplus} using these models and the Poisson likelihood method are shown in Figure 12. We see that for each model, when incorporating the input uncertainties, the hab and hab2 stellar populations yield consistent values of η_{\oplus} . Without using the input uncertainties the hab population yields consistently lower values for η_{\oplus} , though the difference is still within the 68% credible interval. Model 3 with hab2 gives generally higher occurrence rates than model 1. We also see

that the median occurrence rates are about $\sim 10\%$ higher when incorporating input uncertainties, qualitatively consistent with those of Shabram et al. (2020), who also see higher median occurrence rates when incorporating uncertainties. It is not clear what causes this increase in occurrence rates: on the one hand the sum of the inclusion probability, defined in Appendix C, for the planet candidates in Table 8 is 53.6, compared with the 54 planet candidates in the analysis without uncertainty, indicating that, on average, more planets exit the analysis than enter it when incorporating uncertainties. On the other hand, the sum of the inclusion probability times the planet radius is 106.6, compared with 105.0 for the planet candidates in the analysis without uncertainty, indicating that, on average, larger planets are entering the analysis. This may have an impact on the power-law model, leading to higher occurrence rates.

Table 5 gives the occurrence rates for a variety of planet radius and host star effective temperature ranges, computed using the hab2 stellar population and models 1–3. We see that the uncertainties for the 1.5–2.5 R_{\oplus} planets are significantly smaller than those for the 0.5–1.5 R_{\oplus} planets, indicating that the large uncertainties in η_{\oplus} are due to the small number of observed planets in the 0.5–1.5 R_{\oplus} range. The distributions of occurrence for the two bounding extrapolation types are shown in Figure 13. The difference between these two bounding cases is smaller than the uncertainties. Table 6 gives the 95% and 99% intervals for the 0.5–1.5 R_{\oplus} planets using model 1 computed with hab2.

Figure 14 shows the dependence of the HZ occurrence rate on effective temperature for models 1–3 based on the hab2 stellar population and for model 1 for the hab stellar population. For each model, the occurrence using zero and constant extrapolation is shown. Models 1 and 2 show a weak increase in occurrence with increasing effective temperature. Model 3 shows a stronger increase in occurrence with effective temperature, consistent with model 3’s assumption that the

With Uncertainty

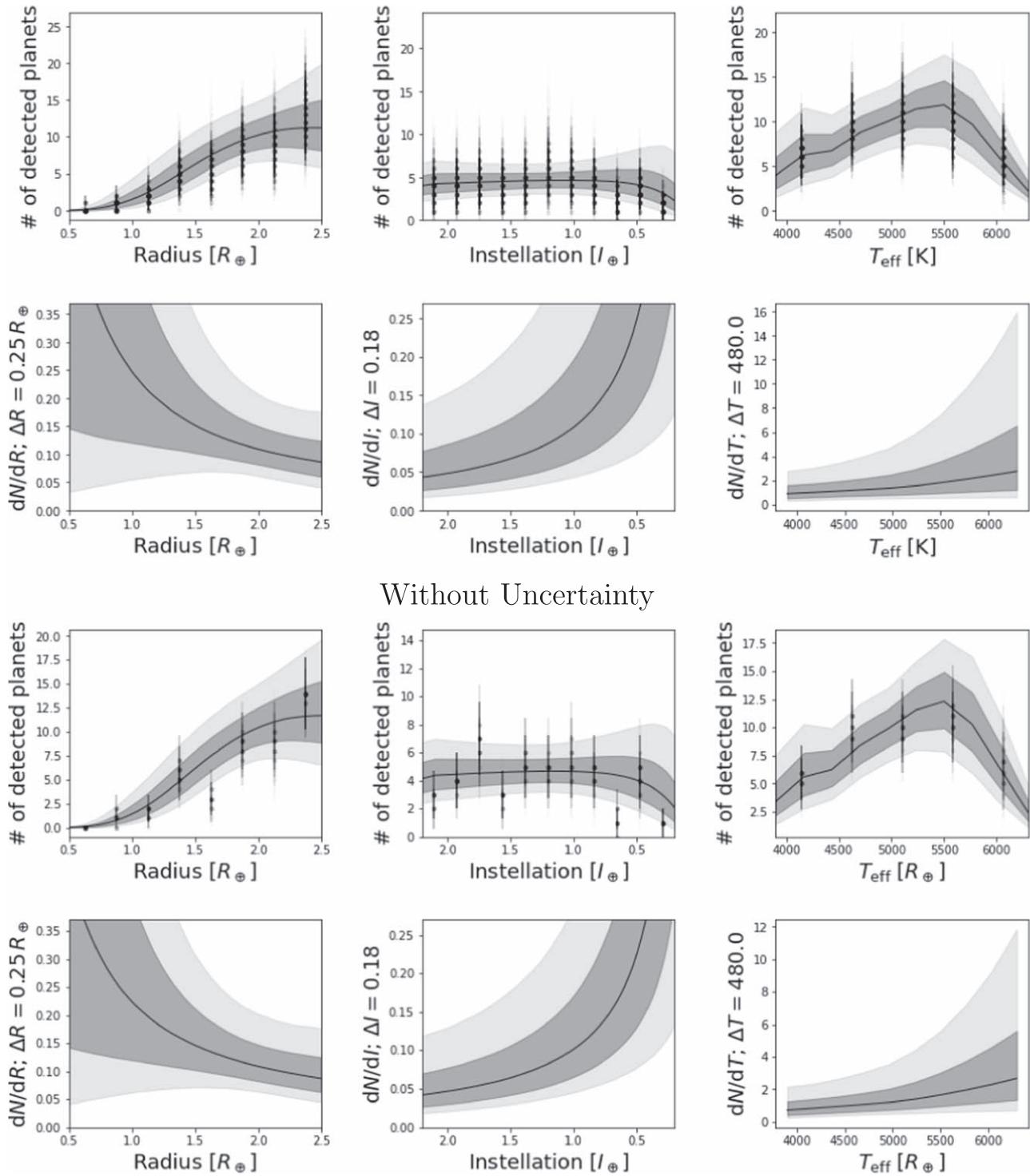


Figure 10. The marginalized population rate of model 1 from Equation (5) for the hab2 stellar population and zero extrapolation completeness, with and without incorporating uncertainties on planet radius, instellation flux, and host star effective temperature. The top row for each case shows the completeness-corrected population model compared with the observed planet population. The bottom row for each case shows the underlying population model. The dark gray regions are the 68% credible intervals, and the light gray regions are the 95% credible intervals.

only temperature dependence is the geometric effect described in Section 2.2. However, as shown in Figure 6, the difference between constant and zero extrapolated completeness is near zero for $T_{\text{eff}} \leq 4500$ K, so we would expect the difference in occurrence rates to be close to zero in that temperature range. This is true for models 1 and 2 but not true for model 3. We

take this as evidence that models 1 and 2 are correctly measuring a T_{eff} dependence beyond the geometric effect. We recognize that the statistical evidence for this T_{eff} dependence is not compelling, since the overlapping 68% credible intervals for the two completeness extrapolations would allow an occurrence rate independent of T_{eff} .

Table 1
Parameter Fits with 68% Confidence Limits for Models 1–3 from Equation (5) for the hab and hab2 Stellar Populations from Section 3.1, Computed with the Poisson Likelihood Method

	With Uncertainty		Without Uncertainty	
	Based on hab Stars Low Bound–High Bound	Based on hab2 Stars Low Bound–High Bound	Based on hab Stars Low Bound–High Bound	Based on hab2 Stars Low Bound–High Bound
F_0	$+1.08^{+1.56}_{-0.57} - +1.97^{+3.73}_{-1.17}$	$+1.11^{+0.88}_{-0.44} - +1.59^{+1.56}_{-0.7}$	$+0.70^{+0.77}_{-0.32} - +1.41^{+2.12}_{-0.77}$	$+1.02^{+0.66}_{-0.37} - +1.46^{+1.18}_{-0.59}$
α	$-1.05^{+1.41}_{-1.2} - -1.09^{+1.36}_{-1.18}$	$-1.08^{+0.94}_{-0.85} - -1.18^{+0.96}_{-0.87}$	$-0.29^{+1.39}_{-1.18} - -0.51^{+1.35}_{-1.15}$	$-0.96^{+0.83}_{-0.74} - -1.03^{+0.83}_{-0.77}$
β	$-0.56^{+0.48}_{-0.42} - -1.18^{+0.6}_{-0.56}$	$-0.84^{+0.32}_{-0.3} - -1.19^{+0.37}_{-0.36}$	$-0.43^{+0.46}_{-0.4} - -1.13^{+0.54}_{-0.5}$	$-0.78^{+0.3}_{-0.28} - -1.15^{+0.34}_{-0.33}$
γ	$-1.84^{+3.33}_{-3.39} - +0.91^{+3.87}_{-3.88}$	$-2.67^{+1.59}_{-1.57} - -1.38^{+1.84}_{-1.78}$	$-2.13^{+3.06}_{-3.13} - +0.25^{+3.39}_{-3.47}$	$-2.33^{+1.47}_{-1.46} - -1.03^{+1.66}_{-1.64}$
F_0	$+1.04^{+1.52}_{-0.55} - +1.96^{+3.72}_{-1.18}$	$+1.13^{+0.92}_{-0.46} - +1.50^{+1.43}_{-0.66}$	$+0.80^{+0.9}_{-0.38} - +1.33^{+1.98}_{-0.71}$	$+1.00^{+0.7}_{-0.38} - +1.43^{+1.21}_{-0.6}$
α	$-1.03^{+1.41}_{-1.23} - -1.05^{+1.45}_{-1.18}$	$-1.13^{+0.97}_{-0.86} - -1.12^{+0.94}_{-0.86}$	$-0.48^{+1.36}_{-1.17} - -0.47^{+1.33}_{-1.16}$	$-0.92^{+0.88}_{-0.78} - -1.01^{+0.88}_{-0.8}$
β	$-0.56^{+0.48}_{-0.42} - -1.20^{+0.6}_{-0.56}$	$-0.85^{+0.32}_{-0.3} - -1.18^{+0.37}_{-0.35}$	$-0.51^{+0.45}_{-0.4} - -1.09^{+0.55}_{-0.51}$	$-0.80^{+0.31}_{-0.28} - -1.18^{+0.34}_{-0.33}$
γ	$+2.60^{+3.56}_{-3.61} - +5.33^{+3.94}_{-4.02}$	$+1.19^{+1.64}_{-1.63} - +2.31^{+1.91}_{-1.85}$	$+2.03^{+3.1}_{-3.23} - +4.66^{+3.43}_{-3.51}$	$+1.26^{+1.54}_{-1.54} - +2.73^{+1.73}_{-1.71}$
F_0	$+1.13^{+1.52}_{-0.58} - +1.83^{+2.76}_{-1.0}$	$+1.41^{+1.14}_{-0.59} - +1.89^{+1.51}_{-0.78}$	$+0.89^{+0.94}_{-0.4} - +1.24^{+1.49}_{-0.6}$	$+1.25^{+0.93}_{-0.5} - +1.75^{+1.26}_{-0.7}$
α	$-1.08^{+1.39}_{-1.18} - -1.06^{+1.38}_{-1.18}$	$-1.37^{+0.91}_{-0.83} - -1.28^{+0.9}_{-0.82}$	$-0.60^{+1.3}_{-1.12} - -0.38^{+1.34}_{-1.18}$	$-1.21^{+0.88}_{-0.78} - -1.17^{+0.85}_{-0.76}$
β	$-0.56^{+0.48}_{-0.41} - -1.16^{+0.55}_{-0.51}$	$-0.89^{+0.32}_{-0.29} - -1.29^{+0.35}_{-0.32}$	$-0.49^{+0.45}_{-0.39} - -1.06^{+0.53}_{-0.48}$	$-0.83^{+0.3}_{-0.28} - -1.26^{+0.33}_{-0.31}$

Notes. The low and high bounds correspond to the constant- and zero-completeness extrapolation of Section 3.3.2. “With Uncertainty” means the planet candidate radius, instellation flux, and host star effective temperature uncertainties are taken into account.

Table 2
Parameter Fits with 68% Confidence Limits for Models 1–3 from Equation (5) for the Four Stellar Populations from Section 3.1, Computed with the ABC Method

	With Uncertainty		Without Uncertainty	
	Based on hab Stars Low Bound–High Bound	Based on hab2 Stars Low Bound—High Bound	Based on hab Stars Low Bound–High Bound	Based on hab2 Stars Low Bound–High Bound
F_0	$+1.18^{+0.95}_{-0.56} - +2.04^{+1.44}_{-0.99}$	$+1.17^{+0.78}_{-0.44} - +1.61^{+1.05}_{-0.65}$	$+0.73^{+0.54}_{-0.29} - +1.37^{+1.08}_{-0.61}$	$+0.94^{+0.45}_{-0.32} - +1.41^{+0.99}_{-0.57}$
α	$-1.14^{+1.02}_{-0.89} - -0.95^{+0.99}_{-0.86}$	$-1.14^{+0.75}_{-0.77} - -1.18^{+0.72}_{-0.67}$	$-0.17^{+1.19}_{-0.97} - -0.11^{+1.17}_{-0.88}$	$-0.71^{+0.67}_{-0.68} - -0.83^{+0.77}_{-0.74}$
β	$-0.69^{+0.41}_{-0.38} - -1.32^{+0.51}_{-0.44}$	$-0.90^{+0.31}_{-0.26} - -1.26^{+0.35}_{-0.31}$	$-0.67^{+0.38}_{-0.35} - -1.30^{+0.43}_{-0.41}$	$-0.89^{+0.25}_{-0.25} - -1.26^{+0.30}_{-0.32}$
γ	$-0.84^{+3.81}_{-4.11} - +2.16^{+3.91}_{-3.68}$	$-2.60^{+1.74}_{-1.84} - -1.14^{+2.15}_{-2.02}$	$-1.93^{+3.54}_{-3.37} - +1.82^{+3.94}_{-3.77}$	$-2.27^{+1.65}_{-1.71} - -0.78^{+2.11}_{-1.91}$
F_0	$+1.06^{+0.90}_{-0.48} - +1.88^{+1.38}_{-0.87}$	$+1.14^{+0.74}_{-0.43} - +1.66^{+1.26}_{-0.70}$	$+0.70^{+0.44}_{-0.28} - +1.22^{+0.87}_{-0.57}$	$+0.96^{+0.45}_{-0.34} - +1.35^{+0.85}_{-0.51}$
α	$-1.07^{+1.11}_{-0.93} - -0.96^{+0.90}_{-0.90}$	$-1.15^{+0.76}_{-0.76} - -1.25^{+0.80}_{-0.72}$	$-0.06^{+1.02}_{-0.88} - -0.09^{+1.07}_{-0.98}$	$-0.75^{+0.70}_{-0.71} - -0.82^{+0.76}_{-0.66}$
β	$-0.67^{+0.41}_{-0.35} - -1.30^{+0.45}_{-0.44}$	$-0.90^{+0.28}_{-0.27} - -1.24^{+0.33}_{-0.36}$	$-0.64^{+0.39}_{-0.36} - -1.19^{+0.43}_{-0.42}$	$-0.89^{+0.26}_{-0.23} - -1.27^{+0.32}_{-0.31}$
γ	$+3.09^{+3.64}_{-3.42} - +5.68^{+2.60}_{-3.42}$	$+1.34^{+1.95}_{-2.04} - +2.85^{+2.21}_{-2.24}$	$+2.72^{+3.19}_{-3.78} - +5.03^{+2.90}_{-3.58}$	$+1.44^{+1.68}_{-1.64} - +3.04^{+2.01}_{-1.97}$
F_0	$+1.20^{+0.93}_{-0.57} - +1.75^{+1.19}_{-0.77}$	$+1.57^{+0.93}_{-0.58} - +1.90^{+1.08}_{-0.69}$	$+0.82^{+0.56}_{-0.31} - +1.21^{+0.82}_{-0.49}$	$+1.25^{+0.69}_{-0.47} - +1.62^{+0.81}_{-0.59}$
α	$-1.16^{+1.14}_{-0.88} - -0.93^{+0.98}_{-0.81}$	$-1.54^{+0.68}_{-0.67} - -1.32^{+0.71}_{-0.65}$	$-0.27^{+1.02}_{-0.98} - -0.09^{+1.06}_{-0.99}$	$-1.14^{+0.76}_{-0.70} - -0.95^{+0.70}_{-0.67}$
β	$-0.71^{+0.41}_{-0.36} - -1.24^{+0.42}_{-0.41}$	$-0.97^{+0.27}_{-0.23} - -1.33^{+0.26}_{-0.27}$	$-0.66^{+0.32}_{-0.34} - -1.24^{+0.42}_{-0.39}$	$-0.96^{+0.25}_{-0.23} - -1.35^{+0.27}_{-0.26}$

Notes. The low and high bounds correspond to the constant- and zero-completeness extrapolation of Section 3.3.2. “With Uncertainty” means the planet candidate radius, instellation flux, and host star effective temperature uncertainties are taken into account.

An issue that arises with zero-completeness extrapolation (= high bound) is that, strictly speaking, planet candidates with orbital periods of >500 days are in a region of zero completeness around their host star and would not contribute to the Poisson likelihood (see Equation (16) in Appendix B). There is one such planet candidate in the hab2 stellar sample with reliability = 0.67. Performing our Poisson inference, we remove this planet with a period of >500 days for model 1 and incorporate input uncertainties, yielding an optimistic $\eta_{\oplus} = 0.70^{+1.01}_{-0.41}$, compared with $0.88^{+1.27}_{-0.51}$ (from Table 3) when including the planet. While this result is well within the 68% credible interval of the result with the planet included, removing this planet has a noticeable impact on the upper bound for the optimistic η_{\oplus} . However this planet is in fact

detected, implying that completeness is not zero for periods of >500 days, at least for this planet’s host star. If the actual completeness is very close to zero, a planet detection implies a large population. We therefore leave this planet in the analysis, thinking of “zero completeness” as a limit of the completeness going to zero when the HZ includes orbital periods of >500 days, summed or averaged over the stellar population for our computations.

5. Discussion

5.1. Instellation Flux versus Orbital Period

We choose to compute our occurrence rates as a function of instellation flux for two major reasons: this allows a more direct

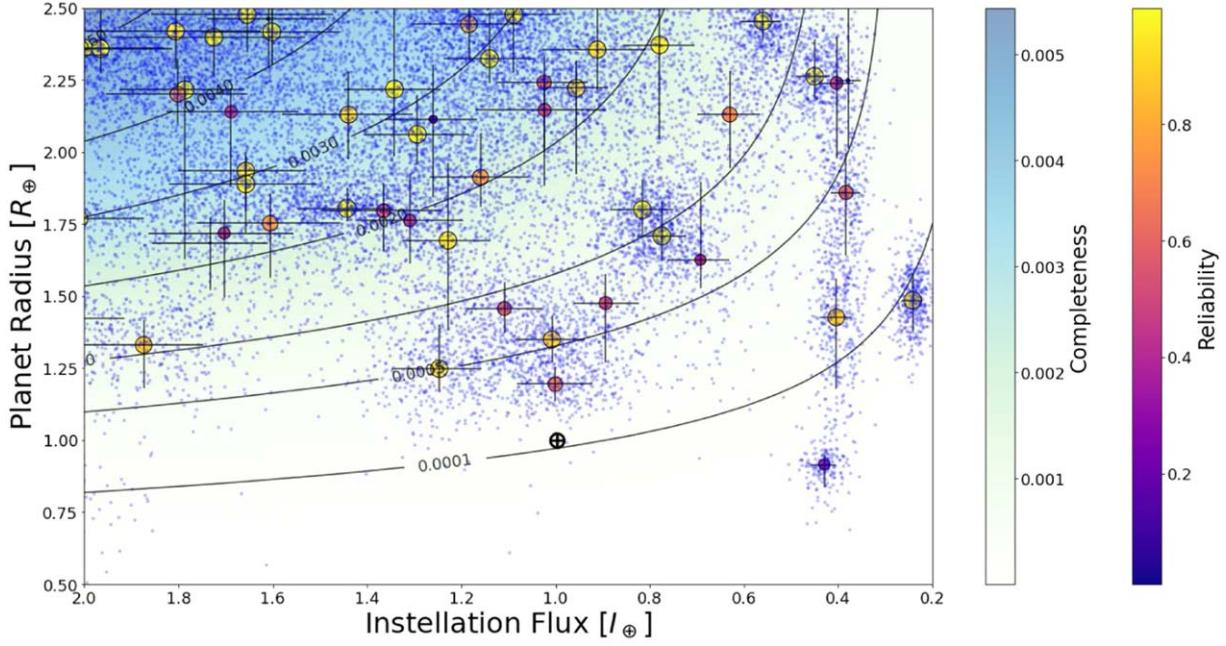


Figure 11. The union of the planet instellation fluxes and radii for the 400 realizations in the uncertainty run shown by blue dots superimposed on the lower panel of Figure 1 for the hab2 stellar population and model 1. See the caption of Figure 1 for an explanation of the other elements of the figure.

Table 3

η_{\oplus} , Computed Using Population Models Based on the hab and hab2 Stellar Populations, with and without Uncertainties for Models 1–3 and Using the Poisson Method

	With Uncertainty		Without Uncertainty	
	Based on hab Stars		Based on hab2 Stars	
	Low Bound–High Bound	Low Bound–High Bound	Low Bound–High Bound	Low Bound–High Bound
Conservative HZ				
Model 1	$0.30^{+0.69}_{-0.21} - 0.54^{+1.46}_{-0.39}$	$0.37^{+0.48}_{-0.21} - 0.60^{+0.90}_{-0.36}$	$0.15^{+0.32}_{-0.11} - 0.34^{+0.83}_{-0.25}$	$0.34^{+0.37}_{-0.18} - 0.54^{+0.69}_{-0.30}$
Model 2	$0.28^{+0.66}_{-0.20} - 0.53^{+1.46}_{-0.39}$	$0.39^{+0.51}_{-0.23} - 0.56^{+0.83}_{-0.33}$	$0.19^{+0.39}_{-0.13} - 0.32^{+0.78}_{-0.23}$	$0.33^{+0.38}_{-0.18} - 0.53^{+0.70}_{-0.30}$
Model 3	$0.31^{+0.69}_{-0.22} - 0.55^{+1.22}_{-0.39}$	$0.59^{+0.74}_{-0.34} - 0.79^{+0.95}_{-0.44}$	$0.21^{+0.42}_{-0.15} - 0.30^{+0.64}_{-0.21}$	$0.50^{+0.60}_{-0.28} - 0.72^{+0.80}_{-0.39}$
Optimistic HZ				
Model 1	$0.50^{+1.09}_{-0.35} - 0.80^{+2.07}_{-0.57}$	$0.58^{+0.73}_{-0.33} - 0.88^{+1.27}_{-0.51}$	$0.26^{+0.52}_{-0.18} - 0.51^{+1.17}_{-0.36}$	$0.54^{+0.56}_{-0.28} - 0.80^{+0.99}_{-0.44}$
Model 2	$0.48^{+1.06}_{-0.33} - 0.78^{+2.05}_{-0.58}$	$0.61^{+0.77}_{-0.35} - 0.83^{+1.17}_{-0.48}$	$0.32^{+0.62}_{-0.22} - 0.48^{+1.11}_{-0.33}$	$0.52^{+0.58}_{-0.28} - 0.78^{+1.00}_{-0.44}$
Model 3	$0.53^{+1.10}_{-0.37} - 0.81^{+1.73}_{-0.57}$	$0.92^{+1.12}_{-0.52} - 1.14^{+1.35}_{-0.63}$	$0.36^{+0.68}_{-0.25} - 0.46^{+0.92}_{-0.31}$	$0.79^{+0.92}_{-0.44} - 1.04^{+1.14}_{-0.56}$

Notes. The low and high bounds correspond to the constant- and zero-completeness extrapolation of Section 3.3.2. “With Uncertainty” means the planet candidate radius, instellation flux, and host star effective temperature uncertainties are taken into account.

Table 4

η_{\oplus} , Computed Using Population Models Based on the hab and hab2 Stellar Populations, with and without Uncertainties for Models 1–3 and Using the ABC Method

	With Uncertainty		Without Uncertainty	
	Based on hab Stars		Based on hab2 Stars	
	Low Bound–High Bound	Low Bound–High Bound	Low Bound–High Bound	Low Bound–High Bound
Conservative HZ				
Model 1	$0.33^{+0.46}_{-0.20} - 0.50^{+0.69}_{-0.31}$	$0.40^{+0.45}_{-0.21} - 0.61^{+0.63}_{-0.33}$	$0.16^{+0.24}_{-0.10} - 0.26^{+0.44}_{-0.16}$	$0.29^{+0.28}_{-0.15} - 0.49^{+0.61}_{-0.27}$
Model 2	$0.30^{+0.41}_{-0.19} - 0.52^{+0.63}_{-0.31}$	$0.40^{+0.43}_{-0.21} - 0.64^{+0.73}_{-0.35}$	$0.14^{+0.19}_{-0.09} - 0.25^{+0.36}_{-0.15}$	$0.31^{+0.28}_{-0.15} - 0.47^{+0.50}_{-0.24}$
Model 3	$0.35^{+0.47}_{-0.22} - 0.50^{+0.56}_{-0.29}$	$0.69^{+0.64}_{-0.35} - 0.81^{+0.69}_{-0.40}$	$0.18^{+0.27}_{-0.11} - 0.27^{+0.36}_{-0.16}$	$0.50^{+0.48}_{-0.26} - 0.62^{+0.55}_{-0.31}$
Optimistic HZ				
Model 1	$0.54^{+0.72}_{-0.33} - 0.73^{+0.98}_{-0.45}$	$0.62^{+0.66}_{-0.32} - 0.89^{+0.89}_{-0.47}$	$0.26^{+0.37}_{-0.16} - 0.39^{+0.62}_{-0.23}$	$0.45^{+0.42}_{-0.23} - 0.71^{+0.86}_{-0.38}$
Model 2	$0.48^{+0.66}_{-0.30} - 0.75^{+0.90}_{-0.44}$	$0.62^{+0.63}_{-0.32} - 0.92^{+1.02}_{-0.49}$	$0.24^{+0.30}_{-0.14} - 0.37^{+0.52}_{-0.22}$	$0.47^{+0.41}_{-0.23} - 0.68^{+0.70}_{-0.34}$
Model 3	$0.56^{+0.72}_{-0.36} - 0.73^{+0.81}_{-0.41}$	$1.05^{+0.96}_{-0.52} - 1.15^{+0.99}_{-0.57}$	$0.30^{+0.43}_{-0.18} - 0.39^{+0.51}_{-0.23}$	$0.75^{+0.72}_{-0.39} - 0.89^{+0.78}_{-0.44}$

Notes. The low and high bounds correspond to the constant- and zero-completeness extrapolation of Section 3.3.2. “With Uncertainty” means the planet candidate radius, instellation flux, and host star effective temperature uncertainties are taken into account.

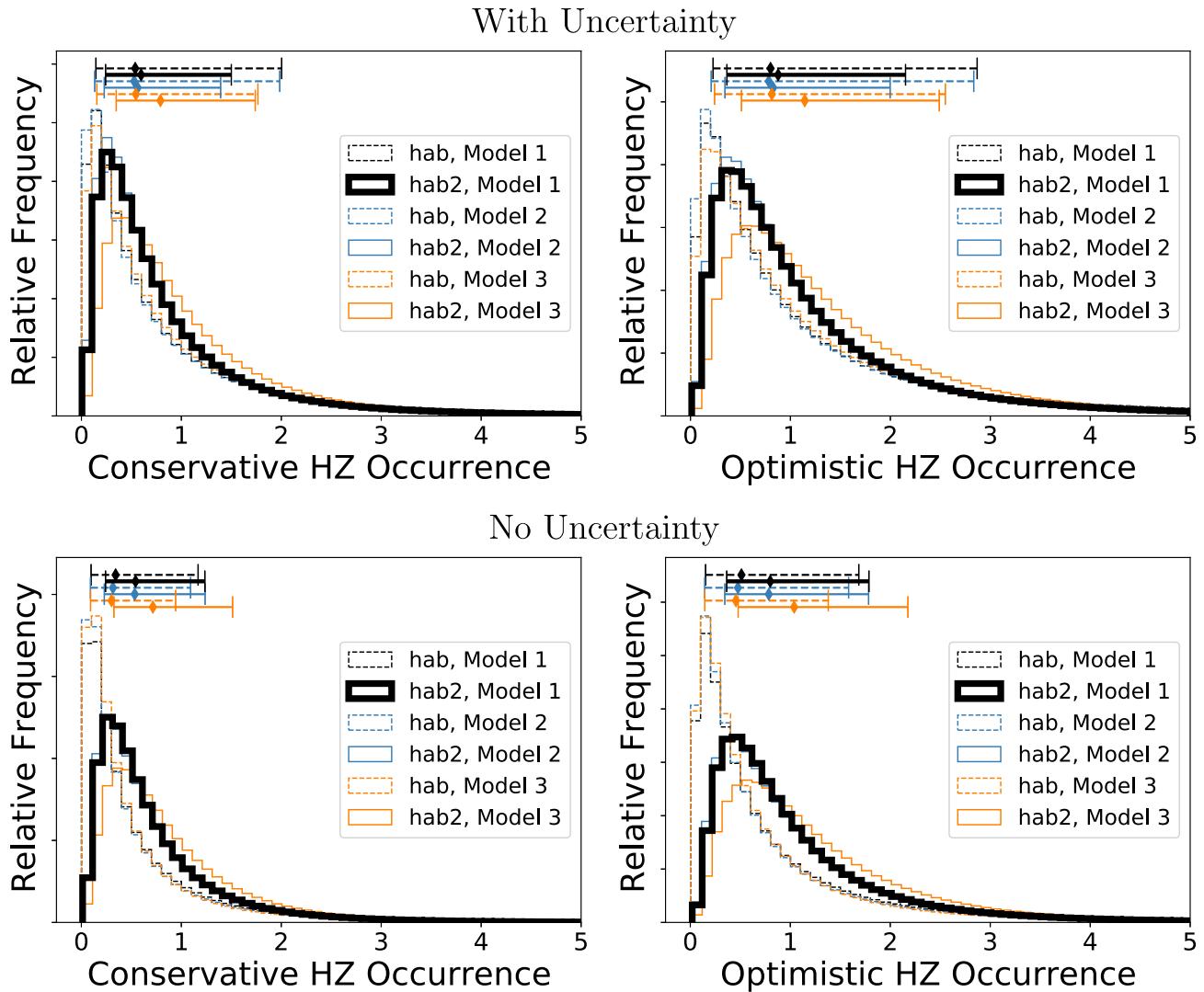


Figure 12. The distribution of η_{\oplus} for population models computed using the hab (dashed lines) and hab2 (solid lines) stellar populations for the three models in Equation (5), demonstrating that we get similar results from models 1 and 2 for both stellar populations. Medians and 68% credible intervals are shown above the distributions. The result from the hab2 population including effective temperature dependence is shown with a thick black line. Top: Incorporating the uncertainty on planet radius, stellar instellation, and stellar effective temperature. Bottom: Without incorporating uncertainties. Left: The conservative HZ. Right: The optimistic HZ.

characterization of each star’s HZ, and it allows a more constrained extrapolation to longer orbital periods than working directly in terms of orbital period.

By considering instellation flux, we can measure HZ occurrence by including observed planets from the HZ of their host stars, which is not possible across a wide range of stellar temperatures when using orbital period (see Figure 2). Instellation flux also allows a direct measurement of the impact of uncertainties in stellar effective temperature and planetary instellation flux.

The HZ of most G and F stars includes orbital periods that are beyond those periods well covered by Kepler observations (see Figures 1 and 2), requiring significant extrapolation of orbital period-based planet population models to long orbital periods. Such extrapolation is poorly known or constrained, leading to possible significant and unbounded inaccuracies. In instellation flux, however, there is planet data throughout those regions of our domain of analysis that have reasonable completeness (see Figure 1), so no extrapolation in instellation flux is required. In this sense, replacing orbital period with instellation flux (determined by the orbital period for each star) moves the problem of

extrapolating the population model to longer orbital periods to extrapolating the completeness data to lower instellation flux. In Section 3.3.2 we argue that completeness, on average, decreases monotonically with decreasing instellation flux. This allows us to bound the extrapolated completeness between no decrease at all (constant extrapolation) and zero completeness for instellation flux for orbital periods beyond the 500 day limit where completeness is measured. We then perform our analysis for the two extrapolation cases and find that their difference in HZ occurrence rates is small relative to our uncertainties. In this way we provide a bounded estimate of HZ occurrence rates using instellation flux, rather than the unbounded extrapolation resulting from using orbital period.

5.2. Comparing the Stellar Population and Rate Function Models

Our approach to measuring η_{\oplus} is to compute the planet population rate model $\lambda(r, I, T_{\text{eff}}) \equiv d^2f(r, I, T_{\text{eff}})/dr dI$, integrate over r and I , and average over T_{eff} . We compute the population model λ using the hab and hab2 stellar populations (Section 3.1) to measure the sensitivity of our results to stellar

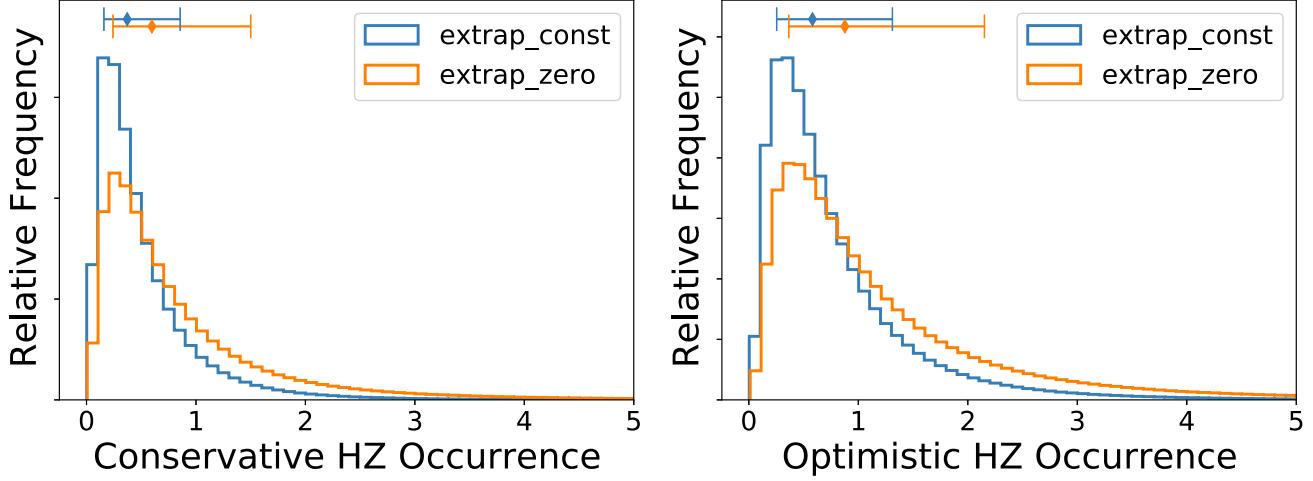


Figure 13. The distribution of η_{\oplus} for the two bounding extrapolation cases, computed with model 1 and hab2 with input uncertainties. Left: The conservative HZ. Right: The optimistic HZ.

Table 5

Number of Habitable Zone Planets per Star for Various Ranges of Planet Radii and Host Star Effective Temperatures, Computed Using the Population Model Based on the hab2 Stellar Population with the Poisson Likelihood Method and Incorporating Uncertainties in Planet Radius, Instellation Flux, and Host Star Effective Temperature

Planet Radius	4800–6300 K		3900–6300 K		3900–5300 K (K Dwarfs)		5300–6000 K (G Dwarfs)	
	Low Bound	High Bound	Low Bound	High Bound	Low Bound	High Bound	Low Bound	High Bound
Conservative HZ								
0.5–1.5 R_{\oplus}	0.37^{+0.48}_{-0.21} – 0.60^{+0.90}_{-0.36}		0.35 ^{+0.43} _{-0.19} – 0.50 ^{+0.73} _{-0.29}		0.32 ^{+0.35} _{-0.17} – 0.42 ^{+0.50} _{-0.23}		0.38 ^{+0.50} _{-0.22} – 0.63 ^{+0.94} _{-0.38}	
1.5–2.5 R_{\oplus}	0.16 ^{+0.07} _{-0.05} – 0.24 ^{+0.14} _{-0.08}		0.15 ^{+0.06} _{-0.05} – 0.20 ^{+0.12} _{-0.07}		0.14 ^{+0.05} _{-0.04} – 0.17 ^{+0.07} _{-0.06}		0.17 ^{+0.06} _{-0.05} – 0.26 ^{+0.13} _{-0.09}	
0.5–2.5 R_{\oplus}	0.54 ^{+0.52} _{-0.24} – 0.85 ^{+0.99} _{-0.42}		0.51 ^{+0.46} _{-0.22} – 0.71 ^{+0.80} _{-0.34}		0.46 ^{+0.37} _{-0.19} – 0.60 ^{+0.52} _{-0.26}		0.56 ^{+0.53} _{-0.25} – 0.90 ^{+1.01} _{-0.45}	
0.5–1.5 R_{\oplus}	0.39^{+0.51}_{-0.23} – 0.56^{+0.83}_{-0.33}		0.36 ^{+0.44} _{-0.20} – 0.47 ^{+0.67} _{-0.27}		0.33 ^{+0.36} _{-0.18} – 0.40 ^{+0.47} _{-0.22}		0.40 ^{+0.53} _{-0.23} – 0.59 ^{+0.86} _{-0.35}	
1.5–2.5 R_{\oplus}	0.16 ^{+0.06} _{-0.05} – 0.24 ^{+0.13} _{-0.08}		0.15 ^{+0.06} _{-0.05} – 0.20 ^{+0.11} _{-0.07}		0.14 ^{+0.05} _{-0.04} – 0.17 ^{+0.07} _{-0.06}		0.17 ^{+0.06} _{-0.05} – 0.25 ^{+0.12} _{-0.08}	
0.5–2.5 R_{\oplus}	0.56 ^{+0.54} _{-0.26} – 0.81 ^{+0.91} _{-0.39}		0.51 ^{+0.47} _{-0.23} – 0.68 ^{+0.75} _{-0.32}		0.47 ^{+0.38} _{-0.20} – 0.58 ^{+0.50} _{-0.25}		0.57 ^{+0.56} _{-0.27} – 0.85 ^{+0.94} _{-0.42}	
0.5–1.5 R_{\oplus}	0.59^{+0.74}_{-0.34} – 0.79^{+0.95}_{-0.44}		0.43 ^{+0.63} _{-0.26} – 0.59 ^{+0.82} _{-0.35}		0.31 ^{+0.37} _{-0.17} – 0.43 ^{+0.51} _{-0.24}		0.64 ^{+0.72} _{-0.35} – 0.85 ^{+0.93} _{-0.46}	
1.5–2.5 R_{\oplus}	0.20 ^{+0.11} _{-0.07} – 0.29 ^{+0.15} _{-0.10}		0.15 ^{+0.13} _{-0.06} – 0.22 ^{+0.17} _{-0.09}		0.11 ^{+0.05} _{-0.04} – 0.16 ^{+0.07} _{-0.05}		0.22 ^{+0.07} _{-0.06} – 0.32 ^{+0.11} _{-0.09}	
0.5–2.5 R_{\oplus}	0.81 ^{+0.79} _{-0.39} – 1.11 ^{+1.02} _{-0.52}		0.60 ^{+0.71} _{-0.32} – 0.84 ^{+0.92} _{-0.44}		0.42 ^{+0.39} _{-0.20} – 0.60 ^{+0.54} _{-0.28}		0.87 ^{+0.75} _{-0.38} – 1.18 ^{+0.97} _{-0.51}	
Optimistic HZ								
0.5–1.5 R_{\oplus}	0.58^{+0.73}_{-0.33} – 0.88^{+1.28}_{-0.51}		0.54 ^{+0.64} _{-0.29} – 0.73 ^{+1.02} _{-0.41}		0.49 ^{+0.51} _{-0.26} – 0.60 ^{+0.69} _{-0.33}		0.60 ^{+0.75} _{-0.34} – 0.93 ^{+1.32} _{-0.55}	
1.5–2.5 R_{\oplus}	0.25 ^{+0.09} _{-0.06} – 0.35 ^{+0.19} _{-0.11}		0.23 ^{+0.09} _{-0.06} – 0.29 ^{+0.17} _{-0.10}		0.21 ^{+0.07} _{-0.06} – 0.25 ^{+0.09} _{-0.08}		0.26 ^{+0.09} _{-0.07} – 0.37 ^{+0.17} _{-0.11}	
0.5–2.5 R_{\oplus}	0.84 ^{+0.78} _{-0.37} – 1.25 ^{+1.39} _{-0.59}		0.78 ^{+0.68} _{-0.33} – 1.04 ^{+1.13} _{-0.47}		0.71 ^{+0.53} _{-0.28} – 0.80 ^{+0.72} _{-0.36}		0.87 ^{+0.79} _{-0.38} – 1.33 ^{+1.42} _{-0.63}	
0.5–1.5 R_{\oplus}	0.61^{+0.77}_{-0.35} – 0.83^{+1.17}_{-0.48}		0.55 ^{+0.66} _{-0.31} – 0.68 ^{+0.95} _{-0.39}		0.50 ^{+0.54} _{-0.27} – 0.57 ^{+0.66} _{-0.31}		0.63 ^{+0.79} _{-0.36} – 0.87 ^{+1.22} _{-0.51}	
1.5–2.5 R_{\oplus}	0.25 ^{+0.09} _{-0.06} – 0.34 ^{+0.17} _{-0.10}		0.23 ^{+0.09} _{-0.07} – 0.29 ^{+0.16} _{-0.10}		0.21 ^{+0.07} _{-0.06} – 0.25 ^{+0.09} _{-0.08}		0.26 ^{+0.08} _{-0.07} – 0.37 ^{+0.16} _{-0.11}	
0.5–2.5 R_{\oplus}	0.87 ^{+0.81} _{-0.39} – 1.19 ^{+1.28} _{-0.56}		0.79 ^{+0.70} _{-0.34} – 0.99 ^{+1.05} _{-0.45}		0.72 ^{+0.55} _{-0.29} – 0.84 ^{+0.69} _{-0.35}		0.90 ^{+0.83} _{-0.40} – 1.25 ^{+1.32} _{-0.59}	
0.5–1.5 R_{\oplus}	0.92^{+1.12}_{-0.52} – 1.14^{+1.35}_{-0.63}		0.67 ^{+0.96} _{-0.40} – 0.85 ^{+1.17} _{-0.50}		0.46 ^{+0.55} _{-0.26} – 0.61 ^{+0.70} _{-0.33}		1.00 ^{+1.08} _{-0.54} – 1.23 ^{+1.32} _{-0.65}	
1.5–2.5 R_{\oplus}	0.31 ^{+0.16} _{-0.11} – 0.42 ^{+0.21} _{-0.15}		0.22 ^{+0.20} _{-0.10} – 0.31 ^{+0.25} _{-0.13}		0.16 ^{+0.07} _{-0.05} – 0.23 ^{+0.10} _{-0.07}		0.34 ^{+0.10} _{-0.08} – 0.46 ^{+0.14} _{-0.12}	
0.5–2.5 R_{\oplus}	1.26 ^{+1.19} _{-0.59} – 1.60 ^{+1.43} _{-0.73}		0.92 ^{+1.07} _{-0.49} – 1.20 ^{+1.30} _{-0.62}		0.63 ^{+0.58} _{-0.29} – 0.85 ^{+0.75} _{-0.38}		1.35 ^{+1.11} _{-0.57} – 1.71 ^{+1.36} _{-0.71}	

Notes. The η_{\oplus} values for model 1 are shown in boldface. The low and high bounds correspond to the constant- and zero-completeness extrapolation of Section 3.3.2. As explained in Section 5.2 we recommend model 1 as the baseline model. Results from the other models are included for comparison.

type, and we consider several possible functional forms for λ (Equation (5)).

5.2.1. Comparing Population Rate Function Models

We believe that we have detected a weak dependence of HZ occurrence on host star effective temperature T_{eff} , with hotter

stars having slightly higher HZ occurrence. Model 2 ($F_0 r^{\alpha} I^{\beta} T_{\text{eff}}^{\gamma}$), which directly measures T_{eff} dependence as a power law with exponent γ , indicates a weak T_{eff} dependence for the zero-completeness extrapolation case, though in the constant extrapolation case γ includes 0 in the 68% credible interval (see Table 1). This is somewhat remarkable given that,

Table 6

Credible Intervals of the Upper and Lower Bounds on HZ Occurrence for Model 1 Computed Using the Population Model Based on the hab2 Stellar Population (see Section 1) and Accounting for Input Uncertainty

	Low	High	Total
95% Credible Interval			
η_{\oplus}^C	[0.07, 1.91]	[0.10, 3.77]	[0.07, 3.77]
η_{\oplus}^O	[0.11, 2.88]	[0.16, 5.29]	[0.11, 5.29]
$\eta_{\oplus,G}^C$	[0.07, 1.92]	[0.10, 3.76]	[0.07, 3.76]
$\eta_{\oplus,G}^O$	[0.11, 2.90]	[0.16, 5.26]	[0.11, 5.26]
$\eta_{\oplus,K}^C$	[0.07, 1.34]	[0.09, 1.92]	[0.07, 1.92]
$\eta_{\oplus,K}^O$	[0.11, 1.96]	[0.13, 2.66]	[0.11, 2.66]
99% Credible Interval			
η_{\oplus}^C	[0.04, 3.19]	[0.06, 6.91]	[0.04, 6.91]
η_{\oplus}^O	[0.06, 4.76]	[0.09, 9.58]	[0.06, 9.58]
$\eta_{\oplus,G}^C$	[0.04, 3.13]	[0.06, 6.57]	[0.04, 6.57]
$\eta_{\oplus,G}^O$	[0.06, 4.65]	[0.10, 9.09]	[0.06, 9.09]
$\eta_{\oplus,K}^C$	[0.04, 2.06]	[0.05, 3.06]	[0.04, 3.06]
$\eta_{\oplus,K}^O$	[0.07, 2.97]	[0.08, 4.20]	[0.07, 4.20]

Note. See Section 1.3 for the definitions of the different types of η_{\oplus} .

as discussed in Section 2.2, the size of the HZ grows as at least T_{eff}^3 . This is consistent with model 1 ($F_0 r^\alpha I^\beta T_{\text{eff}}^\gamma g(T_{\text{eff}})$), which includes a fixed $g(T_{\text{eff}})$ term from Equation (4), reflecting the increase in size of the HZ with increasing temperature, and an additional T_{eff}^γ power law to capture any additional T_{eff} . In Table 1 we see that model 1 yields a very low or negative value for γ , consistent with the weak direct detection of T_{eff} dependence in model 2. The consistency between models 1 and 2 is further indicated by the fact that they yield very similar occurrence rates, as shown in Tables 3–5, as well as in Figure 14.

Model 3 ($F_0 r^\alpha I^\beta g(T_{\text{eff}})$) assumes that the T_{eff} dependence of HZ occurrence is entirely due to the increase in size of the HZ with increasing T_{eff} . When averaged over our η_{\oplus} effective temperature range of 4800–6300 K, model 3 yields somewhat higher occurrence rates than models 2 and 4 (see Tables 3 and 4 and Figure 12).

Models 1 and 2 have the expected behavior of the high and low bounds converging for cooler stars (see Figure 14), consistent with the extrapolation options coinciding for these stars (see Figure 6). Model 3 ($\lambda_3 = F_0 C_3 r^\alpha I^\beta g(T)$) does not have this behavior, but model 3's fixed T_{eff} dependence does not allow such a convergence.

Because models 1 and 2 detect a weaker T_{eff} dependence than would be expected due to the larger HZ for hotter stars (Equation (4)) if planets were uniformly distributed, we do not believe that model 3 is the best model for the data.

Models 1 and 2 yield essentially the same HZ occurrence results, but model 1 separates the geometric effect from intrinsic T_{eff} dependence. We therefore emphasize model 1, but model 2 provides a direct measure of the total T_{eff} dependence.

5.2.2. Comparing the hab and hab2 Stellar Populations

Without input uncertainties, the hab and hab2 stellar populations yield interestingly different values for η_{\oplus} . However, Tables 1 and 2 show that the F_0 parameter fits for hab have significantly larger relative uncertainties than those for hab2, with hab having $\approx 200\%$ positive uncertainties compared

with the $\approx 100\%$ positive uncertainties for the hab2 fits. In addition, the effective temperature exponent γ has larger absolute uncertainties for hab than for hab2. These larger uncertainties propagate to larger relative uncertainties in the occurrence rates in Tables 3, 4, and 5. This can also be seen by comparing hab and hab2 for model 1, as shown in Figure 14. These large uncertainties result in the differences between hab and hab2 being less than the 68% credible interval. With input uncertainties, the results for the hab and hab2 stellar populations are more consistent, being well inside the 68% credible interval, but hab still has larger relative uncertainties.

We believe the larger uncertainties for hab relative to hab2 are due to the hab2 population being less well covered by the Kepler observations than hab. A larger fraction of planet candidates for stars in the hab effective temperature range of 4800 K–6300 K are in a region of lower completeness and reliability (Figure 7) and have poorer observational coverage (Figure 1). The hab2 population, with an effective temperature range of 3900–6300 K, includes regions with better observational coverage and more reliable planet detections.

Basing our occurrence estimates on hab2 covers the full range of K stars without extrapolation, allowing us to produce, for example, GK, G, or K HZ occurrence rates using the same population model. This avoids possible ambiguities that may be due to different population models. Finally, when considering T_{eff} uncertainties, there are several planets close to the lower hab boundary at 4800 K, which leads to a larger impact of T_{eff} uncertainties on the population model because those planets will be in some uncertainty realizations and not in others (see Figure 1). In contrast, the lower T_{eff} boundary for hab2 is outside the 68% credible interval for all detected planets. Therefore, although the hab population exactly matches our effective temperature range for η_{\oplus} , we prefer models computed using the hab2 population.

To summarize, we adopt model 1 based on hab2 for our primary reported result, but we also provide the results for models 1–3 and the hab stellar population.

5.3. Computing η_{\oplus}

We find reasonable consistency in η_{\oplus} across the models for both the hab and hab2 stellar populations as shown in Tables 3 and 4. Table 5 gives the occurrence rates for several planet radius and stellar effective temperature ranges, using the hab2 population model. The uncertainties reflecting our 68% credible intervals for our η_{\oplus} , counting HZ planets with radii of 0.5 – $1.5 R_{\oplus}$, are large, with positive uncertainties being nearly 150% of the value when using input uncertainties. Comparing occurrence rate uncertainties with and without input uncertainties in Tables 3 and 4, we see that the bulk of the uncertainties occur without input uncertainties, while using input uncertainties increases the output uncertainty by nearly 20%. The much smaller uncertainties for larger planets ($1.5 R_{\oplus} \leq r \leq 2.5 R_{\oplus}$) in Table 5 suggest the large uncertainties in η_{\oplus} are driven by the very small number of detections in the η_{\oplus} range, combined with the very low completeness (see Figure 1). Low completeness will cause large completeness corrections, which will magnify the Poisson uncertainty due to few planet detections. There are more planets with larger radii that are in a regime of higher completeness, resulting in lower uncertainties for larger planet occurrence rates.

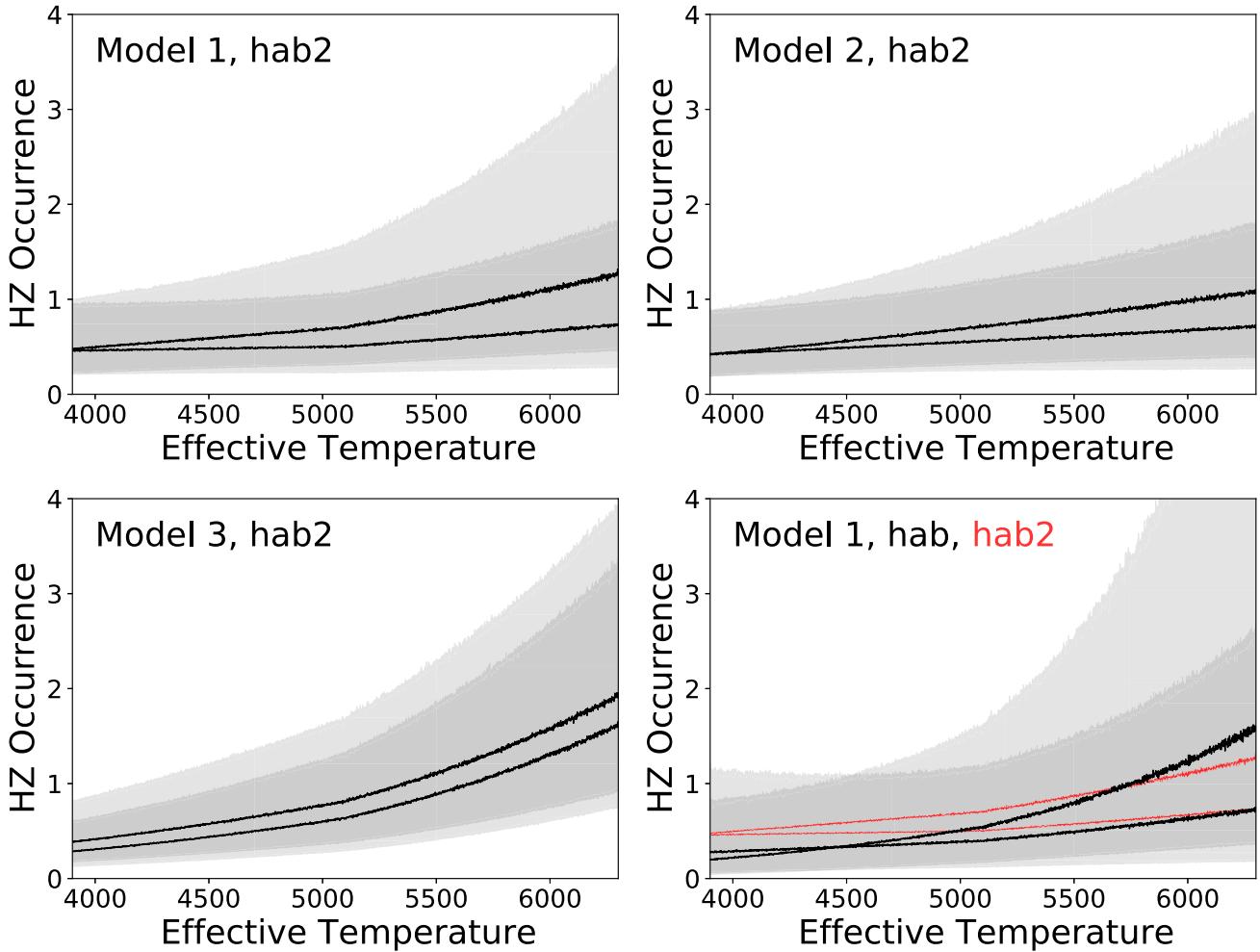


Figure 14. Optimistic HZ occurrence rate for planets with radii between 0.5 and $1.5 R_{\oplus}$ as a function of host star effective temperature. η_{\oplus} is the average over the temperature range $4800 \text{ K} \leq T_{\text{eff}} \leq 6300 \text{ K}$. The black lines show the median occurrence rate when using zero-completeness extrapolation (upper line) and constant-completeness extrapolation (lower line). The gray areas show the 68% confidence limits for the two completeness extrapolation cases, and the darker gray areas are the overlap of the 68% confidence regions. Top left: Model 1 based on hab2. Top right: Model 2 based on hab2. Bottom left: Model 3 based on hab2. Bottom right: Model 1 based on hab, with the medians from model 1 based on hab2 in red.

5.4. Implications of η_{\oplus}

Estimates of η_{\oplus} are useful in calculating exoEarth yields from direct-imaging missions, such as in the flagship concept studies for LUVOIR and HabEx. These mission studies assumed an occurrence rate of $\eta_{\oplus} = 0.24^{+0.46}_{-0.16}$ for Sun-like stars based on the NASA ExoPAG SAG13 meta-analysis of Kepler data (Kopparapu et al. 2018). The expected exoEarth candidate yields from the mission study reports are 54^{+61}_{-34} for LUVOIR-A (15 m), 28^{+30}_{-17} for LUVOIR-B (8 m), and 8 for the HabEx 4 m baseline configuration, which combines a more traditional coronagraphic starlight suppression system with a formation-flying starshade occulter. Table 5 provides the η_{\oplus} values based on the three models for G (Sun-like) and K dwarfs. If we assume the range of $\eta_{\oplus,G}$ from the conservative to the optimistic HZ from Table 5 for planets of $0.5\text{--}1.5 R_{\oplus}$ from the “low” end, say, for model 1, $\eta_{\oplus,G}$ would be between $0.38^{+0.50}_{-0.22}$ and $0.60^{+0.75}_{-0.34}$. While these η_{\oplus} values appear to be larger than the $0.24^{+0.46}_{-0.16}$ occurrence rate assumed by the mission studies, it should be noted that these studies adopted radius range of 0.82 to $1.4 R_{\oplus}$, and a lower limit of $0.8 \times a^{-0.5}$, where a is the HZ-corrected semi-major axis.

This is slightly a smaller HZ region and lower radius than the one used in our study. As a result, it is possible that we might be agreeing with their assumed η_{\oplus} value if we use the same bounding boxes. Computing the conservative HZ as described in Section 3.5 but replacing the planet radius range with $0.82 \leq r \leq 1.4 R_{\oplus}$ gives a lower bound of $0.18^{+0.16}_{-0.09}$ and an upper bound of $0.28^{+0.30}_{-0.14}$, nicely bracketing the value assumed in mission studies.

η_{\oplus} can also be used to estimate, on average, the nearest HZ planet around a G and K dwarf assuming the planets are distributed randomly. Within the solar neighborhood, the stellar number density ranges from 0.0033 to 0.0038 pc^{-3} for G dwarfs and 0.0105 to 0.0153 pc^{-3} for K dwarfs (Mamajek & Hillenbrand 2008; Kirkpatrick et al. 2012).⁵³ For G dwarfs, multiplying with the conservative ($0.38^{+0.50}_{-0.22}$) model 1 “low” end of the $\eta_{\oplus,G}$ values (i.e., the number of planets per star), we get between $0.0013^{+0.0016}_{-0.0007}$ and $0.0014^{+0.0019}_{-0.0008}$ HZ planets pc^{-3} . The nearest HZ planet around a G dwarf would then be expected to be at a distance of $d = (3/(4\pi \times n_p))^{1/3}$, where n_p is the planet number density in pc^{-3} . Substituting, we get a d

⁵³ http://www.pas.rochester.edu/~emamajek/memo_star_dens.html

between $5.9^{+1.76}_{-1.26}$ pc and $5.5^{+1.83}_{-1.34}$ pc, or essentially around ~ 6 pc away. A similar calculation for K dwarfs assuming the model 1 conservative HZ $\eta_{\oplus, \text{K}}$ values from Table 5 indicates that, on average, the nearest HZ planet could be between $4.1^{+1.19}_{-0.90}$ pc and $3.6^{+1.05}_{-0.79}$ pc, or around ~ 4 pc away.

An additional speculative calculation one could do is to take the number of G dwarfs in the solar neighborhood within 10 pc —19 from RECONS⁵⁴—and multiply it with the “low” conservative $\eta_{\oplus, \text{G}}$ value from model 1, $0.38^{+0.50}_{-0.22}$. We then get $7.2^{+9.5}_{-4.2}$ HZ planets around G dwarfs (of all sub-spectral types) within 10 pc. A similar calculation for K dwarfs from the same RECONS data with 44 stars and a model 1 “low” value, conservative HZ $\eta_{\oplus, \text{K}} = 0.32^{+0.35}_{-0.17}$ indicates that there are $14^{+15}_{-7.5}$ HZ planets around K dwarfs within 10 pc. It should be noted that the number for the nearest HZ planet and the number of HZ planets in the solar neighborhood use the “low” end of the rocky planet ($0.5\text{--}1.5 R_{\oplus}$) occurrence rate values from Table 5. As such, these represent the lower-bound estimates. In other words, there may potentially be an HZ rocky planet around a G or K dwarf that is closer, and maybe more HZ planets, than the values quoted above.

This can be quantified from the numbers shown in Table 6, which provides the 95% and 99% credible intervals of the upper and lower bounds on HZ occurrence for model 1 computed with hab2 and accounting for input uncertainty. If we use only the “low” values and the lower end of the conservative HZ occurrence values from this table (0.07 for 95% credible interval, 0.04 for 99%), then the nearest HZ planet around a G or K dwarf star is within ~ 6 pc away with 95% confidence and within ~ 7.5 pc away with 99% confidence. Similarly, there could be ~ 4 HZ planets within 10 pc with 95% confidence and ~ 3 HZ planets with 99% confidence. The lower bound of the 95% credible interval for G dwarfs is also 0.07, so assuming 100 billion stars in our Galaxy and a G-dwarf fraction of 0.041, there are likely at least 287 million rocky conservative HZ planets orbiting G-dwarf stars in our Galaxy.

The numbers provided in this section are first-order estimates to simply show a meaningful application of η_{\oplus} , given its uncertainties. Weiss et al. (2018), Mulders et al. (2018), and He et al. (2020) have shown that there is strong evidence for multiplicity and “clustering” (planets in multiplanet systems are correlated in size to their neighbors). However, the similar-size trend has not been characterized for orbits corresponding to the HZs of G and K dwarfs. If the pattern of similarly sized planets extends to HZ, the nearest sun-like star hosting an Earth-size planet in the HZ would be farther from our sun than we estimated in this paper, but it could have multiple Earth-sized planets in the HZ.

5.5. Comparison with Previous Estimates of η_{\oplus}

Our work is the first to compute HZ occurrence rates using the incident stellar flux for HZs around subsets of FGK stars based on the DR25 catalog and Gaia-based stellar properties and using the HZ definition of Kopparapu et al. (2014). Other works in the literature have produced occurrence rates for orbital periods related to FGK HZs, but as discussed in Section 1.2 and Section 3.5, we measure occurrence between HZ boundaries as a function of T_{eff} . This is a different quantity from occurrence rates based on orbital period. The few occurrence rate estimates in the literature based on instellation

flux, such as that of Petigura et al. (2013), use a rectangular region in terms of radius and instellation flux and only approximate the HZ. Therefore, we do not directly compare our occurrence rates with those in previous literature.

We provide a formal computation of Γ_{\oplus} , which is commonly used to compare η_{\oplus} estimates (see, for example, Figure 14 of Kunimoto & Matthews 2020). For a planet of period p and radius r , we define $\Gamma \equiv d^2f/d \log p d \log r = pr d^2f/dp dr$, and Γ_{\oplus} is Γ evaluated at Earth’s period and radius. We need to express Γ in terms of instellation flux I , which we will do using $d^2f/dp dr = (d^2f/dI dr)(dI/dp)$.

For a particular star, the instellation on a planet at period p in years is given by $I = R_*^2 T^4 M_*^{-2/3} p^{-4/3}$, where R_* is the stellar radius in solar radii, M_* is the stellar mass in solar masses, and $T = T_{\text{eff}}/T_{\odot}$ is the star’s effective temperature divided by the solar effective temperature. Then

$$\begin{aligned} \Gamma &= p r \frac{d^2f}{dI dr} \left(\frac{dI}{dp} \right) \\ &= -\frac{4}{3} \frac{R_*^2 T^4}{M_*^{2/3}} r p^{-\frac{4}{3}} \lambda(I, r, T, \theta) \end{aligned} \quad (10)$$

because $d^2f/dI dr = \lambda(I, r, T, \theta)$, one of the differential population rate models from Equation (5). To compute Γ_{\oplus} for a particular star, we evaluate Equation (10) at $r = 1 R_{\oplus}$, $p = 1$ year, and $I = R_*^2 T^4 M_*^{-2/3}$, the instellation a planet with a 1 year orbital period would have from that star. The result is the Γ_{\oplus} in radius and period implied by our differential population rate function in radius and instellation for that star and may be compared directly with Γ_{\oplus} from period-based occurrence studies.

We compute Γ_{\oplus} using model 1 from Equation (5) with input uncertainty on the hab2 stellar population. For each star in hab2, we evaluate Equation (10) using the posterior θ distribution and concatenate the resulting Γ_{\oplus} distributions from all the stars. We do this for both completeness extrapolations in Section 3.3.2, giving low and high bounds. This results in a Γ_{\oplus} between $0.45^{+0.46}_{-0.24}$ and $0.50^{+0.46}_{-0.26}$. While this is a formal mathematical exercise that has not been demonstrated to be truly equivalent to Γ_{\oplus} defined in period space, the match between this value and our conservative HZ $\eta_{\oplus}^{\text{C}} = 0.37^{+0.48}_{-0.21} - 0.60^{+0.90}_{-0.36}$ for model 1 for hab2 with input uncertainty in Table 3 is remarkable.

Our value of Γ_{\oplus} is somewhat higher than values using post-Gaia stellar and planetary data (see, for example, Figure 14 of Kunimoto & Matthews 2020), but not significantly so. For example, Bryson et al. (2020a) found $\Gamma_{\oplus} = 0.09^{+0.07}_{-0.04}$ when correcting for reliability in the period–radius space. Using the same population model, Bryson et al. (2020a) found a SAG13 η_{\oplus} value of $0.13^{+0.10}_{-0.06}$. It is not clear by how much we should expect Γ_{\oplus} to correspond to η_{\oplus} .

5.6. Effective Temperature Dependence

As described in Section 4.2 and Figure 14, our results indicate a weak, and thus not compelling, increase in HZ planet occurrence with increasing stellar effective temperature. This T_{eff} dependence is weaker than would be expected if planet occurrence were uniformly spaced in the semimajor axis (see Section 2.2) because hotter stars have larger HZs. This can be seen quantitatively in the median differential population rates

⁵⁴ <http://www.recons.org/census.posted.htm>

for models 1 and 2 using the hab2 population in Tables 1 and 2. In model 2 we observe a median T_{eff} exponent $\gamma < 3$, compared with the prediction of $\gamma \approx 3$ to 4.5 due to the larger HZ for hotter stars from Equation (4). This is reflected in model 1, which includes the correction for the larger HZ so if T_{eff} dependence were due only to the larger HZ, then γ would equal 0. The high bound of model 1 finds a median $\gamma < -1$, indicating that we see fewer HZ planets than expected in the larger HZ of hotter stars if the planets were uniformly spaced. However the upper limits of γ 's 68% credible interval in Tables 1 and 2 are consistent with the prediction of uniform planet spacing in larger HZs for hotter stars. For example, the posterior of γ for model 1 (hab2 population, high bound) has $\gamma \geq 0$ 22.3% of the time.

Our detection of a weaker T_{eff} dependence than expected from larger HZs is qualitatively consistent with the increasing planet occurrence with lower T_{eff} found in Garrett et al. (2018) and Mulders et al. (2015). But our uncertainties do not allow us to make definitive conclusions about this T_{eff} dependence.

5.7. Dependence on the Planet Sample

To study the dependence of our result on the planet sample, we perform a bootstrap analysis. We run a Poisson likelihood inference using hab2 and model 1 with zero extrapolation (high bound) 400 times, resampling the planet candidate population with replacement. Each resampled run removes planets according to their reliability as described in Section 3.4.2 but does not consider input uncertainty. The concatenated posterior of these resampled runs gives $F_0 = 1.404^{+1.768}_{-0.680}$, $\alpha = -0.920^{+1.236}_{-1.072}$, $\beta = -1.175^{+0.465}_{-0.444}$, and $\gamma = -1.090^{+2.446}_{-2.217}$. These parameters yield $\eta_{\oplus}^C = 0.483^{+0.997}_{-0.324}$ and $\eta_{\oplus}^O = 0.716^{+1.413}_{-0.472}$. Comparing these with the hab2 model 1 high value without uncertainty in Tables 1 and 3, we see that the central values from the bootstrap study are well within the 68% credible interval of our results and the uncertainties are as much as 50% higher.

A similar study of dependence on the stellar sample is not feasible because each resampled stellar population would require a full recomputation of detection and vetting completeness and reliability. Performing hundreds of these computations is beyond our available resources.

5.8. Impact of Catalog Reliability Correction

All results presented in this paper are computed with corrections for planet catalog completeness and reliability (see Section 3.3). Figure 15 shows an example of what happens when there is no correction for catalog reliability. We compute η_{\oplus}^C , the occurrence in the conservative HZ, with model 1, zero-completeness extrapolation (high value), accounting for input uncertainty and using the hab2 stellar population. With reliability correction, we have $\eta_{\oplus}^C = 0.60^{+0.90}_{-0.36}$, and without reliability correction we have $\eta_{\oplus}^C = 1.25^{+1.40}_{-0.60}$. In this typical case reliability has a factor-of-two impact, consistent with Bryson et al. (2020a), though because of the large uncertainties the difference is less than the 68% credible interval.

5.9. η_{\oplus} based on the GK and FGK Stellar Populations

Our definition of η_{\oplus} , restricted to stars with effective temperatures between 4800 K and 6300 K, varies somewhat from the literature. To connect with other occurrence rate studies we

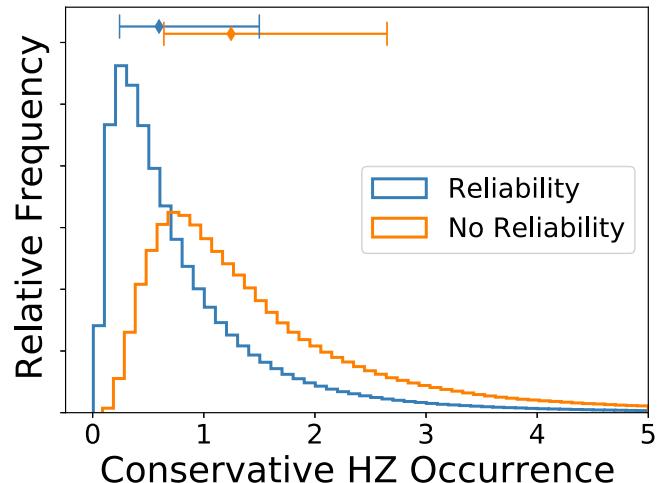


Figure 15. A comparison of the distributions, with and without reliability correction, of the conservative HZ η_{\oplus}^C computed with model 1, zero-completeness extrapolation (high value), accounting for input uncertainty and using the hab2 stellar population.

repeat our analysis using the GK (3900 K $\leq T_{\text{eff}} \leq 6000$ K) and FGK (3900 K $\leq T_{\text{eff}} \leq 7300$ K) stellar populations to compute planet population models, with the results listed in Table 7. We provide our η_{\oplus} derived from these stellar populations as well as the HZ occurrence for the GK and FGK T_{eff} ranges. The values for our definition of η_{\oplus} are consistent with the values in Tables 3 and 4. We caution, however, that the FGK population extends well into stellar effective temperatures where there are no planet detections and very low or zero completeness, so an FGK result necessarily involves extrapolation from cooler stars.

5.10. Caveats

While this study takes care to incorporate detection and vetting completeness and importantly both reliability and observational uncertainty, there are still unresolved issues. We summarize these issues here, each of which can motivate future improvements to our occurrence rate model and methodology.

Power-law assumption. Products of power laws in terms of radius and period are commonly adopted for planet population models in occurrence rate studies, but there is growing evidence that calls their suitability into question. For instance, improvements to stellar radius measurements have revealed that the radius distribution for small, close-in planets is bimodal, rather than a smooth or broken power law (Fulton et al. 2017), which has also been observed in K2 data (Hardegree-Ullman et al. 2020). Power laws are not capable of describing such non-monotonic populations. Looking at the bottom panels of Figure 10 (without uncertainty), we can see some data points in the radius and instellation flux distributions do not lie along our inferred power laws. However, using input uncertainties (top panels of Figure 10) washes out this structure, making it more difficult to discern the failure or success of a power-law model as a descriptor of the data. There is also strong evidence that populations are not well described by products of power laws in terms of radius and period (Petigura et al. 2018; Lopez & Rice 2018) for orbital periods of < 100 days. Therefore a product of power laws such as Equation (5) in terms of radius and instellation flux is unlikely to be a good description of planet populations at high instellation.

Table 7
Parameter Fits and η_{\oplus} with 68% Confidence Limits for Model 1 from Equation (5) Computed Using the Population Model from the Poisson Likelihood Method Applied to the GK and FGK Stellar Populations

	With Uncertainty		Without Uncertainty	
	Based on GK Stars		Based on FGK Stars	
	Low Bound–High Bound	Low Bound–High Bound	Low Bound–High Bound	Low Bound–High Bound
Model 1				
F_0	$+1.01^{+0.82}_{-0.4} - +1.23^{+1.12}_{-0.52}$	$+1.26^{+1.2}_{-0.55} - +2.36^{+2.94}_{-1.2}$	$+0.89^{+0.64}_{-0.33} - +1.12^{+0.83}_{-0.44}$	$+1.24^{+1.03}_{-0.51} - +2.05^{+2.14}_{-0.98}$
α	$-1.00^{+1.0}_{-0.9} - -1.05^{+1.0}_{-0.9}$	$-1.03^{+0.95}_{-0.87} - -1.16^{+0.94}_{-0.85}$	$-0.80^{+0.93}_{-0.83} - -0.90^{+0.87}_{-0.78}$	$-1.00^{+0.84}_{-0.77} - -0.96^{+0.87}_{-0.78}$
β	$-0.90^{+0.34}_{-0.32} - -1.13^{+0.37}_{-0.36}$	$-0.82^{+0.32}_{-0.3} - -1.23^{+0.36}_{-0.34}$	$-0.84^{+0.33}_{-0.3} - -1.08^{+0.36}_{-0.34}$	$-0.82^{+0.3}_{-0.27} - -1.21^{+0.34}_{-0.32}$
γ	$-3.00^{+1.75}_{-1.72} - -2.25^{+1.9}_{-1.86}$	$-2.67^{+1.57}_{-1.58} - -1.07^{+1.8}_{-1.77}$	$-2.76^{+1.61}_{-1.62} - -1.98^{+1.75}_{-1.69}$	$-2.67^{+1.43}_{-1.43} - -1.15^{+1.64}_{-1.66}$
η_{\oplus}^C	$0.34^{+0.48}_{-0.20} - 0.46^{+0.71}_{-0.28}$	$0.36^{+0.46}_{-0.20} - 0.63^{+0.92}_{-0.37}$	$0.29^{+0.37}_{-0.16} - 0.41^{+0.53}_{-0.23}$	$0.35^{+0.40}_{-0.19} - 0.53^{+0.68}_{-0.30}$
$\eta_{\oplus,GK}^C$	$0.32^{+0.41}_{-0.18} - 0.40^{+0.54}_{-0.23}$	$0.32^{+0.38}_{-0.18} - 0.48^{+0.64}_{-0.27}$	$0.26^{+0.31}_{-0.15} - 0.35^{+0.40}_{-0.19}$	$0.32^{+0.33}_{-0.16} - 0.41^{+0.49}_{-0.22}$
$\eta_{\oplus,FGK}^C$	$0.35^{+0.54}_{-0.21} - 0.47^{+0.81}_{-0.29}$	$0.37^{+0.53}_{-0.21} - 0.63^{+1.18}_{-0.39}$	$0.30^{+0.43}_{-0.17} - 0.42^{+0.64}_{-0.24}$	$0.36^{+0.46}_{-0.20} - 0.53^{+0.88}_{-0.31}$
η_{\oplus}^O	$0.52^{+0.72}_{-0.30} - 0.68^{+1.01}_{-0.41}$	$0.56^{+0.70}_{-0.31} - 0.92^{+1.29}_{-0.54}$	$0.45^{+0.56}_{-0.25} - 0.61^{+0.77}_{-0.34}$	$0.55^{+0.60}_{-0.29} - 0.77^{+0.96}_{-0.43}$
$\eta_{\oplus,GK}^O$	$0.48^{+0.59}_{-0.27} - 0.59^{+0.76}_{-0.33}$	$0.50^{+0.56}_{-0.27} - 0.70^{+0.90}_{-0.39}$	$0.41^{+0.46}_{-0.22} - 0.51^{+0.57}_{-0.27}$	$0.49^{+0.49}_{-0.25} - 0.59^{+0.68}_{-0.32}$
$\eta_{\oplus,FGK}^O$	$0.54^{+0.81}_{-0.32} - 0.69^{+1.17}_{-0.42}$	$0.57^{+0.81}_{-0.33} - 0.91^{+1.70}_{-0.55}$	$0.46^{+0.64}_{-0.27} - 0.62^{+0.93}_{-0.35}$	$0.57^{+0.70}_{-0.30} - 0.77^{+1.28}_{-0.45}$

Notes. The low and high bounds correspond to the constant- and zero-completeness extrapolation of Section 3.3.2. The superscripts C and O on η_{\oplus} refer to the conservative and optimistic HZs. $\eta_{\oplus,GK}$ is the HZ occurrence for GK stars ($3900 \text{ K} \leq T_{\text{eff}} \leq 6000 \text{ K}$), and $\eta_{\oplus,FGK}$ is the HZ occurrence for FGK stars ($3900 \text{ K} \leq T_{\text{eff}} \leq 7300 \text{ K}$).

At low instellation of the HZ, however, the observed planet candidate population does not indicate any obvious structure (see Figure 1). Given that our domain of analysis is plagued by few detections, low completeness, and low reliability, more observations are likely needed to determine more appropriate population models. Therefore, because most of our planet population have radii larger than $1.5 R_{\oplus}$, those larger planets are likely driving the population model and may be causing bias in the model in the smaller-planet regime due to our simple product power laws in Equation (5).

Planetary multiplicity. Zink et al. (2019) pointed out that when short-period planets are detected in the Kepler pipeline, data near their transits are removed for subsequent searches, which can suppress the detection of longer-period planets around the same star. They found that for planets with periods greater than 200 days detection completeness can be suppressed by over 15% on average. Our stellar population has removed stars for which more than 30% of the data has been removed due to transit detection via the *dutycycle_post* stellar property from the DR25 stellar property table (for details see Bryson et al. 2020a). We do not attempt to quantify the extent to which this stellar cut mitigates the impact identified in Zink et al. (2019), nor do we account for this effect in our analysis.

Stellar multiplicity contamination. Several authors (Ciardi et al. 2015; Furlan et al. 2017; Furlan & Howell 2017, 2020) have shown that undetected stellar multiplicity can impact occurrence rate studies in at least two ways. Stellar multiplicity can reveal planet candidates to be false positives, reducing the planet population, and valid planet candidates in the presence of unknown stellar multiplicity will have incorrect planet radii due to flux dilution. They estimate that these effects can have an overall impact at the 20% level. Stellar multiplicity can also bias the parent stellar sample because unaccounted-for flux dilution will bias the completeness estimates. Our analysis does not take possible stellar multiplicity into account. However stellar multiplicity has been shown to be associated with poor-quality metrics, specifically the BIN flag of Berger et al. (2018)

and the Gaia RUWE metric (Lindegren 2018). For example, A. Kraus et al. (2021, in preparation) found that few Kepler target stars with $\text{RUWE} > 1.2$ are single stars. As described in Section 3.1, we remove stars from our parent stellar population that have been identified as likely binaries in Berger et al. (2018) or have $\text{RUWE} > 1.2$, which is expected to remove many stars with undetected stellar multiplicity (for details see Bryson et al. 2020a). We do not attempt to quantify the impact of undetected stellar multiplicity for our stellar population after this cut.

Planet radius and HZ limits. There are several stellar, planetary, and climate model-dependent factors that could reduce the occurrence rates calculated in this work. It is quite possible that uncertainties in the stellar radii may alter the planet radii, moving some rocky planets into the mini-Neptune regime of $>1.5 R_{\oplus}$. Or, it is possible that the upper limit of $1.5 R_{\oplus}$ is an overestimate of the rocky planet limit, and the rocky-to-gaseous transition may lie lower than $1.5 R_{\oplus}$. Although, as pointed out in Section 2 following Otegi et al. (2020), the rocky regime can extend to as large as $2.5 R_{\oplus}$, many of these large-radius regime planets are highly irradiated ones, so they may not be relevant to HZ rocky planets.

The HZ limits themselves may be uncertain, as they are model and atmospheric composition dependent. Several studies in recent years have calculated HZ limits with various assumptions (see Kopparapu et al. 2019 for a review). In particular, the inner edge of the HZ could extend further in, closer to the star, due to slow rotation of the planet (Yang et al. 2014; Kopparapu et al. 2016; Way et al. 2016), and the outer edge of the HZ may shrink due to “limit cycling,” a process where the planet near the outer edge of the HZ around FG stars may undergo cycles of globally glaciated and unglaciated periods with no long-term stable climate state (Kadoya & Tajika 2014, 2015; Menou 2015; Haqq-Misra et al. 2016). Consequently, the number of planets truly in the HZ remains uncertain.

5.11. Reducing Uncertainties

Our computation of η_{\oplus} has large uncertainties, with the 68% credible interval spanning factors of 2 (see Tables 3–5). The 99% credible intervals in Table 6 span two orders of magnitude. In Section 5.3 we discuss how comparing occurrence rates with and without input uncertainties in Tables 3 and 4 indicates that these large uncertainties are present before considering the impact of uncertainties in the input data. We also observe in Table 5 that the uncertainties are considerably smaller for planets larger than those contributing to our η_{\oplus} . We conclude that while input uncertainties make a contribution, the dominant cause of our large uncertainties is Poisson uncertainty due to the very small number of HZ planets smaller than $1.5 R_{\oplus}$ in very low completeness regions of the DR25 planet catalog (see Figure 1). Our uncertainties may be close to a noise floor induced by the small number of small HZ planets resulting from the low completeness.

These large Poisson-driven uncertainties are unlikely to be reduced by resolving the issues discussed in Section 5.10. Only by increasing the small-planet catalog’s completeness, resulting in a larger small-planet HZ population, can these uncertainties be reduced.

There are two ways in which a well-characterized catalog with more small planets can be produced:

1. Develop improved planet-vetting metrics that produce a catalog that is both more complete and more reliable than the DR25 catalog. There are several opportunities for such improved metrics, discussed in Bryson et al. (2020a), such as more fully exploiting pixel-level data and existing instrumental flags that can give more accurate reliability characterization than that given using DR25 products. This approach requires new vetting metrics. Bryson et al. (2020b) have shown that varying the DR25 Robovetter thresholds does not significantly change occurrence rates or their uncertainties once completeness and reliability are taken into account. In Appendix D we show that such changes in Robovetter metrics also do not significantly change the occurrence rates we find in this paper.
2. Obtain more data with a quality similar to Kepler’s, likely through more space-based observations. In Section 1 we describe how the unrealized Kepler extended mission, doubling the amount of data relative to DR25, was expected to significantly increase the yield of small planets in the HZ. An additional 4 years of data from observing the same stars as Kepler with similar photometric precision would be sufficient. Eight years of observation on a different stellar population would also suffice. As of this writing, plans for space-based missions such as TESS and PLATO do not include such long stares on a single field. For example, PLATO is currently planning no more than 3 years of continuous observation of a single field.⁵⁵

6. Conclusions

In this paper we compute the occurrence of rocky ($0.5 R_{\oplus} \leq r \leq 1.5 R_{\oplus}$) planets in the HZ for a range of main-sequence dwarf stars from the Kepler DR25 planet

candidate catalog and Gaia-based stellar properties. We base our occurrence rates on differential population models dependent on radius, instellation flux, and host star effective temperature (Section 3.4.1). Our computations are corrected for completeness and reliability, making full use of the DR25 data products. Using instellation flux instead of orbital period allows us to measure occurrence in the HZ even though the HZ boundaries depend on the stellar effective temperature (Section 3.5). Instellation flux also allows us to transfer the unconstrained extrapolation required when extending analysis based on orbital period to a bounded extrapolation of detection completeness (Section 3.3.2), and we present our results in terms of these upper and lower bounds (Section 4). The difference between the upper and lower bounds is smaller than the 68% credible interval on these bounds.

We compute our occurrence rates using a range of models, stellar populations, and computation methods. We propagate uncertainties in the input data, account for detection completeness that depends on the stellar effective temperature, and check the dependence of our result on the population via a bootstrap study. In all cases we find consistent results. We take this as evidence that our occurrence rates are robust.

We find a likely, though not statistically compelling, dependence of our occurrence rates on the stellar host’s effective temperature T_{eff} (Section 4.2, Figure 14). Much of this dependence can be understood as due to the HZ being larger for hotter stars (Section 2.2). But we find that the T_{eff} dependence is weaker than would be expected on purely geometric grounds, implying a decreasing planet occurrence for longer-period orbits.

Our occurrence rates for rocky planets have large uncertainties. Comparing computations with and without input uncertainties, we find that these large uncertainties are not caused by the input uncertainties. Comparing the uncertainties on our rocky planets with the uncertainties on the occurrence of larger planets (Table 5), we find that the larger-planet occurrence has much lower uncertainty. We conclude that the large uncertainties are due to the extremely low completeness of the DR25 catalog for small planets in the HZ, leading to few planet detections. The only way we see to reduce these uncertainties is by generating more complete and reliable catalogs, either through improved analysis of existing data or through obtaining more data with quality comparable to that of Kepler (Section 5.11).

Conservative habitability considerations (Section 2) and the limited coverage of F stars in the Kepler data (Section 1) drive us to define η_{\oplus} as the average number of HZ planets per star with planet radii between $0.5 R_{\oplus}$ and $1.5 R_{\oplus}$ and host star effective temperatures between 4800 and 6300 K. Using this definition, we find that, for the conservative HZ, η_{\oplus} is between $0.37^{+0.48}_{-0.21}$ and $0.60^{+0.90}_{-0.36}$ planets per star, while for the optimistic HZ η_{\oplus} is between $0.58^{+0.73}_{-0.33}$ and $0.88^{+1.28}_{-0.51}$ planets per star. These occurrence rates imply that conservatively, to 95% confidence, the nearest rocky HZ planet around G and K dwarfs is expected to be within ~ 6 pc (Section 5.4). Furthermore, there could, on average, be four HZ rocky planets around G or K dwarfs within 10 pc from the Sun.

This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program. We

⁵⁵ <https://www.cosmos.esa.int/web/plato/observation-concept>

thank our anonymous reviewer for many helpful comments that improved the manuscript. We thank NASA, Kepler management, and the Exoplanet Exploration Office for their continued support of and encouragement for the analysis of Kepler data. R. K.K. acknowledges support from the GSFC Sellers Exoplanet Environments Collaboration, which is supported by the NASA Planetary Science Division's Research Program, and from NASA's NExSS Virtual Planetary Laboratory funded by the NASA Astrobiology Program under grant 80NSSC18K0829. Funding for the Stellar Astrophysics Centre is provided by the Danish National Research Foundation (grant DNRF106). V.S.A. acknowledges support from the Independent Research Fund Denmark (research grant 7027-00096B) and the Carlsberg Foundation (grant agreement CF19-0649). E.B.F. was supported by a grant from the Simons Foundation/SFARI (675601). E.B. F. acknowledges support from the NASA Kepler Participating Scientist Program, through grants #NNX08AR04G, #NNX12AF73G, and #NNX14AN76G. E.B.F. acknowledges support from Penn State's Eberly College of Science, Department of Astronomy & Astrophysics, Institute for Computational & Data Sciences, Center for Exoplanets and Habitable Worlds, and Center for Astrostatistics. E.S.D. acknowledges financial support from the National Council for Scientific and Technological Development (CNPQ) and the IFRJ. D.H. acknowledges support from the Alfred P. Sloan Foundation, the National Aeronautics and Space Administration (80NSSC19K0597), and the National Science Foundation (AST-1717000). S. Mathur acknowledges support from the Spanish Ministry with the Ramon y Cajal fellowship number RYC-2015-17697. L.M.W. is supported by the Beatrice Watson Parrent Fellowship and NASA ADAP Grant \#80NSSC19K0597.

Facility: Kepler.

Software: Python, Jupyter.

Appendix A

Instellation Flux and Effective Temperature Population Rate Dependence from a Period Power Law

We can qualitatively estimate the instellation flux portion of the differential rate function λ by using $df/dI = (df/dp)/(dI/dp)$. From the formula for instellation flux and Kepler's third law, we have $I = R_*^2 T^4 (M_*^2 p^4)^{-\frac{1}{3}}$, where M_* is the stellar mass in solar masses, $T = T_{\text{eff}}/T_{\odot}$ is the effective temperature divided by the solar effective temperature, and p is the orbital period in years. Using the mass-radius relation for main-sequence dwarfs, this becomes $I \approx R_*^\mu T^4 p^{-\frac{4}{3}}$, where $\mu = 2 - \frac{2}{3\xi}$. When $M_* \leq M_{\odot}$, $\xi \approx 0.8$ and $\mu \approx 1.17$, while for $M_* > M_{\odot}$, $\xi \approx 0.57$ and $\mu \approx 0.8$. We make the crude ($\approx 20\%$ error) but convenient approximation that $\mu = 1$. Then using the empirically linear relationship between radius and temperature for the main-sequence dwarfs in our stellar population, $I \approx (\tau T + R_0) T^4 p^{-\frac{4}{3}}$ and, assuming p and T are independent, $dI/dp \approx -\frac{4}{3}(\tau T + R_0) T^4 p^{-\frac{7}{3}}$.

Several studies, such as Burke et al. (2015) and Bryson et al. (2020a), have investigated planet occurrence in terms of the orbital period p and have shown that df/dp is well approximated by a power law Fp^{α} (where F is determined by the radius dependence and normalization). Using this power

law and $p \approx ((\tau T + R_0) T^4 I^{-1})^{\frac{3}{4}}$, we have

$$\begin{aligned} df/dI &= \frac{df/dp}{dI/dp} \\ &\approx \frac{3Fp^{\alpha+\frac{7}{3}}}{4(\tau T + R_0) T^4} \\ &\approx \frac{3F((\tau T + R_0) T^4 I^{-1})^{\frac{3}{4}(\alpha+\frac{7}{3})}}{4(\tau T + R_0) T^4} \\ &\approx CI^{\nu}((\tau T + R_0) T^4)^{\delta} \end{aligned} \quad (11)$$

where $\nu = -\frac{3}{4}(\alpha - \frac{7}{3})$, $\delta = -\nu - 1$, and C is independent of I . Using the value $\alpha \approx -0.8$ from Bryson et al. (2020a), we have $\nu \approx -1.15$ and $\delta \approx 0.15$.

Appendix B

Derivation of the Effective Temperature-Dependent Likelihood

Our observed planet population is described by a point process with an instellation flux, radius, and effective temperature-dependent rate $\lambda(I, r, T)$ and completeness as a function of flux, radius, and effective temperature for each star $\eta_s(I, r, T_s)$. We assume that the probability that n_i planets occur around an individual star in some region B_i (say, a grid cell) of the flux–radius space is given by the Poisson probability

$$P\{N(B_i) = n_i\} = \frac{(\Lambda(B_i))^{n_i}}{n_i!} e^{-\Lambda(B_i)}$$

where

$$\Lambda(B_i) = \int_{B_i} \eta_s(I, r) \lambda(I, r, T_s) dI dr.$$

We do not integrate over T_s because that is fixed to the effective temperature of the star. We now cover our entire flux–radius range D with a sufficiently fine regular grid with spacing Δp and Δr so that each grid cell i centered at the flux and radius (I_i, r_i) contains at most one planet. Then in cell i

$$\begin{aligned} P\{N(B_i) = n_i\} &= \\ &\approx \begin{cases} \eta_s(I_i, r_i) \lambda(I_i, r_i, T_s) \Delta I \Delta r e^{-\Lambda(B_i)} & n_i = 1 \\ e^{-\Lambda(B_i)} & n_i = 0. \end{cases} \end{aligned}$$

We now ask: what is the probability of a specific number n_i of planets in each cell i ? We assume that the probability of a planet in different cells is independent, so

$$\begin{aligned} P\{N(B_i) = n_i, i = 1, \dots, K\} &= \\ &\approx \prod_{i=1}^K \frac{(\Lambda(B_i))^{n_i}}{n_i!} e^{-\Lambda(B_i)} \\ &\approx (\Delta I \Delta r)^{K_1} e^{-\sum_{i=1}^K \Lambda(B_i)} \\ &\quad \times \prod_{i=1}^{K_1} \eta_s(I_i, r_i) \lambda(I_i, r_i, T_s) \\ &= (\Delta I \Delta r)^{K_1} e^{-\int_D \eta_s(I, r) \lambda(I, r, T_s) dI dr} \\ &\quad \times \prod_{i=1}^{K_1} \eta_s(I_i, r_i) \lambda(I_i, r_i, T_s) \end{aligned} \quad (12)$$

because the B_i cover D and are disjoint. Here K is the number of grid cells and K_1 is the number of grid cells that contain a single planet. So the grid has disappeared, and we only need to

evaluate $\lambda(I, r, T_s)$ at the planet locations (I_i, r_i, T_s) and integrate $\eta_s \lambda$ over the entire domain.

We now consider the probability of detecting planets around a set of N_* stars. Assuming that planet detections on different stars are independent of each other, the joint probability of a specific set of detections specified by the set $\{n_i, i = 1, \dots, N_*\}$ in cell i on all stars indexed by s is given by

$$\begin{aligned} P\{N_s(B_i) = n_{s,i}, s = 1, \dots, N_*, i = 1, \dots, K\} \\ = \prod_{s=1}^{N_*} (\Delta I \Delta r)^{K_i} e^{-\int_D \eta_s(I, r) \lambda(I, r, T_s) dI dr} \\ \times \prod_{i=1}^{K_i} \eta_s(I_i, r_i) \lambda(I_i, r_i, T_s). \end{aligned} \quad (13)$$

When λ does not depend on effective temperature, we are able to factor $\prod_{s=1}^{N_*} \exp\left[-\int_D \eta_s(I, r) \lambda(I, r) dI dr\right]$ as $\exp\left[-\int_D \eta(I, r) \lambda(I, r) dI dr\right]$, where $\eta(I, r) = \sum_{s=1}^{N_*} \eta_s(I, r)$ is the sum of the completeness contours over all stars. When λ depends on effective temperature, we partition the stars into effective temperature bins S_k and approximate T_s as the average temperature in each bin \bar{T}_k , so within each bin λ does not depend on the star. Then we can do the factoring within each bin:

$$\begin{aligned} & \prod_{s=1}^{N_*} e^{-\int_D \eta_s(I, r) \lambda(I, r, T_s) dI dr} \\ & \approx \prod_k \prod_{s \in S_k} e^{-\int_D \eta_s(I, r) \lambda(I, r, \bar{T}_k) dI dr} \\ & = \prod_k e^{-\sum_{s \in S_k} \int_D \eta_s(I, r) \lambda(I, r, \bar{T}_k) dI dr} \\ & = \prod_k e^{-\int_D \eta_k(I, r) \lambda(I, r, \bar{T}_k) dI dr} \\ & = e^{-\sum_k \int_D \eta_k(I, r) \lambda(I, r, \bar{T}_k) dI dr} \end{aligned} \quad (14)$$

where $\eta_k(I, r) = \sum_{s \in S_k} \eta_s(I, r)$ is the sum of the completeness contours over the stars in bin S_k . Note that we are not integrating over the effective temperature. Therefore

$$\begin{aligned} P\{N_s(B_i) = n_{s,i}, s = 1, \dots, N_*, i = 1, \dots, K\} \\ = V e^{-\sum_k \int_D \eta_k(I, r) \lambda(I, r, \bar{T}_k) dI dr} \\ \times \prod_{s=1}^{N_*} \prod_{i=1}^{K_i} \eta_s(I_i, r_i) \lambda(I_i, r_i, T_s) \end{aligned} \quad (15)$$

where $V = (\Delta I \Delta r)^{(K_i N_*)}$.

We now let the rate function $\lambda(I, r, T, \theta)$ depend on a parameter vector θ , and consider the problem of finding the θ

that maximizes the likelihood

$$\begin{aligned} P\{N_s(B_i) = n_{s,i}, s = 1, \dots, N_*, i = 1, \dots, K|\theta\} \\ = V e^{-\sum_k \int_D \eta_k(I, r) \lambda(I, r, \bar{T}_k, \theta) dI dr} \\ \times \prod_{s=1}^{N_*} \prod_{i=1}^{K_i} \eta_s(I_i, r_i) \lambda(I_i, r_i, T_s, \theta) \\ = V \left(\prod_{s=1}^{N_*} \eta_s(I_s, r_s) \right) e^{-\sum_k \int_D \eta_k(I, r) \lambda(I, r, \bar{T}_k, \theta) dI dr} \\ \times \prod_{i=1}^{K_i} \lambda(I_i, r_i, T_s, \theta). \end{aligned} \quad (16)$$

Because we are maximizing with respect to θ , we can ignore all terms that do not depend on θ . Therefore, maximizing Equation (16) is equivalent to maximizing

$$\begin{aligned} P\{N_s(B_i) = n_{s,i}, s = 1, \dots, N_*, i = 1, \dots, K|\theta\} \\ = e^{-\sum_k \int_D \eta_k(I, r) \lambda(I, r, \bar{T}_k, \theta) dI dr} \prod_{i=1}^{K_i} \lambda(I_i, r_i, T_s, \theta). \end{aligned} \quad (17)$$

When we neglect the effective temperature dependence of λ and have only one effective temperature partition containing all the stars, Equation (17) reduces to

$$\begin{aligned} P\{N_s(B_i) = n_{s,i}, s = 1, \dots, N_*, i = 1, \dots, K|\theta\} \\ = e^{-\int_D \eta(I, r) \lambda(I, r, \theta) dI dr} \prod_{i=1}^{K_i} \lambda(I_i, r_i, \theta), \end{aligned}$$

used in Bryson et al. (2020a).

Appendix C Planet Candidate Properties

Figure 16 and Table 8 give the properties of the DR25 candidates used in our study. These planet candidates are detected on FGK host stars (of which hab and hab2 are subsets) after the cuts described in Section 3.1. The basic planet candidate population is that within our computation domain of $0.5 R_\oplus \leq r \leq 2.5 R_\oplus$ and $0.2 I_\oplus \leq I \leq 2.2 I_\oplus$, defined by the planet radius and instellation central values. When we account for input uncertainties as described in Section 3.4.2, some of these planets exit our domain and other planets enter the domain. In a particular realization, only those planets in the domain are involved in the computation of population models and occurrence rates. The probability of a planet candidate being in the domain in a particular realization is given by the “Inclusion Probability” column of Table 8. We list planet candidates with an inclusion probability of $>1/4000$, which, if reliability = 1, have a 10% chance of being included in one of the 400 realizations used in the computation with uncertainty.

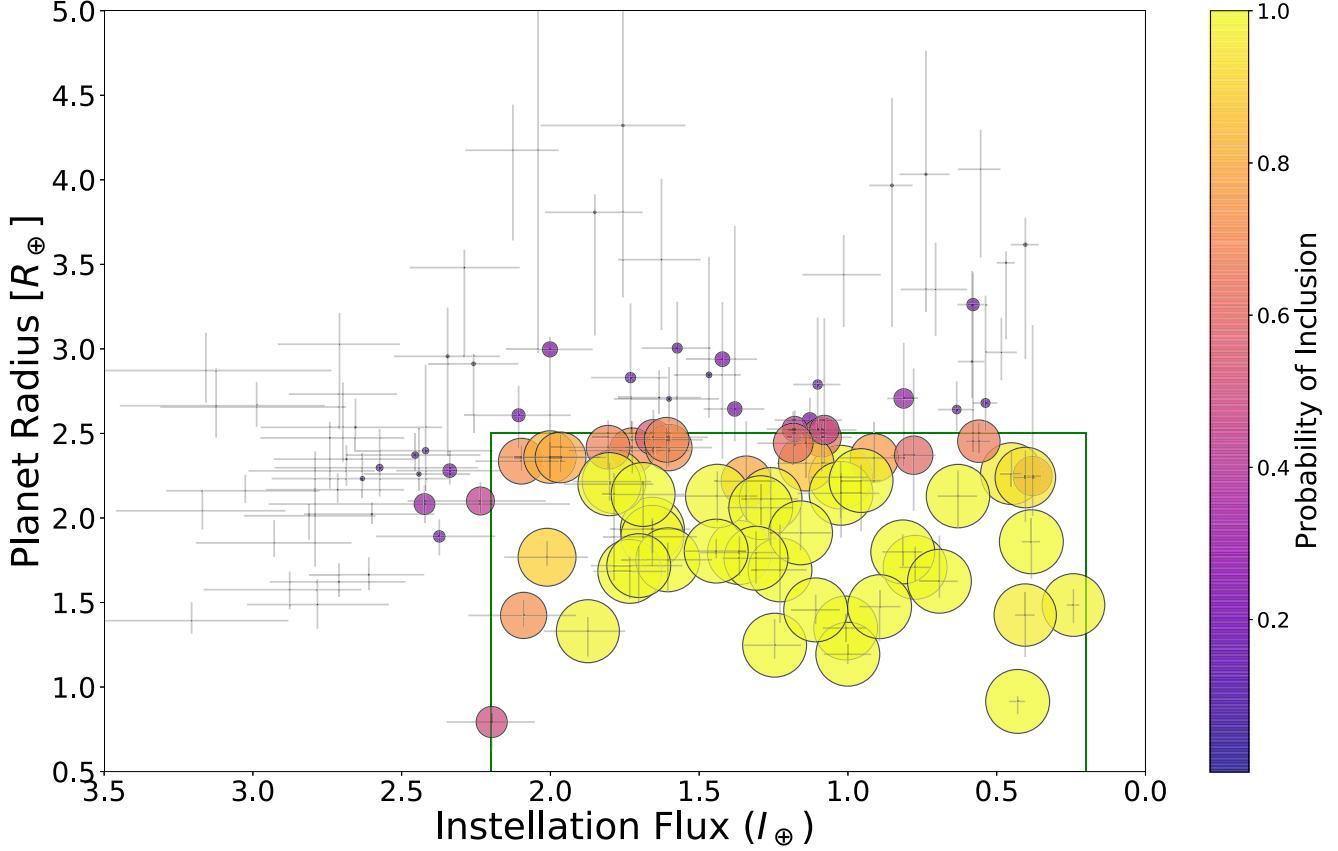


Figure 16. Planet candidates from Table 8, sized and colored by their inclusion probability. The green box shows the computational domain $0.5 R_{\oplus} \leq r \leq 2.5 R_{\oplus}$ and $0.2 I_{\oplus} \leq I \leq 2.2 I_{\oplus}$.

Table 8
Planet Candidate Properties

KOI	Radius (R_{\oplus})	Period (days)	Instellation (I_{\oplus})	Host Star T_{eff} (K)	Reliability	Inclusion Probability
4742.01	$1.35^{+0.08}_{-0.08}$	112.30	$1.01^{+0.08}_{-0.07}$	4602^{+84}_{-76}	0.91	1.00000
8107.01	$1.19^{+0.06}_{-0.06}$	578.89	$1.00^{+0.08}_{-0.08}$	5832^{+102}_{-103}	0.62	1.00000
7016.01	$1.46^{+0.09}_{-0.08}$	384.85	$1.11^{+0.08}_{-0.08}$	5900^{+102}_{-100}	0.68	1.00000
2719.02	$1.25^{+0.15}_{-0.08}$	106.26	$1.25^{+0.10}_{-0.09}$	4601^{+81}_{-76}	0.96	1.00000
701.03	$1.80^{+0.07}_{-0.04}$	122.39	$1.44^{+0.11}_{-0.10}$	4966^{+82}_{-82}	1.00	1.00000
4036.01	$1.71^{+0.12}_{-0.08}$	168.81	$0.77^{+0.05}_{-0.05}$	4697^{+76}_{-68}	1.00	1.00000
2194.03	$1.80^{+0.10}_{-0.14}$	445.22	$1.36^{+0.13}_{-0.12}$	5965^{+122}_{-116}	0.68	1.00000
4087.01	$1.80^{+0.10}_{-0.08}$	101.11	$0.82^{+0.07}_{-0.07}$	4171^{+56}_{-49}	1.00	1.00000
7923.01	$0.91^{+0.03}_{-0.08}$	395.13	$0.43^{+0.03}_{-0.03}$	5064^{+84}_{-73}	0.40	1.00000
8242.01	$1.48^{+0.10}_{-0.21}$	331.55	$0.89^{+0.07}_{-0.07}$	5736^{+105}_{-97}	0.53	1.00000
8047.01	$1.86^{+0.14}_{-0.22}$	302.35	$0.38^{+0.03}_{-0.03}$	4712^{+78}_{-74}	0.72	1.00000
8048.01	$1.76^{+0.16}_{-0.15}$	379.67	$1.31^{+0.11}_{-0.11}$	6058^{+108}_{-107}	0.48	1.00000
7894.01	$1.91^{+0.15}_{-0.10}$	347.98	$1.16^{+0.11}_{-0.10}$	5772^{+108}_{-106}	0.86	0.99996
2184.02	$1.93^{+0.05}_{-0.21}$	95.91	$1.66^{+0.14}_{-0.13}$	4820^{+88}_{-83}	0.97	0.99993
7749.01	$1.68^{+0.09}_{-0.16}$	133.63	$1.73^{+0.12}_{-0.12}$	5098^{+83}_{-78}	0.01	0.99993
7931.01	$1.75^{+0.10}_{-0.19}$	242.04	$1.61^{+0.16}_{-0.15}$	5843^{+106}_{-100}	0.84	0.99992
7915.01	$2.14^{+0.08}_{-0.26}$	382.59	$1.69^{+0.14}_{-0.14}$	6138^{+118}_{-117}	0.47	0.99990
7953.01	$1.63^{+0.27}_{-0.10}$	432.97	$0.69^{+0.07}_{-0.06}$	5421^{+107}_{-95}	0.37	0.99940
4450.01	$2.06^{+0.14}_{-0.10}$	196.44	$1.29^{+0.11}_{-0.11}$	5361^{+91}_{-89}	0.99	0.99932
8246.01	$1.72^{+0.12}_{-0.22}$	425.65	$1.70^{+0.16}_{-0.15}$	6091^{+125}_{-123}	0.36	0.99898
6971.01	$1.69^{+0.27}_{-0.31}$	129.22	$1.23^{+0.10}_{-0.09}$	4921^{+82}_{-83}	0.98	0.99857
87.01	$2.22^{+0.10}_{-0.30}$	289.86	$0.96^{+0.07}_{-0.07}$	5625^{+93}_{-93}	0.96	0.99827
7746.01	$2.15^{+0.12}_{-0.26}$	393.96	$1.02^{+0.15}_{-0.13}$	6135^{+118}_{-114}	0.56	0.99799

Table 8
(Continued)

KOI	Radius (R_{\oplus})	Period (days)	Instellation (I_{\oplus})	Host Star T_{eff} (K)	Reliability	Inclusion Probability
2992.01	$2.24^{+0.09}_{-0.24}$	82.66	$1.02^{+0.10}_{-0.09}$	4166^{+68}_{-57}	0.66	0.99754
2931.01	$2.22^{+0.05}_{-0.59}$	99.25	$1.79^{+0.14}_{-0.13}$	4806^{+84}_{-76}	0.99	0.99682
3344.03	$2.13^{+0.15}_{-0.16}$	208.54	$1.44^{+0.14}_{-0.14}$	5495^{+96}_{-95}	0.97	0.99409
8063.01	$2.13^{+0.15}_{-0.19}$	405.35	$0.63^{+0.07}_{-0.06}$	5455^{+103}_{-102}	0.78	0.99238
8159.02	$2.20^{+0.13}_{-0.10}$	353.02	$1.80^{+0.15}_{-0.15}$	6290^{+121}_{-118}	0.84	0.98719
7882.01	$1.33^{+0.08}_{-0.15}$	65.42	$1.87^{+0.15}_{-0.13}$	4390^{+81}_{-74}	0.90	0.98612
3282.01	$1.89^{+0.11}_{-0.10}$	49.28	$1.66^{+0.16}_{-0.15}$	4050^{+64}_{-69}	1.00	0.98445
4622.01	$1.48^{+0.09}_{-0.11}$	207.25	$0.24^{+0.02}_{-0.02}$	4147^{+67}_{-45}	0.98	0.98121
5067.01	$2.11^{+0.19}_{-0.27}$	219.93	$1.26^{+0.10}_{-0.09}$	5526^{+93}_{-87}	0.21	0.97935
571.05	$1.43^{+0.14}_{-0.25}$	129.95	$0.40^{+0.03}_{-0.03}$	4023^{+58}_{-62}	0.92	0.97528
2770.01	$2.26^{+0.13}_{-0.08}$	205.39	$0.45^{+0.04}_{-0.03}$	4475^{+80}_{-75}	0.99	0.96615
8033.01	$2.24^{+0.16}_{-0.26}$	362.13	$0.40^{+0.04}_{-0.04}$	5035^{+90}_{-83}	0.55	0.94736
2290.01	$1.77^{+0.10}_{-0.05}$	91.50	$2.01^{+0.14}_{-0.14}$	4944^{+75}_{-74}	1.00	0.90532
4084.01	$2.32^{+0.16}_{-0.09}$	214.88	$1.14^{+0.10}_{-0.09}$	5288^{+94}_{-89}	0.99	0.86571
250.04	$2.36^{+0.09}_{-0.13}$	46.83	$2.00^{+0.17}_{-0.17}$	4124^{+43}_{-68}	1.00	0.82052
4005.01	$2.36^{+0.14}_{-0.09}$	178.14	$1.97^{+0.16}_{-0.15}$	5545^{+94}_{-94}	0.99	0.79403
4054.01	$2.22^{+0.35}_{-0.23}$	169.14	$1.34^{+0.11}_{-0.10}$	5216^{+91}_{-86}	1.00	0.78695
5276.01	$2.36^{+0.21}_{-0.14}$	220.72	$0.91^{+0.12}_{-0.10}$	5086^{+95}_{-88}	0.96	0.75601
4015.01	$2.42^{+0.14}_{-0.11}$	133.30	$1.60^{+0.15}_{-0.15}$	5051^{+90}_{-85}	1.00	0.72676
2162.02	$1.42^{+0.09}_{-0.07}$	199.67	$2.09^{+0.19}_{-0.18}$	5814^{+116}_{-112}	0.99	0.72275
1989.01	$2.34^{+0.09}_{-0.06}$	201.12	$2.10^{+0.15}_{-0.15}$	5756^{+97}_{-96}	1.00	0.71591
2028.03	$2.40^{+0.17}_{-0.13}$	142.54	$1.72^{+0.21}_{-0.19}$	5213^{+97}_{-91}	1.00	0.71120
5874.01	$2.46^{+0.07}_{-0.17}$	287.33	$1.61^{+0.13}_{-0.12}$	5432^{+109}_{-102}	0.03	0.70033
5433.01	$2.42^{+0.16}_{-0.13}$	237.82	$1.81^{+0.19}_{-0.18}$	5798^{+112}_{-110}	0.97	0.68318
518.03	$2.45^{+0.11}_{-0.07}$	247.35	$0.56^{+0.04}_{-0.04}$	4918^{+90}_{-88}	1.00	0.66314
7345.01	$2.44^{+0.19}_{-0.13}$	377.50	$1.18^{+0.12}_{-0.11}$	5883^{+113}_{-111}	0.88	0.62068
2834.01	$2.48^{+0.08}_{-0.20}$	136.21	$1.09^{+0.11}_{-0.10}$	4775^{+91}_{-83}	1.00	0.61320
8201.01	$2.25^{+0.89}_{-0.24}$	392.60	$0.38^{+0.03}_{-0.03}$	5141^{+91}_{-88}	0.04	0.61117
4745.01	$2.37^{+0.51}_{-0.33}$	177.67	$0.78^{+0.08}_{-0.07}$	4790^{+84}_{-78}	0.99	0.59915
2841.01	$2.48^{+0.16}_{-0.13}$	159.39	$1.65^{+0.20}_{-0.18}$	5397^{+103}_{-100}	0.99	0.55438
7673.01	$0.79^{+0.05}_{-0.17}$	80.77	$2.20^{+0.15}_{-0.14}$	4747^{+78}_{-70}	0.74	0.48552
4121.01	$2.52^{+0.66}_{-0.19}$	198.09	$1.08^{+0.12}_{-0.11}$	5237^{+90}_{-86}	0.99	0.45819
812.03	$2.10^{+0.11}_{-0.07}$	46.18	$2.24^{+0.24}_{-0.22}$	4293^{+82}_{-90}	1.00	0.43624
2757.01	$2.53^{+0.11}_{-0.10}$	234.64	$1.18^{+0.09}_{-0.09}$	5437^{+96}_{-97}	0.96	0.39568
238.03	$2.08^{+0.11}_{-0.11}$	362.98	$2.42^{+0.52}_{-0.49}$	6572^{+272}_{-320}	0.90	0.32274
4016.01	$2.71^{+0.33}_{-0.40}$	125.41	$0.81^{+0.06}_{-0.05}$	4444^{+78}_{-76}	0.99	0.30347
612.03	$3.00^{+0.07}_{-0.77}$	122.08	$2.00^{+0.15}_{-0.14}$	5192^{+94}_{-89}	0.73	0.23460
1876.01	$2.58^{+0.13}_{-0.11}$	82.53	$1.13^{+0.12}_{-0.11}$	4269^{+81}_{-76}	1.00	0.23173
8156.01	$2.94^{+0.34}_{-0.59}$	364.98	$1.42^{+0.12}_{-0.12}$	6214^{+114}_{-108}	0.41	0.22789
1353.03	$2.64^{+0.08}_{-0.19}$	330.07	$1.38^{+0.10}_{-0.10}$	6081^{+102}_{-101}	0.30	0.22344
427.03	$2.28^{+0.11}_{-0.08}$	117.03	$2.34^{+0.18}_{-0.18}$	5208^{+90}_{-84}	1.00	0.21061
7880.01	$2.61^{+0.17}_{-0.19}$	623.71	$2.11^{+0.18}_{-0.18}$	6753^{+156}_{-139}	0.47	0.20068
8238.01	$3.26^{+0.19}_{-0.87}$	495.66	$0.58^{+0.05}_{-0.05}$	5540^{+108}_{-106}	0.74	0.18981
4076.01	$1.89^{+0.10}_{-0.11}$	124.83	$2.37^{+0.21}_{-0.19}$	5552^{+94}_{-89}	0.97	0.17777
1430.03	$2.83^{+0.44}_{-0.33}$	77.47	$1.73^{+0.13}_{-0.12}$	4543^{+79}_{-75}	1.00	0.15944
1871.01	$3.00^{+0.27}_{-0.49}$	92.73	$1.57^{+0.12}_{-0.11}$	4589^{+75}_{-72}	1.00	0.15384
2762.01	$2.79^{+0.40}_{-0.27}$	133.00	$1.10^{+0.08}_{-0.08}$	4694^{+80}_{-75}	1.00	0.14364
5581.01	$2.68^{+0.64}_{-0.17}$	374.88	$0.54^{+0.04}_{-0.04}$	5311^{+89}_{-89}	0.92	0.13855
4356.01	$2.64^{+0.17}_{-0.13}$	174.51	$0.63^{+0.06}_{-0.06}$	4577^{+85}_{-80}	0.99	0.13305
7889.01	$2.30^{+0.23}_{-0.17}$	130.24	$2.57^{+0.35}_{-0.33}$	5494^{+105}_{-102}	0.94	0.10455
581.02	$2.37^{+0.13}_{-0.09}$	151.86	$2.45^{+0.23}_{-0.22}$	5669^{+100}_{-94}	0.99	0.10319
2529.02	$2.40^{+0.51}_{-0.24}$	64.00	$2.42^{+0.24}_{-0.23}$	4607^{+89}_{-84}	0.96	0.09608
1596.02	$2.85^{+0.70}_{-0.25}$	105.36	$1.47^{+0.12}_{-0.11}$	4626^{+74}_{-69}	0.79	0.08588
3086.01	$2.70^{+0.19}_{-0.14}$	174.73	$1.60^{+0.19}_{-0.17}$	5480^{+104}_{-104}	0.98	0.07218
4636.01	$5.02^{+5964.94}_{-2.05}$	122.75	$2.04^{+0.24}_{-0.22}$	5158^{+92}_{-86}	0.01	0.07088
4009.01	$2.23^{+0.15}_{-0.11}$	175.14	$2.63^{+0.31}_{-0.29}$	5870^{+120}_{-117}	0.99	0.06576
1938.01	$2.26^{+0.27}_{-0.10}$	96.92	$2.44^{+0.19}_{-0.17}$	5086^{+86}_{-82}	1.00	0.06496

Table 8
(Continued)

KOI	Radius (R_{\oplus})	Period (days)	Instellation (I_{\oplus})	Host Star T_{eff} (K)	Reliability	Inclusion Probability
505.05	$2.91^{+0.06}_{-0.41}$	87.09	$2.26^{+0.15}_{-0.15}$	4868^{+75}_{-70}	1.00	0.05449
5622.01	$3.62^{+0.16}_{-0.68}$	469.61	$0.40^{+0.05}_{-0.04}$	5260^{+98}_{-90}	0.83	0.04917
4014.01	$2.96^{+0.29}_{-0.59}$	234.24	$2.35^{+0.18}_{-0.18}$	5993^{+103}_{-101}	0.93	0.04457
5790.01	$3.97^{+0.52}_{-0.84}$	178.27	$0.85^{+0.08}_{-0.07}$	4797^{+80}_{-75}	0.98	0.03965
1527.01	$3.81^{+0.11}_{-0.73}$	192.67	$1.85^{+0.17}_{-0.16}$	5603^{+105}_{-105}	0.82	0.03565
1707.02	$4.32^{+11.24}_{-1.02}$	265.48	$1.76^{+0.28}_{-0.21}$	5766^{+142}_{-130}	0.50	0.03458
8193.01	$4.03^{+0.73}_{-0.82}$	367.95	$0.74^{+0.09}_{-0.08}$	5546^{+95}_{-92}	0.36	0.03023
2210.02	$2.92^{+0.54}_{-0.22}$	210.63	$0.58^{+0.05}_{-0.04}$	4779^{+80}_{-78}	1.00	0.02503
4202.01	$2.54^{+0.17}_{-0.13}$	153.98	$2.66^{+0.32}_{-0.29}$	5741^{+110}_{-104}	1.00	0.02312
947.01	$2.02^{+0.07}_{-0.06}$	28.60	$2.60^{+0.23}_{-0.21}$	3926^{+60}_{-61}	1.00	0.01918
2828.01	$2.35^{+0.08}_{-0.16}$	59.50	$2.69^{+0.24}_{-0.23}$	4629^{+88}_{-80}	1.00	0.01785
4051.01	$2.71^{+0.16}_{-0.10}$	163.69	$1.63^{+0.14}_{-0.14}$	5351^{+100}_{-98}	0.97	0.01523
4242.01	$1.66^{+0.10}_{-0.09}$	145.79	$2.61^{+0.20}_{-0.19}$	5725^{+89}_{-92}	0.90	0.01379
2686.01	$3.51^{+0.07}_{-0.45}$	211.03	$0.47^{+0.03}_{-0.03}$	4475^{+73}_{-69}	1.00	0.01289
3508.01	$1.62^{+0.11}_{-0.09}$	190.80	$2.71^{+0.23}_{-0.22}$	6067^{+106}_{-106}	0.98	0.01120
172.02	$2.17^{+0.10}_{-0.08}$	242.47	$2.71^{+0.24}_{-0.22}$	5890^{+118}_{-114}	0.98	0.01037
8276.01	$3.48^{+0.11}_{-0.53}$	385.86	$2.29^{+0.18}_{-0.19}$	6618^{+128}_{-123}	0.63	0.01028
2172.02	$2.47^{+0.10}_{-0.30}$	116.58	$2.74^{+0.23}_{-0.25}$	5420^{+97}_{-93}	0.98	0.00927
4926.01	$1.49^{+0.15}_{-0.14}$	69.09	$2.78^{+0.24}_{-0.24}$	4831^{+92}_{-86}	0.39	0.00750
1986.01	$3.53^{+0.48}_{-0.42}$	148.46	$1.63^{+0.15}_{-0.13}$	5228^{+107}_{-100}	0.99	0.00673
1608.03	$2.01^{+0.07}_{-0.14}$	232.04	$2.81^{+0.22}_{-0.22}$	6128^{+111}_{-111}	0.89	0.00310
2525.01	$1.85^{+0.13}_{-0.08}$	57.29	$2.93^{+0.26}_{-0.26}$	4617^{+87}_{-79}	1.00	0.00253
8249.01	$1.58^{+0.11}_{-0.12}$	309.19	$2.88^{+0.29}_{-0.24}$	6153^{+129}_{-113}	0.64	0.00249
4385.02	$2.98^{+0.21}_{-0.16}$	386.37	$0.48^{+0.06}_{-0.05}$	5215^{+99}_{-93}	0.90	0.00180
3266.01	$2.28^{+0.12}_{-0.57}$	54.51	$2.79^{+0.22}_{-0.20}$	4459^{+75}_{-70}	1.00	0.00158
8275.01	$4.06^{+0.23}_{-0.52}$	389.88	$0.55^{+0.08}_{-0.07}$	5370^{+113}_{-109}	0.15	0.00138
7982.01	$3.44^{+0.23}_{-0.31}$	376.38	$1.01^{+0.14}_{-0.12}$	5814^{+107}_{-103}	0.56	0.00119
7798.01	$2.66^{+0.22}_{-0.19}$	309.89	$3.12^{+0.32}_{-0.37}$	6258^{+133}_{-127}	0.54	0.00110
6786.01	$3.35^{+0.28}_{-0.28}$	455.62	$0.71^{+0.12}_{-0.11}$	5622^{+118}_{-115}	0.80	0.00099
2650.01	$1.39^{+0.11}_{-0.08}$	34.99	$3.21^{+0.34}_{-0.33}$	4096^{+69}_{-80}	1.00	0.00099
416.02	$3.03^{+0.19}_{-0.50}$	88.26	$2.71^{+0.21}_{-0.20}$	5083^{+91}_{-85}	1.00	0.00091
1980.01	$2.73^{+0.07}_{-0.40}$	122.88	$2.70^{+0.18}_{-0.18}$	5441^{+85}_{-85}	1.00	0.00087
401.02	$4.17^{+0.27}_{-0.53}$	160.02	$2.13^{+0.16}_{-0.15}$	5516^{+90}_{-87}	0.95	0.00057
1078.03	$2.16^{+0.09}_{-0.07}$	28.46	$3.03^{+0.27}_{-0.25}$	4015^{+58}_{-62}	1.00	0.00047
1970.02	$2.66^{+0.15}_{-0.12}$	125.60	$2.99^{+0.32}_{-0.30}$	5585^{+98}_{-96}	1.00	0.00040
4856.01	$2.87^{+0.22}_{-0.19}$	147.39	$3.16^{+0.47}_{-0.42}$	5773^{+108}_{-104}	1.00	0.00029
775.03	$2.04^{+0.12}_{-0.11}$	36.45	$3.17^{+0.29}_{-0.28}$	4164^{+49}_{-62}	1.00	0.00026

Note. Boldfaced KOIs have central values in the computational domain $0.5 R_{\oplus} \leq r \leq 2.5 R_{\oplus}$ and $0.2 I_{\oplus} \leq I \leq 2.2 I_{\oplus}$.

(This table is available in machine-readable form.)

Appendix D Robovetter Variations

Bryson et al. (2020b) provided alternative planet candidate catalogs based on the Kepler data, created by changing automated vetting thresholds. They argued that occurrence rate estimates should roughly agree between these alternative catalogs. Figure 17

shows the distribution of the model parameter F_0 for model 1 for these catalogs, computed without input uncertainties and zero-completeness extrapolation with the hab2 stellar population. The fitted population model coefficients for these catalogs are given in Table 9. We see reasonable agreement between the Robovetter variations when reliability corrections are applied.

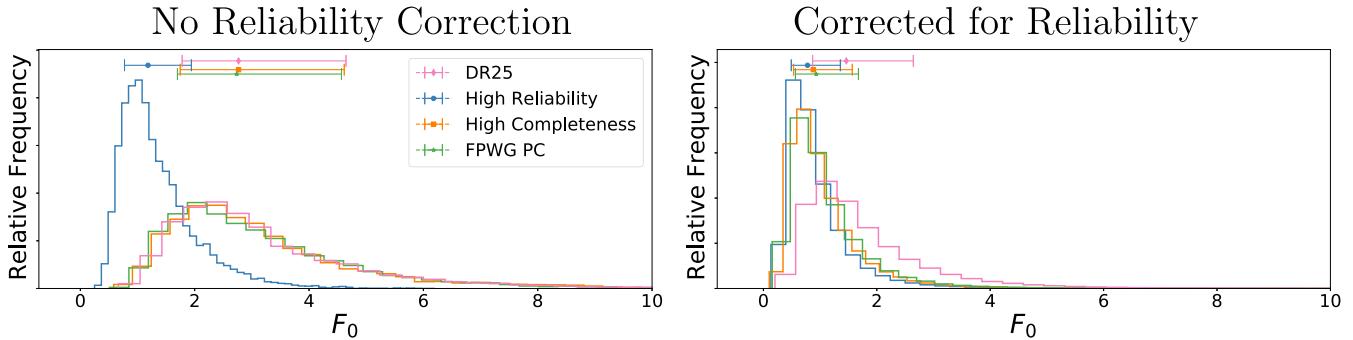


Figure 17. Distributions of the parameter F_0 in model 1 (see Equation (5)), the occurrence for the hab2 stellar population, for $0.5 R_{\oplus} \leq r \leq 2.5 R_{\oplus}$ and instellation flux range $0.2 I_{\oplus} \leq I_{\oplus} \leq 2.2 I_{\oplus}$ for the high-reliability (blue), DR25 (pink), FPWG planet candidate (green), and high-completeness (orange) vetting, computed with the Poisson method. Left: Without correcting for reliability. Right: Corrected for reliability.

Table 9
Fit Coefficients for the Alternative Planet Candidate Catalogs Using the hab2 Population with Zero-completeness Extrapolation

	DR25	High Reliability	High Completeness	FPWG PC	Max Separation (σ)
With Reliability Correction					
F_0	$1.46^{+1.18}_{-0.59}$	$0.78^{+0.58}_{-0.29}$	$0.88^{+0.69}_{-0.35}$	$0.93^{+0.74}_{-0.37}$	0.82
α	$-1.03^{+0.83}_{-0.77}$	$0.10^{+1.08}_{-0.98}$	$-0.65^{+0.97}_{-0.87}$	$-0.66^{+0.94}_{-0.86}$	0.88
β	$-1.15^{+0.34}_{-0.33}$	$-1.17^{+0.36}_{-0.34}$	$-0.95^{+0.38}_{-0.35}$	$-1.00^{+0.37}_{-0.35}$	0.44
γ	$-1.03^{+1.66}_{-1.64}$	$-2.12^{+1.77}_{-1.77}$	$-2.83^{+1.73}_{-1.68}$	$-2.42^{+1.72}_{-1.73}$	0.76
No Reliability Correction					
F_0	$2.77^{+1.88}_{-0.99}$	$1.18^{+0.76}_{-0.41}$	$2.77^{+1.85}_{-1.02}$	$2.74^{+1.83}_{-1.04}$	1.27
α	$-1.35^{+0.59}_{-0.63}$	$-0.36^{+0.79}_{-0.80}$	$-1.39^{+0.62}_{-0.62}$	$-1.36^{+0.64}_{-0.62}$	1.01
β	$-1.39^{+0.28}_{-0.28}$	$-1.26^{+0.32}_{-0.30}$	$-1.38^{+0.28}_{-0.27}$	$-1.39^{+0.28}_{-0.28}$	0.32
γ	$0.54^{+1.43}_{-1.34}$	$-1.18^{+1.56}_{-1.55}$	$0.44^{+1.38}_{-1.34}$	$0.46^{+1.45}_{-1.46}$	0.84

Notes. These are the posteriors of model 1 for the DR25, high-reliability, high-completeness, and FPWG planet candidate catalogs from Bryson et al. (2020b) using the hab2 stellar population and the Poisson likelihood method with zero-completeness extrapolation. The maximum separation in each row is the maximum over each row of the difference in medians divided by the propagated uncertainty of that distance.

ORCID iDs

- Steve Bryson <https://orcid.org/0000-0003-0081-1797>
- Michelle Kunimoto <https://orcid.org/0000-0001-9269-8060>
- Ravi K. Kopparapu <https://orcid.org/0000-0002-5893-2471>
- Jeffrey L. Coughlin <https://orcid.org/0000-0003-1634-9672>
- Victor Silva Aguirre <https://orcid.org/0000-0002-6137-903X>
- Geert Barentsen <https://orcid.org/0000-0002-3306-3484>
- Natalie M. Batalha <https://orcid.org/0000-0002-7030-9519>
- Travis Berger <https://orcid.org/0000-0002-2580-3614>
- Alan Boss <https://orcid.org/0000-0001-6974-9990>
- Lars A. Buchhave <https://orcid.org/0000-0003-1605-5666>
- Christopher J. Burke <https://orcid.org/0000-0002-7754-9486>
- Douglas A. Caldwell <https://orcid.org/0000-0003-1963-9616>
- Joseph Catanzarite <https://orcid.org/0000-0001-7980-4658>
- Hema Chandrasekaran <https://orcid.org/0000-0002-9760-7735>
- William J. Chaplin <https://orcid.org/0000-0002-5714-8618>
- Jessie L. Christiansen <https://orcid.org/0000-0002-8035-4778>
- Jørgen Christensen-Dalsgaard <https://orcid.org/0000-0001-5137-0966>

- David R. Ciardi <https://orcid.org/0000-0002-5741-3047>
- William D. Cochran <https://orcid.org/0000-0001-9662-3496>
- Jessie L. Dotson <https://orcid.org/0000-0003-4206-5649>
- Andrea K. Dupree <https://orcid.org/0000-0002-8985-8489>
- Michael Endl <https://orcid.org/0000-0002-7714-6310>
- Eric B. Ford <https://orcid.org/0000-0001-6545-639X>
- John C. Geary <https://orcid.org/0000-0003-3630-3735>
- Ronald L. Gilliland <https://orcid.org/0000-0002-1554-5578>
- Michael R. Haas <https://orcid.org/0000-0003-2397-0045>
- Matthew J. Holman <https://orcid.org/0000-0002-1139-4880>
- Andrew W. Howard <https://orcid.org/0000-0001-8638-0320>
- Steve B. Howell <https://orcid.org/0000-0002-2532-2853>
- Daniel Huber <https://orcid.org/0000-0001-8832-4488>
- Jon M. Jenkins <https://orcid.org/0000-0002-4715-9460>
- Hans Kjeldsen <https://orcid.org/0000-0002-9037-0018>
- David W. Latham <https://orcid.org/0000-0001-9911-7388>
- Robert L. Morris <https://orcid.org/0000-0001-9303-3204>
- Timothy D. Morton <https://orcid.org/0000-0002-8537-5711>
- Susan E. Mullally <https://orcid.org/0000-0001-7106-4683>
- David Pletcher <https://orcid.org/0000-0001-7691-2194>
- Andrej Prsa <https://orcid.org/0000-0002-1913-0281>
- Samuel N. Quinn <https://orcid.org/0000-0002-8964-8377>
- Elisa V. Quintana <https://orcid.org/0000-0003-1309-2904>
- Darin Ragozzine <https://orcid.org/0000-0003-1080-9770>
- Dimitar Sasselov <https://orcid.org/0000-0001-7014-1771>
- Shawn E. Seader <https://orcid.org/0000-0002-6905-7676>
- Megan Shabram <https://orcid.org/0000-0003-1179-3125>
- Avi Shporer <https://orcid.org/0000-0002-1836-3120>
- Jeffrey C. Smith <https://orcid.org/0000-0002-6148-7903>
- Jason H. Steffen <https://orcid.org/0000-0003-2202-3847>
- Guillermo Torres <https://orcid.org/0000-0002-5286-0251>
- Joseph D. Twicken <https://orcid.org/0000-0002-6778-7552>
- Jeffrey E. Van Cleve <https://orcid.org/0000-0002-4534-3969>
- Lauren M. Weiss <https://orcid.org/0000-0002-3725-3058>
- William F. Welsh <https://orcid.org/0000-0003-2381-5301>
- Bill Wohler <https://orcid.org/0000-0002-5402-9613>

References

- Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, *ApJS*, 204, 24
- Berger, T. A., Huber, D., Gaidos, E., & van Saders, J. L. 2018, *ApJ*, 866, 99
- Berger, T. A., Huber, D., Gaidos, E., van Saders, J. L., & Weiss, L. M. 2020a, *AJ*, 160, 108
- Berger, T. A., Huber, D., van Saders, J. L., et al. 2020b, *AJ*, 159, 280
- Bin, J., Tian, F., & Liu, L. 2018, *E&PSL*, 492, 121

- Borucki, W. J. 2016, *RPPh*, **79**, 036901
- Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Sci*, **327**, 977
- Borucki, W. J., Koch, D. G., Basri, G., et al. 2011, *ApJ*, **736**, 19
- Bryson, S., Coughlin, J., Batalha, N. M., et al. 2020a, *AJ*, **159**, 279
- Bryson, S., Coughlin, J., Kunitomo, M., & Mullally, S. E. 2020b, *AJ*, **160**, 200
- Burke, C. J., & Catanzarite, J. 2017, Planet Detection Metrics: Per-Target Detection Contours for Data Release 25, Kepler Science Document, [KSCI-19111-002](#)
- Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, *ApJ*, **809**, 8
- Catanzarite, J., & Shao, M. 2011, *ApJ*, **738**, 151
- Chen, J., & Kipping, D. 2017, *ApJ*, **834**, 17
- Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2020, *AJ*, **160**, 159
- Ciardi, D. R., Beichman, C. A., Horch, E. P., & Howell, S. B. 2015, *ApJ*, **805**, 16
- Coughlin, J. L. 2017, Planet Detection Metrics: Robovetter Completeness and Effectiveness for Data Release 25, Kepler Science Document, [KSCI-19114-002](#)
- Coughlin, J. L., Mullally, F., Thompson, S. E., et al. 2016, *ApJS*, **224**, 12
- Dressing, C. D., & Charbonneau, D. 2013, *ApJ*, **767**, 95
- Dressing, C. D., & Charbonneau, D. 2015, *ApJ*, **807**, 45
- Engmann, S., & Cousineau, D. 2011, JAQM, **6**, 1, http://jaqm.ro/issues/volume-6/issue-3/1_engmann_cousineau.php
- Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, *ApJ*, **795**, 64
- Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, *AJ*, **154**, 109
- Furlan, E., Ciardi, D. R., Everett, M. E., et al. 2017, *AJ*, **153**, 71
- Furlan, E., & Howell, S. B. 2017, *AJ*, **154**, 66
- Furlan, E., & Howell, S. B. 2020, *ApJ*, **898**, 47
- Garrett, D., Savransky, D., & Belikov, R. 2018, *PASP*, **130**, 114403
- Gaudi, B. S., Seager, S., Mennesson, B., et al. 2020, arXiv:2001.06683
- Gilliland, R. L., Chaplin, W. J., Dunham, E. W., et al. 2011, *ApJS*, **197**, 6
- Godolt, M., Grenfell, J. L., Hamann-Reinhus, A., et al. 2015, *P&SS*, **111**, 62
- Haqq-Misra, J., Kopparapu, R. K., Batalha, N. E., Harman, C. E., & Kasting, J. F. 2016, *ApJ*, **827**, 120
- Hardegree-Ullman, K. K., Zink, J. K., Christiansen, J. L., et al. 2020, *ApJS*, **247**, 28
- He, M. Y., Ford, E. B., & Ragozzine, D. 2020, *AJ*, **161**, 16
- Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, *ApJS*, **201**, 15
- Howell, S. B., Sobeck, C., Haas, M., et al. 2014, *PASP*, **126**, 398
- Hsu, D. C., Ford, E. B., Ragozzine, D., & Ashby, K. 2019, *AJ*, **158**, 109
- Hsu, D. C., Ford, E. B., Ragozzine, D., & Morehead, R. C. 2018, *AJ*, **155**, 205
- Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010, *ApJL*, **713**, L87
- Johnson, J. A., Petigura, E. A., Fulton, B. J., et al. 2017, *AJ*, **154**, 108
- Kadoya, S., & Tajika, E. 2014, *ApJ*, **790**, 107
- Kadoya, S., & Tajika, E. 2015, *ApJL*, **815**, L7
- Kane, S. R., Hill, M. L., Kasting, J. F., et al. 2016, *ApJ*, **830**, 1
- Kirkpatrick, J. D., Gelino, C. R., Cushing, M. C., et al. 2012, *ApJ*, **753**, 156
- Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, *ApJ*, **713**, L79
- Kopparapu, R. K. 2018, in Handbook of Exoplanets, ed. H. Deeg & J. Belmonte (Cham: Springer), 58
- Kopparapu, R. K., Hébrard, E., Belikov, R., et al. 2018, *ApJ*, **856**, 122
- Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, *ApJ*, **765**, 131
- Kopparapu, R. K., Ramirez, R. M., SchottelKotte, J., et al. 2014, *ApJL*, **787**, L29
- Kopparapu, R. K., Wolf, E. T., Arney, G., et al. 2017, *ApJ*, **845**, 5
- Kopparapu, R. K., Wolf, E. T., Haqq-Misra, J., et al. 2016, *ApJ*, **819**, 84
- Kopparapu, R. K., Wolf, E. T., & Meadows, V. S. 2019, arXiv:1911.04441
- Kunitomo, M., & Bryson, S. 2020, *RNAAS*, **4**, 83
- Kunitomo, M., & Matthews, J. M. 2020, *AJ*, **159**, 248
- Lindgren, L. 2018, Re-normalising the astrometric chi-square in Gaia DR2, GAIA-C3-TN-LU-LL-124-01, http://www.rssd.esa.int/doc_fetch.php?id=3757412
- Lopez, E. D., & Rice, K. 2018, *MNRAS*, **479**, 5303
- Mamajek, E. E., & Hillenbrand, L. A. 2008, *ApJ*, **687**, 1264
- Meixner, M., Cooray, A., Leisawitz, D., et al. 2019, arXiv:1912.06213
- Menou, K. 2015, *E&PSL*, **429**, 20
- Mulders, G. D., Pascucci, I., & Apai, D. 2015, *ApJ*, **798**, 112
- Mulders, G. D., Pascucci, I., Apai, D., & Ciesla, F. J. 2018, *AJ*, **156**, 24
- Mullally, F., Coughlin, J. L., Thompson, S. E., et al. 2016, *PASP*, **128**, 074502
- Otegi, J. F., Bouchy, F., & Helled, R. 2020, *A&A*, **634**, A43
- Pascucci, I., Mulders, G. D., & Lopez, E. 2019, *ApJL*, **883**, L15
- Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, *PNAS*, **110**, 19273
- Petigura, E. A., Howard, A. W., Marcy, G. W., et al. 2017, *AJ*, **154**, 107
- Petigura, E. A., Marcy, G. W., Winn, J. N., et al. 2018, *AJ*, **155**, 89
- Prangle, D. 2017, *BayAn*, **12**, 289
- Rogers, L. A. 2015, *ApJ*, **801**, 41
- Shabram, M. I., Batalha, N., Thompson, S. E., et al. 2020, *AJ*, **160**, 16
- Silburt, A., Gaidos, E., & Wu, Y. 2015, *ApJ*, **799**, 180
- Stark, C. C., Roberge, A., Mandell, A., et al. 2015, *ApJ*, **808**, 149
- The LUVOIR Team 2019, arXiv:1912.06219
- Thompson, S. E., Coughlin, J. L., Hoffman, K., et al. 2018, *ApJS*, **235**, 38
- Traub, W. A. 2012, *ApJ*, **745**, 20
- Trusina, J., Franc, J., & Novotny, A. 2020, *JPhCS*, **1525**, 012109
- Twicken, J. D., Jenkins, J. M., Seader, S. E., et al. 2016, *AJ*, **152**, 158
- Way, M. J., Del Genio, A. D., Kelley, M., Aleinov, I., & Clune, T. 2015, AGU Meeting, **2015**, P34C-07
- Way, M. J., Del Genio, A. D., Kiang, N. Y., et al. 2016, *GeoRL*, **43**, 8376
- Weiss, L. M., Marcy, G. W., Petigura, E. A., et al. 2018, *AJ*, **155**, 48
- Wolf, E. T., & Toon, O. B. 2015, *JRGD*, **120**, 5775
- Wolfgang, A., Rogers, L. A., & Ford, E. B. 2016, *ApJ*, **825**, 19
- Yang, J., Boué, G., Fabrycky, D. C., & Abbot, D. S. 2014, *ApJL*, **787**, L2
- Yang, J., Cowan, N. B., & Abbot, D. S. 2013, *ApJL*, **771**, L45
- Youdin, A. N. 2011, *ApJ*, **742**, 38
- Zink, J. K., Christiansen, J. L., & Hansen, B. M. S. 2019, *MNRAS*, **483**, 4479