The Astrophysical Journal, 866:99 (13pp), 2018 October 20 https://doi.org/10.3847/1538-4357/aada83

Revised Radii of *Kepler* Stars and Planets Using *Gaia* Data Release 2

Travis A. Berger[1], Daniel Huber[1,2,3,4], Eric Gaidos[5], and Jennifer L. van Saders[1]
[1] Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA; taberger@hawaii.edu
[2] Sydney Institute for Astronomy (SIfA), School of Physics, University of Sydney, NSW 2006, Australia
[3] SETI Institute, 189 Bernardo Avenue, Mountain View, CA 94043, USA
[4] Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark
[5] Department of Geology & Geophysics, University of Hawaii at Mānoa, Honolulu, HI 96822, USA
Received 2018 May 1; revised 2018 August 10; accepted 2018 August 13; published 2018 October 16

Abstract

One bottleneck for the exploitation of data from the *Kepler* mission for stellar astrophysics and exoplanet research has been the lack of precise radii and evolutionary states for most of the observed stars. We report revised radii of 177,911 *Kepler* stars derived by combining parallaxes from the *Gaia* Data Release 2 with the DR25 *Kepler* Stellar Properties Catalog. The median radius precision is ≈8%, a typical improvement by a factor of 4–5 over previous estimates for typical *Kepler* stars. We find that ≈67% (≈120,000) of all *Kepler* targets are main-sequence stars, ≈21% (≈37,000) are subgiants, and ≈12% (≈21,000) are red giants, demonstrating that subgiant contamination is less severe than some previous estimates and that *Kepler* targets are mostly main-sequence stars. Using the revised stellar radii, we recalculate the radii for 2123 confirmed and 1922 candidate exoplanets. We confirm the presence of a gap in the radius distribution of small, close-in planets, but find that the gap is mostly limited to incident fluxes $>200\,F_{\oplus}$, and its location may be at a slightly larger radius (closer to $\approx 2\,R_{\oplus}$) when compared to previous results. Furthermore, we find several confirmed exoplanets occupying a previously described "hot super-Earth desert" at high irradiance, show the relation between a gas-giant planet's radius and its incident flux, and establish a bona fide sample of eight confirmed planets and 30 planet candidates with $R_{\rm p} < 2\,R_{\oplus}$ in circumstellar "habitable zones" (incident fluxes between 0.25 and $1.50\,F_{\oplus}$). The results presented here demonstrate the potential for transformative characterization of stellar and exoplanet populations using *Gaia* data.

Key words: catalogs – planetary systems – stars: fundamental parameters – techniques: photometric

Supporting material: machine-readable tables

1. Introduction

Precise estimates of exoplanet properties, such as radius, mass, and density, inevitably require precise characterization of the host stars. Precise stellar classifications are also required to study the dynamics and evolution of planetary orbits (Kane et al. 2012; Sliski & Kipping 2014; Van Eylen & Albrecht 2015; Shabram et al. 2016) and to derive an accurate planet occurrence (e.g., Howard et al. 2012; Burke et al. 2015).

Traditional methods used to classify the target stars of exoplanet surveys include broadband colors and proper motions, which efficiently separate dwarfs from giants but cannot resolve intermediate evolutionary states, with typical uncertainties of ≈0.3–0.4 dex in $\log g$ (Brown et al. 2011; Huber et al. 2016). High-resolution spectroscopy delivers typical precisions of ≈0.15 dex in $\log g$ (Torres et al. 2012) for solar-type stars, while methods calibrated to benchmark stars can achieve precisions down to ≈0.07 dex (Brewer et al. 2015; Petigura 2015). Finally, time-domain variability of stars offers currently the highest precision $\log g$ values for field stars, for example, by measuring amplitudes or timescales of stellar granulation (≈0.1 dex, Bastien et al. 2013; ≈0.03 dex, Kallinger et al. 2016) or stellar oscillations (≈0.01 dex, Huber et al. 2013).

Despite this progress, most of these methods are only applicable to a subset of the large samples of stars that are typically observed in exoplanet transit surveys (190,000 stars for *Kepler*, >200,000 stars for *K2*, and >500,000 stars for the *Transiting Exoplanet Survey Satellite* (*TESS*)). As a result, 70% of the overall *Kepler* population in the latest version of the *Kepler* Stellar Properties Catalog (KSPC DR25; Mathur et al. 2017) still have $\log g$ values determined from photometry. This translates into 30%–40% uncertainties in stellar radii that severely limit our understanding of the stellar and planet population probed by *Kepler*.

Improved stellar radii of *Kepler* hosts have recently led to several important results for our understanding of exoplanets, such as the discovery of a gap in the distribution of small planets by the California-*Kepler* Survey (CKS; Fulton et al. 2017; Johnson et al. 2017; Petigura et al. 2017) and evidence of a dearth of hot super-Earths (Lundkvist et al. 2016). Both results have been tied to processes, such as photo-evaporation (Lopez et al. 2012; Owen & Wu 2017), but are limited to subsamples consisting of less than half of planet candidates.

The bottleneck caused by imprecise stellar radii of *Kepler* stars can now be relieved thanks to precise parallaxes from *Gaia* Data Release 2 (DR2) for over one billion stars in the galaxy (Gaia Collaboration et al. 2018; Lindegren et al. 2018). In this paper, we rederive radii for 177,911 *Kepler* stars using *Gaia* DR2 parallaxes and investigate the stellar and exoplanet radius distributions of *Kepler* targets.

2. Methodology

2.1. Kepler–Gaia *DR2 Crossmatching*

First, we crossmatched the positions of all stars from the KSPC DR25 (Mathur et al. 2017) by utilizing the X-match service of the Centre de Données astronomiques de Strasbourg (CDS). This provided a table of *Gaia* DR2 source matches within three arcseconds of each *Kepler* star. To determine bona

fide *Kepler–Gaia* source matches, we first removed all matches with distances greater than 1.5 arcsec from the *Kepler*-determined position. We chose 1.5 arcsec because the distribution of separations displayed a minimum there, and the increase of matches at greater angular separations indicates the inclusion of spurious background sources.

Next, we imposed a variety of magnitude cuts, depending on the available photometry, to ensure our *Kepler–Gaia* matches were of similar brightnesses. Unfortunately, not all *Kepler* stars had similar quality photometry to compare to the measured *Gaia* *G*-band magnitudes, so we had to utilize the American Association of Variable Star Observers' Photometric All-Sky Survey (APASS) *g*, *r*, and/or *i* photometry for instances where KSPC stars did not have *g*-, *r*-, or *i*- band photometry from the *Kepler* Input Catalog (KIC; Brown et al. 2011). For stars that were still missing any *g*, *r*, or *i* photometry, we used *Kepler* magnitudes (*Kp*) for comparisons with *G* magnitudes.

To compute our predicted *G* magnitudes, we utilized the *g*, *r*, and *i* color–color polynomial fits in Table 7 of Jordi et al. (2010). After inspecting the distribution of G_{Gaia}–$G_{\rm pred}$, we chose to remove all stars with absolute differences greater than two magnitudes. For the remaining sample of stars with only *Kp* magnitudes, we compared G_{Gaia}–*Kp* and again removed all stars with absolute differences greater than two magnitudes.

For stars with multiple matches that satisfied these criteria, we decided to keep those with the smallest angular separations. Of the 197,104 stars present in the KSPC, we identified *Gaia* DR2 source matches for 195,710. Stars with poorly determined parallaxes ($\sigma_\pi/\pi > 0.2$), low effective temperatures based on our adopted values ($T_{\rm eff} < 3000$ K, see Section 2.2), extremely low $\log g$ (<0.1 dex), and/or non-"AAA"-quality Two Micron All Sky Survey (2MASS) photometry were rejected from our sample.

Additionally, we made astrometric cuts similar to those described in Appendix C of Lindegren et al. (2018) and Section 4.1 of Arenou et al. (2018). In particular, we used Equation (1) (unit weight error compared to a function of the *G* magnitude of the source that helps filter contamination from binaries and calibration problems) and Equation (3) (greater than eight groups of observations separated by at least 4 days) of Arenou et al. (2018) to remove stars with bad astrometric solutions. We did not use the astrometric excess noise values provided by *Gaia* DR2 to filter stars because they were less discriminating for stars with $G < 15$ due to the "degree of freedom bug" (see Appendices A and C of Lindegren et al. 2018). We did not use Equation (2) of Arenou et al. (2018), a cut ensuring that *Gaia* has clean photometry of the included sources, because we utilized separate 2MASS photometry in our analysis. As discussed in Lindegren et al. (2018), our imposed cuts removed many stars that appear in unphysical areas of the radius–$T_{\rm eff}$ parameter space, such as the "subdwarfs" between the stellar main sequence and the white-dwarf branch. Excluding these stars reduced our final sample to 177,911 *Kepler* stars.

2.2. Stellar Radii Determination

To calculate stellar radii, we employed the stellar classification code `isoclassify` (Huber et al. 2017) in its "direct method," using the *Gaia* DR2 parallax (Lindegren et al. 2018), the 2MASS *K*-band magnitude, $T_{\rm eff}$, $\log g$, and [Fe/H] values from the KSPC DR25 (Mathur et al. 2017) as the input. We replaced the input values given in the KSPC for two sets of stars: stars in the CKS, for which we adopted spectroscopic parameters from Petigura et al. (2017); and stars with $T_{\rm eff} < 4000$ K with $T_{\rm eff}$ provenances from the KIC, for which we adopted revised $T_{\rm eff}$ values from Gaidos et al. (2016).

For each star, we first converted parallaxes into distances using an exponentially decreasing volume density prior with a length scale of 1.35 kpc (Bailer-Jones 2015; Astraatmadja & Bailer-Jones 2016) and included a systematic parallax offset of 0.03 mas (Lindegren et al. 2018). We note that *Gaia* DR2 has systematic parallax offsets that vary with the position, angular scale, and color (Arenou et al. 2018; Lindegren et al. 2018). Zinn et al. (2018) used asteroseismology to compare distances to those derived from *Gaia* parallaxes and found a systematic offset of 0.05 mas within the *Kepler* field. Although this measurement applies to the *Kepler* field, we still used the Lindegren et al. (2018) value of 0.03 mas derived from quasars because of potential systematics in asteroseismic scaling relations and poorly constrained color dependencies in the parallax offset. In addition, the 0.02 mas offset was small compared to the median parallax of 0.66 mas in our sample.

We then combined the 2MASS *K*-band magnitude with extinctions A_V derived from the 3D reddening map and interpolated the reddening vectors in Table 1 of Green et al. (2018). We also added the gray component of the extinction curve $b = 0.063$, computed from $A_H/A_K = 1.74$ (Nishiyama et al. 2006) by Green et al. (2018), to our extinction values. Next, we added these extinction values to our magnitudes, which we then combined with distances to calculate absolute magnitudes. We derived bolometric corrections by linearly interpolating $T_{\rm eff}$, $\log g$, [Fe/H], and A_V in the bolometric correction tables from the Modules for Experiments in Stellar Astrophysics (MESA) Isochrones & Stellar Tracks (MIST; Choi et al. 2016) grids (MIST/C3K; C. Conroy et al. 2018, in preparation[6]), which we combined with our absolute magnitudes to compute luminosities. Finally, we combined the derived luminosities with $T_{\rm eff}$ from Mathur et al. (2017; or other sources as indicated above), and *Gaia* parallaxes in the Stefan–Boltzmann relation to calculate stellar radii. The procedure is implemented as a Monte Carlo sampling scheme, and the resulting distributions were used to calculate the median and 1σ confidence interval for the radius of each star. Table 1 lists our revised radii for all 177,911 *Kepler* stars analyzed here.

The above method produced systematically overestimated radii for M dwarfs due to inaccuracies in bolometric corrections in `isoclassify`, which are based on ATLAS model stellar atmospheres (Kurucz 1993). Therefore, we used an empirical relationship between the absolute *K* magnitude (M_{K_s}) and the stellar radius described by Equation (4) and Table 1 of Mann et al. (2015) to compute stellar radii and hence luminosities for stars with $T_{\rm eff} < 4100$ K and $M_{K_s} > 3.0$ mag. We added 2.7%, corresponding to the uncertainty of the relation, to uncertainties in the radii of these stars. Although the M_{K_s}–radius relation only applies for $M_{K_s} > 4.0$ mag, we have confirmed that an extrapolation to $M_{K_s} = 3.0$ mag produces radii that are approximately compatible with those predicted by MIST isochrones.

Figure 1 shows a histogram of fractional radius uncertainties for 177,702 of 177,911 *Kepler* stars with radii derived in this work. The remaining 209 stars have higher fractional radius uncertainties and are likely some of the most distant stars in the *Kepler* field. The typical uncertainty is ≈8%, a factor of 4–5

[6] http://waps.cfa.harvard.edu/MIST/model_grids.html

Table 1
Revised Parameters of *Kepler* Stars

KIC ID	*Gaia* DR2 ID	$T_{\rm eff}$ (K)	$\sigma_{T_{\rm eff}}$ (K)	d (pc)	σ_{d+} (pc)	σ_{d-} (pc)	$R_\star$ ($R_\odot$)	$\sigma_{R_{\star+}}$ ($R_\odot$)	$\sigma_{R_{\star-}}$ ($R_\odot$)	A_V (mag)	Evol. Flag	Bin. Flag
757076	2050233807328471424	5164	181	658.465	21.419	20.163	3.986	0.324	0.293	0.273	1	0
757099	2050233601176543104	5521	193	369.374	3.708	3.645	1.053	0.078	0.071	0.120	0	0
757137	2050230543159814656	4751	166	570.715	8.271	8.060	13.406	1.004	0.916	0.230	2	0
757280	2050230611879323904	6543	229	824.791	15.079	14.586	2.687	0.205	0.186	0.323	0	0
757450	2050231848829944320	5306	106	835.371	18.423	17.692	0.962	0.047	0.044	0.298	0	0
892010	2050234975566082176	4834	169	1856.534	86.437	79.285	14.826	1.302	1.178	0.258	2	0
892107	2050234696381511808	5086	178	941.305	20.518	19.713	4.334	0.334	0.303	0.186	2	0
892195	2050234735047928320	5521	193	480.822	3.850	3.800	0.983	0.073	0.066	0.141	0	0
892203	2050236521754360832	5945	208	555.165	4.828	4.759	1.022	0.076	0.069	0.124	0	0
892667	2050232329866306176	6604	231	1175.938	21.455	20.754	2.207	0.168	0.153	0.352	0	0
892675	2050232329866320512	6312	221	584.442	4.837	4.772	1.052	0.078	0.071	0.175	0	0

Note. KIC ID, *Gaia* DR2 ID, $T_{\rm eff}$, distance, stellar radii, extinction, evolutionary flag, and binary flag (and errors, where reported) for our sample of 177,911 *Kepler* stars. The evolutionary flags are as follows: 0 = main-sequence dwarf, 1 = subgiant, and 2 = red giant. The binary flags are as follows: 0 = no indication of binary, 1 = binary candidate based on *Gaia* radius only, 2 = AO-detected binary only (Ziegler et al. 2018), and 3 = binary candidate based on the *Gaia* radius and the presence of an AO-detected binary. See Figure 5 for stars with evolutionary state flags = 0–2 (black, green, and red, respectively) and binary flags = 1 or 3 (blue). A slice of our derived parameters is provided here to illustrate the form and format.

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

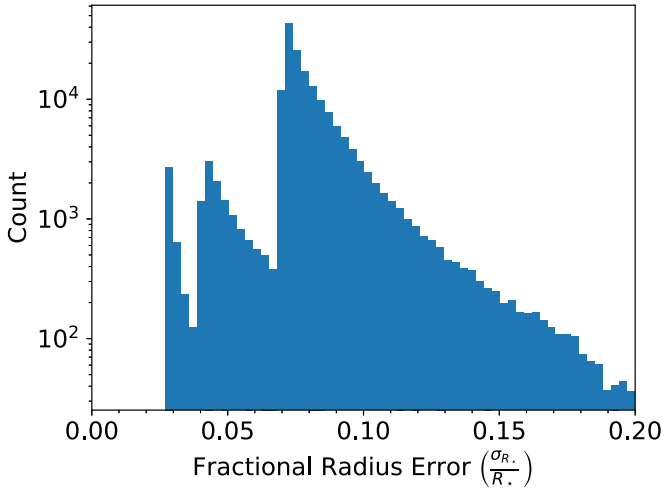

Figure 1. Histogram of the fractional radius uncertainty for 177,702 *Kepler* stars derived in this work. The sample of 209 stars with fractional radius uncertainties >0.2 are some of the most distant stars in the *Kepler* field. The typical radius uncertainty pre-*Gaia* DR2 was ≈30%. The peaks at ≈3%, ≈4.5%, and ≈8% errors correspond to M dwarfs with radii determined from M_{K_s}–radius relations, stars with spectroscopic constraints on $T_{\rm eff}$, and stars with photometric $T_{\rm eff}$, respectively.

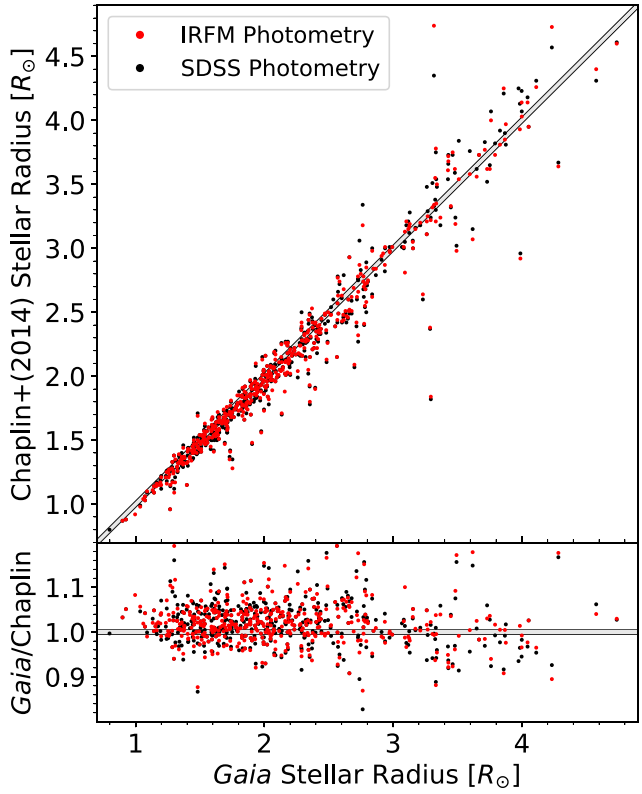

Figure 2. Comparison of our derived stellar radii using *Gaia* parallaxes to asteroseismic radii from Chaplin et al. (2014). Chaplin et al. (2014) radii derived from $T_{\rm eff}$ determined through the InfraRed Flux Method (IRFM) and SDSS photometry are shown in red and black, respectively. The top panel plots, Chaplin et al. (2014) radii vs. those derived in this work, while the bottom panel plots the ratio (our radii divided by the Chaplin radii) vs. our radii.

improvement over the KSPC. The radius uncertainty is dominated by $T_{\rm eff}$, which for a typical *Kepler* target is ≈3.5% based on broadband photometry (Huber et al. 2014). The peak at ≈3% fractional radius uncertainty corresponds to M-dwarf radii computed through the M_{K_s}–radius relation (not dependent on $T_{\rm eff}$; Mann et al. 2015), while the peak at ≈4.5% fractional radius uncertainty represents stars with spectroscopic temperatures (2% errors in $T_{\rm eff}$). Our error budget also included uncertainties of 0.04 mag in A_V and 0.02 mag in bolometric corrections, which are typical values for the *Kepler* field (Huber et al. 2017). To compute the uncertainty in A_V, which carries into the error in the stellar radius, we combined both the distance uncertainty, which translates into an uncertainty in A_V along the line of sight (minimal) and uncertainties in the reddening model itself (dominant). We determined the latter by comparing the Green et al. (2015) map with the Green et al. (2018) map for our sample, yielding a median absolute deviation of ∼22%, which we adopt as a fractional uncertainty for our reported extinction values. Therefore, for our typical $A_V = 0.18$ mag, we report a typical uncertainty of 0.04 mag. This corresponds to $A_K = 0.013 \pm 0.003$ mag, which factors into the absolute *K*-band magnitude uncertainty and hence our stellar radius uncertainty. We emphasize that the above routine uses $\log g$ from the KSPC only to determine bolometric corrections, which are only mildly dependent on $\log g$, and hence the derived radii are mostly insensitive to inaccurate $\log g$ values.

The 3.5% and 2% uncertainties in $T_{\rm eff}$ (≈200 K and ≈115 K at solar $T_{\rm eff}$) were conservative but large enough to have encompassed systematic differences between $T_{\rm eff}$ scales and covariances between extinction and color–$T_{\rm eff}$ relations (Pinsonneault et al. 2012). Future revisions of the $T_{\rm eff}$ scale for *Kepler* stars, taking into account revised reddening maps based on *Gaia* DR2, can be expected to improve the typical radius precision to ∼5% or better.

The *Gaia* Collaboration released radii and effective temperatures for 178,706 *Kepler* targets based on *Gaia* photometry and parallaxes (Andrae et al. 2018; Gaia Collaboration et al. 2018; Lindegren et al. 2018). However, these parameters are optimized for >160 million stars across the sky. In contrast, the *Kepler* field is one of the most well-studied samples of stars due to its relevance to exoplanet science, and the KSPC includes information from the vast amount of photometric, spectroscopic and asteroseismic analyses that have been performed over the past ten years. Therefore, we expect the stellar radii derived in this work to be more accurate than those reported by the *Gaia* Collaboration.

2.3. Validation of Stellar Radii

2.3.1. Comparison to Asteroseismic Radii

To test the precision of our radii, we compared them to radii derived using asteroseismology as given in Chaplin et al. (2014) (Figure 2). Red and black points represent radii from Chaplin et al. (2014), which were determined by combining asteroseismology with photometric temperatures derived from the InfraRed Flux Method (IRFM) and SDSS photometry, respectively. Only a small fraction of this asteroseismic sample has been observed by CKS (which targeted planet host stars), and hence the temperatures adopted in our catalog come from spectroscopic measurements by Buchhave & Latham (2015), as adopted by Mathur et al. (2017). Overall, we find that the scatter is on the order of ≈4%, which is fully consistent with the typical ≈4% uncertainties of our radii for stars with spectroscopic constraints (see Figure 1). We also identify a ≈3% systematic offset in the subgiant range (1.5–3.0 $R_\odot$),

where the Chaplin et al. (2014) radii are systematically smaller. Part of this offset can be explained by the use of different effective temperature scales, as discussed in Huber et al. (2017), which identifies a similar offset based on a comparison of asteroseismology with *Gaia* DR1. Ultimately, this comparison with independent measurements supports the precision of the radii reported in our catalog.

2.3.2. Systematic Error Sources

A variety of factors can affect the accuracy of our reported stellar radii. Offsets in the effective temperature, in most cases, have the largest effect on our reported radii (>60% of the error budget for a typical star with either spectroscopic (2%) or photometric (3.5%) fractional errors in $T_{\rm eff}$). We used conservative errors on our $T_{\rm eff}$ values because of the inhomogeneity of the KSPC $T_{\rm eff}$ sources. These uncertainties contained $T_{\rm eff}$ offsets between different methods, which are typically less than 150 K (see Table 7 and $T_{\rm eff}$ comparison plots in Petigura et al. 2017).

2MASS reports typical errors of 0.03 mag in K-band photometry. Therefore, any systematic offset in the zero point of 2MASS photometry would, at most, result in a ≈1.5% error in our computed stellar radius. *Gaia* DR2 parallaxes in the *Kepler* field may be systematically underestimated by about 0.02 mas, a figure smaller than typical formal error, as well as the global systematic value of 0.03 mas (Lindegren et al. 2018; Zinn et al. 2018). This offset would produce an overestimation of stellar radii of ≈1% for nearby stars and up to ≈5% for stars as far as 5 kpc. We included a 0.02 mag uncertainty ($\lesssim$1% error in the stellar radius) to account for uncertainty in our MIST/C3K bolometric correction grid, but that does not account for issues in the models. Although the grid appears to work well for most stars, it fails for M dwarfs, where, in some cases, radii were overestimated by $\sim$20%. We therefore computed M-dwarf radii using the Mann et al. (2015) relation detailed in Section 2.2. Finally, we considered systematic errors in our extinction values. As we discussed in Section 2.2, the bulk of the uncertainty in the extinction will come from intrinsic inaccuracies in the reddening map. Taking our typical extinction uncertainty of 0.003 mag in A_K, this translates into a $\ll$1% underestimation of the stellar radius. Even when we considered the worst-case scenario from Green et al. (2018), where their map significantly underestimates reddening by 0.25 mag in A_V ($A_K = 0.02$ mag) compared to the (Planck Collaboration et al. 2014) map, this only corresponds to a ≈1% underestimation of the stellar radius.

In summary, we expect that individual systematic errors are well within our quoted uncertainties. Since some of the error sources are independent (e.g., temperature and parallax, photometric zero-point offsets, and bolometric corrections), we consider it unlikely that they would be linearly additive, in which case radius systematics would exceed our quoted uncertainties.

2.3.3. Stellar and Exoplanet Radius Dilution

2MASS photometry in some cases includes flux from unresolved stellar companions, which affects both stellar and exoplanet radii. To minimize the number of stars with problematic 2MASS photometry, we only used sources with an "AAA" photometry quality, which removed 5000 (≈3%) sources from our catalog. 2MASS photometry has an effective resolution of 4″ (Skrutskie et al. 2006), which is similar to the size of *Kepler* pixels. Ziegler et al. (2018) showed that, of the companions within 4″ from their hosts, the contrasts (Δm) range over 0–6 mag in the LP600 bandpass (a long-pass filter that begins to transmit at 600 nm and that roughly matches the *Kepler* passband; Law et al. 2014). This corresponds to $\Delta m \approx$ 0–3 mag for 2MASS K-band photometry for a G-type main-sequence star and its companion, which results in a ≈41%–3% overestimated stellar radius for the primary star. This is significantly larger than our radius uncertainties in some cases, but, lacking adaptive optics (AO) followup for all stars in the *Kepler* field, we did not amend our radii. Ziegler et al. (2018) found that ≈14.5% of *Kepler* stars with candidate planets have close-in (<4″) detected companions. However, only ≈7% of stars in the Ziegler et al. (2018) sample had $\Delta K < 2$%. Thus, only these low-contrast companions could dilute measured fluxes enough to exceed our reported 8% uncertainties. Companions more widely separated than 4″ should be resolved by 2MASS, and in these cases, the amount of dilution and the effect on planet radius should be small.

If the stellar radius is actually smaller, then any transiting planet radius is smaller too. However, unresolved companions also affect the transit signal in the *Kepler* light curve, and there is a net effect only to the extent that surface brightnesses of the stars are different. For the *Kepler* bandpass, differences in $T_{\rm eff}$ between the primary and companion will dominate, while differences in $\log g$ and [Fe/H] will have a minimal effect. We flagged stars identified as multiples by Ziegler et al. (2018) as AO binaries in Table 1 (binary flag = 1 or 3). We caution that these flags are not complete as there may be companions unresolved by Robo-AO, they are restricted to the *Kepler* Objects of Interest (KOIs), and not all detections are physical companions.

3. Revised Radii of *Kepler* Stars

3.1. The Gaia H–R Diagram of Kepler Stars

Figure 3 shows stellar radius versus the effective temperature for the *Kepler* stars with radii revised by this work. This diagram is the first nearly model-independent H–R diagram of the *Kepler* field. We see a clear main sequence, from M dwarfs at $T_{\rm eff} = 3000$ K and $R \approx 0.2\,R_\odot$, through A stars at $T_{\rm eff} \lesssim$ 9000 K and $R \approx 2\,R_\odot$. The main-sequence turnoff at $T_{\rm eff} \approx$ 6000 K and $R \approx 2\,R_\odot$ is visible along with the giant branch. We identify the "red clump" as the concentration of stars surrounding $T_{\rm eff} \approx$ 4900 K and $R \approx 11\,R_\odot$. As expected, the *Kepler* targets are heavily dominated by FG-type stars as a result of the target selection focusing on solar-type stars to detect transiting exoplanets (Batalha et al. 2010).

The distribution in Figure 3 contains artifacts: most prominently, the gap in the main sequence around 4000 K. This gap is the result of the combination of two photometric $T_{\rm eff}$ scales in the KSPC (Mathur et al. 2017), namely $T_{\rm eff}$ values from Pinsonneault et al. (2012) for FGK stars and the classification of M dwarfs by Dressing & Charbonneau (2013). An accurate recalibration of the $T_{\rm eff}$ scale for all *Kepler* targets is beyond the scope of this paper, but the use of the DR25 ensures the inclusion of the best available values for $T_{\rm eff}$ and [Fe/H] on a star-by-star basis. A number of stars below the main sequence that may be white dwarfs ($T_{\rm eff} = 6500$–$10{,}000$ K and $R = 0.02\,R_\odot$, not shown in Figure 3) and subdwarfs ($T_{\rm eff} = 3600$–5400 K and $R_\star$

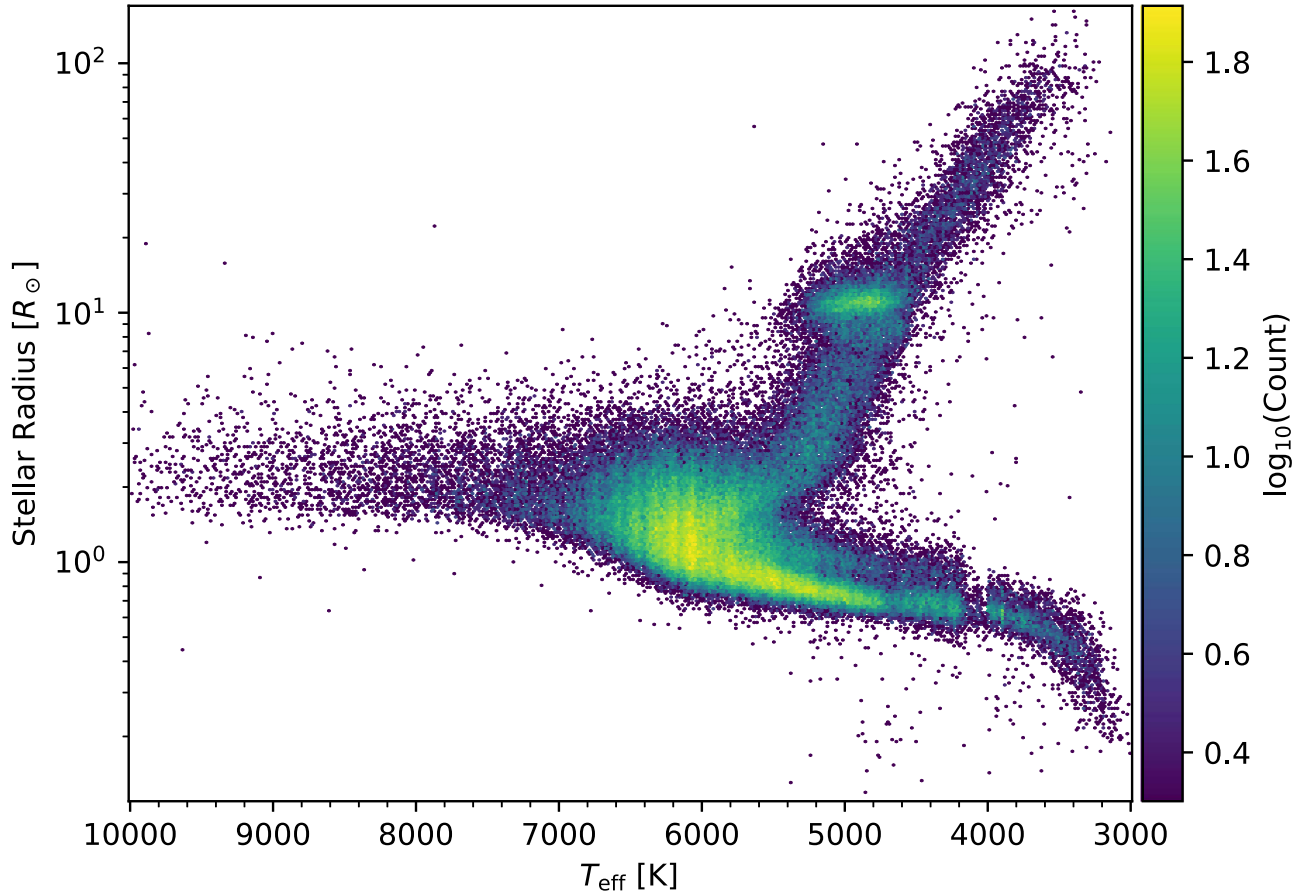

Figure 3. Radius vs. effective temperature for 177,722 *Kepler* stars with radii based on *Gaia* DR2 parallaxes presented in this work. A sample of 189 stars falling off the plot limits shown here includes hot stars ($T_{\rm eff} > 10{,}000$ K) and white dwarfs. Color-coding represents logarithmic number density. Note that the discontinuity in $T_{\rm eff}$ near 4000 K is an artifact due to systematic shifts in $T_{\rm eff}$ scales in the DR25 *Kepler* Stellar Properties Catalog.

$< 0.6\,R_\odot$), as well as in other extreme parameter regimes, could be catalog mismatches or have erroneous $T_{\rm eff}$ values (Table 1).

Figure 3 contains an apparent second sequence above the main sequence for dwarfs with $T_{\rm eff} < 5200$ K. Because K stars are less massive than their hotter main-sequence counterparts, we do not expect these stars to have evolved significantly over a Hubble time, and the intrinsic spread in metallicity is not expected to be asymmetric enough to produce this feature. Rather, this feature likely contains nearly equal-mass binaries; the luminosities and radii of these stars will be overestimated by our methods, but also indicates that *Gaia* DR2 parallaxes can be used to identify readily cool main-sequence binaries.

3.2. Comparison to the DR25 Kepler Stellar Properties Catalog

Figure 4 shows a comparison of stellar radii in the DR25 stellar properties catalog (Mathur et al. 2017) to those derived in this paper. The distribution approximately tracks the 1:1 line, but there is large scatter and strong systematic offsets caused by large uncertainties in the DR25 radii, which were mostly based on photometric $\log g$ values from the KIC. We measure an overall median offset and scatter in the *Gaia*/DR25 residuals of 12% and 34% for all stars, 14% and 32% for unevolved stars ($<3\,R_\odot$), and –7% and 35% for red giants ($>3\,R_\odot$), where positive offsets indicate a larger *Gaia* radius. The residuals clearly demonstrate that a substantial fraction of *Kepler* stars are more evolved than implied in the KSPC.

We also identify 975 giants that were misclassified as dwarfs and 483 dwarfs that were misclassified as giants (bottom right and top left areas in the top panel of Figure 4, respectively). The revised classifications presented here will thus aid in increasing cool dwarf samples for studies of stellar rotation and activity (e.g., McQuillan et al. 2014; Angus et al. 2016; Davenport 2016) and red giants for asteroseismology (e.g., Hekker et al. 2011; Mosser et al. 2012; Stello et al. 2013; Yu et al. 2018).

3.3. Evolutionary States of Kepler Stars

Since the initial *Kepler* target selection (Batalha et al. 2010), there has been growing evidence that the number of subgiants in the KIC (Brown et al. 2011) and subsequent KSPC revisions (Huber et al. 2014; Mathur et al. 2017) has been significantly underestimated due to the Malmquist bias (Gaidos & Mann 2013) and the insensitivity of broadband photometry to determine surface gravities. For example, Verner et al. (2011) show that radii in the KIC are underestimated by up to 50% for a sample of subgiants with asteroseismic detections. Everett et al. (2013) used medium resolution spectroscopy to arrive at a similar conclusion for faint *Kepler* exoplanet host stars, while surface gravities derived from granulation noise ("flicker") suggested that nearly 50% of all bright exoplanet host stars are subgiants (Bastien et al. 2014).

The revised radii using *Gaia* DR2 parallaxes presented in this work allow the first definite classification of the

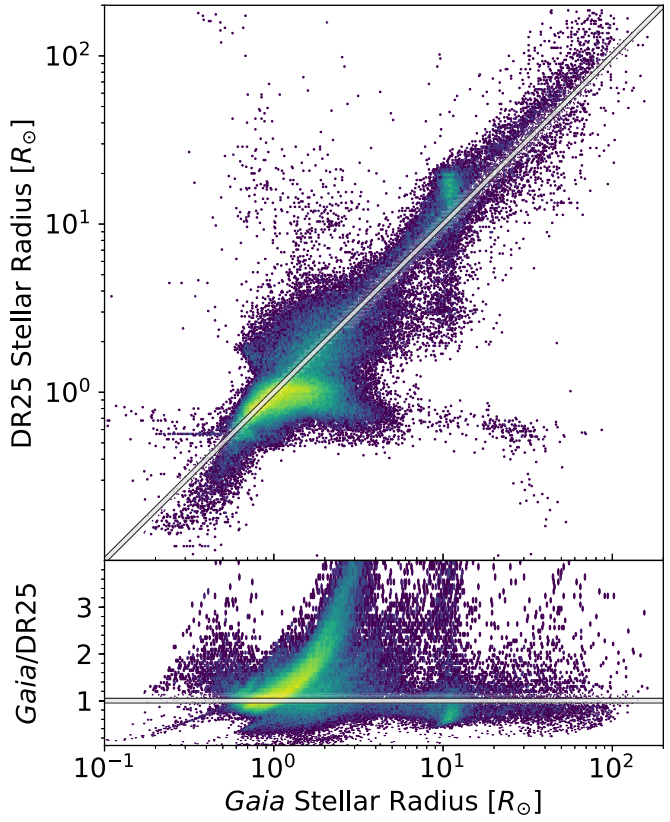

Figure 4. Comparison of radii in the DR25 *Kepler* Stellar Properties Catalog (Mathur et al. 2017) and the radii derived in this paper. The colors represent the density of points. The white and black line is the 1:1 comparison between DR25 radii and our derived radii. The bottom panel shows the ratio between DR25 stellar radii and our stellar radii.

evolutionary states of nearly all *Kepler* targets. To do this, we used solar metallicity parsec evolutionary tracks (Bressan et al. 2012) to define the terminal age main sequence and base of the red-giant branch in the temperature–radius plane, as shown in Figure 5. Assuming that solar metallicity means that the classifications will be only statistically accurate, spectroscopic surveys of the *Kepler* field, such as the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) (De Cat et al. 2015), have confirmed that the average metallicity of *Kepler* targets is solar (Dong et al. 2014).

To classify cool main-sequence stars affected by binarity, we combined a 15 Gyr isochrone at [Fe/H] = 0.5 dex (for warmer stars) with an empirical cut-off determined from a fiducial main sequence (for cooler stars). The latter was determined by fitting Gaussians to radius distributions in fixed $T_{\rm eff}$ bins and fitting a fourth order polynomial to the centroid values, yielding:

$$\log L = -0.69772909 + 2.1574491x + 1.9520690x^2 + 16.041470x^3 - 37.341466x^4 \quad (1)$$

where $x = T_{\rm eff}/4633.78 - 1$. Based on the observed bimodality at a given temperature we choose a cut-off of 1.4 L to define candidate cool main-sequence binaries (blue points in Figure 5). Based on the classifications shown in Figure 5, we find that ≈67% (120,000) of all *Kepler* targets are main-sequence stars, ≈21% (37,000) are subgiants, and ≈12% (21,000) are red giants. Approximately 3100 *Kepler* targets are cool main-sequence binary candidates (blue). Restricting the sample to $T_{\rm eff}$ = 5100–6300 K yields a subgiant fraction of ≈31%, and we confirmed that this fraction is relatively insensitive to apparent magnitude. While this confirms that a substantial fraction of *Kepler* stars are more evolved than previously thought (see also Figure 4), it also demonstrates that some earlier estimates of subgiant contamination rates in the KIC and KSPC were too high, and that *Kepler* did mostly target main-sequence stars. Indeed, the subgiant fractions stated above are upper limits since some stars will be affected by binarity similar to the cool main-sequence stars. The new classifications provided here will provide valuable input for planet occurrence studies, which rely on accurate stellar parameters of the parent sample (e.g., Burke et al. 2015).

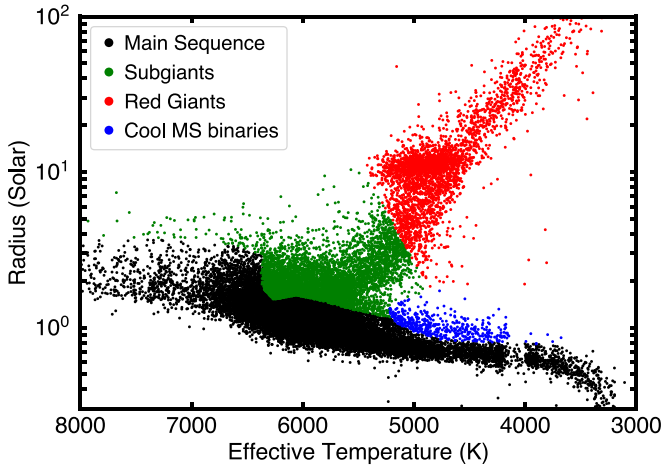

Figure 5. Evolutionary state classifications of all *Kepler* targets based on physically motivated boundaries for evolutionary states (see the text). We find that ≈67% (120,000) of all *Kepler* targets are main-sequence stars (black), ≈21% (37,000) are subgiants (green), and ≈12% (21,000) are red giants (red). Approximately 3100 cool main-sequence stars are affected by binarity (blue).

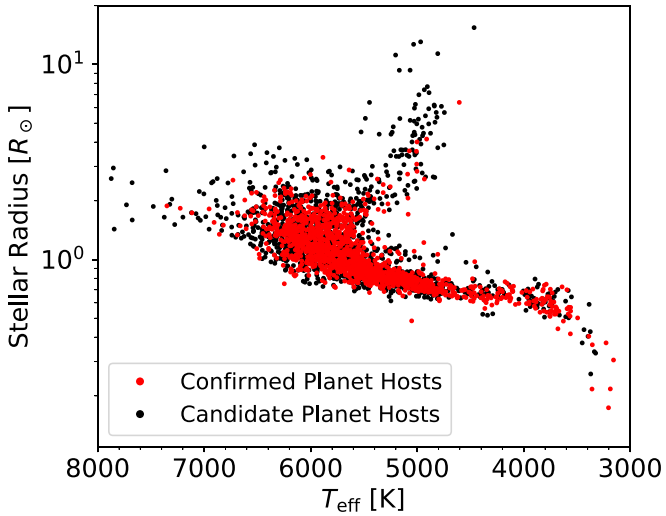

Figure 6. Hertzprung–Russell diagram displaying 1470 *Kepler* confirmed planet hosts (in red) and 1524 *Kepler* candidate planet host stars (in black).

4. Revised Properties of *Kepler* Exoplanets

4.1. The Gaia *H–R Diagram of* Kepler *Planet Host Stars*

Figure 6 displays the stellar radii and $T_{\rm eff}$ distribution of *Kepler* planet host stars, which mostly tracks the overall *Kepler*

Table 2
Revised Parameters of *Kepler* Exoplanets

KIC ID	KOI ID	R_p ($R_\oplus$)	σ_{R_p+} ($R_\oplus$)	σ_{R_p-} ($R_\oplus$)	F_p ($F_\oplus$)	σ_{F+} ($F_\oplus$)	σ_{F-} ($F_\oplus$)	Binary Flag
10797460	K00752.01	2.316	0.156	0.134	104.641	8.011	7.367	0
10797460	K00752.02	2.898	0.955	0.207	10.186	0.780	0.717	0
10854555	K00755.01	2.308	0.410	0.230	652.435	55.222	50.242	0
10872983	K00756.01	4.600	0.721	0.334	122.778	14.831	13.014	0
10872983	K00756.02	3.268	0.341	0.337	457.318	55.241	48.473	0
10872983	K00756.03	1.874	0.510	0.217	863.776	104.338	91.555	0
10910878	K00757.01	4.879	0.284	0.260	21.637	1.680	1.541	1
10910878	K00757.02	3.272	0.207	0.180	6.162	0.478	0.439	1
10910878	K00757.03	2.245	0.150	0.125	76.217	5.917	5.428	1
11446443	K00001.01	14.186	0.602	0.570	903.864	20.395	20.041	1

Note. KIC ID, KOI ID, planetary radii, incident fluxes (and errors where reported), and AO-detected companion flags (Ziegler et al. 2018) of our sample of 4045 *Kepler* confirmed/candidate planets. A slice of our derived parameters is provided here to illustrate the form and format.

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

population in Figure 3. While there are a similar number of confirmed (1470, red) and candidate (1524, black) planet hosts, a larger proportion of the hosts stars are more evolved. This is consistent with the expected larger number of false positives around more evolved stars, which display larger correlated noise due to granulation (Sliski & Kipping 2014; Barclay et al. 2015). Several confirmed and candidate host stars fall below the main sequence and may be metal-poor subdwarfs.

4.2. Comparison to Previous Planet Radii

From the stellar radii derived above, we computed updated planet radii by utilizing the planet–star radius ratio reported in the cumulative KOI table of the NASA Exoplanet Archive (Akeson et al. 2013; Thompson et al. 2018) and then multiplying this ratio by our computed stellar radius. Our revised planet radii and uncertainties are given in Table 2 along with a binary flag for stars with detected companions (binary flag = 1, Ziegler et al. 2018). All of our data products (Tables 1 and 2 and additional parameters) are available at the Mikulski Archive for Space Telescopes (MAST).[7] In an attempt to quantify how much the corrections to stellar radii affect planet radii, we compare planet radii that were calculated using the stellar radii in the KSPC DR25 and in this work in Figure 7. We can see from the top panel that some planets radii change significantly with the stellar radius corrections initiated by *Gaia* DR2. The bottom panel reveals a slight systematic offset from 1 to 5 $R_\oplus$, with our revised planet radii being larger. We expect that this discrepancy arises because most of these planets orbit subgiant stars that were previously misclassified as dwarfs.

In Figure 8, we plot histograms of planet radii, separating candidate (gray) from confirmed (red) planets. Figure 8(a) includes the entire sample of 3959 planets with computed $R_p < 30\ R_\oplus$. Even from this (likely contaminated) sample, we readily recover the previously reported gap in the radius distribution at $\sim 2\ R_\oplus$(Lopez & Fortney 2013; Owen & Wu 2013). Utilizing the precise radii of the CKS (Johnson et al. 2017; Petigura et al. 2017), Fulton et al. (2017) confirmed a dearth of planets with radii $\approx 1.8\ R_\oplus$. In addition, Van Eylen et al. (2017) used asteroseismic radii to investigate the distribution of sizes of smaller planets and found a similar feature. Interestingly, our gap appears to occur at slightly larger planet radii compared to Fulton et al. (2017), and that the intrinsic width of the gap is not visibly increased by the more precise planet radii made possible by *Gaia* DR2 (i.e., the width of the gap is not primarily controlled by a measurement error).

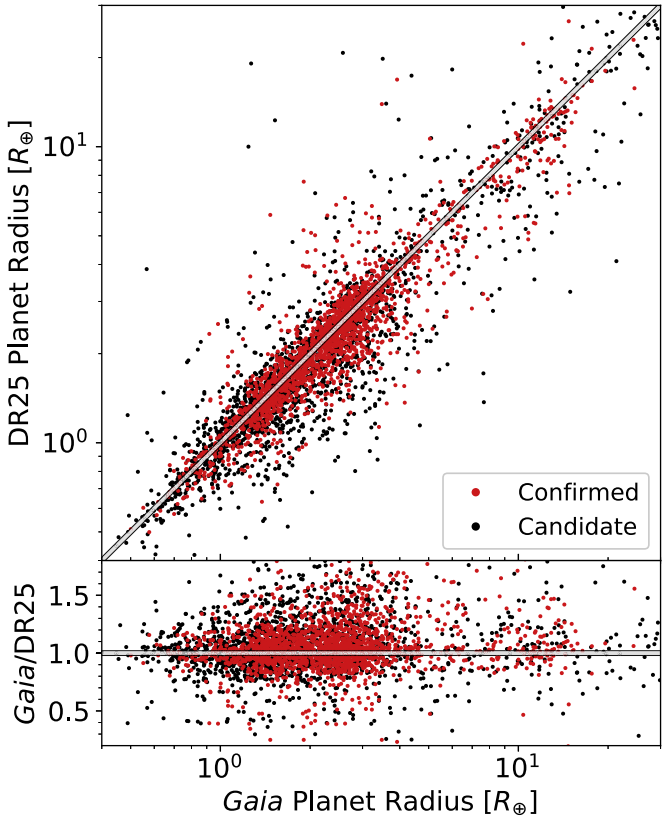

Figure 7. Planet radii calculated from stellar radii derived in this work compared to those based on stellar radii in the *Kepler* DR25 Stellar Properties Catalog (Mathur et al. 2017). The red points are confirmed planets, while the black points are planet candidates. The white and black line is the 1:1 comparison between DR25 planet radii and our derived planet radii. The bottom panel shows the ratio between DR25 radii and our radii.

Next, we implemented the same filters as in Fulton et al. (2017) to ensure a complete, well-defined population of parent stars and planets. Figure 8(b), which includes 503 confirmed and 260 candidate planets, shows our "clean" sample after making the cuts of Fulton et al. (2017): $K_p < 14.2$ mag, $4700 < T_{\rm eff} < 6500$ K, $b < 0.7$, $P < 100$ days, remove all giants and subgiants, and ignore all planets with current dispositions

[7] doi:10.17909/t9-bspb-b780

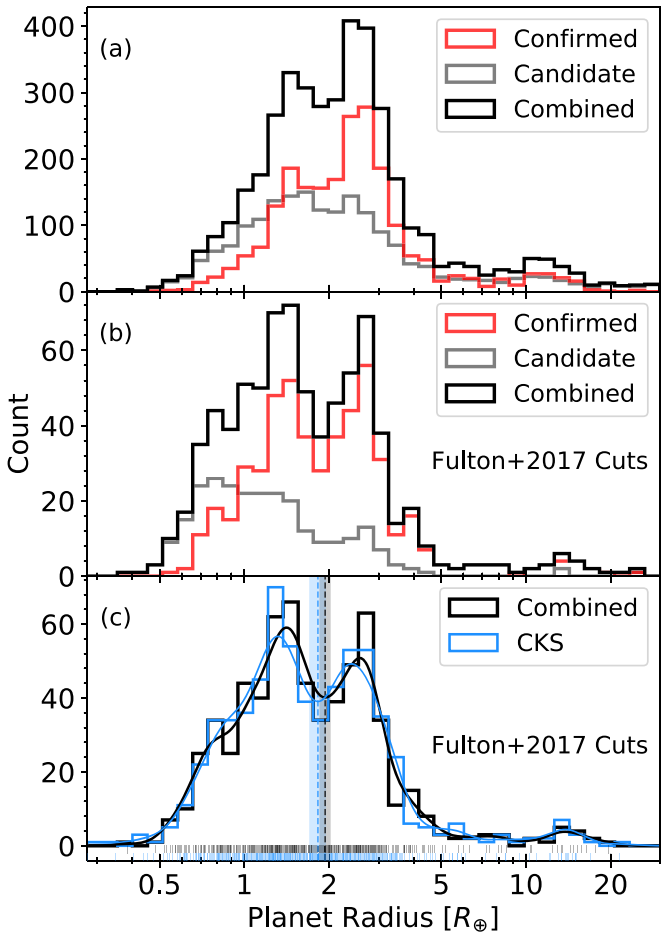

Figure 8. Distribution of *Kepler* exoplanet radii computed in this work. Panel (a): the red, gray, and black histograms contain the confirmed (2120 planets), candidate (1839 planets), and combined samples of 3959 *Kepler* planets, respectively. Panel (b): the same as panel (a) but after performing the sample cuts described in Fulton et al. (2017). Panel (c): the same as panel (b) but using only stars in the CKS sample and overplotting the CKS-derived radii in blue. Smooth lines show KDE distributions, normalized to the total number of planets. The gap locations derived from the KDE distributions (uncorrected for occurrence rates) are $1.94 \pm 0.09\,R_\oplus$ (this work) and $1.83 \pm 0.13\,R_\oplus$ (CKS).

as false positives according to the NASA Exoplanet Archive. We see a significantly deeper gap in the confirmed sample compared to the candidate sample, and it appears to occur at the same location as the combined sample displayed in Figure 8(a). Figure 8(b) also shows a number of very small candidate planets ($R_p < 1.0\,R_\oplus$), although we expect at least some of these planet candidates will be flagged as false positives in the future due to their low signal-to-noise ratio transits.

Figure 8(c) provides a comparison of planet radii for the CKS sample of planets. The black histogram represents the planet radii computed in this work, while the blue histogram comprises those computed by the CKS team (Johnson et al. 2017; Petigura et al. 2017), after applying the Fulton et al. (2017) filters to both. There are 641 planets in each histogram. We also plot a kernel density estimate (KDE) normalized to the total number of planets within each histogram. We use a Gaussian kernel for our KDEs. Below the curves, the vertical, colored ticks are the exact planet radius values that produce the color-matched curves. Finally, the dashed, colored vertical lines and the shaded regions indicate the gap location and the uncertainties, respectively, for each matched KDE.

We calculated the gap location and uncertainties by randomly drawing a planet radius value from a Gaussian distribution with a mean of its actual value and a standard deviation corresponding to its uncertainty. We then produced a KDE out of the simulated planet radii, from which we could identify the gap by finding the relative minimum between the two peaks in the simulated KDE. We repeated this process 100 times and then computed the standard deviation of the distribution of gap locations. We find the location of the gap in our distribution to be at $1.94 \pm 0.09\,R_\oplus$, compared to $1.83 \pm 0.13\,R_\oplus$ for the CKS radii, where both distributions are uncorrected for occurrence rates. We thus find that the gap location derived from our radius values is slightly larger but is still consistent to within 1σ of previously reported values.

In addition, we quantified the effect that occurrence corrections have on the location of the gap. We did this by multiplying both of the KDEs, from 1 to $3\,R_\oplus$, by a linear function so that the relative heights of the "corrected" KDEs match those of the super-Earth and sub-Neptune peaks in Figure 7 of Fulton et al. (2017). The resulting changes ($\approx$–$0.07\,R_\oplus$ or smaller) shift both gap locations to smaller values, but both are within our reported uncertainties.

4.3. Distributions of Planets with Radius and Stellar Irradiation

Figure 9 plots planet radii versus orbit-averaged incident stellar irradiation F in Earth units, using the revised host star parameters and assuming the semimajor axes reported in the NASA Exoplanet Archive and in circular orbits. Planets with slight eccentricities, or near-circular orbits, do not experience a large difference in their incident fluxes compared to planets on perfectly circular orbits since $F \propto 1/\sqrt{1-e^2}$ (Méndez & Rivera-Valentín 2017). We do not account for possible differences in host star masses derived from pre-*Gaia* DR2 stellar radius values and those reported here, as those effects will be much smaller than the change in luminosity and would require isochrone fitting. Several features in this diagram that have been previously described in the literature become more distinct with the improved precision in stellar and planet properties enabled by *Gaia*.

4.3.1. The Small Planet Radius Gap

As shown in Figures 8 and 9, our revised parameters confirm the bimodal distribution of planet radii, with a gap or an "evaporation valley" between the two peaks. The depth of the gap depends on stellar irradiance, with a clear gap just above $2\,R_\oplus$ for $F > 200\,F_\oplus$, the absence of an obvious gap at 30–$200\,F_\oplus$, and a less distinct deficit of planets in this size range at $F < 30\,F_\oplus$. Similar to the integrated value reported in Section 4.2, the gap in the high-irradiance regime appears at slightly larger planet radii than in Fulton et al. (2017; see their Figure 8). We suspect that this difference arises from the sample selection and systematically smaller CKS stellar radii compared to *Gaia* radii for slightly evolved stars, as we find a $\sim$5% systematic underestimation in CKS stellar radii (see also Figure 3 in Fulton & Petigura 2018).

The gap is predicted by models in which photoevaporation due to X-ray and ultraviolet (XUV) radiation, which is more common early in a star's lifetime, removes the light molecular weight envelopes of planets. The relationship between the planet mass, surface gravity, and loss rate means that the

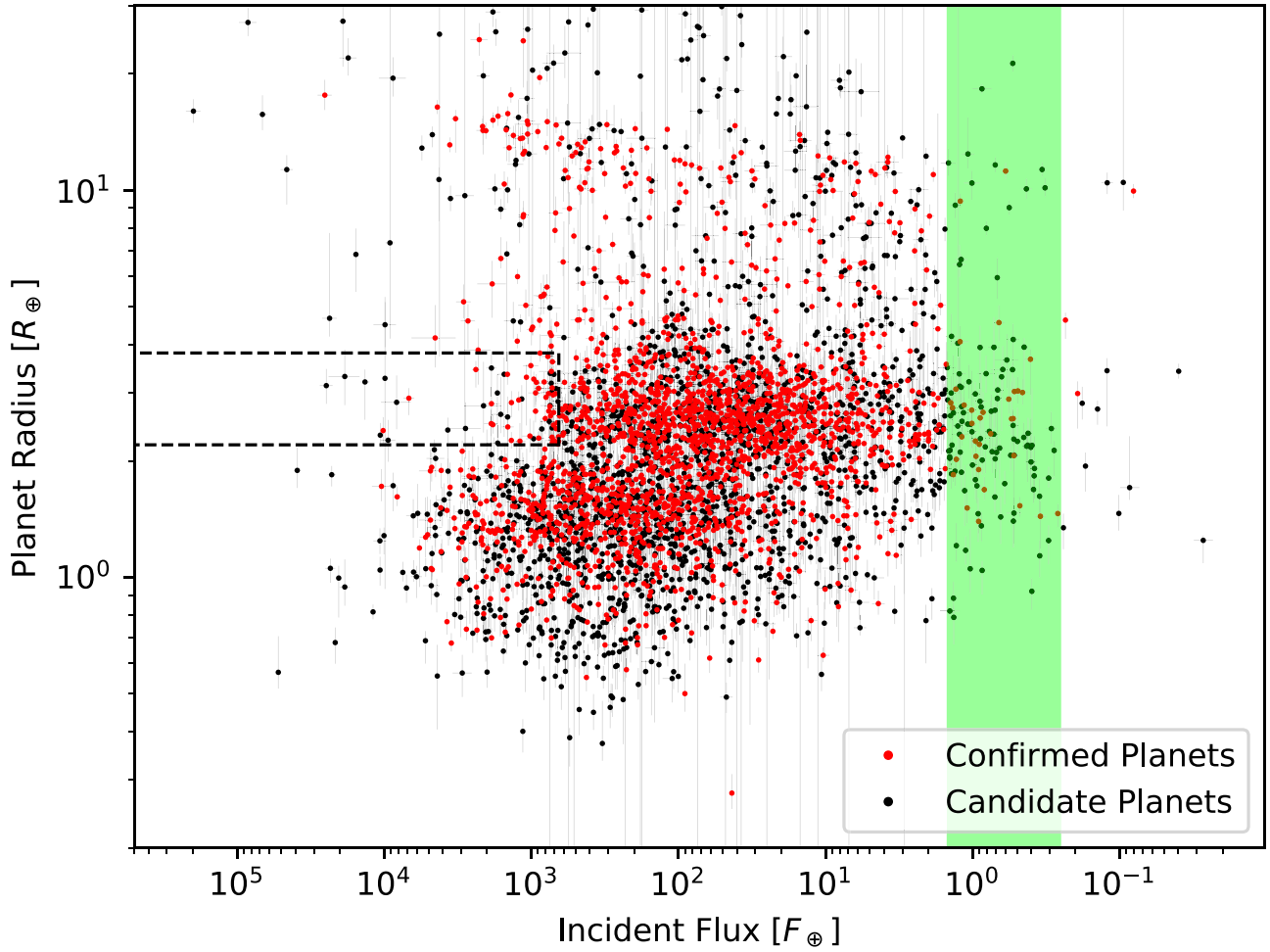

Figure 9. Planet radius vs. incident flux for *Kepler* exoplanets. Red and black dots are confirmed and candidate exoplanets, respectively. We also plot our asymmetric error bars in transparent gray. The dashed line box represents the extension of the super-Earth desert identified in Lundkvist et al. (2016), while the green bar indicates the approximate optimistic habitable zone for FGK stars, as detailed in Kane et al. (2016).

envelopes of intermediate-size planets are efficiently stripped, producing distinct populations of rocky planets and more massive planets that retain their envelopes (Owen & Wu 2017; Jin & Mordasini 2018). This process is more efficient at high irradiance, which explains the prominence of the gap in that regime. Also according to models, the location of the gap constrains the composition of the residual planet "cores." A gap at a larger radius would mean a greater contribution by lower-density ices. For example, Jin & Mordasini (2018) finds that an "evaporation valley" at $1.6\,R_\oplus$ corresponds to an Earth-like composition of silicates and metals, so a valley at a larger radius implies a significant ice component.

Recent investigations have revealed that the location of the radius gap depends on the host star's mass (Fulton & Petigura 2018; Wu 2018). Fulton & Petigura (2018) also investigated the distribution of planets in radius-orbital period space and did not find a strong dependence of the orbital period distribution on the stellar mass (and thus the main-sequence luminosity). This supports XUV-driven photoevaporation as the dominant mechanism sculpting the radii of the exoplanet population, while other mechanisms, such as core-powered mass loss (Ginzburg et al. 2016, 2018), seem less important. Deriving stellar masses for the entire *Kepler* sample will be left for future work, but we note that some differences in the radius distributions may be due to the fact that our sample includes host stars spanning all spectral types (including M dwarfs).

4.3.2. Hot Super-Earth "Desert"

Our revised radius and irradiance values confirm the existence of a deficit or "desert" of super-Earth- to Neptune-size planets at high irradiance (Owen & Wu 2013), i.e., with $2.2 < R_p < 3.8\,R_\oplus$ and $F > 650\,F_\oplus$ (Lundkvist et al. 2016). This desert could be a consequence of photoevaporation of the hydrogen–helium envelopes of sub-Neptune-size planets at stellar irradiance levels more extreme than those which produced the gap (Owen & Wu 2016; Lehmer & Catling 2017), but Ionov et al. (2018) suggests that some other mechanism must be present. Alternatively, the desert could be explained if only rocky planets, not mini-Neptunes, form close to stars because the inner disk is depleted in gas and volatiles (Lopez & Rice 2016). For these two mechanisms, the underlying important variable is the irradiation by the host star and the orbital period/semimajor axis, respectively. These variables are weakly related at the population level because of the wide range of luminosities (five orders of magnitude) of the host stars in the *Kepler* sample. In a plot of the radius versus the orbital period (Figure 10), the boundaries of the desert are also apparent. However, the transition to the desert at short orbital periods for sub-Neptunes is not as abrupt compared

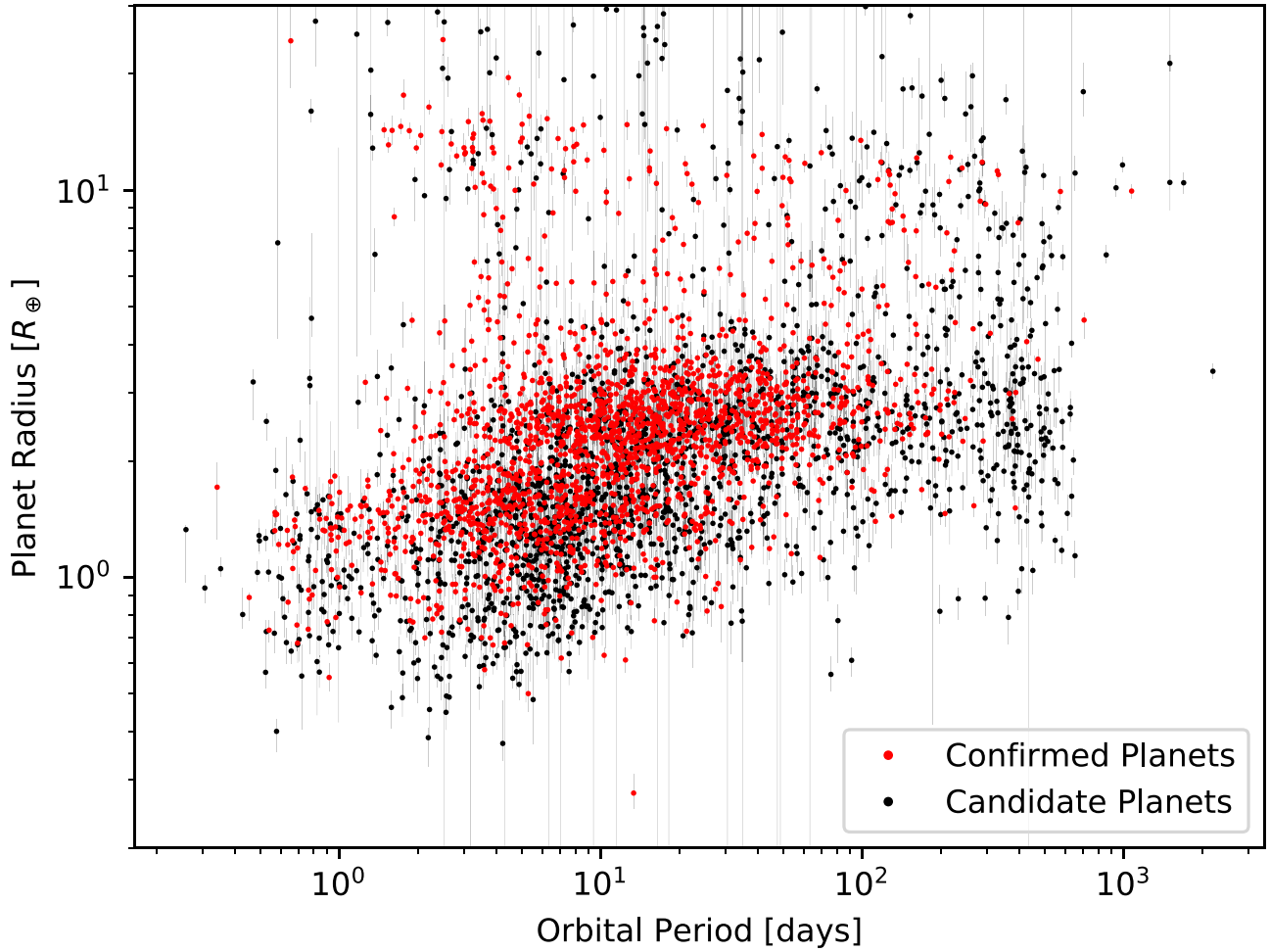

Figure 10. Same as Figure 9, but with the orbital period in place of the incident flux as the *x*-axis.

to the marginally sharper drop-off in planets at $F > 650\,F_\oplus$ in Figure 9, indicating that orbital period is not the underlying "master" variable.

Additionally, we find that the "hot desert" (Lundkvist et al. 2016) is not so empty after all. Forty-six confirmed and 28 candidate planets fall within this range. About half are close to the $650\,F_\oplus$ boundary, and our refined parameters suggest that a distinct edge exists at $\approx 10^3\,F_\oplus$, but 13 confirmed and 2 candidate planets are more than 2σ interior to all the edges of the desert. The host stars of these desert dwellers are almost exclusively subgiant stars that are more massive than the Sun and are evolving toward or at the red-giant branch. This is in contrast with the smaller planets in this irradiance range, which orbit both evolved and main-sequence stars, and larger (sub-Jovian and Jovian) hot planets, which are found around subgiants with a range of masses. A transit detection bias can explain the large number of smaller hot planets around dwarf stars, but not the absence of mini-Neptunes. If the hot mini-Neptunes were the transient remnant of a depleted population, we would expect their host stars to be younger than average, but their evolutionary state suggests that they are older.

Lopez (2017) finds that the absence of sub-Neptunes in the "desert" can be explained if planets of this size have hydrogen–helium envelopes, but *not* substantial envelopes of high molecular weight volatiles (e.g., H_2O) which would be retained. The exceptions here suggest that at least some of these objects do have high molecular weight envelopes, and/or that they have evolved from a different planet population. One explanation for these interlopers is that they are the product of evaporation of still larger objects, i.e., sub-Jovian or even Jupiter-size planets that have lost much of their envelopes. Dong et al. (2018) find that the metallicities of host stars of hot Neptunes are distributed similarly to that of the host stars of hot Jupiters, suggesting a relationship between the two populations. One long-standing idea is that hot Neptunes are the product of massive photoevaporation of a giant planet's envelope (Baraffe et al. 2005).

Another potential explanation for the presence of planets within the "desert" is guided by the theory discussed in Owen & Wu (2017). Because the hosts of these desert-dwelling planets are probably more massive, which is why they have subsequently evolved into subgiants, the integrated XUV radiation from the main-sequence progenitors was lower due to the shorter-main-sequence lifetime and inefficient dynamo operation in star without a convective-radiative boundary ($M > 1.3\ M_\odot$). The dearth of XUV irradiation from these stars allowed their planets to retain low-molecular weight envelopes.

4.3.3. *Inflated Hot Jupiters*

Another feature revealed by Figure 9 is the well-known trend of increasing giant-planet radii with increasing stellar irradiance (e.g., Burrows et al. 2000; Demory & Seager 2011; Laughlin et al. 2011). Confirmed planets with inflated ($>1.2\,R_J$) radii are numerous at $F > 150\,F_\oplus$, consistent with previous work and

planet inflation theory (Lopez & Fortney 2016). These include giant planets orbiting subgiants and low-luminosity red-giants hosts, including previously discovered examples (Grunblatt et al. 2016, 2017). Giant-planet inflation by irradiation could arise from different mechanisms of the transport of heat to the planet's interior or suppression of cooling (Lopez & Fortney 2016). We identified four confirmed inflated giant planets at low ($<150\,F_\oplus$) irradiation: Kepler-447b, Kepler-470b, Kepler-706b, and Kepler-950b, but of these only Kepler-470b satisfy the "cool" inflated planet at more than two sigma significance. Despite the disposition listed in the DR25, Kepler-470b was identified by Santerne et al. (2016) to be an eclipsing binary based on radial velocities.

4.3.4. Habitable Zone Planet Candidates

Finally, we identify candidate and confirmed planets within the circumstellar "habitable zone" where surface temperatures on an Earth-size planet with an Earth-like composition, geology, and geochemistry would permit liquid water. Following Kane et al. (2016), we adopt the "optimistic" definition of $0.25 < F < 1.50\,F_\oplus$ and illustrate this as the green bar in Figure 9. In this habitable zone, we identify 34 confirmed planets and 109 candidate planets. Of these, 30 planet candidates and 8 confirmed planets have $R_p < 2\,R_\oplus$: Kepler-62e, Kepler-62f, Kepler-186f, Kepler-440b, Kepler-441b, Kepler-442b, Kepler-452b (but see also Mullally et al. 2018), and Kepler-1544b. These candidate planets should be priority targets for follow-up observations to vet the planets and to better characterize the host stars, so as to better establish the occurrence of potential Earth-like planets, $\eta_\oplus$.

5. Summary and Conclusions

We presented a reclassification of stellar radii for 177,911 observed by the *Kepler* Mission by combining *Gaia* DR2 parallaxes with the KSPC DR25 (Huber et al. 2014; Mathur et al. 2017). The typical precision of stellar radii is ∼8%, a factor of 4–5 better than previous estimates in the KSPC. Based on the revised stellar radii, we have furthermore rederived radii for 2123 confirmed planets and 1922 planet candidates discovered by *Kepler*. Our main conclusions are as follows:

1. We find that 67% (120,000) of all *Kepler* targets are main-sequence stars, 21% (37,000) are subgiants, and 12% (21,000) are red giants. While many radii are revised to larger values, this demonstrates that previous findings of large subgiant contaminations in the KIC and KSPC were likely overestimated, and that the *Kepler* parent population indeed consists mostly of main-sequence stars.
2. We find evidence for binarity in 3100 cool main-sequence stars (∼2% of the overall sample) based on their inflated radii in the H–R diagram. This demonstrates that *Gaia* parallaxes can be used to efficiently identify binary stars, and we encourage follow-up observations of the binary candidates identified in our work (see Table 1).
3. We confirm the gap in the radius distribution of small *Kepler* planets (Fulton et al. 2017). Our observed gap for the Fulton et al. (2017) sample of $1.94 \pm 0.09\,R_\oplus$ (without occurrence rate corrections, which would shift the value by $\approx$–$0.07\,R_\oplus$) is at a slightly larger radius but is consistent within 1σ with previously reported planet radius distributions. The planet radius–incident flux plot reveals the gap over a wide range of incident fluxes, with the largest gap occurring at $200\,F_\oplus$. The location of the gap has important implications for planet formation and evolution theory, as it can constrain planetary core compositions.
4. Planets do reside in a region of radius-irradiance space previously referred to as the "hot super-Earth desert" (Lundkvist et al. 2016). We identify 74 stars hosting 46 confirmed planets and 28 planet candidates that receive $>650\,F_\oplus$ and have radii between 2.2 and $3.8\,R_\oplus$. However, we confirm that there is a clear paucity of super-Earths in the desert regime, especially at incident fluxes $>1000\,F_\oplus$.
5. We observe a clear inflation trend for hot Jupiters, where inflated planets become numerous at an irradiation level $>150\,F_\oplus$. We identify a few confirmed planets that may be inflated Jupiters at incident fluxes $<150\,F_\oplus$ (Kepler-447b, Kepler-470b, Kepler-706b, and Kepler-950b), but find that the most promising case (Kepler-470b) was previously reported as an eclipsing binary.
6. We identify 34 confirmed planets and 109 planet candidates within the habitable zone. Of these planets, 30 planet candidates and 8 confirmed planets have $R_p < 2\,R_\oplus$: Kepler-62e, Kepler-62f, Kepler-186f, Kepler-440b, Kepler-441b, Kepler-442b, Kepler-452b (but see also Mullally et al. 2018), and Kepler-1544b. These systems in particular represent a high-priority sample for ground-based follow-up observations.

We have applied *Gaia* DR2 measurements to *Kepler* stars and their planets and identified several patterns in the distribution of both stars and planet properties that suggest avenues of future investigation. In this work, we have restricted our refinement of stellar properties to their radii and luminosities, but future work will exploit precise *Gaia* parallaxes by applying stellar evolution models to infer surface gravities, densities, masses, and ages. Planet populations are expected to evolve with time as a result of cooling and contraction of envelopes, photoevaporation of atmospheres, and mutual dynamical scattering. It may also be possible to observe this evolution with sufficiently well-selected and characterized samples of old and young stars and planetary systems (e.g., Mann et al. 2017; Berger et al. 2018). The unprecedented parallaxes provided by *Gaia* will continue to reveal new and interesting information about stars and their companions, and more in-depth analyses of singular systems will inevitably lead to some unpredicted discoveries.

We gratefully acknowledge everyone involved in the *Gaia* and *Kepler* missions for their tireless efforts that have made this paper possible. We also thank the reviewer for helpful feedback that improved this paper, and Erik Petigura and BJ Fulton for helpful discussions. T.A.B. and D.H. thank Savita Mathur for providing supplementary material for the DR25 stellar properties catalog. T.A.B. and D.H. acknowledge support by the National Science Foundation (AST-1717000) and the National Aeronautics and Space Administration under Grants NNX14AB92G issued through the *Kepler* Participating Scientist Program. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the

institutions participating in the *Gaia* Multilateral Agreement. This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation. This research has made use of NASA's Astrophysics Data System. This research was made possible through the use of the AAVSO Photometric All-Sky Survey (APASS), funded by the Robert Martin Ayers Sciences Fund. This research made use of the crossmatch service provided by CDS, Strasbourg. This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

Software: `astropy` (Astropy Collaboration et al. 2013), `dustmaps` (Green et al. 2018), `GNU Parallel` (Tange 2018), `isoclassify` (Huber et al. 2017), `Matplotlib` (Hunter 2007), `mwdust` (Bovy et al. 2016), `Pandas` (McKinney 2010), `SciPy` (Jones et al. 2001).

ORCID iDs

Travis A. Berger https://orcid.org/0000-0002-2580-3614
Daniel Huber https://orcid.org/0000-0001-8832-4488
Eric Gaidos https://orcid.org/0000-0002-5258-6846
Jennifer L. van Saders https://orcid.org/0000-0002-4284-8638

References

Akeson, R. L., Chen, X., Ciardi, D., et al. 2013, PASP, 125, 989
Andrae, R., Fouesneau, M., Creevey, O., et al. 2018, arXiv:1804.09374
Angus, R., Foreman-Mackey, D., & Johnson, J. A. 2016, ApJ, 818, 109
Arenou, F., Luri, X., Babusiaux, C., et al. 2018, arXiv:1804.09375
Astraatmadja, T. L., & Bailer-Jones, C. A. L. 2016, ApJ, 832, 137
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Bailer-Jones, C. A. L. 2015, PASP, 127, 994
Baraffe, I., Chabrier, G., Barman, T. S., et al. 2005, A&A, 436, L47
Barclay, T., Endl, M., Huber, D., et al. 2015, ApJ, 800, 46
Bastien, F. A., Stassun, K. G., Basri, G., & Pepper, J. 2013, Natur, 500, 427
Bastien, F. A., Stassun, K. G., & Pepper, J. 2014, ApJL, 788, L9
Batalha, N. M., Borucki, W. J., Koch, D. G., et al. 2010, ApJL, 713, L109
Berger, T. A., Howard, A. W., & Boesgaard, A. M. 2018, ApJ, 855, 115
Bovy, J., Rix, H.-W., Green, G. M., Schlafly, E. F., & Finkbeiner, D. P. 2016, ApJ, 818, 130
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Brewer, J. M., Fischer, D. A., Basu, S., Valenti, J. A., & Piskunov, N. 2015, ApJ, 805, 126
Brown, T. M., Latham, D. W., Everett, M. E., & Esquerdo, G. A. 2011, AJ, 142, 112
Buchhave, L. A., & Latham, D. W. 2015, ApJ, 808, 187
Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, ApJ, 809, 8
Burrows, A., Guillot, T., Hubbard, W. B., et al. 2000, ApJL, 534, L97
Chaplin, W. J., Basu, S., Huber, D., et al. 2014, ApJS, 210, 1
Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102
Davenport, J. R. A. 2016, ApJ, 829, 23
De Cat, P., Fu, J. N., Ren, A. B., et al. 2015, ApJS, 220, 19
Demory, B.-O., & Seager, S. 2011, ApJS, 197, 12
Dong, S., Xie, J.-W., Zhou, J.-L., Zheng, Z., & Luo, A. 2018, PNAS, 115, 266
Dong, S., Zheng, Z., Zhu, Z., et al. 2014, ApJL, 789, L3
Dressing, C. D., & Charbonneau, D. 2013, ApJ, 767, 95
Everett, M. E., Howell, S. B., Silva, D. R., & Szkody, P. 2013, ApJ, 771, 107
Fulton, B. J., & Petigura, E. A. 2018, arXiv:1805.01453
Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, AJ, 154, 109
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, arXiv:1804.09365
Gaidos, E., & Mann, A. W. 2013, ApJ, 762, 41
Gaidos, E., Mann, A. W., Kraus, A. L., & Ireland, M. 2016, MNRAS, 457, 2877
Ginzburg, S., Schlichting, H. E., & Sari, R. 2016, ApJ, 825, 29
Ginzburg, S., Schlichting, H. E., & Sari, R. 2018, MNRAS, 476, 759
Green, G. M., Schlafly, E. F., Finkbeiner, D., et al. 2018, MNRAS, 478, 651
Green, G. M., Schlafly, E. F., Finkbeiner, D. P., et al. 2015, ApJ, 810, 25
Grunblatt, S. K., Huber, D., Gaidos, E., et al. 2017, AJ, 154, 254
Grunblatt, S. K., Huber, D., Gaidos, E. J., et al. 2016, AJ, 152, 185
Hekker, S., Gilliland, R. L., Elsworth, Y., et al. 2011, MNRAS, 414, 2594
Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, ApJS, 201, 15
Huber, D., Bryson, S. T., Haas, M. R., et al. 2016, ApJS, 224, 2
Huber, D., Chaplin, W. J., Christensen-Dalsgaard, J., et al. 2013, ApJ, 767, 127
Huber, D., Silva Aguirre, V., Matthews, J. M., et al. 2014, ApJS, 211, 2
Huber, D., Zinn, J., Bojsen-Hansen, M., et al. 2017, ApJ, 844, 102
Hunter, J. D. 2007, CSE, 9, 90
Ionov, D. E., Pavlyuchenkov, Y. N., & Shematovich, V. I. 2018, MNRAS, 476, 5639
Jin, S., & Mordasini, C. 2018, ApJ, 853, 163
Johnson, J. A., Petigura, E. A., Fulton, B. J., et al. 2017, AJ, 154, 108
Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python, [Online; accessed 2018-01-19]. http://www.scipy.org/
Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48
Kallinger, T., Hekker, S., Garcia, R. A., Huber, D., & Matthews, J. M. 2016, SciA, 2, 1500654
Kane, S. R., Ciardi, D. R., Gelino, D. M., & von Braun, K. 2012, MNRAS, 425, 757
Kane, S. R., Hill, M. L., Kasting, J. F., et al. 2016, ApJ, 830, 1
Kurucz, R. L. 1993, SYNTHE Spectrum Synthesis Programs and Line Data (Cambridge, MA: Smithsonian Astrophysical Observatory)
Laughlin, G., Crismani, M., & Adams, F. C. 2011, ApJL, 729, L7
Law, N. M., Morton, T., Baranec, C., et al. 2014, ApJ, 791, 35
Lehmer, O. R., & Catling, D. C. 2017, ApJ, 845, 130
Lindegren, L., Hernandez, J., Bombrun, A., et al. 2018, arXiv:1804.09366
Lopez, E. D. 2017, MNRAS, 472, 245
Lopez, E. D., & Fortney, J. J. 2013, ApJ, 776, 2
Lopez, E. D., & Fortney, J. J. 2016, ApJ, 818, 4
Lopez, E. D., Fortney, J. J., & Miller, N. 2012, ApJ, 761, 59
Lopez, E. D., & Rice, K. 2016, arXiv:1610.09390
Lundkvist, M. S., Kjeldsen, H., Albrecht, S., et al. 2016, NatCo, 7, 11201
Mann, A. W., Feiden, G. A., Gaidos, E., Boyajian, T., & von Braun, K. 2015, ApJ, 804, 64
Mann, A. W., Gaidos, E., Vanderburg, A., et al. 2017, AJ, 153, 64
Mathur, S., Huber, D., Batalha, N. M., et al. 2017, ApJS, 229, 30
McKinney, W. 2010, in Proc. 9th Python in Science Conf., ed. S. van der Walt & J. Millman, 51
McQuillan, A., Mazeh, T., & Aigrain, S. 2014, ApJS, 211, 24
Méndez, A., & Rivera-Valentín, E. G. 2017, ApJL, 837, L1
Mosser, B., Elsworth, Y., Hekker, S., et al. 2012, A&A, 537, A30
Mullally, F., Thompson, S. E., Coughlin, J. L., Burke, C. J., & Rowe, J. F. 2018, AJ, 155, 210
Nishiyama, S., Nagata, T., Kusakabe, N., et al. 2006, ApJ, 638, 839
Owen, J. E., & Wu, Y. 2013, ApJ, 775, 105
Owen, J. E., & Wu, Y. 2016, ApJ, 817, 107
Owen, J. E., & Wu, Y. 2017, ApJ, 847, 29
Petigura, E. 2015, PhD thesis, Univ. California
Petigura, E. A., Howard, A. W., Marcy, G. W., et al. 2017, AJ, 154, 107
Pinsonneault, M. H., An, D., Molenda-Żakowicz, J., et al. 2012, ApJS, 199, 30
Planck Collaboration, Abergel, A., Ade, P. A. R., et al. 2014, A&A, 571, A11
Santerne, A., Moutou, C., Tsantaki, M., et al. 2016, A&A, 587, 64
Shabram, M., Demory, B.-O., Cisewski, J., Ford, E. B., & Rogers, L. 2016, ApJ, 820, 93
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Sliski, D. H., & Kipping, D. M. 2014, ApJ, 788, 148
Stello, D., Huber, D., Bedding, T. R., et al. 2013, ApJL, 765, L41
Tange, O. 2018, GNU Parallel 2018 (Ole Tange), Zenodo, doi:10.5281/zenodo.1146014
Thompson, S. E., Coughlin, J. L., Hoffman, K., et al. 2018, ApJS, 235, 38
Torres, G., Fischer, D. A., Sozzetti, A., et al. 2012, ApJ, 757, 161
Van Eylen, V., Agentoft, C., Lundkvist, M. S., et al. 2017, arXiv:1710.05398
Van Eylen, V., & Albrecht, S. 2015, ApJ, 808, 126
Verner, G. A., Elsworth, Y., Chaplin, W. J., et al. 2011, MNRAS, 415, 3539
Wu, Y. 2018, arXiv:1806.04693
Yu, J., Huber, D., Bedding, T. R., et al. 2018, arXiv:1802.04455
Ziegler, C., Law, N. M., Baranec, C., et al. 2018, AJ, 155, 161
Zinn, J. C., Pinsonneault, M. H., Huber, D., & Stello, D. 2018, arXiv:1805.02650