Title

The California-Kepler Survey. III. A Gap in the Radius Distribution of Small Planets* * Based on observations obtained at the W. M. Keck Observatory, which is operated jointly by the University of California and the California Institute of Technology...

Journal

The Astronomical Journal, 154(3)

ISSN

0004-6256

Authors

Fulton, Benjamin J
Petigura, Erik A
Howard, Andrew W
et al.

Publication Date

2017-09-01

DOI

10.3847/1538-3881/aa80eb

Accepted for publication in the Astronomical Journal
Preprint typeset using LaTeX style emulateapj v. 12/16/11

THE CALIFORNIA-KEPLER SURVEY. III. A GAP IN THE RADIUS DISTRIBUTION OF SMALL PLANETS[1]

Benjamin J. Fulton[2,3,12,*], Erik A. Petigura[3,15], Andrew W. Howard[3], Howard Isaacson[4], Geoffrey W. Marcy[4], Phillip A. Cargile[5], Leslie Hebb[6], Lauren M. Weiss[7,13], John Asher Johnson[5], Timothy D. Morton[8], Evan Sinukoff[2,3,14], Ian J. M. Crossfield[9,16], Lea A. Hirsch[3]

Accepted for publication in the Astronomical Journal

ABSTRACT

The size of a planet is an observable property directly connected to the physics of its formation and evolution. We used precise radius measurements from the California-Kepler Survey (CKS) to study the size distribution of 2025 *Kepler* planets in fine detail. We detect a factor of $\geq$2 deficit in the occurrence rate distribution at 1.5–2.0 $R_\oplus$. This gap splits the population of close-in ($P < 100$ d) small planets into two size regimes: $R_P < 1.5$ $R_\oplus$ and $R_P = 2.0$–3.0 $R_\oplus$, with few planets in between. Planets in these two regimes have nearly the same intrinsic frequency based on occurrence measurements that account for planet detection efficiencies. The paucity of planets between 1.5 and 2.0 $R_\oplus$ supports the emerging picture that close-in planets smaller than Neptune are composed of rocky cores measuring 1.5 $R_\oplus$ or smaller with varying amounts of low-density gas that determine their total sizes.

Subject headings: planetary systems, *Kepler*

1. INTRODUCTION

NASA's *Kepler* space telescope enabled the discovery of over 4000 transiting planet candidates[16,17] opened the door to detailed studies of exoplanet demographics. One of the first surprises to arise from studies of the newly revealed sample of planets was the multitude of planets with radii smaller than Neptune but larger than Earth (R_P=1.0–3.9 $R_\oplus$, Batalha et al. 2013). Our solar system has no example of these intermediate planets, yet they are by far the most common in the *Kepler* sample (Howard et al. 2012; Fressin et al. 2013; Petigura et al. 2013b; Youdin 2011; Christiansen et al. 2015; Dressing & Charbonneau 2015; Morton & Swift 2014).

A key early question of the *Kepler* mission was whether these sub-Neptune-size planets are predominantly rocky or possess low-density envelopes that contribute significantly to the planet's overall size. The radial velocity (RV) follow-up effort of the *Kepler* project focused on 22 stars hosting one or more sub-Neptunes (Marcy et al. 2014). In addition, detailed modeling of transit timing variations (TTVs) provided mass constraints for a large number of systems in specific architectures (e.g., Wu & Lithwick 2013; Hadden & Lithwick 2014, 2016). The resulting mass measurements revealed that most planets larger than 1.6 $R_\oplus$ have low densities that were inconsistent with purely rocky compositions, and instead required gaseous envelopes (Weiss & Marcy 2014; Rogers 2015).

The distinction between rocky and gaseous planets reflects the typical core sizes of planets as well as the physical mechanisms by which planets acquire (and lose) gaseous envelopes. The densities of planets with radii smaller than $\sim$1.6 $R_\oplus$ are generally consistent with a purely rocky composition (Weiss & Marcy 2014; Rogers 2015) and their radius distribution likely reflects their initial core sizes. However, a small amount of H/He gas added to a roughly Earth-size rocky core can substantially increase planet size, without significantly increasing planet mass. For this reason, it has been suggested that the radii of sub-Neptune-size planets, along with knowledge of the irradiation history, would be sufficient to estimate bulk composition without additional information (Lopez & Fortney 2013; Wolfgang & Lopez 2015).

The large number of planets smaller than Neptune discovered by the *Kepler* mission was unexpected given prevailing theories of planet formation, which were developed to explain the distribution of giant planets (Ida & Lin 2004; Mordasini et al. 2009). These theories predicted that planets should either fail to accrete enough material to become super-Earths, or they would grow quickly, accreting all of the gas in their feeding zones

[1] Based on observations obtained at the W. M. Keck Observatory, which is operated jointly by the University of California and the California Institute of Technology. Keck time was granted for this project by the University of California, and California Institute of Technology, the University of Hawaii, and NASA.

[2] Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
[3] California Institute of Technology, Pasadena, California, U.S.A.
[4] Department of Astronomy, University of California, Berkeley, CA 94720, USA
[5] Harvard-Smithsonian Center for Astrophysics, 60 Garden St, Cambridge, MA 02138, USA
[6] Hobart and William Smith Colleges, Geneva, NY 14456, USA
[7] Institut de Recherche sur les Exoplanètes, Université de Montréal, Montréal, QC, Canada
[8] Department of Astrophysical Sciences, Peyton Hall, 4 Ivy Lane, Princeton, NJ 08540 USA
[9] Astronomy and Astrophysics Department, University of California, Santa Cruz, CA, USA
[12] National Science Foundation Graduate Research Fellow
[13] Trottier Fellow
[14] Natural Sciences and Engineering Research Council of Canada Graduate Student Fellow
[15] Hubble Fellow
[16] NASA Sagan Fellow
[*] bfulton@hawaii.edu

[16] NASA Exoplanet Archive, 2/27/2017

[17] The false positive probability for the majority of the *Kepler* candidates is 5–10% (Morton & Johnson 2011).

growing to massive, gas-rich giant planets. Modern formation models are now able to reproduce the observed population of super-Earths (Hansen & Murray 2012; Mordasini et al. 2012; Alibert et al. 2013; Chiang & Laughlin 2013; Lee et al. 2014; Chatterjee & Tan 2014; Coleman & Nelson 2014; Raymond & Cossou 2014; Lee & Chiang 2016). Many of these new models can be corroborated by measuring the bulk properties of individual planets and the typical properties of the population.

As formation models continue to be refined, the role of atmospheric erosion on these short-period planets is becoming more apparent. Several authors have predicted the existence of a "photoevaporation valley" in the distribution of planet radii (e.g., Owen & Wu 2013; Lopez & Fortney 2014; Jin et al. 2014; Chen & Rogers 2016; Lopez & Rice 2016).

Photoevaporation models predict that there should be a dearth of intermediate sub-Neptune size planets orbiting in highly irradiated environments. The mass of H/He in the envelope must be finely tuned to produce a planet in this intermediate size range. Planets with too little gas in their envelopes are stripped to bare, rocky cores by the radiation from their host stars. In general, the radii of bare, rocky cores versus planets with a few percent by mass H/He envelopes depend on many uncertain variables such as the initial core mass distribution and the insolation flux received by the planet. A rift in the distribution of small planet radii is a common result of the planet formation models that include photoevaporation.

Owen & Wu (2013) provided tentative observational evidence for such a feature in the radius distribution of *Kepler* planets. They observed a bimodal structure in the planet radius distribution, particularly when the planet sample was split into subsamples with low and high integrated X-ray exposure histories. However, the relatively large planet radius uncertainties in Owen & Wu (2013) diluted the gap and reduced its statistical significance. Their study also considered the number distribution of planets, and was not corrected for completeness as we do below. Such corrections mitigate sample bias and allow for the recovery of the underlying planet distribution from the observed one.

Here, we examine a sample of planets orbiting stars with precisely measured radii from the California-Kepler Survey (CKS; see Petigura et al. (2017) and Johnson et al. (2017)). We use the precise stellar radii to update the planet radii, bringing the distribution of planet radii into sharper focus and revealing a gap between 1.5 and 2.0 $R_\oplus$.

This paper is structured as follows. In §2 we discuss our stellar and planetary samples. We describe our methods for correcting for pipeline search sensitivity and transit probabilities in §3. In §4 we examine the one-dimensional marginalized radius distribution and also two-dimensional distributions of planet radius as a function of orbital period, stellar radius, and insolation flux. We discuss potential explanations for the observed planet radius gap in §5 and finish with some concluding remarks in §6.

2. SAMPLE OF PLANETS

2.1. *California Kepler Survey*

For this work we adopt the stellar sample and the measured stellar parameters from the CKS program (Petigura et al. 2017, hereafter Paper I). The measured values of $T_{\rm eff}$, $\log g$, and [Fe/H] are based on a detailed spectroscopic characterization of 1305 *Kepler* Object of Interest (KOI) host stars using observations from Keck/HIRES (Vogt et al. 1994). In Johnson et al. (2017, hereafter Paper II), we associated those stellar parameters from Paper I to Dartmouth isochrones (Dotter et al. 2008) to derive improved stellar radii and masses, allowing us to recalculate planetary radii using the light curve parameters from Mullally et al. (2015), hereafter "Q16". Median uncertainties in stellar radius improve from 25% (Huber et al. 2014) to 11% after our CKS spectroscopic analysis. Stellar mass uncertainties improve from 14% to 4% in the Paper II catalog. This leads to median uncertainties in planet radii of 12% which enable the detection of finer structures in the planet radius distribution.

2.2. *Sample Selection*

The CKS stellar sample was constructed to address a variety of science topics (Paper I). The core sample is a magnitude-limited set of KOIs ($Kp < 14.2$). Additional fainter stars were added to include habitable zone planets, ultra-short-period planets, and multi-planet systems. Here, we enumerate a list of cuts in parameter space designed to create a sample of planets with well-measured radii and with well-quantified detection completeness. The primary goal is to determine anew the occurrence of planets as a function of planet radius, with greater reliability than was previously possible.

We start by removing planet candidates deemed false positives in Paper I. The Paper I false positive designations were determined using the false positive probabilities calculated by Morton & Johnson (2011); Morton (2012); Morton et al. (2016), the *Kepler* team's designation available on the NASA Exoplanet Archive, and a search for secondary lines in the HIRES spectra (Kolbl et al. 2015) as well as any other information available in the literature for individual KOIs. Next, we restrict our sample to only the magnitude-limited portion of the larger CKS sample ($Kp < 14.2$).

The planet-to-star radius ratio ($R_P/R_\star$) becomes uncertain at high impact parameters (b) due to degeneracies with limb-darkening. We excluded KOIs with $b > 0.7$ to minimize the impact of grazing geometries. We experimented other thresholds in b and found that our results are relatively insensititve to $b < 0.6$, 0.7, or 0.8, with the trade-off of smaller sample size with decreasing threshold in b.

We removed planets with orbital periods longer than 100 days in order to avoid domains of low completeness (especially for planets smaller than about 4 $R_\oplus$) and low transit probability.

We also excised planets orbiting evolved stars since they have somewhat lower detectability and less certain radii. This was implemented using an *ad hoc* temperature-dependent stellar radius filter,

$$\frac{R_\star}{R_\odot} > 10^{0.00025(T_{\rm eff}/{\rm K}-5500)+0.20}, \quad (1)$$

which is plotted in Figure 1. We also restricted our sample to planets orbiting stars within the temperature range

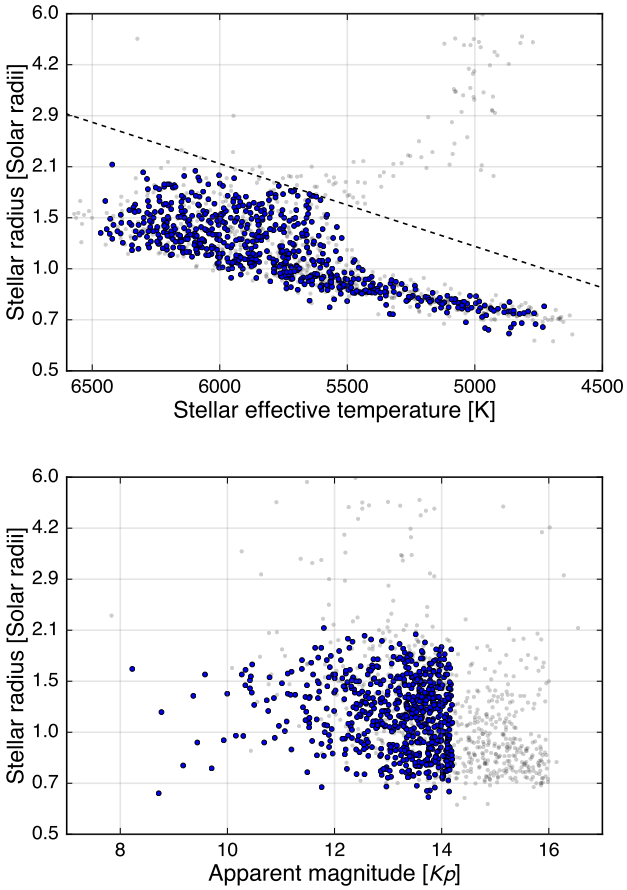

FIG. 1.— *Top:* HR diagram of the sample of stars selected for analysis. The full Paper II sample is plotted in light grey points and the sample selected for analysis after applying the filters discussed in Section 2.2 are plotted as blue squares. Giant planet hosting stars that fall above the dashed line given by Equation 1 are omitted from the final sample. *Bottom:* Stellar radius of CKS stars as a function of *Kepler* magnitude (*Kp*). We note that stars fainter than 14.2 do not follow the same stellar radius distribution. We omit stars fainter than $Kp = 14.2$ to avoid biasing our planet radius distribution. The point colors are the same as in the *top* panel.

TABLE 1
DEPTH OF THE GAP

Filter	V_A
Full CKS sample	0.746
False positives removed	0.742
$Kp < 14.2$	0.686
$b < 0.7$	0.572
$P < 100$ d	0.498
Giant stars removed	0.507
$T_{\rm eff}$ = 4700–6500 K	0.483

where we can extract precise stellar parameters from our high resolution optical spectra (6500–4700 K). Finally, we accounted for uncertainties in the completeness corrections caused by systematic and random measurement errors in the simulations, described in Appendix C.

The multiple filters purify the CKS sample of stars and planets and are summarized in Figure 2. We assessed the impact of filters on the depth of the planet radius valley using an *ad hoc* metric V_A. This quantity is defined as the ratio of the number of planets with radii of 1.64–1.97 $R_\oplus$ (the bottom of the valley) to the average number of planets with radii of 1.2–1.44 $R_\oplus$ or 2.16–2.62 $R_\oplus$(the peaks of the distrubtion immediately outside of the valley). The radius limits for the calculation of V_A were chosen so that $V_A = 1$ for a log-uniform distribution of planets with radii between 1.2 $R_\oplus$ and 2.62 $R_\oplus$. Smaller values of V_A denote a deeper valley. The values of V_A after applying each successive filter are tabulated in Table 1.

Furlan et al. (2017) compiled a catalog of KOI host stars that were observed using a collection of high-resolution imaging facilities (Lillo-Box et al. 2012, 2014; Horch et al. 2012, 2014; Everett et al. 2015; Gilliland et al. 2015; Cartier et al. 2015; Wang et al. 2015a,b; Adams et al. 2012, 2013; Dressing et al. 2014; Law et al. 2014; Baranec et al. 2016; Howell et al. 2011). Many of the 1902 KOIs in the Furlan et al. (2017) catalog also appear in our sample. We investigated removing KOI hosts with known companions or large dilution corrections but found no significant changes to the shape of the distribution. Since only a subset of our KOIs were observed by Furlan et al. (2017) and it is difficult to determine the binarity of the parent stellar population for occurrence calculations, we chose not to filter our planet catalog using the results of high-resolution imaging. However, many of these stars may have already been identified as false positives in the Paper I catalog and therefore removed from our final sample of planets.

We investigated the impact of our apparent magnitude cut by examinging the size distribution for three ranges of *Kp* (Figure 3). For these tests we applied all of the filters described in this section except the $Kp < 14.2$ magnitude cut. We found that the planet radius distribution for $Kp < 13.5$ is statistically indistinguishable from the radius distribution for planets orbiting stars with $13.5 < Kp \leq 14.2$. An Anderson-Darling test (Anderson & Darling 1952; Scholz & Stephens 1987) predicts that the two distributions were drawn from the same parent population with a p-value of 0.6. However, the radius distribution of planets orbiting host stars with $Kp \geq 14.2$ is visually and statistically different (p-value < 0.0004). This is somewhat expected given the non-systematic target selection for both the initial *Kepler* target stars and the stars observed in the CKS survey. Stars with $Kp > 14.2$ were only observed in the CKS program because they were hosts to multi-planet systems, habitable-zone candidates, ultra-short period planets, or other special cases. Targets fainter than $Kp > 14.0$ were observed by *Kepler* only if their stellar and noise properties indicated that there was a high probability of the detection of small planets (Batalha et al. 2010). These non-uniform *Kepler* target selection effects motivate our choice to exclude faint stars. The final distributions of planet radii do not depend on the $Kp < 14.2$ or $Kp < 14.0$ (p-value > 0.95) choice. But there are 153 planet candidates with $14.0 < Kp < 14.2$ so we choose to include those additional candidates to maximize the statistical power of the final sample.

The two distinct peaks separated by a valley (Figure 2) are apparent in the initial number distribution of planet radii and the final distribution after the filters are ap-

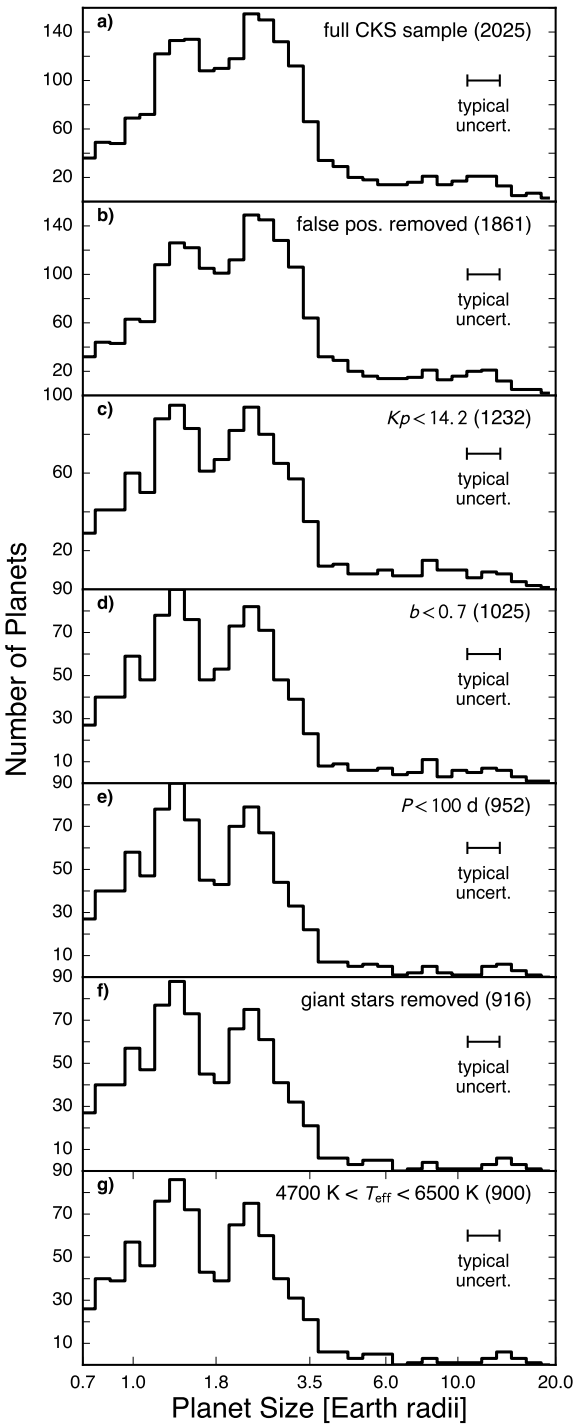

FIG. 2.— **(a)** Size distribution of all planet candidates in the CKS planet sample. Panels **(b)**–**(g)** show the radius distribution after applying several successive cuts to **(b)**: remove known false positives, **(c)**: keep candidates orbiting bright stars ($Kp < 14.2$), **(d)**: retain candidates with low impact parameters ($b < 0.7$), **(e)**: keep candidates with orbital periods shorter than 100 days, **(f)**: remove candidates orbiting giant host stars, and **(g)**: include only candidates orbiting stars within our adopted $T_{\rm eff}$ range (4700 K < $T_{\rm eff}$ < 6500 K). The number of planets remaining after applying each successive filter is annotated in the upper right portion of each panel. Our filters produce a reliable sample of accurate planet radii and accentuate the deficit of planets at 1.8 $R_\oplus$.

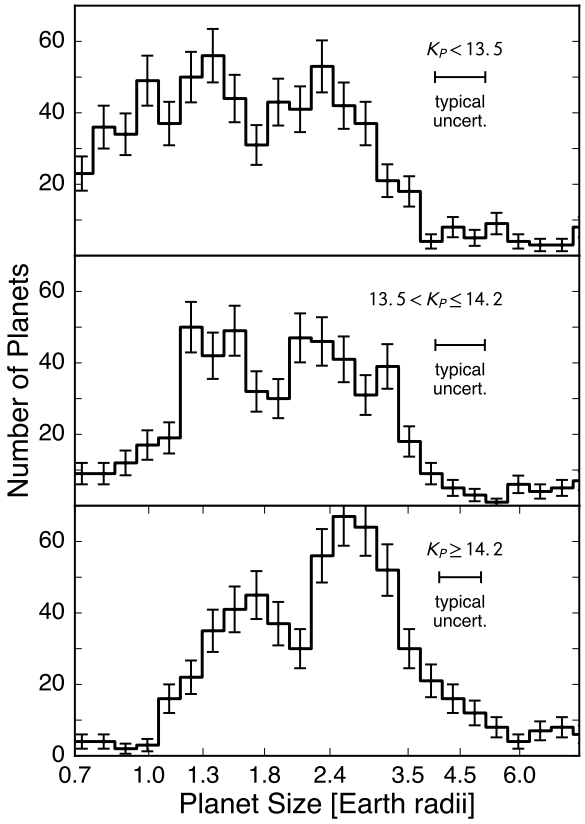

FIG. 3.— Histograms of planet radii broken up into the three magnitude ranges annotated in each panel. All of the filters have been applied to the sample as described in §2.2. The gap is apparent in all magnitude ranges. The distribution of planet radii in the two brightest magnitude ranges are indistinguishable (p-value = 0.6). However, the planets orbiting stars with $Kp > 14.2$ are statistically different (p-value = 0.0004) when compared to the Kp = 13.5–14.2 magnitude range. This is expected due to the non-systematic nature of the target selection for CKS and KIC stars fainter than $Kp = 14.2$. This motivates our removal of planets with hosts fainter than $Kp = 14.2$.

plied. The depth of the valley increases as we apply these filters, suggesting that the purity of the planet sample improves with filter application. Note that the filters act on the stellar characteristics and are agnostic to planet radius.

Figure 4 shows histograms of the stellar radii and planet-to-star radius ratios ($R_P/R_\star$) for the filtered sample stars. These two distributions are both unimodel. This demonstrates that the bimodality of the planet radius distribution is not an artifact of the stellar sample or the light curve fitting used to measure $R_P/R_\star$.

3. COMPLETENESS CORRECTIONS

To recover the underlying planet radius distribution from the observed distribution we made completeness corrections to compensate for decreasing detectability of planets with small radii and/or long orbital periods.

An additional complication associated with the completeness corrections in this work is that the stellar properties of the planet-hosting stars come from a different source and have higher precision than the stellar properties for the full set of *Kepler* target stars. We explore the additional uncertainties introduced by this fact by running a suite of simulated transit surveys described

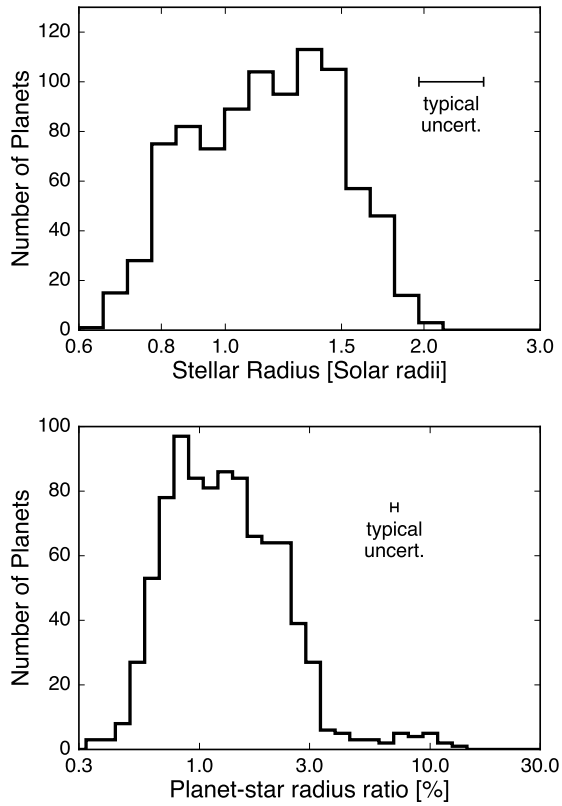

FIG. 4.— *Top:* Histogram of stellar radii derived in Paper II and used to update planet radii in this work after the filters described in Section 2.2 are applied. *Bottom:* Histogram of planet-to-star radii ratios for the stars remaining after the filters described in Section 2.2 are applied to the full Paper II sample of planet candidates. In both cases, the median measurement uncertainties are plotted in the upper right. Neither of these two histograms shows the same bimodal feature that is observed in the planet radius distribution, which demonstrates that the feature is not an artifact of our stellar sample or transit fitting.

in Appendix C. We inflate the uncertainties on the histogram bin heights by the scaling factors listed in Table C.1 to account for these effects.

3.1. *Pipeline Efficiency*

We followed the procedure described in Christiansen et al. (2016) using the results from their injection-recovery experiments (Christiansen et al. 2015). They injected about ten-thousand transit signals into the raw pixel data and processed the results with version 9.1 of the official *Kepler* pipeline (Jenkins et al. 2010). These completeness tests were used to identify combinations of transit light curve parameters that could be recovered by the *Kepler* pipeline for a given sample of target stars. They injected signals onto both target stars and neighboring pixels to quantify the pipeline's ability to identify astrophysical false positives. We assumed that our sample is free of the vast majority of false positives so we only considered injections of transits onto the target stars. We only considered injections on stars that would have been included in the CKS sample and would not be removed by the filters described in §2.2. Namely, we considered injected impact parameters less than 0.7, injected periods shorter than 100 days, $Kp \leq 14.2$, 4700 K $< T_{\rm eff} <$ 6500 K, and stellar radii compatible with Equation 1 based on the values in the Stellar17 catalog[18] prepared by the *Kepler* stellar parameters working group (Mathur et al. 2016). This leaves a total of 3840 synthetic transit signals injected onto the target pixels of 3840 stars observed by *Kepler*. We also apply these same filters to the stars in the Stellar17 catalog. The number of stars remaining after the filters are applied is the number of stars observed by *Kepler* that could have led to detections of planets that would be present in our filtered planet catalog ($N_\star = 36{,}075$). We calculated the fraction of injected signals recovered as a function of injected signal-to-noise as

$$m_i = \left(\frac{R_P}{R_{\star,i}}\right)^2 \sqrt{\frac{T_{{\rm obs},i}}{P}} \left(\frac{1}{{\rm CDPP}_{{\rm dur},i}}\right), \qquad (2)$$

where R_P and P are the radius and period of the particular injected planet. $R_{\star,i}$ is the stellar radius for the $i^{\rm th}$ star in the Stellar17 catalog, $T_{{\rm obs},i}$ is the amount of time that the particular star was observed, and ${\rm CDPP}_{{\rm dur},i}$ is the Combined Differential Photometric Precision (CDDP, Koch et al. 2010) value for each star extrapolated to the transit duration for each injection. We fit a 2$^{\rm nd}$ order polynomial in $1/\sqrt{d}$ to the $d = 3$, 6, and 12-hour CDPP values for each star to perform the extrapolation (Sinukoff et al. 2013).

We fit a Γ cumulative distribution function (CDF) to the recovery fraction vs. injected (m_i) of the form

$$C(m_i; k, \theta, l) = \Gamma(k) \int_0^{\frac{m_i - l}{\theta}} t^{k-1} e^{-t} dt, \qquad (3)$$

to derive the average pipeline efficiency. $C(m_i)$ is the probability that a signal with a given value of m_i would actually be detected by the *Kepler* transit search pipeline. In practice we used the `scipy.stats.gammacdf`(t, k, l, θ) function in SciPy version 0.18.1. Using the `lmfit` Python package (Newville et al. 2014) to minimize the residuals we found best-fit values of $k = 17.56$, $l = 1.00$ (fixed), and $\theta = 0.49$. Figure 5 shows the fraction of injections recovered as a function of m_i and our model for pipeline efficiency.

Our pipeline efficiency curve is ∼15-25% lower than the efficiency as a function of the *Kepler* multi-event statistic (MES) derived in (Christiansen et al. 2015) for their FGK subsample. The difference can be explained by the fact that the MES is estimated in the *Kepler* pipeline during a multidimensional grid search. In most cases, the search grid is not fine enough to find the exact period and transit time for a given planet candidate. Since the grid search doesn't find the best-fit transit model it generally underestimates the SNR (m_i) by a factor of ∼25% (Petigura et al., in preparation).

3.2. *Survey Sensitivity*

For each planet detection there are a number of similar planets that would not have been detected due to a lack of sensitivity or unfavorable geometric transit probability. To compensate, we weighted each planet detection by the inverse of these probabilities,

$$w_i = \frac{1}{(p_{\rm det} \cdot p_{\rm tr})}, \qquad (4)$$

[18] https://archive.stsci.edu/kepler/stellar17/search.php

TABLE 2
Planet Detection Statistics

Planet candidate	P d	R_P $R_\oplus$	SNR m_i	Detection probability $p_{\rm det}$	Transit probability $p_{\rm det}$	Weight $1/w_i$
K00002.01	2.20	13.41	750.22	1.00	0.14	6.94
K00003.01	4.89	5.11	877.10	1.00	0.05	20.14
K00007.01	3.21	4.13	146.38	1.00	0.11	8.88
K00010.01	3.52	13.39	914.62	1.00	0.09	11.06
K00017.01	3.23	15.04	1212.38	1.00	0.11	9.40
K00018.01	3.55	13.94	820.96	1.00	0.10	9.58
K00020.01	4.44	21.41	1469.42	1.00	0.10	10.15
K00022.01	7.89	14.20	1085.97	1.00	0.06	17.98
K00041.01	12.82	2.37	37.15	0.98	0.05	22.37
K00041.02	6.89	1.35	15.04	0.91	0.07	15.98

Note. — Table 2 is available in its entirety in machine-readable format, which also includes period and radius uncertainties. A portion is shown here for guidance regarding its form and content. Refer to Paper II for the CKS stellar parameters associated with each KOI. This table contains only the subset of planet detections that passed the filters described in §2.2. The full sample of planet candidates orbiting CKS target stars can be found in Paper II.

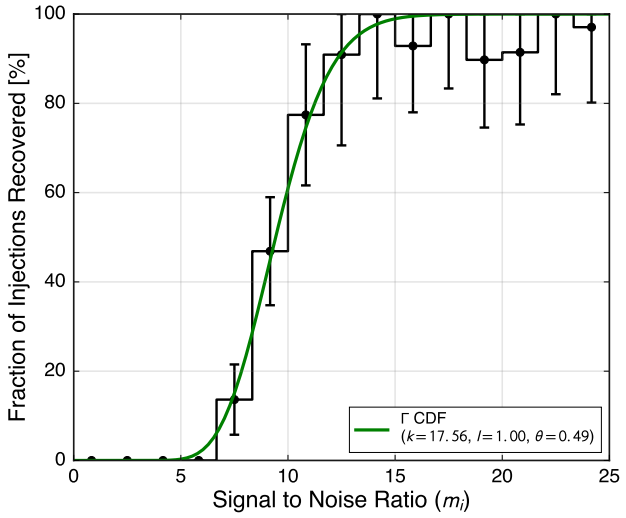

Fig. 5.— Fraction of injected transit signals recovered as a function of signal to noise ratio (m_i, Equation 2) in our subsample of the *Kepler* target stars using the injection recovery tests from Christiansen et al. (2015). We fit a Γ CDF (Equation 3) and plot the best-fit model in green.

where $p_{\rm det}$ is the fraction of stars in our sample where a transiting planet with a given signal to noise ratio given by Equation 2 could be detected:

$$p_{\rm det} = \frac{1}{N_\star} \sum_{i}^{N_\star} C(m_i). \quad (5)$$

The geometric transit probability is $p_{\rm tr} = 0.7 R_\star / a$. The factor of 0.7 compensates for our omission of planet detections with $b > 0.7$ from the planet catalog. Figure 6 shows the mean pipeline completeness ($p_{\rm det}$) and mean total search completeness ($1/w_i$) as a function of planet radius and orbital period for the filtered Stellar17 sample of *Kepler* target stars. The detection probabilites, transit probabilities, and weights (w_i) for each planet in our final catalog are listed in Table 2.

3.3. *Occurrence Calculation*

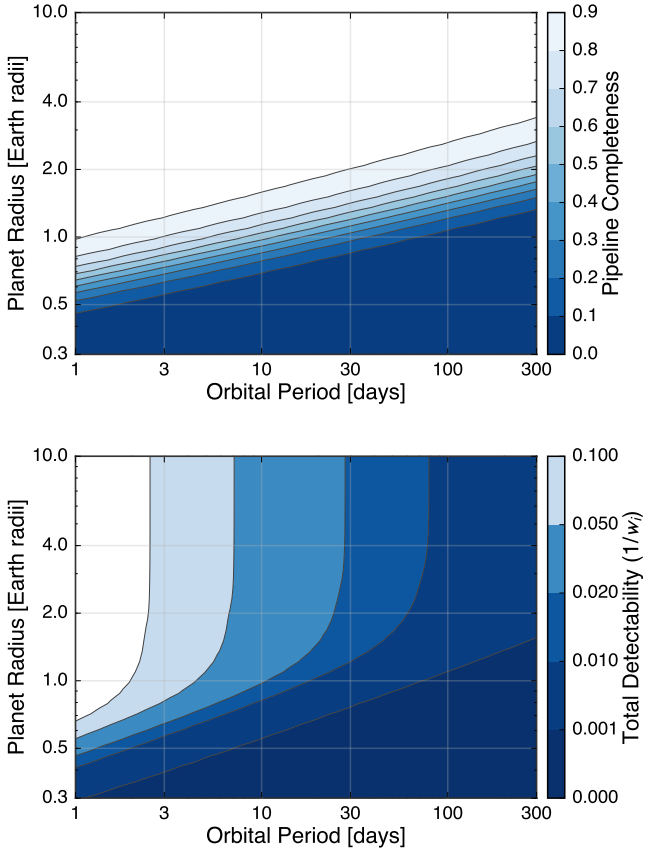

Fig. 6.— *Top:* Mean survey completeness for transiting planets orbiting the stars in our sample ($p_{\rm det}$). *Bottom:* Mean survey completeness for all planets orbiting stars in our sample ($p_{\rm det} \cdot p_{\rm tr}$).

Following the definitions in Petigura et al. (2013a), the average planet occurrence rate (number of planets per star) for any discrete bin in planet radius or orbital period is the sum of these weights divided by the total number of stars in the sample ($N_\star$):

$$f_{\rm bin} = \frac{1}{N_\star} \sum_{i=1}^{n_{\rm pl,bin}} w_i. \quad (6)$$

TABLE 3
Planet Occurrence

Radius bin $R_\oplus$	Number of planets per star $f_{\rm bin}$ for $P < 100$ d
1.16–1.29	0.078 ± 0.017
1.29–1.43	0.08 ± 0.013
1.43–1.59	0.053 ± 0.011
1.59–1.77	0.0334 ± 0.0092
1.77–1.97	0.05 ± 0.01
1.97–2.19	0.086 ± 0.016
2.19–2.43	0.098 ± 0.016
2.43–2.70	0.077 ± 0.016
2.70–3.00	0.053 ± 0.012
3.00–3.33	0.0316 ± 0.0089
3.33–3.70	0.0242 ± 0.0066
3.70–4.12	0.0094 ± 0.0057
4.12–4.57	0.0056 ± 0.0034
4.57–5.08	0.0037 ± 0.0031
5.08–5.65	0.0066 ± 0.0048
5.65–6.27	0.005 ± 0.003
6.27–6.97	$0.0 \pm inf$
6.97–7.75	0.0019 ± 0.0029
7.75–8.61	0.0044 ± 0.0034
8.61–9.56	0.00022 ± 0.00032
9.56–10.63	0.001 ± 0.0015
10.63–11.81	0.00035 ± 0.00053
11.81–13.12	0.00104 ± 0.00094
13.12–14.58	0.0038 ± 0.0021
14.58–16.20	0.00084 ± 0.00066
16.20–18.00	0.0003 ± 0.0004

Again, $N_\star = 36,075$ is the total number of dwarf stars in the Stellar17 catalog that pass the same filters on stellar parameters that were applied to the planet catalog: no giant stars (selected using Equation 1), 4700 K $< T_{\rm eff} <$ 6500 K, and $Kp \leq 14.2$.

4. THE PLANET RADIUS GAP

Figure 7 shows the completeness-corrected distribution of planet radii for the filtered sample of 900 planets and the corresponding occurrence values are tabulated in Table 3. Uncertainties on the bin heights are calculated using Poisson statistics on the number of detections within the bin, scaled by the size of the completeness correction in each bin, and scaled again by a correction factor determined from a collection of simulated transit surveys as described in Section C. The completeness corrections are generally small. We are sensitive to $> 80\%$ of 2.0 $R_\oplus$ planets out to orbital periods of 100 days, and $> 50\%$ of 1.0 $R_\oplus$ planets out to 30 days (Figure 6). The transit probability term in Equation 4 dominates the corrections in most of the parameter space explored. Somewhat surprisingly, the larger, sub-Neptunes receive a completeness boost that is larger than the boost received by the smaller, super-Earths (compare the dotted grey line in Figure 7 to the solid black line) because the sub-Neptunes tend to orbit at larger orbital distances where transit probabilities are smaller. The mean transit probability ($p_{\rm tr}$) for planets with radii of 1.0–1.75 $R_\oplus$ in our sample is 6% while the transit probability for planets with radii of 1.75–3.5 $R_\oplus$ is a factor of two lower (3%). However, the mean detectability ($p_{\rm det}$) for those same two classes of planets are both very high at 86% and 96% respectively.

4.1. *Comparison with Log-Uniform Distribution*

TABLE 4
Spline Fit

Node Location $R_\oplus$	Best-fit Value ($f_{\rm bin}$)	1 σ Credible Interval ($f_{\rm bin}$)
1.3	0.078	fixed
1.5	0.051	0.05 ± 0.02
1.9	0.030	0.03 ± 0.02
2.4	0.116	0.11 ± 0.01
3.0	0.043	0.044 ± 0.005
4.5	0.0050	0.005 ± 0.002
11.0	0.00050	0.0005 ± 0.0003

We performed several tests to quantify the significance of the gap in the planet radius distribution. First, we performed a two-sided Kolmogorov-Smirnov (K-S, Kolmogorov 1933; Smirnov 1948) test to assess the probability that the planet radius number distribution for radii in the range 1–3 $R_\oplus$ is drawn from a log-uniform distribution. This test returns a probability of 0.003 that the planet radii between 1–3 $R_\oplus$ are drawn from a log-uniform distribution. However, we note that blind interpretation of p-values from K-S tests can often lead to overestimates of significance (Babu & Feigelson 2006). Similarly, an Anderson-Darling test also rejects the hypothesis that the planet radii between 1–3 $R_\oplus$ were drawn from a log-uniform distribution with a p-value of 0.012.

4.2. *Dip Test of Multimodality*

Hartigan's dip test is a statistical tool used to estimate the probability that a sample was drawn from a unimodal distribution or a multi-modal distribution with $\geq$2 modes (Hartigan & Hartigan 1985). It is similar to the K-S statistic in that it measures the maximum distance between an empirical distribution and a unimodal distribution. Applying this test to the number distribution of $\log R_P$ for planet radii in the range 1–3 $R_\oplus$ returns a p-value of 1.4×10^{-3} that the distribution was drawn from a unimodal distribution. This strongly suggests that the planet radius distribution is multi-modal.

4.3. *Spline Model*

Modeling the planet radius distribution with splines having nodes at fixed values gives a good fit for a range of planet sizes. Virtues of this model are the small number of free parameters and model flexibility, particularly in asymptotic regions where others models (e.g. Gaussians) force the distribution to zero. We fit a second-order spline with seven node points fixed at specific radii to the weighted histogram of planet occurrence. We excluded from the fit bins for radii smaller than 1.14 $R_\oplus$ where the pipeline completeness at $P = 100$ days is less than 25%. The model was adjusted by varying the amplitudes of the spline nodes, then convolving with a Gaussian kernel whose width is the median fractional planet radius uncertainty (12%). The convolved model is averaged over each of the histogram bins before performing the χ^2 comparison. This allows us to separate the smearing of the observed distribution due to measurement uncertainties from a "deconvolved" view of the underlying distribution. Again we found the best-fit solution using the `lmfit` package to minimize the normalized residuals of the histogram bins relative to the convolved model.

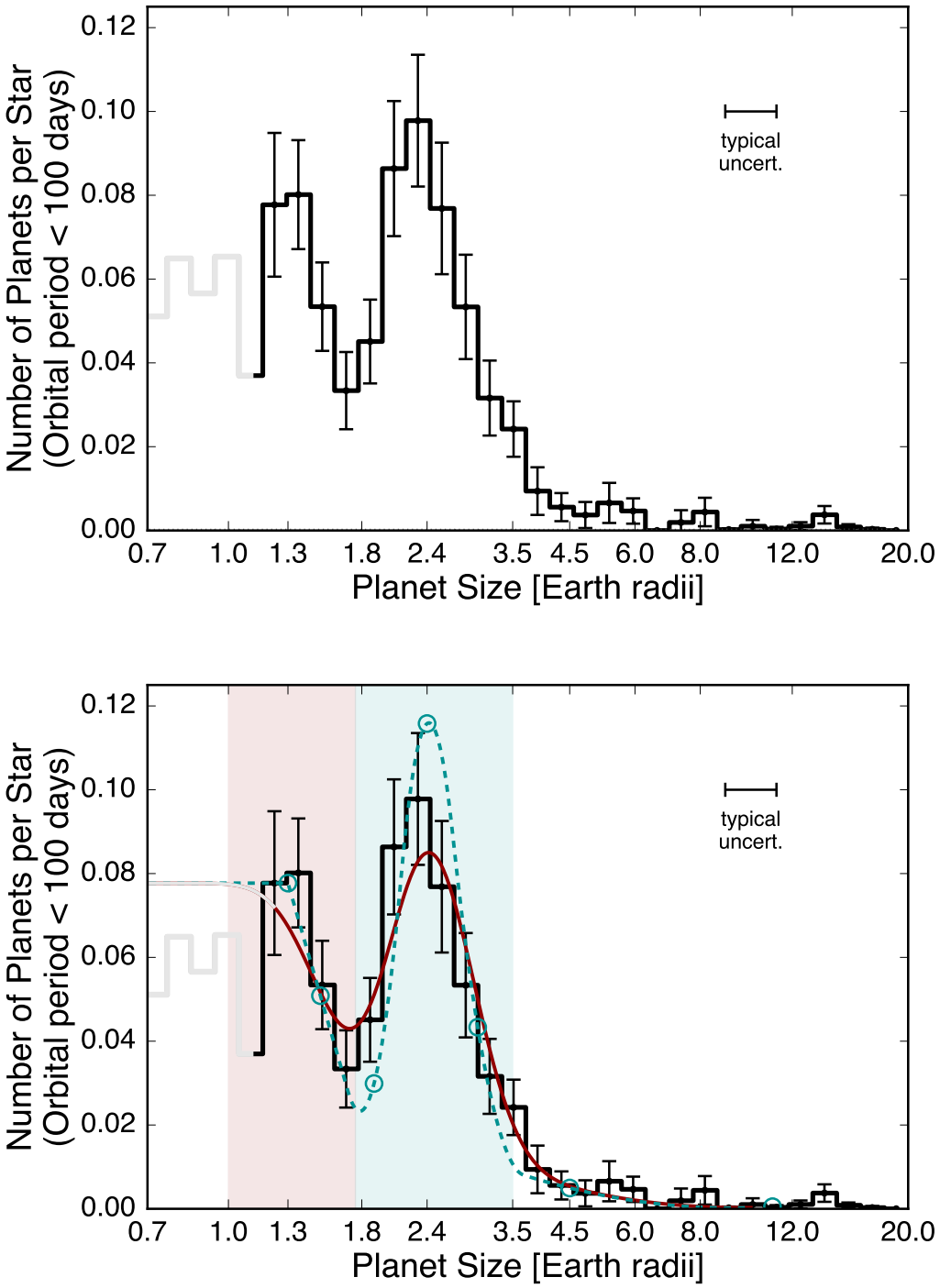

FIG. 7.— *Top:* Completeness-corrected histogram of planet radii for planets with orbital periods shorter than 100 days. Uncertainties in the bin amplitudes are calculated using the suite of simulated surveys described in Section C. The light gray region of the histogram for radii smaller than 1.14 $R_\oplus$ suffers from low completeness. The histogram plotted in the dotted grey line is the same distribution of planet radii uncorrected for completeness. The median radius uncertainty is plotted in the upper right portion of the plot. *Bottom:* Same as top panel with the best-fit spline model over-plotted in the solid dark red line. The region of the histogram plotted in light grey is not included in the fit due to low completeness. Lightly shaded regions encompass our definitions of "super-Earths" (light red) and "sub-Neptunes" (light cyan). The dashed cyan line is a plausible model for the underlying occurrence distribution after removing the smearing caused by uncertainties on the planet radii measurements. The cyan circles on the dashed cyan line mark the node positions and values from the spline fit described in §4.3.

We used the `emcee` (Foreman-Mackey et al. 2013) interface built into `lmfit` to estimate the uncertainties on the node values. We performed the fits working in $\log(R_P)$ and required positive occurrence values for the deconvolved model. For radii outside of the range spanned by our node locations, we extrapolated assuming constant (log-uniform) occurrence.

Deconvolution is an inherently unstable process and we caution against over-interpretation of the deconvolved model. Our best-fit deconvolved model is not the only solution that could produce an equivalent convolved model. The deconvolved model is also somewhat sensitive to the choice of the node locations, while the convolved model is insensitive to those choices. However, the deconvolved model suggests that the gap is likely deeper than observed. This motivates detailed follow-up and characterization of the planets that fall within the gap. The best-fit model (red line) and deconvolved model (dashed cyan line) are both over-plotted on the completeness-corrected planet radius distribution in Figure 7. Table 4 lists the locations, best-fit values, and 1σ credible intervals for the spline nodes.

4.4. *Relative Frequency of Super-Earths and Sub-Neptunes*

Many authors use the terms "super-Earth" and "sub-Neptune" interchangeably, or draw arbitrary distinctions in mass or radius between these two classes. The observed gap in the radius distribution of small planets suggests a less arbitrary division. In the text below we define a "super-Earth" as a planet with a radius of 1-1.75 $R_\oplus$, and a "sub-Neptune" as having a radius of 1.75–3.5 $R_\oplus$.

We calculated the occurrence ratio of super-Earths to sub-Neptunes to be 0.8 ± 0.2. The uncertainty is determined using a suite of simulated surveys described in Appendix C. The nearly equal occurrence of super-Earths and sub-Neptunes with $P < 100$ days provides an important constraint for planet formation models. This is likely a lower limit on this ratio since the super-Earth domain likely extends to sizes smaller than 1.1 $R_\oplus$.

4.5. *Two-Dimensional Weighted Kernel Density Estimation*

In the following subsections we present and discuss several contour plots. The contours were derived using the Weighted Kernel Density Estimation (wKDE) technique described in Appendix B and have all been corrected for completeness (with the exception of Figure 9). We calculated bi-variate Gaussians for each pair of planet parameters over a fixed high-resolution grid in the two parameters, sum these Gaussians over all planets, and divide by the total number of stars in the sample ($N_\star$=36,075). Each bi-variate Gaussian is normalized to have a maximum value of 1.0, then multiplied by the weight associated with the given planet detection (w_i, Equation 4). The points plotted are the CKS parameters.

4.5.1. *Planet Radius vs. Orbital Period*

We first look at the distribution of planet radii as a function of orbital period (P). Figure 8 shows the distribution of planet radii as a function of orbital period for planet and stellar parameters from the Q16 catalog (*top panel*). It also offers a comparison with the same distribution derived from the CKS parameters (*bottom panel*).

There is a declining number of small planet detections going toward long orbital periods. However the underlying completeness-corrected contours suggest that the occurrence rate of these planets does not fall off with the number of detections. Instead, the lack of detections is likely an artifact of decreasing transit detectability and probability.

Figure 8 shows that small planets are significantly more common than large planets. The fact that planets smaller than Neptune (4 $R_\oplus$) are much more common than Jovian-size planets has been well documented in the literature (e.g. Howard et al. (2010); Mayor et al. (2011); Howard et al. (2012); Fressin et al. (2013); Dong & Zhu (2013); Petigura et al. (2013a); Dressing & Charbonneau (2015); Burke et al. (2015)). However, the increase in occurrence with decreasing planet size is evidently more rapid than was apparent in previous studies.

There is another feature in the R_P vs. P occurrence distribution that motivates a closer examination of the planet radius distribution along other axes. There are very few planets larger than 2 $R_\oplus$ with orbital periods shorter than about 10 days while planets with radii smaller than 1.8 $R_\oplus$ remain quite common down to orbital periods of about 3 days. A sharp decline in the occurrence rate of planets larger than approximately 1.6 $R_\oplus$ with orbital periods shorter than 10 days has been previously observed (Howard et al. 2012; Dong & Zhu 2013; Sanchis-Ojeda et al. 2014).

4.5.2. *Planet Radius vs. Stellar Radius*

Figure 9 shows the distribution of planet size as a function of host star size. This distribution shows two distinct populations of planets with a gap separating them. Planets appear to preferentially fall into two classes, one with radii of $\sim$2.4 $R_\oplus$ and another with radii of $\sim$1.3 $R_\oplus$. Planets with intermediate radii of 1.5-2.0 $R_\oplus$ are comparatively rare. The gap occurs at the same planet radius for all stellar sizes in our sample. The bimodal planet size distribution holds true for planets orbiting stars with radii ranging from 0.7 $R_\odot$ to 2.0 $R_\odot$.

4.5.3. *Planet Radius vs. Incident Flux*

Figure 10 shows the planet radius distribution as a function of incident flux. The two planet populations shear apart in this domain. There is a dearth of sub-Neptunes orbiting in high incident flux environments. This trend is also visible in one-dimensional histograms of planet radii when broken up into groups based on $S_{\rm inc}$ (Figure 11). Most of the planets that contribute to the peak in the marginalized radius distribution at 1.3 $R_\oplus$ are orbiting in environments with $S_{\rm inc} > 200$ $S_\oplus$, while the planets that contribute to the peak at 2.4 $R_\oplus$ experience $S_{\rm inc} < 80$ $S_\oplus$. It is clear that the gap is present even at low incident fluxes and the two-dimensional $S_{\rm inc}$ and period distributions show a potential deepening and/or widening of the gap toward lower incident fluxes. However, we can not determine if the gap radius is dependent on incident flux, or if the break radius is constant as a function of incident flux due to lack of completeness for small planets orbiting in cool environments.

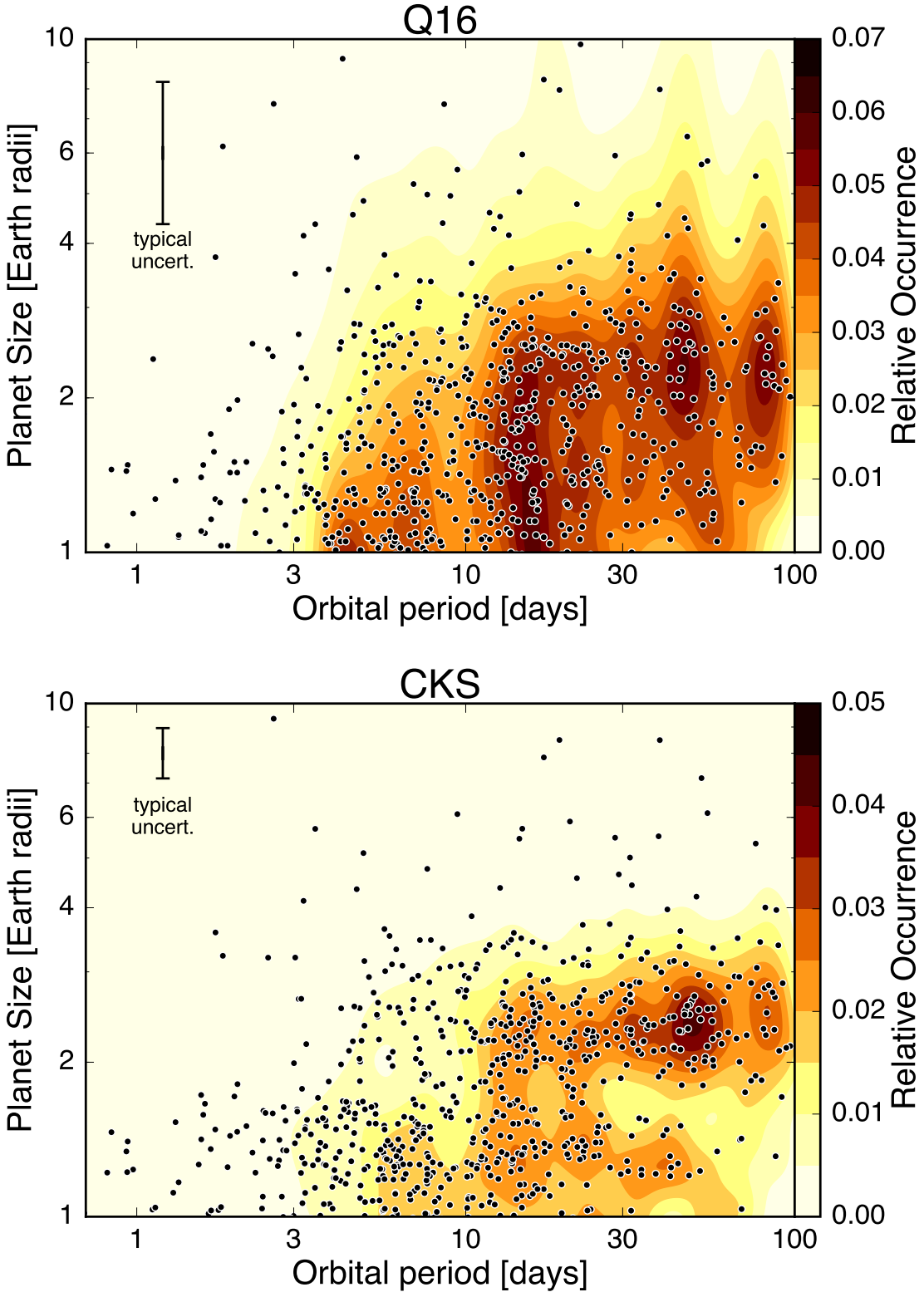

FIG. 8.— *Top:* Two-dimensional planet radius distribution as a function of orbital period using stellar parameters from the Q16 catalog. *Bottom:* Two-dimensional planet radius distribution as a function of orbital period using updated planet parameters from Paper II. In both cases the median uncertainty is plotted in the upper left. Individual planet detections are plotted as black points. The contours are corrected for completeness using the wKDE technique.

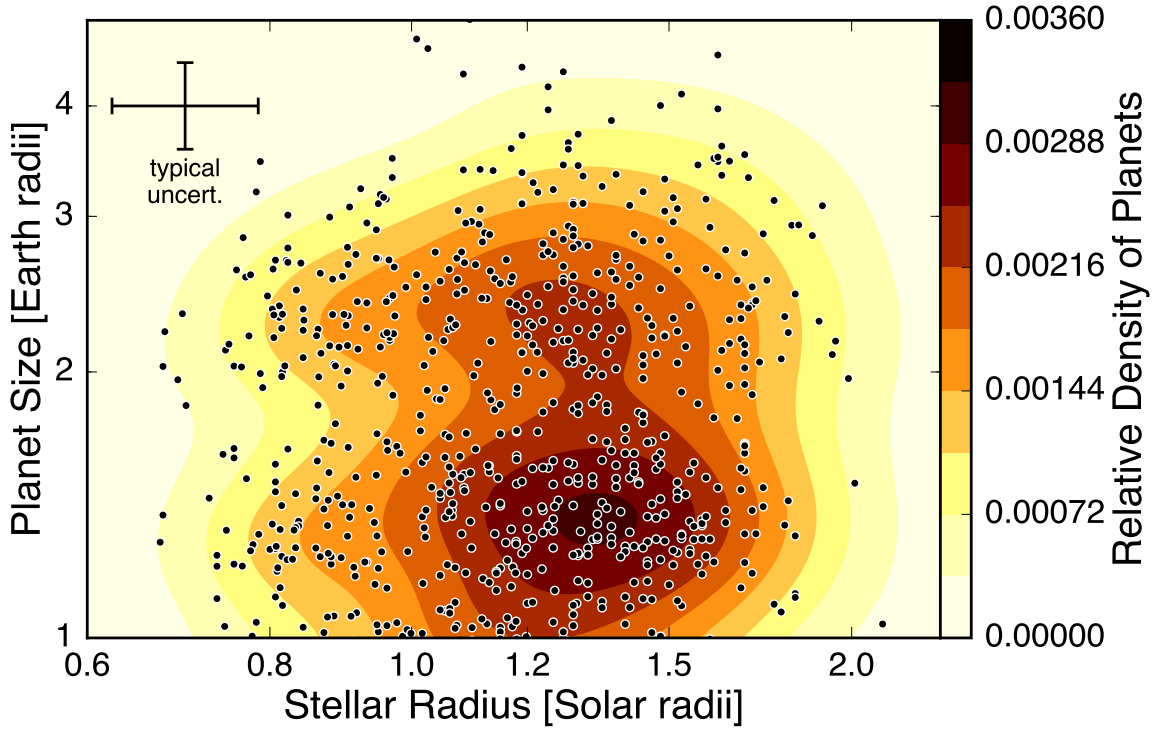

FIG. 9.— Two-dimensional planet radius distribution as a function of stellar radius using updated planet parameters from Paper II. The median uncertainty is plotted in the upper left. Individual planet detections are plotted as black points. The underlying contours are not corrected for completeness. The bifurcation of planet radii is independent of the size of the host star.

There is also an upper envelope of planet size which decreases as a function of incident flux. Although there are a few exceptions, there is a clear dearth of planets in the upper left quadrant of Figure 10. These should be some of the easiest planets to detect yet they do not appear in our sample of planets. This feature has been previously observed (e.g., Howard et al. 2012; Mazeh et al. 2016; Lundkvist et al. 2016) but our larger sample of planets with high-precision host star properties sharpens the boundary. The lack of planet detections in the lower right region of Figure 10 is the result of low survey completeness for small, long-period planets.

5. DISCUSSION

We have provided observational evidence that the distribution of planet sizes is not smooth (Figure 7). Small planets have characteristic sizes of $\sim$1.3 $R_\oplus$ (super-Earths) and $\sim$2.4 $R_\oplus$ (sub-Neptunes). These two planet populations each have intrinsic widths in their size distributions, but there is a gap that separates them. Intermediate-size planets with radii of $\sim$1.5–2.0 $R_\oplus$ are comparatively rare.

5.1. *Previous Studies of the Radius Distribution*

Many studies have examined the planet radius distribution using the *Kepler* sample. To date, none have shown statistically significant evidence for a gap in the distribution at 1.5–2.0 $R_\oplus$.

The pioneering study of Owen & Wu (2013) pointed out a marginally-significant gap at $\sim$1.5–2 $R_\oplus$ in the observed radius distribution and interpreted it as connected to the high-energy irradiation history of the planets. They did not have a large set of accurate planet radii and they did not perform the completeness corrections necessary to confirm the feature. Here, we firmly detect a gap in the planet radius distribution between two peaks at 2.4 $R_\oplus$ and $\leq$ 1.3 $R_\oplus$.

Based on the initial *Kepler* planet catalog, Howard et al. (2012) investigated the domain of planets with $R_P > 2R_\oplus$ and $P < 50$ days. They demonstrated that small planets are common. However, they did not examine the detailed shape of the small planet occurrence function, due to the severe lack of completeness to small planets with the early *Kepler* data releases, and large uncertainties in the planetary radii. At that time, the planetary radii were based on the relatively coarse estimates of the stellar radii from the KIC.

Follow-on studies (Youdin 2011; Catanzarite & Shao 2011; Traub 2012) were similarly limited. Dong & Zhu (2013) benefited from a larger dataset. They focused on the orbital period distribution, with large (factor of two) bins in planet radius. Petigura et al. (2013a) utilized a much longer photometric time series (lasting 15 of 17 *Kepler* quarters), and a custom planet detection pipeline enabling completeness corrections, but the sample was only large enough to allow for three bins in the radius range 1.0–2.8 $R_\oplus$. Silburt et al. (2015) measured occurrence for planets with radii between 1.0 and 4.0 $R_\oplus$ and orbital periods between 20 and 200 days. They found a peak in the distribution near 2.4 $R_\oplus$ and a slight decline in the frequency of smaller planets. More recently, Burke et al. (2015) studied the occurrence of small, long-period planets. With 1σ significance, they observed a diminution in planet occurrence in the 1.5–2.0 $R_\oplus$ interval for

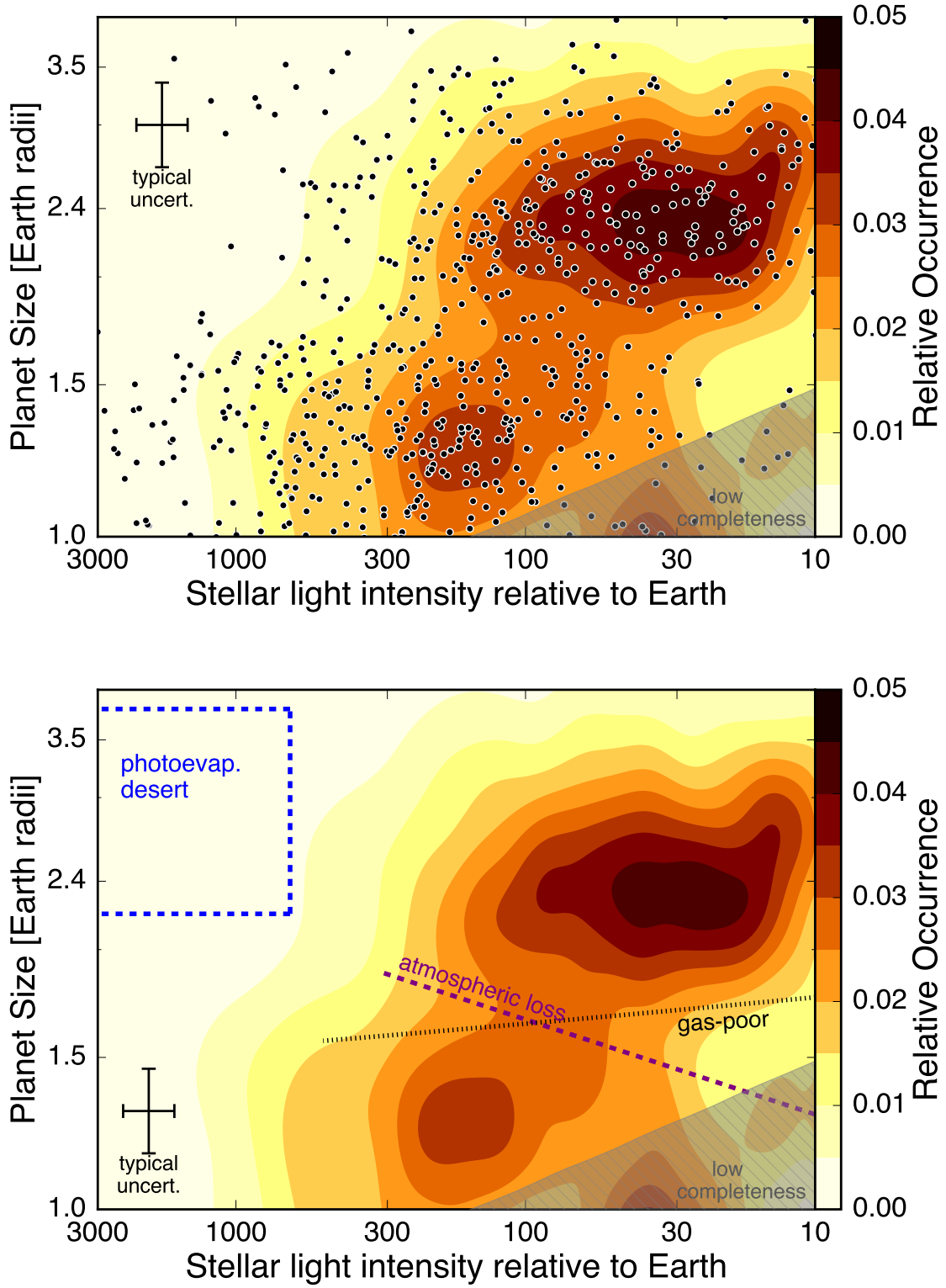

FIG. 10.— *Top:* Two-dimensional distribution of planet size and incident stellar flux. The median uncertainty is plotted in the upper left. There are at least two peaks in the distribution. One class of planets has typical radii of ∼1.3 $R_\oplus$ and generally orbit in environments with $S_{\rm inc}$ >100 $S_\oplus$, while another class of slightly larger planets with typical radii of ∼2.4 $R_\oplus$ orbit in less irradiated environments with $S_{\rm inc}$ < 200 $S_\oplus$. *Bottom:* Same as top panel with individual planet detection points removed, annotations added, and vertical axis scaling changed. The region enclosed by the dashed blue lines marks the photoevaporation desert, or hot-Super Earth desert as defined by Lundkvist et al. (2016). The shaded region in the lower right indicates low completeness. Pipeline completeness in this region is less than 25%. The purple and black lines show the scaling relations for the photoevaproation valley predicted by Lopez & Rice (2016) for scenarios where these planets are the remnant cores of photoevaporated Neptune size planets (dashed purple line) or that these planets are formed at late times in a gas-poor disk (dotted black line).

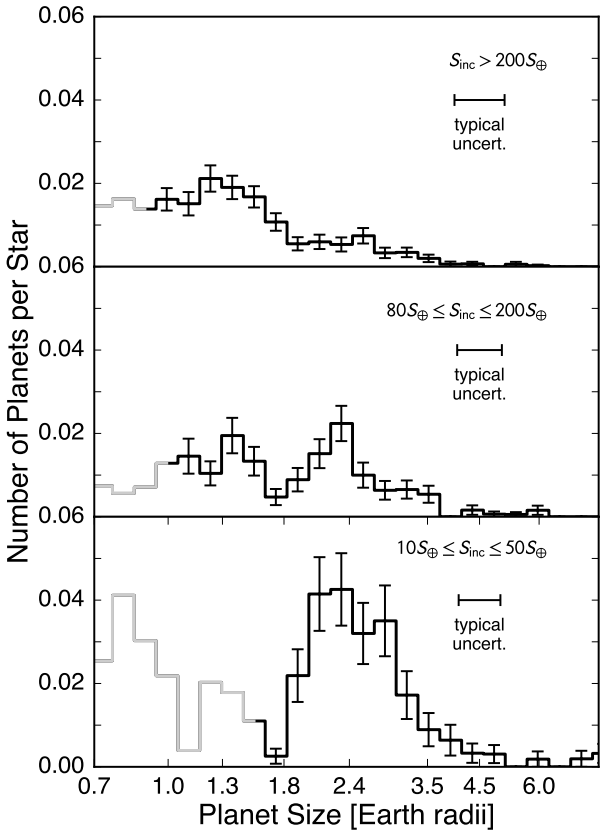

FIG. 11.— Histograms of planet radii broken up into the ranges of incident flux ($S_{\rm inc}$) annotated in the upper right region of each panel. Planets orbiting in environments of higher $S_{\rm inc}$ tend to be smaller than those in low $S_{\rm inc}$ environments. Regions of the histograms plotted in light grey are highly uncertain due to pipeline completeness (<25%).

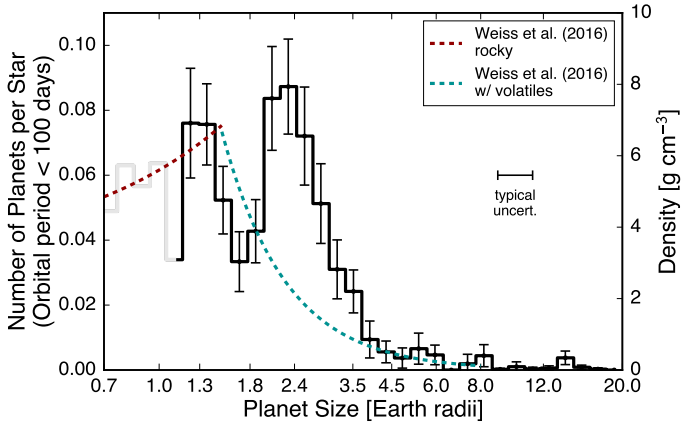

FIG. 12.— An empirical fit to planet radius and mass measurements from (Weiss et al. 2016) over-plotted on the completeness-corrected planet radius distribution derived in this work. The maximum in the planet density fit peaks near the gap in the planet radius distribution.

planets having P = 300–700 days.

5.1.1. *Occurrence Rate Comparisons*

Table 5 compares the occurrence rates measured in this work to those of several touchstone studies from the literature: Howard et al. (2012, H12), Petigura et al. (2013b, P13), Fressin et al. (2013, F13), and Mulders et al. (2015, M15). These works all analyzed Kepler planets, but used catalogs constructed from different amounts of Kepler photometry. In addition, these studies applied different treatments of pipeline completeness, adopted different false positive rates, analyzed different sub-samples of Kepler stars, and accounted for multi-planet systems in different ways. All of these differences can significantly affect the derived occurrence (Burke et al. 2015). However, the relative occurrence rates between bins are insensitive to most of these issues and potential discrepancies in the absolute occurrence rates do not affect the presence or shape of the gap in the radius distribution.

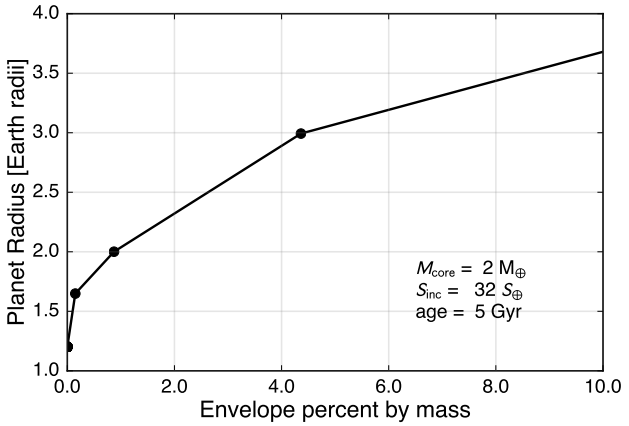

FIG. 13.— Model for planet radius as a function of envelope size from Lopez & Fortney (2014). The final planet radius is plotted for a simulated planet with a 2 $M_\oplus$ core mass that has been irradiated by 32 times the incident flux received by Earth for a period of 5 Gyr. A bare 2 $M_\oplus$ core has a radius of 1.2 $R_\oplus$. Adding an envelope of H/He which is less than 0.2% of the planet's mass inflates the planet to over 1.6 $R_\oplus$. An additional 0.7% envelope by mass inflates the planet to 2 $R_\oplus$.

We choose to closely compare our occurrence values in this work to those of P13, because they used a nearly complete photometric dataset (43/48 months) [19] and corrected for pipeline completeness through direct injection and recovery. Our occurrence rates are typically 50% higher than those of P13. However, P13 (and H12) measured the fraction of stars with planets as opposed to the number of stars per planet measured in this work (and in F13 and M15). The number of planets per star will always be larger than the fraction of stars with planets due to multi-planet systems. P13 estimated that their occurrence rates would have been 25–45% higher if they had included multi-planet systems (depending on period and radius limits), which can reconcile much of differences between the two studies.

In comparing to previous results, we find that 2–2.8 $R_\oplus$ are more common than 1.4–2 $R_\oplus$ planets, in a relative sense. For example, we find that P13 found that 18.6% of stars had a 2–2.8 $R_\oplus$ planet with P < 100 d vs. 14.2% of stars with a 1.4–2 $R_\oplus$ planet in the same period range. This corresponds to a ratio of 18.6/14.2 = 1.3. In this work, that ratio is 16.1% / 27.0% = 0.6. We can understand this difference in terms of the gap between 1.5 and 2.0 $R_\oplus$ and the peak between 2.0 and

[19] H12, F13, P13, and M15 used 4, 16, 43, and 22 months of photometry, respectively.

2.4 $R_\oplus$ that emerged after we refined the host star radii through spectroscopy. Planets with true sizes between 2.0 and 2.8 $R_\oplus$ were often scattered to the 1.4–2.0 $R_\oplus$ bin due to the 40% radius uncertainties from photometry. Thus the peak from 2.0–2.8 $R_\oplus$ was diminished, while the gap from 1.4–2.0 $R_\oplus$ was filled in. In summary, the integrated occurrence rates presented are largely consistent with previous works, with differences in the detailed radius distribution, owing to improved stellar radii.

5.2. *Rocky to Gaseous Transition*

Studies of the relationship between planet density and radius suggest that planet core sizes reach a maximum of about 1.6 $R_\oplus$. Planets with larger radii and measured masses are mostly low-density and require an extended atmosphere to simultaneously explain their masses and radii (Marcy et al. 2014; Weiss & Marcy 2014; Rogers 2015; Wolfgang & Lopez 2015). Figure 12 shows the radius distribution derived in this work and an empirical fit to the densities and radii of small planets (Weiss et al. 2016). This fit to a sample of planets with measured densities peaks near our observed gap in the planet radius distribution. This suggests that the majority of planets smaller than the minimum in the occurrence distribution are rocky while larger planets likely contain enough volatiles to contribute significantly to the planets' radii.

Additionally, ultra-short-period planets (USPs, having $P < 1$ day) present a clean sample of stripped, rocky planet cores. It is unlikely that H/He atmospheres could survive on small planets bathed in the intense irradiation experienced by USPs. These planets must be bare, rocky cores, stripped of any significant atmosphere. Sanchis-Ojeda et al. (2014) found that the occurrence of ultra-short period planets falls off sharply for $R_P > 1.6$ $R_\oplus$. The apparent lack of rocky cores larger than 1.6 $R_\oplus$ also suggests that planets larger than that must have non-negligible volatile envelopes.

5.3. *Potential Explanations for the Gap*

5.3.1. *Photoevaporation*

Photoevaporation provides a possible mechanism to produce a gap in the radius distribution, even if the initial radius distribution was continuous (Owen & Wu 2013). Lopez & Rice (2016) modeled the masses and radii of planets with various gas envelope fractions. A bare, rocky planet (no envelope) with a mass of 2 $M_\oplus$ has a radius of 1.2 $R_\oplus$ in their models. Adding an H/He envelope with a mass of 0.002 $M_\oplus$ (0.1% mass fraction) increases the planet size to 1.5 $R_\oplus$, a large change in size for a small change in mass. Adding an additional 0.7% by mass of H/He swells the planet to 2.0 $R_\oplus$(see Figure 13). This non-linear mass-radius dependence on volatile fraction has two effects. First, making a planet with a thin atmosphere requires a finely tuned amount of H/He. Second, photoevaporating a planet's envelope significantly changes its size. Our observation of two peaks in the planet size distribution is consistent with super-Earths being rocky planets with atmospheres that contribute negligibly to their size, while sub-Neptunes are planets that retain envelopes with mass fractions of a few percent.

5.3.2. *Gas-poor formation*

Accretion of a modest gas envelope poses a theoretical challenge because fine-tuning is required to end up with an appreciable atmosphere that does not trigger runaway gas accretion and giant planet formation. Lee et al. (2014) proposed a mechanism that produces small planets with low envelope fractions by delaying gas accretion until the gas in the protoplanetary disk is nearly dissipated. They also proposed that small planets could form in very metal-rich disks where high opacity slows cooling and accretion.

In addition, a few-percent-by-mass secondary atmosphere can be outgassed during planet formation and evolution (Adams et al. 2008). Our observed gap in the planet radius distribution could be explained by a mechanism that causes the creation of a secondary atmosphere during the formation of only ∼50% of terrestrial planets.

5.3.3. *Impact Erosion*

Impacts can also provide a way to sculpt the atmospheric properties of small planets and strip large primordial envelopes down to a few percent by mass (e.g., Schlichting et al. 2015; Liu et al. 2015; Inamdar & Schlichting 2016). It is unclear whether a gap in the radius distribution could arise from impacts alone since impact erosion is a highly stochastic process. However, the atmospheric heating initiated by an impact can cause the envelope to expand, making it more susceptible to photoevaporation.

5.3.4. *Signatures of Atmospheric Sculpting*

Lopez & Rice (2016) considered two scenarios for the formation of sub-Neptunes/super-Earths. In one scenario, super-Earths are the remnant cores of photoevaporated, Neptune-size planets. In the other scenario super-Earths form late in the evolution of the protoplanetary disk, just as the gas dissipates (Lee et al. 2014). They predict that the transition radius between these two populations (the gap that we observed) should be a function of semi-major axis. If super-Earths are evaporated cores then the transition radius should be larger at lower incident flux. However, if super-Earths form in a gas-poor disk, or lose gas during the late stages of formation due to giant impacts, then the transition radius should decrease with increasing orbital distance. The distribution of planet radii as a function of insolation flux (Figure 10) does not show a clearly increasing or decreasing transition radius.

If photoevaporation is the dominant mechanism driving the distribution of planet sizes at short orbital periods, then we might expect that closely-spaced planets within multi-planet systems which experience similar irradiation histories would have similar sizes. Kepler-36 is one example to the contrary with both a sub-Neptune and super-Earth orbiting the same star at very similar orbital distances (Carter et al. 2012). A detailed analysis of the statistical properties of multi-planet systems utilizing the CKS stellar parameters is currently ongoing (Weiss et al. (in prepration)).

5.4. *Core Mass Distribution*

The masses of planets smaller than Neptune are dominated by the solid core. Thus, measuring the distribution of core masses provides a valuable constraint on their

TABLE 5
Occurrence Rate Comparison

Radius Interval $R_\oplus$	Period Interval (days)	**This Work**[1] ($f_{\rm bin}$ %)	H12[2,6,7] ($f_{\rm bin}$ %)	P13[3,6] ($f_{\rm bin}$ %)	F13[4] ($f_{\rm bin}$ %)	M15[5] ($f_{\rm bin}$ %)
1.4–2.8	< 100	43.1 ± 2.2	···	32.8 ± 1.4	35.0 ± 2.8[8]	26.7 ± 1.7[8]
2–2.8	< 50	19.4 ± 1.4	9.0 ± 1.5	18.6 ± 1.6	17.5 ± 1.6	12.8 ± 0.5
2–4	< 50	25.4 ± 1.6	13.0 ± 0.8	16.6 ± 1.8	18.3 ± 1.3	18.6 ± 0.6
2–4	< 100	36.6 ± 2.2	···	24.1 ± 2.3	24.0 ± 2.2[8]	22.9 ± 0.8[8]

Note. — Each occurrence rate study focused on different stellar samples, planet detection pipelines, period limits, etc. This table is not meant to be an exact comparison of the results from each study, but instead a rough comparison to show general agreement or highlight large disagreements.

[1] Uncertainties do not include the scaling factors derived in Appendix C
[2] Howard et al. (2012)
[3] Petigura et al. (2013b)
[4] Fressin et al. (2013)
[5] Mulders et al. (2015)
[6] Measured fraction of stars with planets instead of number of planets per star
[7] Only studied planets with periods shorter than 50 days and larger than 2 $R_\oplus$
[8] Periods shorter than 85 days

formation histories. The precise location and depth of the photo-evaporation valley likely depends on the underlying core mass distribution. Planet masses can be constrained using TTVs (Holman & Murray 2005; Agol et al. 2005), but only in specific architectures that may probe different underlying populations. Most of the *Kepler* systems studied in this work are faint and out of reach of the current generation of RV instruments. And the number of RV mass measurements for small planets is too small to map out the core mass distribution in fine detail (Howard et al. 2010; Mayor et al. 2011). Teasing out this distrubtion will require a large sample of low-mass planets amenable to mass measurements. Ongoing and upcoming surveys such as the APF-50 survey (Fulton et al. 2016), the HARPS-N rocky planet search (Motalebi et al. 2015), MINERVA (Swift et al. 2015), and TESS (Ricker et al. 2014) are working to achieve this goal.

6. CONCLUSION

Using precise planet radii for 2025 *Kepler* planets from the CKS Survey, we examined the planet radius distribution at high-resolution. We find evidence for a bimodal distribution of small planet sizes. Sub-Neptunes and super-Earths appear to be two distinct planet classes. Planets tend to prefer radii of either $\sim$1.3 $R_\oplus$ or $\sim$2.4 $R_\oplus$, with relatively few planets having radii of 1.5–2.0 $R_\oplus$. Planets in the gap have the maximum size for a rocky core, as seen in previous studies of bulk planet density and of ultra-short period planets. We posit that the bimodal planet radius distribution stems from differences in the envelope masses of small planets. While our current dataset is insufficient to distinguish between theoretical models that produce the gap, it charts a path forward to unraveling further details of the properties of the galaxy's most abundant planets.

Keck:I (HIRES), Kepler

The CKS project was conceived, planned, and initiated by AWH, GWM, JAJ, HTI, and TDM. AWH, GWM, JAJ acquired Keck telescope time to conduct the magnitude-limited survey. Keck time for the other stellar samples was acquired by JNW, LAR, and GWM. The observations were coordinated by HTI and AWH and carried out by AWH, HTI, GWM, JAJ, TDM, BJF, LMW, EAP, ES, and LAH. AWH secured CKS project funding. SpecMatch was developed and run by EAP and SME@XSEDE was developed and run by LH and PAC. Downstream data products were developed by EAP, HTI, and BJF. Results from the two pipelines were consolidated and the integrity of the parameters were verified by AWH, HTI, EAP, GWM, with assistance from BJF, LMW, ES, LAH, and IJMC. EAP computed derived planetary and stellar properties with assistance from BJF. BJF performed the analysis in this paper, with assistance from EAP, AWH, and GWM. This manuscript was largely written by BJF with assistance from EAP, AWH, GWM, JNW, and LMW.

We thank Josh Winn, Jason Rowe, Eric Lopez, Jeff Valenti, Daniel Huber, and Leslie Rogers for contributing insight during many helpful conversations and providing comments on early drafts of the manuscript. Most of the data presented here were determined directly from observations at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California, and NASA. We are grateful to the time assignment committees of the University of Hawaii, the University of California, the California Institute of Technology, and NASA for their generous allocations of observing time that enabled this large project. Kepler was competitively selected as the tenth NASA Discovery mission. Funding for this mission is provided by the NASA Science Mission Directorate. BJF acknowledges that this material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 2014184874. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. EAP acknowledges support from Hubble Fellowship grant HST-HF2-51365.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc. for NASA under contract NAS 5-26555. AWH acknowledges NASA grant NNX12AJ23G. TDM acknowledges NASA

grant NNX14AE11G. PAC acknowledges National Science Foundation grant AST-1109612. LH acknowledges National Science Foundation grant AST-1009810. LMW acknowledges support from Gloria and Ken Levy and from the Trottier Family. ES is supported by a postgraduate scholarship from the Natural Sciences and Engineering Research Council of Canada. Finally, the authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

REFERENCES

Adams, E. R., Ciardi, D. R., Dupree, A. K., et al. 2012, AJ, 144, 42
Adams, E. R., Dupree, A. K., Kulesa, C., & McCarthy, D. 2013, AJ, 146, 9
Adams, E. R., Seager, S., & Elkins-Tanton, L. 2008, ApJ, 673, 1160
Agol, E., Steffen, J., Sari, R., & Clarkson, W. 2005, MNRAS, 359, 567
Alibert, Y., Carron, F., Fortier, A., et al. 2013, A&A, 558, A109
Anderson, T. W., & Darling, D. A. 1952, Ann. Math. Statist., 23, 193. http://dx.doi.org/10.1214/aoms/1177729437
Babu, G. J., & Feigelson, E. D. 2006, in Astronomical Society of the Pacific Conference Series, Vol. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, 127
Baranec, C., Ziegler, C., Law, N. M., et al. 2016, AJ, 152, 18
Batalha, N. M., Borucki, W. J., Koch, D. G., et al. 2010, ApJ, 713, L109
Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, ApJS, 204, 24
Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, ApJ, 809, 8
Carter, J. A., Agol, E., Chaplin, W. J., et al. 2012, Science, 337, 556
Cartier, K. M. S., Gilliland, R. L., Wright, J. T., & Ciardi, D. R. 2015, ApJ, 804, 97
Catanzarite, J., & Shao, M. 2011, ApJ, 738, 151
Chatterjee, S., & Tan, J. C. 2014, ApJ, 780, 53
Chen, H., & Rogers, L. A. 2016, ApJ, 831, 180
Chiang, E., & Laughlin, G. 2013, MNRAS, 431, 3444
Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2015, ApJ, 810, 95
—. 2016, ApJ, 828, 99
Coleman, G. A. L., & Nelson, R. P. 2014, MNRAS, 445, 479
Dong, S., & Zhu, Z. 2013, ApJ, 778, 53
Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, ApJS, 178, 89
Dressing, C. D., Adams, E. R., Dupree, A. K., Kulesa, C., & McCarthy, D. 2014, AJ, 148, 78
Dressing, C. D., & Charbonneau, D. 2015, ApJ, 807, 45
Everett, M. E., Barclay, T., Ciardi, D. R., et al. 2015, AJ, 149, 55
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Fressin, F., Torres, G., Charbonneau, D., et al. 2013, ApJ, 766, 81
Fulton, B. J., Howard, A. W., Weiss, L. M., et al. 2016, ApJ, 830, 46
Furlan, E., Ciardi, D. R., Everett, M. E., et al. 2017, The Astronomical Journal, 153, 71
Gilliland, R. L., Cartier, K. M. S., Adams, E. R., et al. 2015, AJ, 149, 24
Hadden, S., & Lithwick, Y. 2014, ApJ, 787, 80
—. 2016, ArXiv e-prints, arXiv:1611.03516
Hansen, B. M. S., & Murray, N. 2012, ApJ, 751, 158
Hartigan, J. A., & Hartigan, P. M. 1985, Ann. Statist., 13, 70. http://dx.doi.org/10.1214/aos/1176346577
Holman, M. J., & Murray, N. W. 2005, Science, 307, 1288
Horch, E. P., Howell, S. B., Everett, M. E., & Ciardi, D. R. 2012, AJ, 144, 165
—. 2014, ApJ, 795, 60
Howard, A. W., Marcy, G. W., Johnson, J. A., et al. 2010, Science, 330, 653
Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, ApJS, 201, 15
Howell, S. B., Everett, M. E., Sherry, W., Horch, E., & Ciardi, D. R. 2011, AJ, 142, 19
Huber, D., Silva Aguirre, V., Matthews, J. M., et al. 2014, ApJS, 211, 2
Ida, S., & Lin, D. N. C. 2004, ApJ, 604, 388
Inamdar, N. K., & Schlichting, H. E. 2016, ApJ, 817, L13
Jenkins, J. M., Chandrasekaran, H., McCauliff, S. D., et al. 2010, in Proc. SPIE, Vol. 7740, Software and Cyberinfrastructure for Astronomy, 77400D
Jin, S., Mordasini, C., Parmentier, V., et al. 2014, ApJ, 795, 65
Johnson, J. A., Petigura, E. A., Fulton, B. J., et al. 2017, ArXiv e-prints, arXiv:1703.10402
Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, ApJ, 713, L79
Kolbl, R., Marcy, G. W., Isaacson, H., & Howard, A. W. 2015, AJ, 149, 18
Kolmogorov, A. 1933, Giornale dell' Istituto Italiano degli Attuari, 4, 83
Law, N. M., Morton, T., Baranec, C., et al. 2014, ApJ, 791, 35
Lee, E. J., & Chiang, E. 2016, ApJ, 817, 90
Lee, E. J., Chiang, E., & Ormel, C. W. 2014, ApJ, 797, 95
Lillo-Box, J., Barrado, D., & Bouy, H. 2012, A&A, 546, A10
—. 2014, A&A, 566, A103
Liu, S.-F., Hori, Y., Lin, D. N. C., & Asphaug, E. 2015, ApJ, 812, 164
Lopez, E. D., & Fortney, J. J. 2013, The Astrophysical Journal, 776, 2
Lopez, E. D., & Fortney, J. J. 2014, ApJ, 792, 1
Lopez, E. D., & Rice, K. 2016, ArXiv e-prints, arXiv:1610.09390
Lundkvist, M. S., Kjeldsen, H., Albrecht, S., et al. 2016, Nature Communications, 7, 11201
Marcy, G. W., Isaacson, H., Howard, A. W., et al. 2014, ApJS, 210, 20
Mathur, S., Huber, D., Batalha, N. M., et al. 2016, ArXiv e-prints, arXiv:1609.04128
Mayor, M., Marmier, M., Lovis, C., et al. 2011, arXiv:1109.2497, arXiv:1109.2497
Mazeh, T., Holczer, T., & Faigler, S. 2016, A&A, 589, A75
Mordasini, C., Alibert, Y., & Benz, W. 2009, A&A, 501, 1139
Mordasini, C., Alibert, Y., Georgy, C., et al. 2012, A&A, 547, A112
Morton, T. D. 2012, ApJ, 761, 6
Morton, T. D., Bryson, S. T., Coughlin, J. L., et al. 2016, ApJ, 822, 86
Morton, T. D., & Johnson, J. A. 2011, ApJ, 738, 170
Morton, T. D., & Swift, J. 2014, ApJ, 791, 10
Motalebi, F., Udry, S., Gillon, M., et al. 2015, A&A, 584, A72
Mulders, G. D., Pascucci, I., & Apai, D. 2015, ApJ, 798, 112
Mullally, F., Coughlin, J. L., Thompson, S. E., et al. 2015, ApJS, 217, 31
Newville, M., Stensitzki, T., Allen, D. B., & Ingargiola, A. 2014, LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python, , , doi:10.5281/zenodo.11813. https://doi.org/10.5281/zenodo.11813
Owen, J. E., & Wu, Y. 2013, ApJ, 775, 105
Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013a, Proceedings of the National Academy of Science, 110, 19273
Petigura, E. A., Marcy, G. W., & Howard, A. W. 2013b, ApJ, 770, 69
Petigura, E. A., Howard, A. W., Marcy, G. W., et al. 2017, ArXiv e-prints, arXiv:1703.10400
Raymond, S. N., & Cossou, C. 2014, MNRAS, 440, L11
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9143, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 20
Rogers, L. A. 2015, ApJ, 801, 41
Sanchis-Ojeda, R., Rappaport, S., Winn, J. N., et al. 2014, ApJ, 787, 47

Schlichting, H. E., Sari, R., & Yalinewich, A. 2015, Icarus, 247, 81
Scholz, F. W., & Stephens, M. A. 1987, Journal of the American Statistical Association, 82, 918. http://dx.doi.org/10.1080/01621459.1987.10478517
Silburt, A., Gaidos, E., & Wu, Y. 2015, ApJ, 799, 180
Sinukoff, E., Fulton, B., Scuderi, L., & Gaidos, E. 2013, Space Sci. Rev., 180, 71
Smirnov, N. 1948, Ann. Math. Statist., 19, 279. http://dx.doi.org/10.1214/aoms/1177730256
Swift, J. J., Bottom, M., Johnson, J. A., et al. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 027002
Traub, W. A. 2012, ApJ, 745, 20
Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, in Proc. SPIE Instrumentation in Astronomy VIII, David L. Crawford; Eric R. Craine; Eds., 2198, 362
Wang, J., Fischer, D. A., Horch, E. P., & Xie, J.-W. 2015a, ApJ, 806, 248
Wang, J., Fischer, D. A., Xie, J.-W., & Ciardi, D. R. 2015b, ApJ, 813, 130
Weiss, L. M., & Marcy, G. W. 2014, ApJ, 783, L6
Weiss, L. M., Deck, K., Sinukoff, E., et al. 2016, ArXiv e-prints, arXiv:1612.04856
Wolfgang, A., & Lopez, E. 2015, ApJ, 806, 183
Wu, Y., & Lithwick, Y. 2013, ApJ, 772, 74
Youdin, A. N. 2011, ApJ, 742, 38

APPENDIX

NON-CUMMULATIVE FILTERS

We investigate the impact of each individual filter on the planet catalog by producing a figure similar to Figure 2. However, instead of plotting the distribution after all successive filters are applied to the original sample we plot the distributions after applying *only* the filter specified in the annotations and the figure caption (Figure A.1). The magnitude and impact parameter cuts have the greatest impact on the final sample since they subtract the greatest number of planets. However, no filter preferentially removes planets in the gap or preferentially preserves planets just outside the gap.

WEIGHTED KERNEL DENSITY ESTIMATION

The weights calculated in Section 3 can be used to estimate the occurrence rate distribution of any planet property using weighted kernel density estimation as an alternative to binned histograms (wKDE, Morton & Swift 2014). We calculate the kernel density estimate as:

$$\phi(x) = \frac{1}{N_\star} \sum_{i=1}^{n_{\rm pl}} w_i \cdot K(x - x_i, \sigma_{x,i}). \tag{B1}$$

K is the "kernel" and, in general, it can be any non-negative function that integrates to one and has a centroid of zero. x_i are the individual measurements for a given planet property and $\sigma_{x,i}$ are the uncertainties on those measurements. We treat double-sided uncertainties as symmetric Gaussian uncertainties by taking the mean of the reported upper and lower 1-sigma uncertainties. We adopt a standard Gaussian kernel to calculate the one-dimensional distributions of planet properties, and a bivariate Gaussian for two-dimensional distributions. In order to ensure smooth distributions and contours we limit fractional measurement uncertainty to $\geq$5% in the calculation of the 2D wKDEs. Orbital period is the only parameter that is subject to this limit.

To investigate the possibility that the gap in the planet radius distribution is an artifact of binning we calculate the planet radius distribution using wKDE (Figure B.2). We choose a Gaussian kernel and a variable bandwidth that matches the radius uncertainty for each individual measurement. Again, there are two peaks in the radius distribution separated by a gap. The wKDE demonstrates that the presence or location of the gap does not depend on the particular choice of bin size. The contrast between the bottom of the gap and the top of the peaks is reduced in the wKDE-derived planet radius distribution. However, as shown in the simulations described in Appendix C, this is an artifact of the wKDE technique and probably not a good representation of the underlying radius distribution. The planet radius uncertainties are effectively being counted twice in both the scatter of the median values and the width of the Gaussians summed to create the wKDE. The simulations described in Appendix C show the same dilution of the gap depth when using the wKDE to recover known distributions of simulated planets. Quantifying the valley depth from the wKDE radius distribution may require a careful exploration and justification of the kernel bandwidth selection. Our simulations show that the histograms better reproduce the known input distributions, so we choose leave this bandwith tuning for future studies and conclude that the histogram gives a more accurate picture of the planet radius distribution over this particular application and implementation of the wKDE.

VALIDATION OF THE COMPLETENESS CORRECTIONS

We validate our occurrence calculations and estimate uncertainties by constructing a suite of 100 simulated transit surveys. For each simulation, we draw a distribution of 45000 planet radii and orbital periods from two lognormal distributions then sum those distributions together to create a bimodal distribution similar to the distribution observed in our real planet detections (Figure C.3). We assign each simulated planet to a star in our filtered sample of KOI hosts and calculate detection probabilities and weights as described in §3.3. These detection probabilities are used to decide which planets would have been detected in our real survey. The number of simulated planets (45000) was chosen such that the mean number of planets in the 100 simulated planet detection catalogues is equal to the total number of planets in our filtered KOI catalogue (900).

The stellar radii for the stars in the Stellar17 sample, which are used in the completeness corrections, are perturbed in two different ways in each simulation. We multiply all of the Stellar17 stellar radii by a common constant drawn from a normal distribution centered at 1.0 with a width of 0.25 to simulate potential systematic offsets between the stellar radii in the Stellar17 catalogue and the stellar radii in the CKS catalog. We also add Gaussian noise to the stellar radii for all stars with distribution widths determined from their individual measurement uncertainties. The uncertainties in our final bin heights and occurrence ratios estimated from these simulations account for both systematic and Gaussian random errors in the stellar parameters in the Stellar17 catalog.

We produce histograms for each simulation and correct them for completeness as described in §3.3. The standard deviation of the values in each histogram bin become the uncertainty on the bin values. When compared with uncertainties calculated using Poisson statistics on the number of simulated detections in each bin we find that the Poisson uncertainties are underestimated by a factor of 1.5–2.9 depending on the radius bin. In order to avoid small number statistics for the histogram bins where the simulated distribution approaches zero we repeat the simulations with an input distribution of planets that is log-uniform in radius from 0.5–20.0 $R_\oplus$ and log-uniform in period from

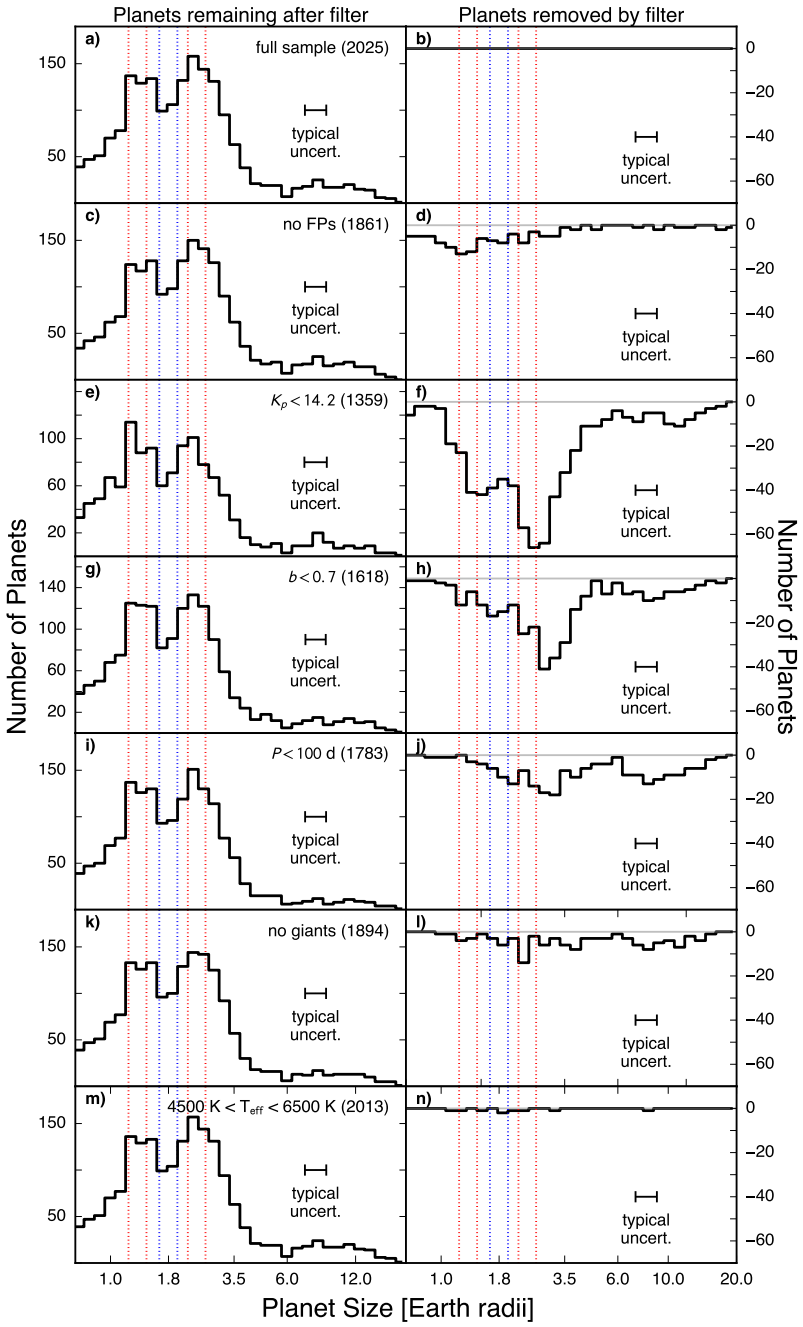

FIG. A.1.— **(a)** size distribution of planet candidates from the CKS sample. **(b)** planets removed by the specified filter. Panels **(c)**–**(n)** show the radius distribution and planets removed from the full sample after applying only a single cut removing known false positives **(c)**, planets orbiting faint host stars **(e)**, planets with grazing transits **(g)**, planets with orbital periods longer than 100 days **(i)**, planets orbiting giant host stars **(k)**, and planets orbiting host stars cooler than 4700 K or hotter than 6500 K **(m)**. No completeness corrections have been applied. The $b < 0.7$ cut and the $Kp < 14.2$ cut remove the most planet candidates, but no filter preferentially removes planets in the gap (between blue dotted lines).

1–200 days solely for the purpose of calculating the uncertainty scaling factors for each radius bin. We adopt the scaling factors listed in Table C.1 in the calculation of all completeness-corrected planet radius histograms and for fitting the distribution described in §4.

We calculate the occurrence ratio of super-Earths to sub-Neptunes in the same way as we do for the real planet catalogue in §4.4. The mean occurrence ratio is consistent with the same ratio for the input distribution of simulated planets and the standard deviation as a fraction of the ratio is 33%. We adopt this fractional uncertainty for the occurrence ratio calculation on the real planet catalogue.

We also calculate the radius distribution for each simulation using the wKDE technique described in Appendix B. We find that the wKDE slightly underestimates the contrast between the peaks of the radius distribution and the bottom of the gap. This is likely due to the fact that there is scatter in the radii measurements due to uncertainties. Those

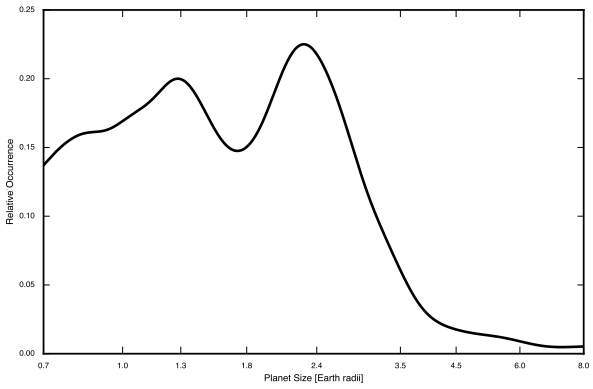

FIG. B.2.— Bin-free view of the planet radius distribution calculated using wKDE (Equation B1). The 1-sigma uncertainty region is shaded in red and calculated using a suite of simulated transit surveys as described in Appendix B.

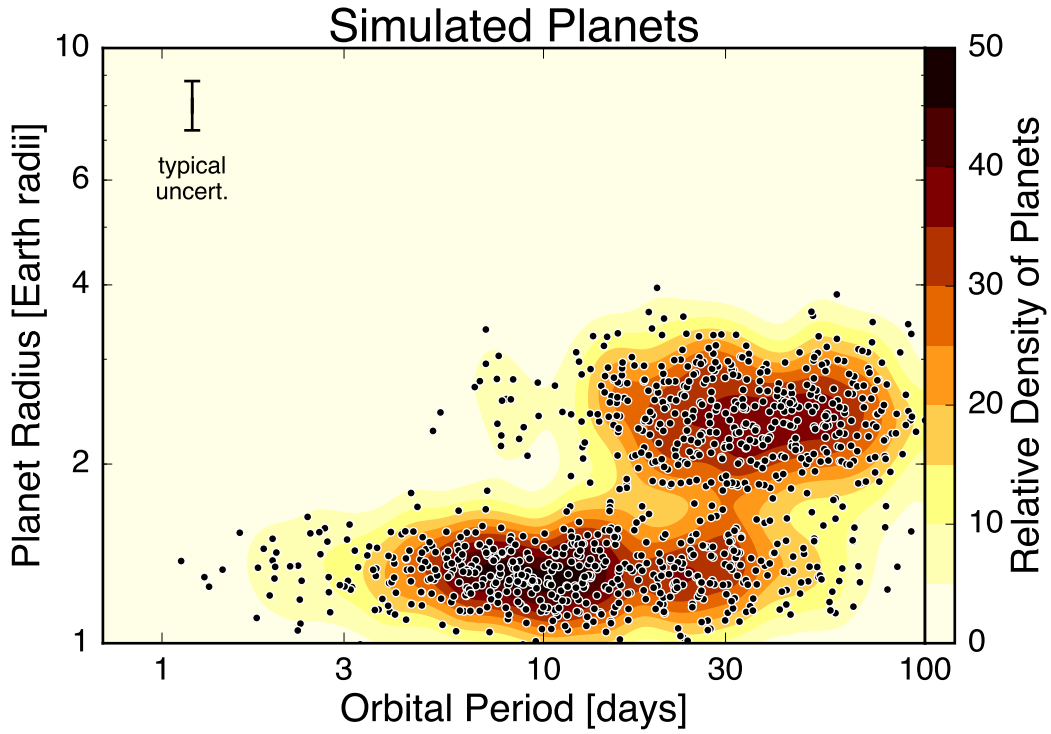

FIG. C.3.— Radius vs. period distribution for simulated sample of planets. For plotting clarity and speed we plot only 1,000 randomly chosen simulated planets out of the 45,000 simulated planets.

uncertainties are also being included as the widths of the Gaussians used to calculate the wKDE, in effect counting the uncertainty twice. Since we do not perform any quantitative analysis on the wKDE we choose not to "de-bias" the wKDE as described in Morton & Swift (2014), but instead limit our quantitative analysis to the histograms that seem to be a more accurate representation of the underlying distributions in our simulations. We use the resulting wKDEs from the simulated surveys in order to estimate the fractional uncertainty as a function of planet radius for the wKDE calculated from the real planet catalog (Figure B.2).

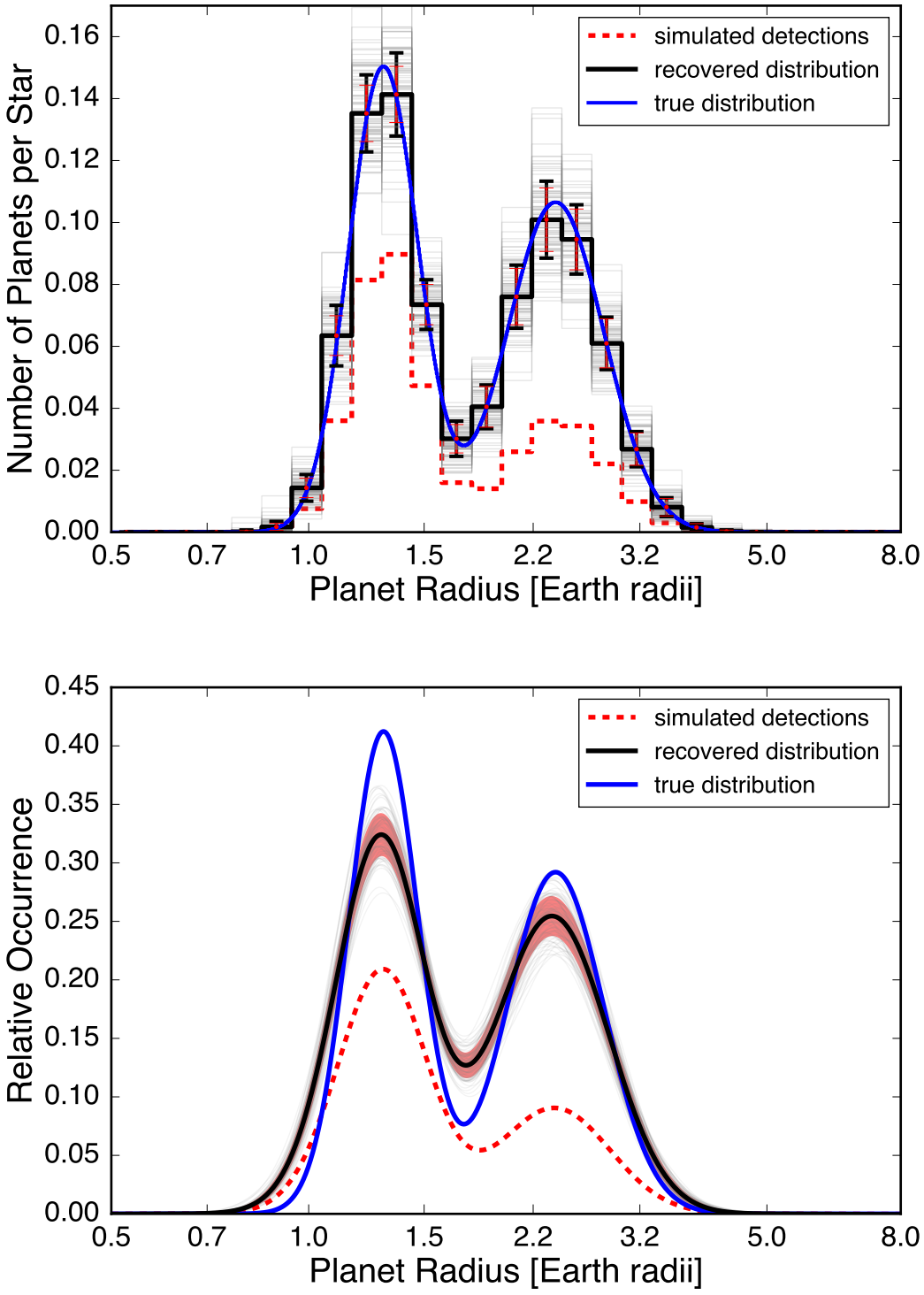

FIG. C.4.— *Top:* Results from simulating 100 transit surveys with a known input distribution of planets. The input distribution of simulated planets is plotted in blue, and the simulated detections are plotted in a red dashed line. The completeness-corrected distributions measured from each of the simulations are plotted as thin grey lines and the median of those recovered distributions is plotted in a thick black line. The thick black error bars are the standard deviation of all of the simulations in each bin and the thin red error bars are poisson uncertainties on the number of detections in each bin scaled by the completeness correction for that bin. *Bottom:* Same as *top* panel but calculated using the wKDE technique described in §3.3. The shaded red area encompasses the standard deviation of the resulting wKDEs over all 100 simulations. We adopt this fractional uncertainty for the one-dimensional KDE plotted in Figure B.2.

TABLE C.1
Bin Uncertainty Scaling Factors

Radius bin $R_\oplus$	Scaling Factor
0.50–0.56	2.82
0.56–0.62	2.50
0.62–0.69	2.30
0.69–0.76	2.54
0.76–0.85	2.35
0.85–0.94	2.09
0.94–1.05	1.92
1.05–1.16	1.95
1.16–1.29	1.89
1.29–1.43	1.46
1.43–1.59	1.65
1.59–1.77	1.81
1.77–1.97	1.38
1.97–2.19	1.50
2.19–2.43	1.39
2.43–2.70	1.58
2.70–3.00	1.48
3.00–3.33	1.58
3.33–3.70	1.25
3.70–4.12	1.48
4.12–4.57	1.47
4.57–5.08	1.46
5.08–5.65	1.63
5.65–6.27	1.45
6.27–6.97	1.50
6.97–7.75	1.52
7.75–8.61	1.34
8.61–9.56	1.44
9.56–10.63	1.46
10.63–11.81	1.52
11.81–13.12	1.57
13.12–14.58	1.36
14.58–16.20	1.35
16.20–18.00	1.45
18.00–20.00	1.44