Astro2020 Science White Paper

Characterizing Exoplanet Populations as a Constraint on Planet Formation and Input for Future NASA Missions

Thematic Area: Planetary Systems

Principal Author:
Name: Eric B. Ford
Institution: The Pennsylvania State University
Email: ebf11-at-psu.edu
Phone: 814-863-5558

Co-authors:
Darin Ragozzine, Brigham Young University, darin_ragozzine-at-byu.edu

Co-signers:
Matthias He, Penn State, myh7-at-psu.edu

White Paper Description:
The scientific impact of future exoplanet discoveries can be amplified if they are part of survey that have been designed and characterized to enable statistically valid characterization of exoplanet populations. As an example, we show recent characterization of exoplanet occurrence rates based on Kepler and Gaia data products.

Characterizing Populations Requires More than Individual Objects

Inevitably, young fields initially focus on discovering and characterizing individual objects, e.g., exoplanets, debris disks, or particularly fascinating planetary systems. As the sensitivity of surveys improves, it becomes practical to design and execute surveys that can meaningly constrain the intrinsic population of objects, and not just individual objects. The field of exoplanets is currently undergoing such a transition.

Modeling a population requires a model for both the intrinsic population and the experimental/observation process. Therefore, rigorously interpreting results of Doppler surveys in terms of an intrinsic population has not been practical due to the complex set of decisions that went into decisions of what object to observe each night. Similarly, interpreting the discoveries of ground-based transit surveys was difficult, both due to the complexities of the transit search process, but also the multi-stage vetting process for transiting planet candidates. Exoplanet population modeling got a humongous boost thanks to NASA's Kepler mission.

NASA's Kepler Mission

The most obvious contributors to Kepler's success were its unprecedented precision, accuracy, and duty cycle. Those enabled the detection of thousands of exoplanets (e.g., Batalha et al. 2013). However, accurately modeling the intrinsic population of exoplanets also required foresight, deliberate planning, extensive additional calculations to characterize the detection pipeline, and the automation of the vetting process. For example, Kepler has provided valuable constraints on the occurence rates of planets per sun-like star. Given the enormous interest, many early studies did what they could to estimate of the planet candidate occurence rates (e.g., Youdin 2011, Howard et al. 2012, Dressing & Charbonneau 2013, Fressin et al. 2013, Petigura et al. 2013, Foreman-Mackey et al. 2014, Burke et al. 2015). However, it was only once the Kepler science team performed a fully automated planet search (Thompson et al. 2018), automated vetting of planet candidates (Coughlin 2017), *and* provided extensive data products that enabled researchers to characterize the efficiency and completeness of the detection and vetting process (e.g., Christiansen et al. 2017, Burke & Catanzarite 2017) that it been possible to perform statistically rigorous inferences about the exoplanet population.

Hsu et al. (2019) recently applied a hierarchical model to estimate the rate of strong planet candidates (i.e., those that passed Kepler's robo-vetter) around Kepler's FGK main-sequence stars (see Fig 1.). This analysis accounts for most important sources of uncertainty (e.g., measurement noise, stellar radii, finite number of detections, geometric transit probability, detection efficiency, vetting efficiency, correlated noise in stellar photometry, window function). There are still unmodeled effects (e.g., pipeline timeouts, reduced detection efficiency for multiple planet systems) that are estimated to have only a modest effect on the total occurence rate (i.e., upper limits unlikely to increase by 10% due to these effects). However,

there is still considerable uncertainty in the occurence rate of long-period planets due to the potential of false alarms being included in the planet candidates. (Note that this doesn't affect the Hsu et al. (2019) estimates of occurence rate for such planets less than 1.5 $R_\oplus$, since no such planet remain after updating and filtering on stellar properties using Gaia DR2; Gaia Collaboration 2018). In principle, one could include false alarms in a hierarchical model. However, the currently available Kepler data products are not sufficient to rigorously characterize the rate of false alarms as a function of all the potentially relevant parameters. This demonstrates the importance of thinking through the entire data lifecycle, from basic reduction to final population analyses, during the early stages of project design and budget formulation.

Of particular interest for mission planning is the rate of with sizes of 1-1.75 $R_\oplus$ and orbital periods of 237-500 days. In the baseline model of Hsu et al. (2019), the {5, 15.87, 50, 84.13, 84.13, 95} quantiles for the rate of such planets per star are {0.09, 0.14, 0.24, 0.35, 0.46}. The corresponding percentiles for the differential rates are $\Gamma_\oplus \equiv (d^2f)/[d(\ln P)\, d(\ln R_p)] = \{0.22, 0.34, 0.57, 0.84, 1.1\}$ for the same range. While the data clearly inform the upper limits, the lower quantiles of this occurrence rate are heavily influenced by the choice of prior (see Figure 2). That is, Kepler data do not yet provide a direct lower limit on the rate of such planets. Due to the above mentioned concerns about potential unreliability of planet candidates with P = 237-500 and $R_p > 1.75\ R_\oplus$, it may be wise to allow for a potential factor of ~2 reduction in the upper quantiles of this occurrence rate (Thompson et al. 2018). Ten years after Kepler's launch, the exoplanet community is still working hard to perform the challenging statistical analyses necessary to provide robust constraints on planet formation models.

While characterizing the distribution of exoplanets is challenging, characterizing the distribution of exoplanetary architectures is even more challenging (Lissauer et al. 2011, Fabrycky et al. 2014, Mulders et al. 2018, Zhu et al. 2018, He et al. in prep.). For examples, we observe significant correlations in the orbital periods, radii and masses of exoplanets (Millholand et al. 2017, Weiss et al. 2017). This clustering effect means that the fraction of of stars with planets (within a given range of sizes and orbital period) is likely significantly less than the average number of such planets per star (e.g., Ragozzine & Holman 2010, Zhu et al. 2018, He et al. in prep). Our own preliminary results suggests that allowing for correlations in planet sizes and orbital periods likely results in the fraction of stars with planets being a factor of ~2 less than the average number of planets per star (He et al. in prep).

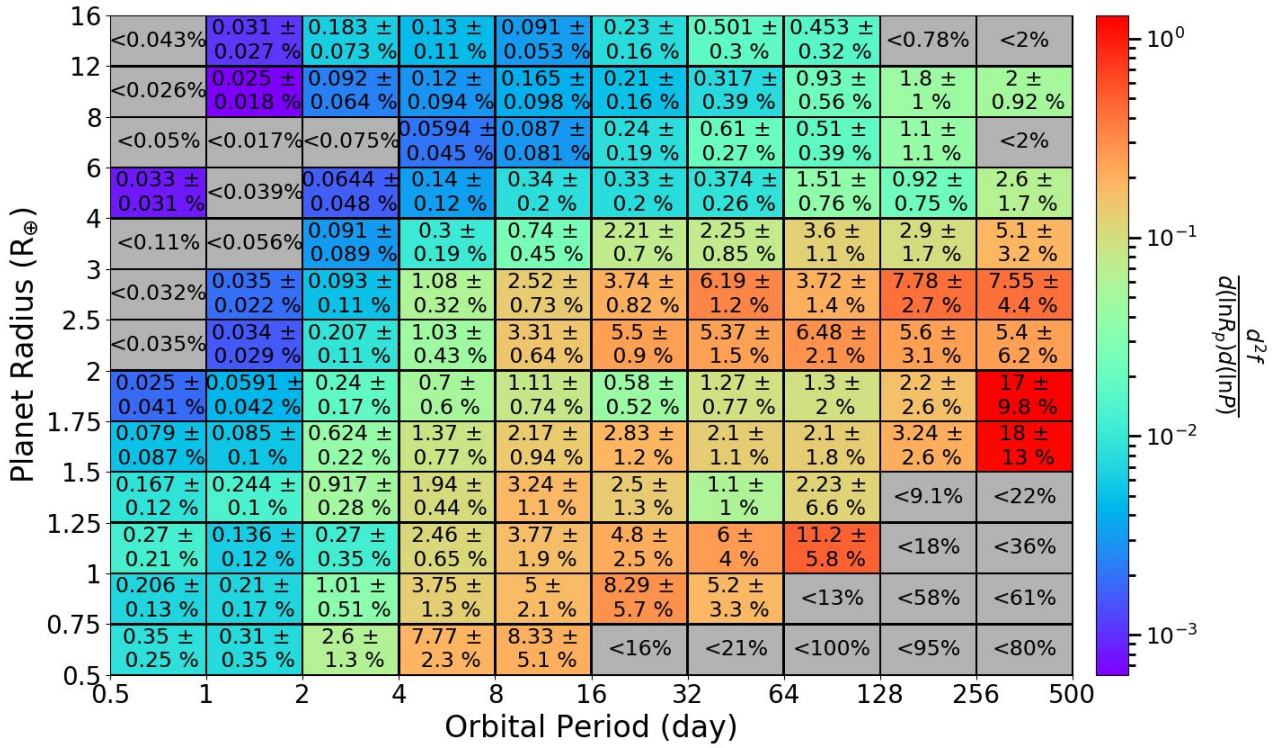

Figure 1: Planet occurrence rates based on Kepler's DR25 catalog of strong planet candidates, Gaia DR2 stellar properties and the hierarchical model of Hsu et al. (2019). The number and uncertainties in each cell corresponds to the 16, 50 and 84 percentiles of the ABC posterior distribution for the average number of planet candidates per high-quality FGK target star. The color code corresponds to the differential rate of planets per logarithmic interval in orbital period and planet radius. (The bin sizes are not constant.) Cells with zero or one planet candidates are include only the 84th percentile and are greyed out. Rates in the right-most column may be significantly influenced by false alarms that are not included in the model.

Conclusions

Future exoplanet surveys and NASA missions may reasonably include a component that is focused on detecting and characterizing a small number of particularly interesting planets. Mission concepts aiming to characterize potentially rocky planets in or near the habitable zone of sun-like stars should prepare compelling science programs that would be robust for a wide range of planet occurrence rates. If the decadal committee would value an occurrence rate calculated for a specific range of periods and radii, they are welcome to contact the primary author's research group.

More generally, the same surveys and missions have enormous potential for characterizing exoplanet populations, if this are made a priority during the survey/mission concepcion, design, planning, execution and analysis. Simulation studies have shown that simulation studies can improve upon astronomer intuition for how to design a survey or optimize follow-up observations (e.g., Ford 2008, Burt et al. 2018). Therefore, it is important that researchers with significant expertise in experimental design and the statistical analysis of anticipated data products be involved throughout the planning of surveys. Survey design and development and development of large statistical models to account for survey design, sensitivity and completeness should not be regarded as afterthought or as something that can be done by the community "for free", but should be integral parts of survey/project plans, in order to maximize the science return of future exoplanet surveys and future space missions.

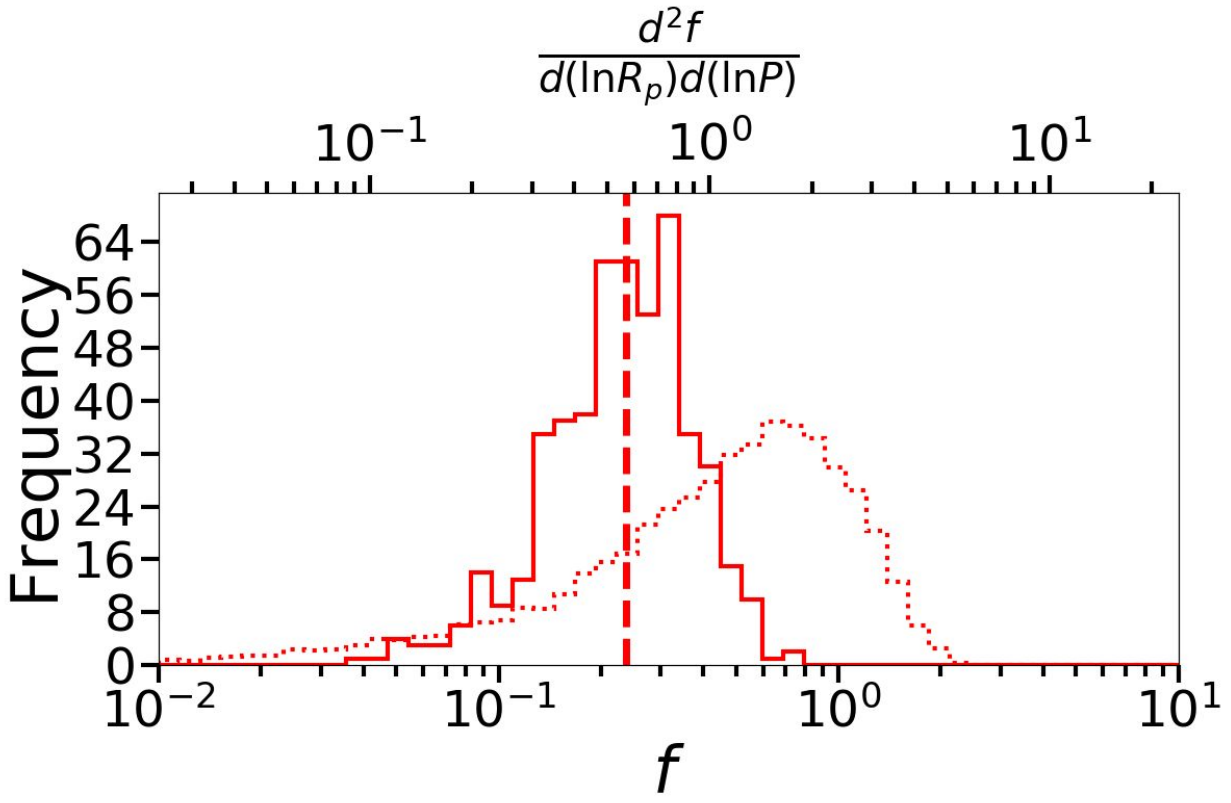

Figure 2: Posterior distribution for occurence rate of strong Kepler DR25 planet candidates per high-quality FGK target star integrated over a planet radii of 1-1.75$R_\oplus$ and orbital periods of 237-500 days. The lower horizontal axis is the integrated rate (f), while the top axis is the differential rate $\Gamma_\oplus \equiv (d^2f)/[d(\ln P)\, d(\ln R_p)]$. The solid histogram is the posterior distribution using a hierarchical Bayesian model and Approximate Bayesian Computing (ABC). The dotted curve shows the prior distribution. Results from Hsu et al. (2019).