





A Spectroscopic Analysis of the California-Kepler Survey Sample. I. Stellar Parameters, Planetary Radii, and a Slope in the Radius Gap

Cintia F. Martinez¹ , Katia Cunha^{1,2}, Luan Ghezzi¹ , and Verne V. Smith³

¹ Observatório Nacional, Rua General José Cristino, 77, 20921-400 São Cristóvão, Rio de Janeiro, RJ, Brazil; cmartinez@on.br

² Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA

³ National Optical Astronomy Observatory, 950 North Cherry Avenue, Tucson, AZ 85719, USA

Received 2018 October 19; revised 2019 February 27; accepted 2019 March 5; published 2019 April 10

Abstract

We present results from a quantitative spectroscopic analysis conducted on archival Keck/HIRES high-resolution spectra from the California-Kepler Survey (CKS) sample of transiting planetary host stars identified from the *Kepler* mission. The spectroscopic analysis was based on a carefully selected set of Fe I and Fe II lines, resulting in precise values for the stellar parameters of effective temperature (T_{eff}) and surface gravity ($\log g$). Combining the stellar parameters with *Gaia* DR2 parallaxes and precise distances, we derived both stellar and planetary radii for our sample, with a median internal uncertainty of 2.8% in the stellar radii and 3.7% in the planetary radii. An investigation into the distribution of planetary radii confirmed the bimodal nature of this distribution for the small-radius planets found in previous studies, with peaks at $\sim 1.47 \pm 0.05$ and $\sim 2.72 \pm 0.10 R_{\oplus}$ with a gap at $\sim 1.9 R_{\oplus}$. Previous studies that modeled planetary formation that is dominated by photoevaporation predicted this bimodal radii distribution and the presence of a radius gap, or photoevaporation valley. Our results are in overall agreement with these models, as well as core powered mass-loss models. The high internal precision achieved here in the derived planetary radii clearly reveal the presence of a slope in the photoevaporation valley for the CKS sample, indicating that the position of the radius gap decreases with orbital period; this decrease was fit by a power law of the form $R_{\text{pl}} \propto P^{-0.11}$, which is consistent with both photoevaporation and core powered mass-loss models of planet formation, with Earth-like core compositions.

Key words: planetary systems – stars: fundamental parameters – techniques: spectroscopic

Supporting material: machine-readable tables

1. Introduction

One well-known axiom in exoplanetary studies connects the properties of exoplanets with the properties of their host stars through the expression that “one can only know the planet to the level that the host star is known.” To know certain exoplanetary physical properties (such as radius, mass, and mean density) requires the knowledge of those same physical properties for the host star: for transiting exoplanets, it is possible to determine the planetary radius relative to the stellar radius (R_p/R_*) from the analysis of the transit light curve, while the planetary mass depends on the host stellar mass ($M_p \propto M_*^{2/3}$) and is derived from the radial velocity curve.

Although several thousand exoplanet candidates have been discovered by the *Kepler* mission (Borucki et al. 2010; Koch et al. 2010; Borucki 2016), the initial stellar parameters, as derived from the *Kepler* Input Catalog (KIC), were limited in accuracy, as the KIC provided seven-band photometry (g , r , i , and z , plus J , H , and K from 2MASS) along with a narrow filter centered on the Mg I b lines used as a luminosity indicator (Borucki 2016). The deduced stellar radii were found to have a scatter of about 30%–40% (Huber et al. 2014), with the errors for late-type dwarfs being even larger (Dressing & Charbonneau 2013). Errors of this size result in large uncertainties in exoplanetary properties and could mask correlations and trends in exoplanet properties or the types of systems they inhabit.

Improvements in the derived stellar radii can be accomplished by conducting precise, quantitative spectroscopic analyses of the host stars, in particular, using high-quality, high-resolution spectra. In addition to high-resolution spectroscopic data, *Gaia* (Gaia Collaboration et al. 2018) has now provided precise parallaxes

for a large number of *Kepler* exoplanet host stars. The *Gaia* parallaxes, combined with tightly constrained stellar parameters derived from spectroscopy, result in stellar and exoplanetary radii with accuracies of 3%–5% (Stassun et al. 2017).

A recent example of how improved measurements for stellar radii can reveal new characteristics in exoplanet populations can be found from the California-Kepler Survey (CKS; Petigura et al. 2017), where Fulton et al. (2017) discovered a bimodal distribution for small-planet radii with peaks at ~ 1.3 and $\sim 2.4 R_{\oplus}$ and a gap in between that points to a transition radius separating super-Earths from sub-Neptunes. An earlier detection of this small-planet gap was prevented due to the large uncertainties in the exoplanetary radii, although a small-planet gap was originally predicted by several formation models (Owen & Wu 2013; Jin et al. 2014; Lopez & Fortney 2014; Chen & Rogers 2016; Lopez & Rice 2016), which predicted that gaseous planets may suffer photoevaporation of their envelopes by radiation coming from their host stars. The presence of the small-planet gap has now been confirmed by other studies (Berger et al. 2018; Fulton & Petigura 2018; Van Eylen et al. 2018). In addition to photoevaporation, Ginzburg et al. (2016) and, more recently, Ginzburg et al. (2018) showed that the small-planet gap can also be produced by a young, hot planetary core, whose energy can drive atmospheric mass loss, with the ability to retain an atmosphere depending on the mass of the planet. Other processes, such as internal planetary outgassing (Dorn et al. 2018) or large impacts on young planets (Inamdar & Schlichting 2016), can both produce a planetary atmosphere and remove it.

In this study, a homogeneous spectroscopic analysis has been carried out in order to derive precise stellar parameters

(effective temperatures T_{eff} and surface gravity) for a sample of *Kepler* hosts using a homogeneous set of high-resolution optical Keck/High Resolution Echelle Spectrometer (HIRES) spectra made public by the CKS team (Fulton et al. 2017; Johnson et al. 2017; Petigura et al. 2017). Stellar radii are derived using the *Gaia* DR2 parallaxes (Bailer-Jones et al. 2018; Lindgren et al. 2018), with the improvements in the stellar radii given by the precise parallaxes and better distances from *Gaia*. Based on our analysis, we independently derive planetary radii and confirm the presence of a small-planet gap in the distribution of exoplanetary sizes.

2. Observations

The high-resolution spectra analyzed in this study were obtained as part of the CKS (Fulton et al. 2017; Johnson et al. 2017; Petigura et al. 2017), a large observational campaign targeting stars identified as *Kepler* objects of interest (KOIs). The CKS campaign was conducted between 2012 and 2014 using HIRES (Vogt et al. 1994) at the Keck telescope. All CKS spectra analyzed here were reduced by Petigura et al. (2017) and are publicly available in the Keck Observatory Archive. The spectra were obtained from <https://california-planet-search.github.io/cks-website/>. (Sample HIRES spectra of the CKS are shown in Figure 2 of Petigura et al. 2017.)

From the full CKS sample of 1305 stars, we removed a small sample of 20 stars that had low-quality spectra (all having signal-to-noise ratios (S/Ns) lower than ~ 30). The remaining sample containing 1285 stars was analyzed spectroscopically in this study; most of these stars have spectra with S/N ratios between 40 and 70, while some $\sim 8\%$ of the spectra have excellent quality with S/Ns higher than 100.

3. Analysis

3.1. Spectroscopic Stellar Parameters

We derived stellar parameters (T_{eff} , $\log g$, and microturbulent velocities), as well as metallicities (taken to be represented by $[\text{Fe}/\text{H}]$), for the studied stars using standard techniques employed in quantitative stellar spectroscopy, which relies on equivalent-width (EW) measurements of selected samples of Fe I and Fe II lines.

All abundance calculations were done under the assumption of local thermodynamic equilibrium (LTE) using 1D model atmospheres. The atmospheric parameters were obtained by iterating until the line-by-line Fe I abundances, $A(\text{Fe I})$, exhibit no dependence with the excitation potential (EP) of the transitions (excitation equilibrium), and, at the same time, the values of $A(\text{Fe I})$ show no trend with the reduced EWs ($\log(\text{EW}/\lambda)$), while finally, the mean Fe I and Fe II abundances reach agreement (ionization equilibrium). These three conditions define the stellar T_{eff} , $\log g$, and microturbulent velocity (ξ). Figure 1 shows an example of the iterated solution for the effective temperature, $\log g$, and microturbulent velocity, as well as the abundance of iron for a sample star, KOI-1.

In order to analyze the large number of stars in our sample in an efficient and homogeneous way, we used the automated stellar parameter and metallicity pipeline described in detail in Ghezzi et al. (2010, 2018). This code uses the updated version of the routine ARES (Sousa et al. 2015) to measure EWs of Fe I and Fe II lines automatically, the abundance analysis code MOOG (Snedden 1973) to compute the iron abundances, and model atmospheres from the Kurucz ATLAS9 ODFNEW grid

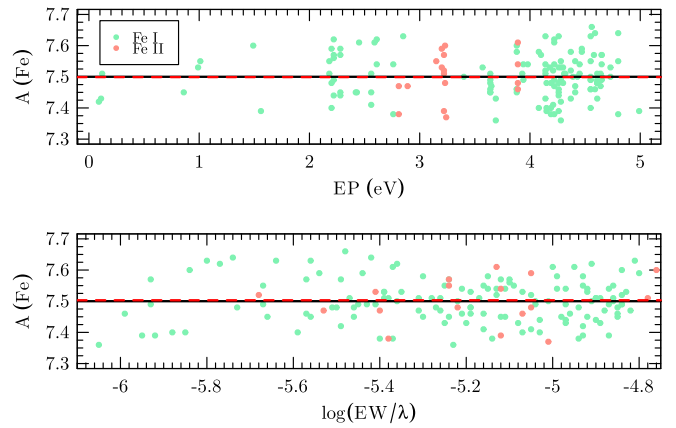


Figure 1. Example of the applied methodology to KOI-1, one of the stars in our sample. Iron abundances ($A(\text{Fe I})$) are shown as a function of the EP of the Fe I and Fe II transitions (top panel) and the reduced EW of the lines (EW/λ ; bottom panel). The atmospheric parameters, iron abundances, and microturbulent velocities are obtained once the correlation coefficients of the dashed lines show no dependence with the line parameters.

(Castelli & Kurucz 2004). In summary, the code starts with a model atmosphere calculated assuming solar values for T_{eff} , $\log g$, and metallicity and then iterates until obtaining a final adjusted value for the spectroscopic parameters of each star.

The adopted line list in this study was taken from Ghezzi et al. (2018) and consists of 158 Fe I and 18 Fe II isolated and unblended lines. The $\log gf$ values of the Fe I and II lines were obtained in Ghezzi et al. (2018) from an inverse solar analysis using a Kurucz ATLAS9 ODFNEW model atmosphere for the Sun ($T_{\text{eff}} = 5777 \text{ K}$, $\log g = 4.44$, $[\text{Fe}/\text{H}] = 0.00$, and $\xi = 1.00 \text{ km s}^{-1}$) and an adopted solar abundance ($A(\text{Fe}_{\odot}) = 7.50$) from Asplund et al. (2009).

As a consistency check, we also manually measured the EWs of a total of 540 Fe I and Fe II lines in four sample stars (KOIs 64, 268, 280, and 5782, with a mean S/N of 70 in their spectra) using the IRAF package *splot*. A comparison of our manual EW measurements with the automatic ones using the ARES code is presented in Figure 2. Despite the fact that there were some lines with discrepant EW measurements (these are not an issue in the final solution because the pipeline performs two rounds of σ clipping to remove lines with abundances that are too discrepant from the average values), the EWs compared well, showing, on average, a small offset of 1.25 m\AA (in the sense of the ARES EW being larger than ours) and an rms scatter of 3.18 m\AA .

Table 1 presents the resulting effective temperatures, surface gravities, microturbulent velocities, and stellar radii (these will be discussed in Section 3.2) and their respective uncertainties for all stars in our sample. The metallicities obtained for the CKS sample will be presented and discussed in a forthcoming paper (C. F. Martinez et al. 2019, in preparation hereafter Paper II).

3.2. Stellar and Planetary Radii

Precise stellar radii result in precise planetary radii, a crucial parameter necessary to unveil planetary composition and ultimately define the transition from rocky to gaseous planets.

To calculate the stellar radii (R_{\star}), we used the Stefan–Boltzmann law, which depends on the Stefan–Boltzmann constant (σ_{sb}), stellar effective temperature (T_{eff}), and

Table 1
Stellar Parameters and Radii

KOI No.	T_{eff} (K)	δT_{eff} (K)	$\log g$ (dex)	$\delta \log g$ (dex)	ξ (km s $^{-1}$)	$\delta \xi$ (km s $^{-1}$)	R_{\star} (R_{\odot})	δR_{\star} (R_{\odot})
K00245	5410	11	4.55	0.02	0.83	0.04	0.78	0.01
K01925	5418	18	4.48	0.06	0.82	0.04	0.90	0.02
K01612	6167	35	4.32	0.09	1.43	0.06	1.24	0.03
K00069	5646	21	4.50	0.04	0.91	0.03	0.94	0.02
K03167	5467	29	4.48	0.07	1.03	0.06	0.88	0.02
K00082	4849	49	4.51	0.08	0.62	0.12	0.73	0.01
K00975	6227	42	4.00	0.03	1.50	0.06	1.95	0.08
K02687	5800	22	4.55	0.06	1.02	0.03	0.94	0.02
K01924	5973	28	3.86	0.06	1.53	0.03	2.57	0.14
...

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

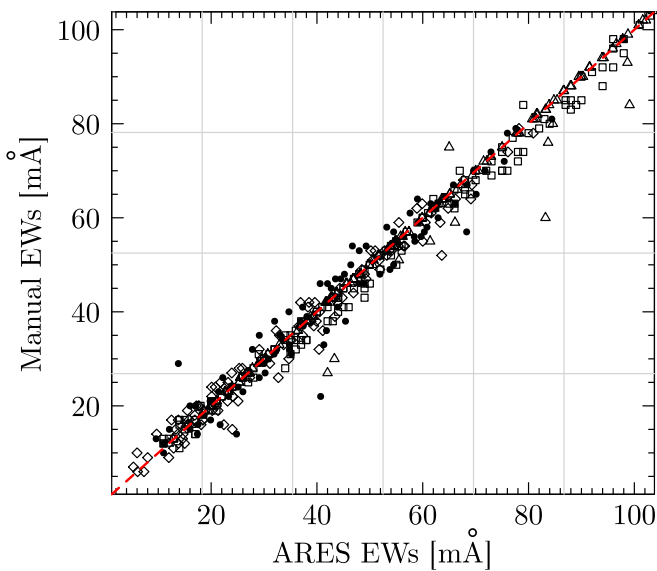


Figure 2. Comparison between automatic (using the ARES code) and manual (using the *splot* task from IRAF) EW measurements for a sample of 540 Fe I and Fe II lines in four target stars: KOI-64 (open squares), KOI-268 (filled circles), KOI-280 (open diamonds), and KOI-5782 (open triangles). The red dashed line represents perfect agreement.

luminosity (L_{\star}),

$$R_{\star} = \left(\frac{L_{\star}}{4\pi\sigma_{\text{sb}}T_{\text{eff}}^4} \right)^{1/2},$$

where

$$L_{\star} = L_0 10^{-0.4M_{\text{bol}}}.$$

Here, L_0 is the zero-point of the bolometric magnitude scale (Mamajek et al. 2015) and M_{bol} is the bolometric magnitude related to photometric apparent magnitude (m_k), extinction (A_k), and bolometric correction (BC) in the same band and distance modulus (μ) via

$$M_{\text{bol}} = m_k - A_k - \mu + \text{BC}_k.$$

For each target star, we used the distances estimated by Bailer-Jones et al. (2018), when available. Bailer-Jones et al. (2018) adopted a Bayesian approach and geometric priors to obtain the distances from the *Gaia* DR2 parallaxes and considered the

systematic parallax offsets determined from *Gaia*'s observations of quasars (Lindegren et al. 2018; Zinn et al. 2018). We used the 2MASS K band, combined with the reddening $E(B - V)$ derived from the 3D dust map of Green et al. (2018) and transformed into A_k extinction using the relations in Bilir et al. (2008), and the *isoclassify* package (Huber et al. 2017) in its “direct mode” using T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, and A_v extinction as inputs to interpolate BCs from MIST grids (Choi et al. 2016) and calculate absolute magnitudes and stellar luminosities.

In order to calculate the stellar radii, we combined all of these parameters with our derived T_{eff} ; the results are presented in Table 1.

Finally, we used the stellar radii to determine the planet radii using the values of transit depth, ΔF , which are the fraction of stellar flux lost at the minimum of the planetary transit, cataloged in Thompson et al. (2018), and the equation from Seager & Mallén-Ornelas (2003):

$$R_{\text{pl}} = 109.1979 \times (\Delta F \times 10^{-6})^{1/2} \times R_{\star}. \quad (1)$$

To assure a higher level of reliability in the computed planetary radii, we only included in our sample those planets whose host stars have $\leq 10\%$ error in their derived stellar radius (see Figure 4). In addition, we removed KOIs that have been classified as “false positives” (adopting the same dispositions as in Thompson et al. 2018). The derived radii for the final planetary sample in this study are presented in Table 2.

3.3. Uncertainties in the Derived Parameters

The internal errors in the derived effective temperatures and microturbulent velocities were calculated by changing these parameters until the slopes of both $A(\text{Fe I})$ versus EP and $A(\text{Fe I})$ versus $\log(\text{EW}/\lambda)$, respectively, reach the same values as the errors in the slopes from the converged solution. The uncertainties in $\log g$ were estimated by varying this parameter until the Fe I and Fe II mean abundances differed by one standard deviation of the mean of $A(\text{Fe I})$.

In order to estimate the total error budget in the derived stellar and planetary radii for our sample, we consider the individual contributions of the errors in each one of the parameters used in the computation of the radii. (See also the discussion in Fulton & Petigura 2018.)

The error in the K -band stellar magnitudes (m_k) contributes 1% to the stellar radius error if we consider 0.022 mag to be the median error in m_k from 2MASS (Skrutskie et al. 2006) for our

Table 2
Planetary Radii

Planet	R_{pl} (R_{\oplus})	δR_{pl} (R_{\oplus})
K00001.01	13.62	0.30
K00002.01	17.93	1.07
K00007.01	4.49	0.13
K00010.01	16.62	1.14
K00017.01	14.31	0.35
K00018.01	16.78	0.80
K00020.01	21.19	0.63
K00022.01	14.06	0.34
K00041.01	2.48	0.07
K00041.02	1.38	0.04
...

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

target stars. The contribution due to errors in the extinction A_k is even smaller, given that the A_k values for the target stars are quite small, ranging between 0.001 and 0.052 mag, with a median A_k of 0.009 mag. If we were to completely neglect extinction, this would result in an error of 0.6% in the stellar radii (using the median A_k in the estimate). The errors in the K -band BCs (BC_k) are mainly dominated by uncertainties in the effective temperatures. We estimate an error in BC_k in the same manner as in Fulton & Petigura (2018), by changing the effective temperature of a solar-type star by the median error in our T_{eff} values (40 K) and investigating the corresponding change in BC_k . (We also investigated the effect of the errors in $\log g$ and metallicities, but these were found to be negligible.) Taking the error obtained in BC_k for a test solar-type star as typical, we estimate a change of 0.03 mag in BC_k (Huber et al. 2017 also estimated the error in BC_k to be 0.03 mag) and an error of 0.3% for the stellar radius.

The distances and respective errors for the target stars were taken directly from Bailer-Jones et al. (2018) and correspond to a median error of 0.006 mag for the “distance modulus” (μ) and 0.08% error in R_* . The internal precision (median error) in the effective temperatures in this study is 40 K, corresponding to 2% in the R_* error.

Combining all of the errors in the parameters discussed above in quadrature and propagating the errors, we obtain a median internal uncertainty in our derived stellar radii distribution of $\sim 2.8\%$. This uncertainty in the stellar radii has a direct impact, along with the transit depth (ΔF) errors, on the determination of the planetary radii errors. We adopted the transit depth values ΔF and respective errors from Thompson et al. (2018), which, for the planets in our sample, result in 4% internal precision in ΔF and correspond to a 3.7% internal precision for the R_{pl} error budget. A summary of the contributions to the error budgets in the R_* and R_{pl} determinations is presented in Table 3.

4. Results

The atmospheric parameter distributions for the studied sample are shown in the different histograms of Figure 3; most of the target stars have effective temperatures roughly between 4800 and 6500 K, having a peak around the solar effective temperature ($T_{\text{eff}} = 5777$ K) and a smaller peak corresponding to cooler stars at $T_{\text{eff}} \sim 5000$ K. The $\log g$ distribution for the studied sample corresponds mostly to unevolved stars, with the

Table 3
Error Budget

Parameter	Median Uncertainty
m_k	0.022 mag
A_k	0.009 mag
BC_k	0.03 mag
μ	0.006 mag
T_{eff}	40 K
R_{star}	2.8%
ΔF	4%
R_{pl}	3.7%

distribution having a peak between roughly $\log g = 4.3$ and 4.5 dex, but it also contains a tail with more evolved stars having $\log g < \sim 4.2$. Most of the target stars (about 80%) are from the solar neighborhood having distances (from Bailer-Jones et al. 2018) generally within ~ 1 kpc of the Sun (Figure 3, panel (c)).

As discussed in the previous section, the stellar radii in this study were calculated using distances based on *Gaia* DR2 parallaxes. In Figure 4, we present an H-R diagram showing the derived effective temperatures and stellar radii, color-coded by the errors in the stellar radii. The median internal uncertainty in the radius distribution in this study is $\sim 2.8\%$ (Section 3.3), and most target stars follow the expected sequences in the H-R diagram, showing a densely populated main sequence and the presence of some evolved stars (as expected from the T_{eff} and $\log g$ distributions obtained for the targets and shown in Figure 3). There are, however, clear outliers that occupy unexpected loci in the diagram corresponding to at least a 10% error in the stellar radii (represented by green, yellow, and red points in Figure 4).

4.1. Previous Results from the Literature

4.1.1. Stellar Parameter Comparisons for the CKS Sample

As previously mentioned, the study by Petigura et al. (2017) was the first to present a spectroscopic analysis of the CKS sample. That study derived spectroscopic parameters using two different spectral synthesis techniques in LTE: SpecMatch and SME@XSEDE. This methodology is very distinct from the one used in our study. SpecMatch was specifically designed for the CKS project. In summary, SpecMatch fits model spectra (computed with Kurucz models by Coelho et al. 2005) individually to five different wavelength segments in the observed spectra and averages the resulting sets of parameters T_{eff} , $\log g$, metallicity, and $v \sin i$ of each segment. The results from the Brewer et al. (2016) catalog were used to calibrate the SpecMatch results.

The spectral synthesis technique SME@XSEDE is an automated version of the spectral synthesis code Spectroscopy Made Easy (SME; Valenti & Piskunov 1996). It uses a line list with atomic parameters taken from the VALD database to interpolate between a grid of plane-parallel MARCS model atmospheres (Gustafsson et al. 2008) until the optimal solution is found, using a χ^2 minimization. The final results in Petigura et al. (2017) were obtained by applying linear corrections to the raw SME@XSEDE results to put them on the SpecMatch scale (originally calibrated to the Brewer et al. 2016 scale), while for those target stars with consistent results between SpecMatch

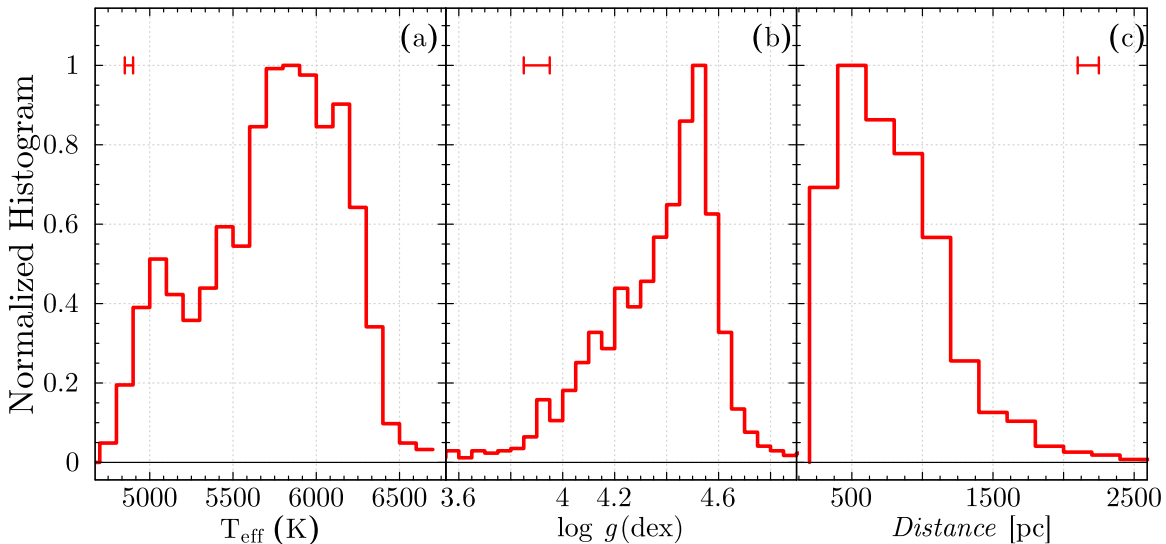


Figure 3. Effective temperature (panel (a)) and surface gravity (panel (b)) distributions for the sample stars. Panel (c) has the distribution of the stellar distances for the sample, with distances obtained from *Gaia* DR2 parallaxes and taken from Bailer-Jones et al. (2018). The median uncertainties in the parameters are plotted in the upper corner of each panel.

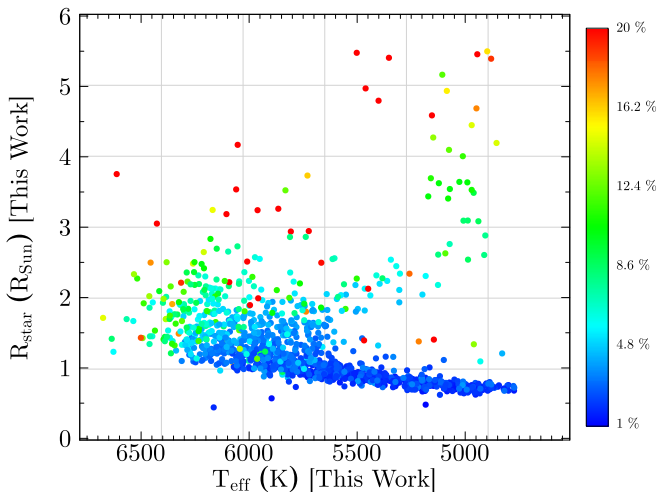


Figure 4. Derived stellar radii vs. the effective temperatures for the target stars. The stellar radii are based on *Gaia* DR2 parallaxes and shown color-coded by the stellar radii errors, from an $\sim 1\%$ error in blue to a 20% error or larger (up to $\sim 200\%$) in red.

and SME@XSEDE, the authors adopted the mean value from the two methods.

The top panels of Figure 5 show a comparison of the derived atmospheric parameters T_{eff} and $\log g$ with results from Petigura et al. (2017). In general, there is very good agreement (within the uncertainties) with the stellar parameters derived by Petigura et al. (2017), although there is a small systematic difference of about ~ 60 K in the effective temperatures (our T_{eff} scale being hotter than that of Petigura et al. 2017). When considering mean differences for $\log g$, there is only a small offset with our derived $\log g$ values ($\langle \text{Petigura et al. 2017} - \text{this study} \rangle = -0.035$ dex; rms = 0.14 dex), but it should be noted that 12 stars in our sample have $\log g > \sim 4.7$; although their $\log g$ errors are within the expected uncertainties (the mean of the $\log g$ errors is 0.12 dex), these results are all systematically higher than those of Petigura et al. (2017).

More recently, Brewer & Fischer (2018) also analyzed the CKS data set. They adopted the SME semi-automated spectral

synthesis code (also used in Brewer et al. 2016) to fit the observed spectra and constrain the stellar parameters. The synthesis code uses as input an atomic and molecular line list and a grid of plane-parallel model atmospheres (Castelli & Kurucz 2004) and finds the best solution for global parameters such as T_{eff} , $\log g$, $[M/H]$, or v_{macro} . The authors applied an inverse solar analysis to adjust the oscillator strengths, $\log gfs$, of the transitions and performed a combination of spectral synthesis and asteroseismic techniques to obtain their final results.

A comparison of our derived stellar parameters with those obtained by Brewer & Fischer (2018) is shown in the bottom panels of Figure 5 for 847 stars in common. The conclusions are similar to those found with Petigura et al. (2017), which is expected, given that these authors effectively calibrated their results to be on the Brewer & Fischer (2018) and, consequently, the Brewer et al. (2016) scale. For the effective temperatures, the mean difference ($\langle \text{BF18} - \text{this study} \rangle$) is -69 ± 3 K and the rms = 77 K, again indicating a small systematic offset in the two T_{eff} scales. For $\log g$, there is a small systematic difference of -0.044 dex, which is not significant.

4.1.2. Stellar Parameter Comparisons for Other Samples of Kepler Stars

Buchhave et al. (2012) used multiple observations from high-resolution spectrographs on several telescopes to derive stellar parameters for 152 planet-hosting stars discovered by the *Kepler* mission. They derived the stellar parameters, T_{eff} , $\log g$, $[m/H]$, and v_{rot} , using the spectral synthesis code Stellar Parameter Classification (SPC), which uses a library of model atmospheres (Kurucz 1992) to synthesize the spectrum between 5050 and 5360 Å and measures a cross-correlation function peak that indicates how well the synthetic data reproduce the observed ones.

The top panels of Figure 6 show that there is a small trend in the comparison of our effective temperature scale with that of Buchhave et al. (2012): at higher temperatures ($T_{\text{eff}} > 5750$ K), our effective temperatures are systematically larger by

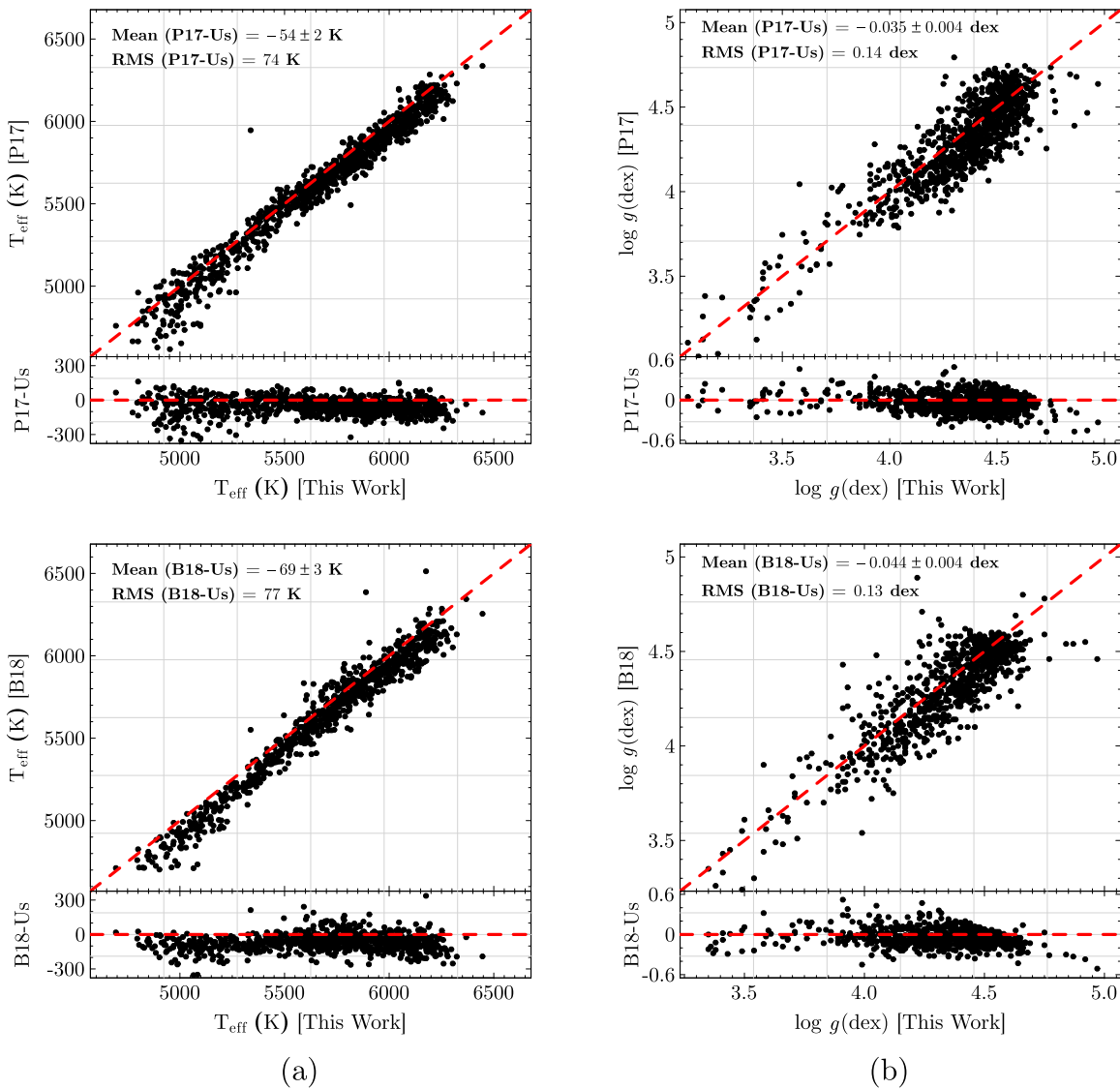


Figure 5. Comparison of the effective temperatures (panels (a)) and $\log g$ values (panels (b)) derived in this work and in Petigura et al. (2017) for 1013 stars in common (top panels) and for 847 stars in common with Brewer & Fischer (2018) (bottom panels). The corresponding rms scatters are indicated in each case. The red dashed lines represent the equality.

82 ± 9 K than those of Buchhave et al. (2012), while for the range between 5200 and 5750 K, our temperatures are systematically lower by 21 ± 10 K; for T_{eff} values lower than 5200 K, there is no systematic trend but a larger scatter. The agreement with $\log g$ is good, with an insignificant systematic difference of -0.02 dex but a higher rms of ~ 0.18 dex.

The middle panels of Figure 6 compare our results with those from Everett et al. (2013) for a sample of 268 faint candidate exoplanet-hosting stars discovered by the *Kepler* mission. Everett et al. (2013) obtained low-resolution spectra ($R = 3000$) using the RCSpec long-slit spectrograph on the 4 m telescope at Kitt Peak Observatory and determined T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$ by fitting the observed spectra with synthetic ones computed from stellar model atmospheres (Castelli & Kurucz 2003). Effective temperatures from the Everett et al. (2013) results show a more significant mean offset of ~ 100 K relative to ours, with our T_{eff} scale being hotter, with a similar value for the rms scatter. It is also noticeable that there is a negative trend in the $\log g$ differences: for $\log g > \sim 4.3$ dex, our derived $\log g$ values are systematically larger

than theirs (with a mean difference of 0.14 ± 0.01 dex), while for $\log g$ smaller than ~ 4.3 dex, our results are systematically lower (with a mean difference of 0.05 ± 0.04 dex). The bottom panels of Figure 6 compare results for 343 KOIs in common that were observed with the near-infrared (λ 1.5–1.7 μm), high-resolution spectroscopic ($R \sim 22,500$) Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2017), which is a survey in SDSS-IV.

The stellar parameters shown are part of APOGEE Data Release 14 (DR14; Holtzman et al. 2018) and were derived automatically using the APOGEE Stellar Parameters and Chemical Abundances pipeline (ASPCAP; García Pérez et al. 2016), which fits the observed spectra to grids of synthetic spectra using χ^2 minimization. There is a significant systematic offset in effective temperatures, with our T_{eff} scale being hotter than that of APOGEE DR14 by 160 K in the mean, with an rms scatter of 126 K. There is also an offset in $\log g$ of -0.06 dex (rms = 0.17 dex), with a negative trend in the mean difference $\langle \text{APOGEE} - \text{this study} \rangle$ as a function of $\log g$, in particular

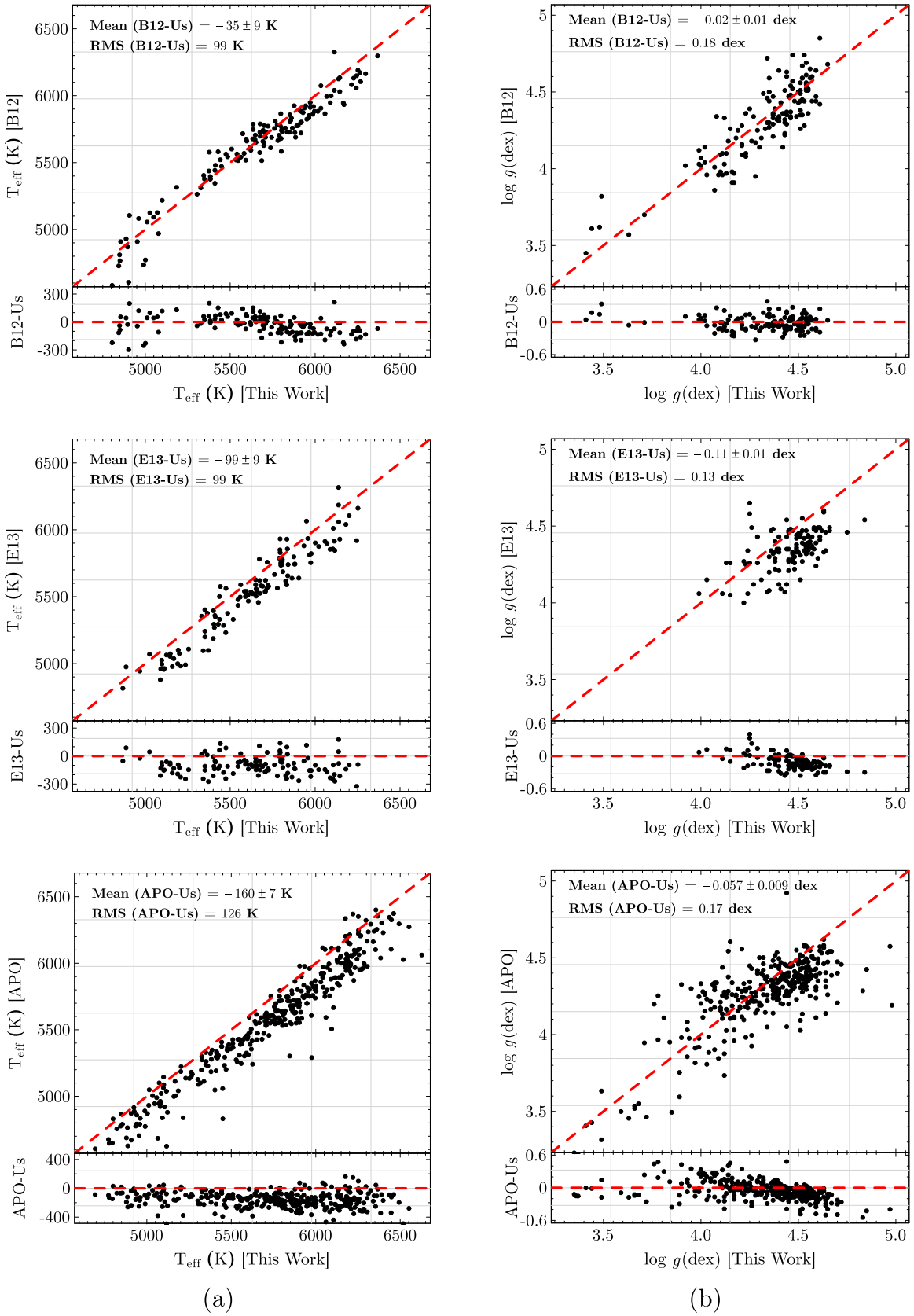


Figure 6. Comparison of the effective temperatures (column (a)) and $\log g$ values (column (b)) derived in this work and in Buchhave et al. (2012) for 135 stars in common (top panels), 119 stars in common with Everett et al. (2013) (middle panels), and 343 stars in common in the APOGEE DR14 (bottom panels). The mean differences between the parameters and the corresponding rms scatters are indicated in each case. The red dashed lines represent the equality.

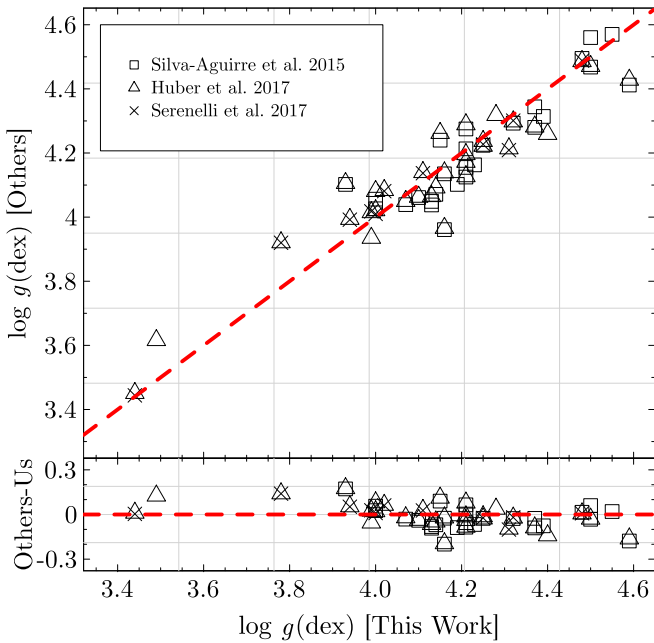


Figure 7. Our derived spectroscopic surface gravities in comparison with asteroseismic surface gravities from Silva Aguirre et al. (2015), Huber et al. (2017), and Serenelli et al. (2017) for 40 stars in common.

for $\log g$ values larger than ~ 4.0 . It is well known, however, that the surface gravities from ASPCAP DR14 have systematic offsets for red giants, as well as for dwarfs (Holtzman et al. 2018; Jönsson et al. 2018).

4.1.3. Asteroseismic versus Spectroscopic Surface Gravities

In this study, we use the Fe I and Fe II lines in order to derive surface gravities for the stars, with the $\log g$ derivation being done concomitantly with the determinations of effective temperatures, microturbulent velocities, and iron abundances (Section 3.1). Correlations that exist between these parameters can lead to systematic errors in the derived parameters that can be investigated. In fact, one of the stellar parameters that is typically not very well constrained via spectroscopy is the surface gravity. Asteroseismology, on the other hand, can provide accurate $\log g$ values (to 0.05 dex; Pinsonneault et al. 2018) that can serve as valuable benchmarks to investigate the possible systematic offsets in spectroscopic determinations of $\log g$ (Chaplin et al. 2014; Silva Aguirre et al. 2015; Huber et al. 2017; Lundkvist et al. 2018 and references therein).

Figure 7 shows the comparison between our derived $\log g$ values with those determined via asteroseismology. The asteroseismic data for 40 sample stars were collected from Silva Aguirre et al. (2015), Huber et al. (2017), and Serenelli et al. (2017). It should be noted that both T_{eff} and $[\text{Fe}/\text{H}]$ are essential to constrain the asteroseismic $\log g$. Huber et al. (2017) and Serenelli et al. (2017) used the APOGEE results for effective temperatures and metallicities in deriving $\log g$, while Silva Aguirre et al. (2015) used SME (Valenti & Piskunov 1996) and SPC (Buchhave et al. 2012), two spectral synthesis techniques, to fit optical high-resolution spectra to synthetic ones.

The seismic results presented in Figure 7 indicate good agreement between the $\log g$ determinations in the three different asteroseismic studies (Silva Aguirre et al. 2015, Huber et al. 2017, and Serenelli et al. 2017) for the few stars in

common. Our derived $\log g$ values (based on the Fe I and Fe II lines) also compare well with the asteroseismic ones, resulting in an insignificant mean offset in $\log g$ ($<$ asteroseismic $-$ this study $>$) of -0.01 dex and a reasonable rms of 0.08 dex. Such an agreement is completely within the expected uncertainties in spectroscopic $\log g$ determinations.

We note that the coverage in the range $\log g = 3.4\text{--}3.8$ is rather limited, and the comparison is based only on three stars, two of them with seismic $\log g$ values higher than the ones derived here by 0.13 dex. More stars with seismic $\log g$ values would be needed in order to reach a firmer conclusion; based on the current data, there is no indication of significant offsets in our spectroscopic $\log g$ determinations.

4.1.4. Stellar and Planetary Radii Comparisons for the CKS Sample

Stellar radii for the CKS sample were previously derived in Johnson et al. (2017) by converting the spectroscopic T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$ from Petigura et al. (2017) into stellar masses, radii, and ages using the Dartmouth Stellar Evolution Program models (Dotter et al. 2008) interpolated with the *isochrones* package (Morton 2015). Johnson et al. (2017) also used their derived stellar radii to determine the planetary radii. Fulton et al. (2017) used the planetary radii computed by Johnson et al. (2017) to make completeness corrections and calculate the resulting radius distribution. As *Gaia* DR2 became available, Fulton & Petigura (2018) computed stellar radii using distances from the *Gaia* DR2 inverted parallaxes and also derived the planetary radii for the CKS sample.

Stellar radii computed by taking into account the parallaxes are, in principle, more precise. Fulton & Petigura (2018) estimated that the errors in their stellar and planetary radii, using distances from the inversion of the *Gaia* parallaxes, are at the level of 2% and 5%, respectively. Although Fulton & Petigura (2018) obtained a scatter of 13% in the ratios of stellar radii when compared with those from Johnson et al. (2017), their distribution of planet sizes remained basically the same.

Figure 8 shows the comparison between our derived stellar radii with those from Fulton & Petigura (2018); there is an overall good agreement between the results for the vast majority of the targets. The mean stellar radii ratio between Fulton & Petigura (2018) and this study (“F&P18/Us”) is 0.9851 ± 0.0004 with a rms scatter of 0.013. A closer look at the results in Figure 8 indicates that the systematic differences are slightly larger for $R_{\star} > \sim 2.5 R_{\odot}$; if we compute the mean stellar radii ratio and the rms for those stars, we obtain 0.973 ± 0.002 and 0.010, respectively.

Figure 9 shows the comparison of the planetary radii. The results show good agreement, but there are some outliers for which Fulton & Petigura (2018) compute unrealistically large planetary radii when compared to ours ($R_{\text{pl}} > 23 R_{\oplus}$ in Fulton & Petigura 2018). In addition, Fulton & Petigura (2018) have a few planets with R_{pl} ranging roughly between 1000 and 24,000 R_{\oplus} (e.g., for K03891.01, $R_{\text{pl}} = 23,732 R_{\oplus}$; planets with $R_{\text{pl}} > 200 R_{\oplus}$ are off scale in Figure 9(a)). It is worth noting that the errors associated with these unrealistic radii in Fulton & Petigura (2018) are as large as the planetary radii themselves. We note that the discrepant planetary radii shown in the comparison of the results in Figure 9 are not due to differences in the stellar radii, as the comparison between our respective stellar radii values agrees quite well.

If we remove the outliers from the comparison or the planets with $R_{\text{pl}} > 23 R_{\oplus}$ from the Fulton & Petigura (2018) sample,

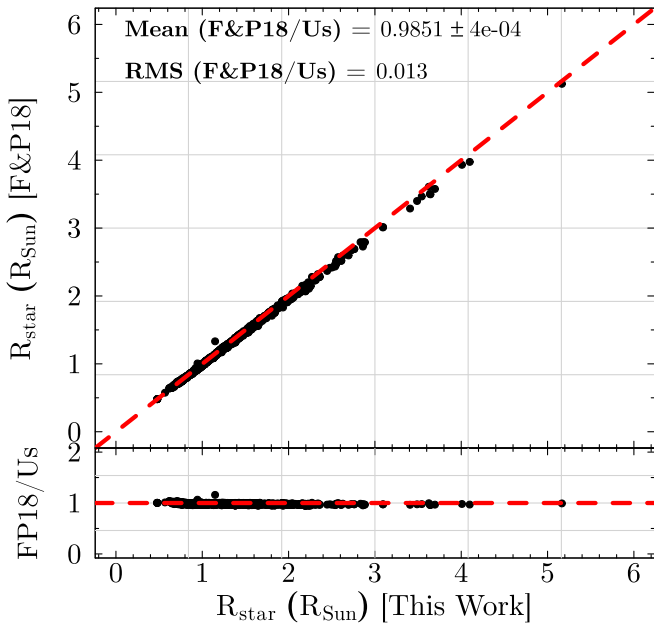


Figure 8. Comparison of the stellar radii derived from Fulton & Petigura (2018) with those derived in this work. Our stellar radii are just slightly larger than those of Fulton & Petigura (2018). The mean ratio between the radii in the studies and the corresponding rms scatter is shown. There is a tendency to find larger differences for $R_* > 2.5 R_\odot$.

we obtain 0.957 ± 0.007 for the mean ratio “F&P18/Us” and an rms of 3%. It is also found that the differences in planetary radii between this study and those from Fulton & Petigura (2018) increase somewhat for the smallest planets ($R_{pl} < 1 R_\oplus$, with a mean planetary ratio of 0.925), where Figure 9(b) shows the comparison of our planetary radii $\leq 1 R_\oplus$ with those derived by Fulton & Petigura (2018) for planets in common.

Figure 10 shows comparisons of the stellar radii derived in this study with results from asteroseismology. Such comparisons are particularly valuable because asteroseismology can deliver precise stellar radii, providing the basis for an assessment of both the accuracy and precision of our derived stellar radii. The three panels in Figure 10 show results for samples of stars in common with three asteroseismic studies (these same studies were used for the log g comparisons in Figure 7): Serenelli et al. (2017, panel (a)), Silva Aguirre et al. (2015, panel (b)), and Huber et al. (2017, panel (c)). The mean values for the ratios “other studies/this study,” along with the corresponding rms, are presented in each panel of Figure 10; it is clear that our stellar radii compare very well with those from asteroseismology.

It should be kept in mind, however, that the selected asteroseismic studies use different methodologies (scaling relations versus modeling of individual frequencies), as well as different sets of stellar parameters (effective temperatures and metallicities), which are needed for their determinations of stellar radii. The internal precisions estimated for their derived radii are 2.7% for Serenelli et al. (2017), 1.2% for Silva Aguirre et al. (2015), and 2.7% for Huber et al. (2017). The internal precision inferred from our determinations of stellar radii (Table 3) is also at a similar level: $\sim 2.8\%$. Concerning systematic offsets, we find that our derived radii are just slightly higher than the asteroseismic ones, being on average 4.1% larger than those in Serenelli et al. (2017), 1.7% larger than those in Silva Aguirre et al. (2015), and 0.8% larger than

those in Huber et al. (2017). The rms values obtained for the three studies are also quite small, indicating a small scatter of 0.02–0.03. All in all, we can conclude that our derived stellar radii from precise parallaxes and spectroscopic determinations of the effective temperatures achieve a comparable precision against stellar radii derived from asteroseismology and do not contain systematic offsets that are much larger than the typical offsets seen between the different asteroseismic studies themselves.

5. Discussion

5.1. Exoplanetary Radii, Orbital Periods, and Incident Fluxes

5.1.1. The Small-planet Radius Gap

Until recently, the detailed properties for the unprecedented variety of planets discovered by the *Kepler* mission (Borucki et al. 2010; Koch et al. 2010; Borucki 2016) have been ambiguous due in part to uncertainties in the planetary radii that stem from uncertainties in the stellar radii, which have been for the most part estimated from broadband photometry.

Based on improved measurements for the stellar radii for the CKS sample by Johnson et al. (2017); using stellar and planetary parameters from Petigura et al. 2017 and Johnson et al. 2017), Fulton et al. (2017) found a bimodal distribution for the small-planet radii having peaks at ~ 1.3 and $\sim 2.4 R_\oplus$, with a gap between 1.5 and $2.0 R_\oplus$. Van Eylen et al. (2018), using asteroseismic radii for a small subsample of CKS targets with very precise radii, also detected a bimodal distribution with a clear gap around $2 R_\oplus$. A similar gap in planetary radius was confirmed by Fulton & Petigura (2018) using *Gaia* DR2 parallaxes. The dearth of planets at $R_{pl} \sim 1.8 R_\oplus$ has been predicted by theoretical models and is interpreted as a transition radius separating planets with masses large enough to retain their gas envelope and those that have lost their atmospheres and consist of their remnant cores (Owen & Wu 2013; Lopez & Fortney 2014; Chen & Rogers 2016; Lopez & Rice 2016; Owen & Wu 2017; Ginzburg et al. 2018).

The distribution of planetary radii derived in this study is shown in the different histograms in Figure 11. Although the methodology used in the determination of stellar parameters (needed to derive the stellar and planetary radii) for the CKS sample is completely independent of previous studies in the literature analyzing the same data set, the planetary radii distributions (presented in all panels of Figure 11) are found to be bimodal, showing the clear presence of a valley in the small-sized planet radius distribution. These independent results give support to the fact that the lack of planets around $2.0 R_\oplus$ is not an analysis artifact but rather represents a real transition between rocky planets and those with extensive atmospheres (Weiss & Marcy 2014; Rogers 2015).

The full sample analyzed in this study (composed of 1633 planets), without any cuts, is shown in panel (a) of Figure 11. It should be kept in mind, however, that the full sample does not include KOIs with planets deemed “false positives” (see Section 2). The uncorrected-completeness radius distribution shows two clear peaks: one at $\sim 1.6 R_\oplus$ and another at $\sim 2.8 R_\oplus$. The radius gap is seen roughly between 1.8 and $2.2 R_\oplus$.

It is of interest to investigate the positions of the peaks and the planetary radius gap using only precise planetary radii, as it is clear that the derived radii have different levels of precision depending, for example, on the errors in the *Gaia* DR2 parallaxes, which are folded into the distance error estimates by

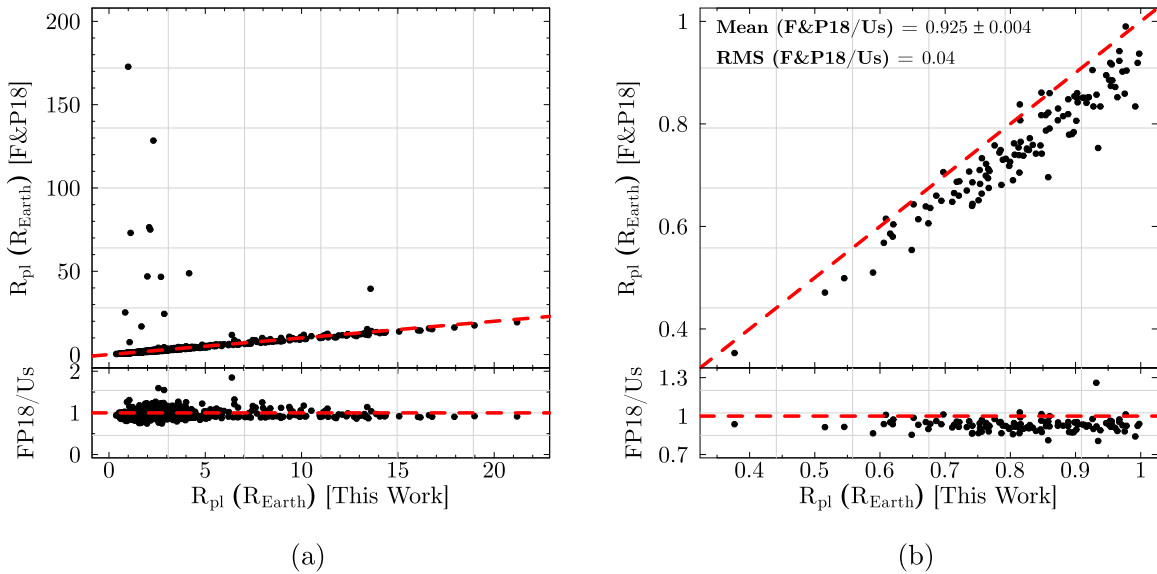


Figure 9. Panel (a) presents the comparison of our planetary radii with those derived by Fulton & Petigura (2018) for planets in common having planetary radii less than $200 R_{\oplus}$ in Fulton & Petigura (2018). Panel (b) shows the same comparison shown in panel (a) but for planetary radii $\leq 1 R_{\oplus}$. The offset between the results is clear. The mean ratio between the radii in the studies and the corresponding rms scatter is also shown. The red dashed lines represent the equality line.

Bailer-Jones et al. (2018). In panel (b), we show a similar histogram as in panel (a), but in this case, we restricted our sample to consider only those planets with uncertainties in the derived radii of less than 8% (corresponding to $\sim 2\times$ the median uncertainty; Section 3.3). In panel (c), we applied the same selection as in panel (b), but, similarly to Fulton et al. (2017), we removed from the sample any planets with $P > 100$ days and in a transit configuration corresponding to the impact parameter, b , being larger than 0.7; this is our “clean” sample, which is composed of 965 planets.

The radii histograms shown in Figure 11 panels (a) (for the entire sample), (b) (for planets with precise radii), and (c) (for the “clean” sample) are in general quite similar: in all distributions, there is a dearth of planets at $\sim 2.0 R_{\oplus}$, and the location of the two peaks in the distributions is similar as well. The full sample has a small excess of planets at $R_{\text{pl}} \sim 10 R_{\oplus}$, and this is mostly due to the presence of planets with $P > 100$ days that were removed in the “clean” sample. There is also a large population of small planets with $R_{\text{pl}} < 1.0 R_{\oplus}$, but, as pointed out by Berger et al. (2018), it is expected that some of these small planets will be classified as “false positives” in the future.

5.1.2. Completeness Corrections

To assure that the trend described in the previous section is not an artifact of completeness and affected by the lack of detectability of planets with small radii and/or long orbital periods by *Kepler*, we reconstruct the planet occurrence rate of the *Kepler* sample after applying completeness corrections of Christiansen et al. (2016, 2015), Fulton et al. (2017), Fulton & Petigura (2018), and Mulders et al. (2018, 2016).

We used the injection-recovery experiments described in Christiansen et al. (2015, 2016). They measured the *Kepler* pipeline detection efficiency by injecting simulated transiting planets into the raw pixel data and analyzed the recovery rate after processing them with the *Kepler* pipeline to reconstruct the planet occurrence rate of the *Kepler* sample.

Considering injections in the transit signals for the stars included in our sample, we used our derived stellar and planetary radii to calculate the reliability of the transit measure (m) following the procedure described in Fulton et al. (2017),

$$m = \left(\frac{R_{\text{pl}}}{R_{\star}} \right)^2 \frac{1}{\text{CDPP}_{\text{dur}}} \sqrt{\frac{t_{\text{obs}}}{P}},$$

where t_{obs} is the time that a star of radius R_{\star} , harboring a planet of radius R_{pl} and orbital period P , was observed, while the combined differential photometric precision (CDPP_{dur} ; Koch et al. 2010) is the noise of a transit signal interpolated to the transit duration time. The transit parameters necessary to do the calculations were taken from the *Kepler* database (DR25; Thompson et al. 2018).

To account for the pipeline efficiency, Fulton et al. (2017) fit a Γ cumulative distribution function of the form

$$C(m; k, \theta, l) = \int_0^{\frac{m-l}{\theta}} x^{k-1} e^{-x} \cdot dx$$

to model the distribution of values from the injection-recovery transit signals (m); we use the k , l , and θ values as determined by Fulton et al. (2017).

We determined the detection probability, p_{det} , obtained by using $C(m)$ values, and the geometric transit probability, p_{tr} , to check for the survey sensitivity.

The transit probability, p_{tr} , defined as the geometric probability that a planet with radius R_{pl} transiting a host star of radius R_{\star} at a distance of a could be detected, is $p_{\text{tr}} = 0.7 R_{\star}/a$, where the factor 0.7 corresponds to the imposed limit in the impact parameter of the planet candidates, according to the work of Fulton et al.

To compensate for incompleteness due to a lack of detection efficiency, p_{det} , or a low probability of transit detection, p_{tr} , we weighted each planet by the inverse of these probabilities:

$$w_i = \frac{1}{p_{\text{det}} \times p_{\text{tr}}}.$$

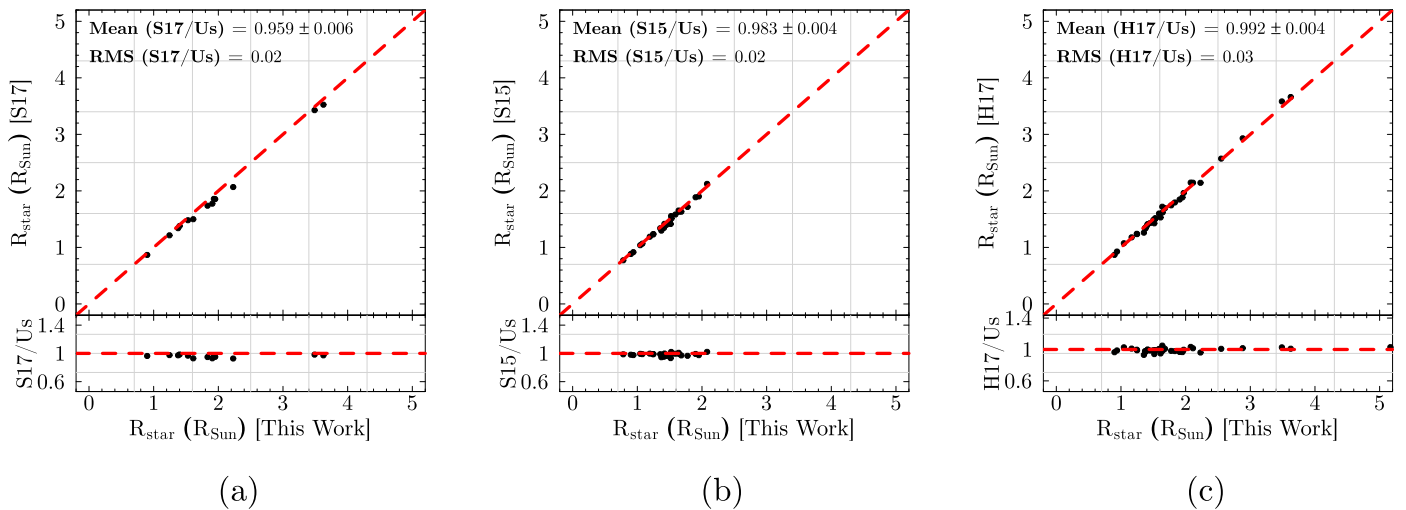


Figure 10. Comparison of the derived stellar radii in this work against the precise stellar radii from asteroseismology from (a) Serenelli et al. (2017), (b) Silva Aguirre et al. (2015), and (c) Huber et al. (2017).

The true measure of the occurrence rate, f_{bin} , the number of planets per star in any orbital period or radius bin, is then given by

$$f_{\text{bin}} = \frac{1}{N_{\star}} \sum_{i=1}^{n_{\text{pl,bin}}} w_i.$$

Figure 11(d) presents the same distribution of planetary radii for our “clean sample” (shown in panel (c)) in comparison with the completeness-corrected distribution, shown as the dashed line. It is noticeable that the location of the gap and the peaks in the completeness-corrected distribution are shifted to slightly smaller radii when compared with the uncorrected distribution, similar to what has been found by Fulton & Petigura (2018).

The location of the peaks of the completeness-corrected planetary radius distribution can be estimated from the kernel density estimate for a Gaussian distribution (shown in Figure 11(e)); the peak positions of the distribution are found at 1.47 ± 0.05 and $2.72 \pm 0.10 R_{\oplus}$, and there is a radius gap at $1.89 \pm 0.07 R_{\oplus}$ (Figure 11(e)). The completeness-corrected distribution from Fulton & Petigura (2018) is also shown for comparison as the gray line in Figure 11(e). There is a marginal shift in our planetary distribution relative to Fulton & Petigura (2018). Van Eylen et al. (2018) also investigated the planetary radii distributions in their, albeit smaller, sample but with precise radii from asteroseismic parameters; they find the peaks to be at 1.5 and $2.5 R_{\oplus}$, respectively, with the radius gap minimum falling at $2.0 R_{\oplus}$.

5.1.3. A Slope in the Planetary Radius Gap

As discussed in Section 5.1.1, one mechanism suggested to explain the bimodal small-planet radius distribution is photoevaporation: X-ray and UV fluxes from the young planet-hosting star evaporate the envelopes of the H–He rich sub-Neptunes, exposing their stripped rocky cores. In addition to photoevaporation, Ginzburg et al. (2018) modeled the radius valley as being caused by the energy from young, hot planetary cores driving planetary-mass-dependent atmospheric mass loss (or “core-powered mass loss”). Several theoretical models also predict the shape and slope of the evaporation valley

(e.g., Owen & Wu 2013, 2017; Jin & Mordasini 2018); the planetary radii for the CKS derived in this study are precise enough to investigate the presence of such signatures in the planetary radius as a function of orbital period or incident flux.

Figures 12 and 13 present the derived exoplanetary radii as a function of their orbital periods (with *Kepler* DR25 periods taken from Thompson et al. 2018) and insolation fluxes, respectively. The stellar fluxes were calculated from the derived T_{eff} and R_{\star} , while the semimajor axes of the planetary orbits, assuming an orbital eccentricity equal to zero, were used to compute the incident flux at the planet.

In both figures, panel (a) shows the full sample with 1633 planets, without any cuts, and panel (b) shows a subset of the “clean” sample for the region containing the smaller exoplanets (radii less than $\sim 4 R_{\oplus}$). Visual inspection of these figures clearly indicates the presence of two clouds representing the population density of the distributions of super-Earths and sub-Neptunes in the log radius – log period/incident flux planes, with a low density of exoplanets at radii between the two clouds, with $R_{\text{pl}} \sim 2 R_{\oplus}$. This “evaporation valley” corresponds to the separation between the super-Earth and sub-Neptune regimes.

The exoplanet population plotted in Figures 12 and 13 also shows an overall lack of sub-Neptunes with short orbital periods ($P < \sim 3$ days) and incident fluxes (relative to Earth) $> \sim 700$, which is likely related to the photoevaporation of the atmospheres of sub-Neptune-sized exoplanets that are very close to their parent stars and suffer high stellar incident flux levels (Ikoma & Hori 2012; Lopez et al. 2012; Ciardi et al. 2013; Owen & Wu 2013; Wu & Lithwick 2013). It should be pointed out, however, that some exoplanets have been detected in the sub-Neptune “desert” region (see West et al. 2018).

Close inspection of Figure 12 gives some hints that the value of R_{pl} in the evaporation valley minimum overall decreases with increasing orbital period. The change in the radius gap minimum as a function of orbital period can be quantified in a simple way by dividing our “clean” sample of planets (with radii errors $\leq 8\%$, $P \leq 100$ days, and $b \leq 0.7$) into 10 orbital period bins containing an equal number of planets. Within each orbital period bin, the minimum value in the radius gap was measured, resulting in a linear relation defined by $\log(R_{\text{pl}}) = (-0.11 \pm 0.02) \log(P) +$

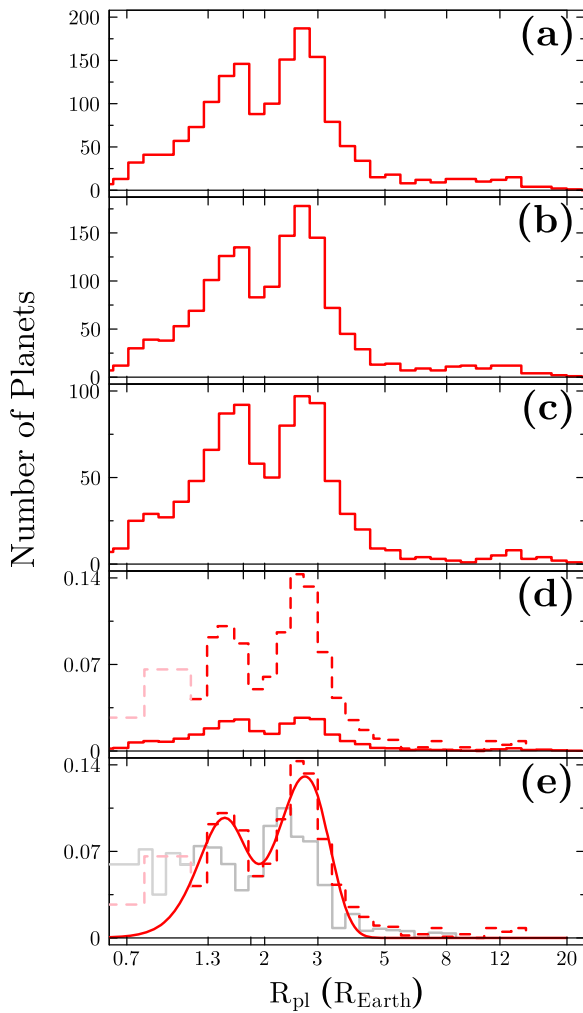


Figure 11. Radius distribution. (a) Considering the entire planet sample (1633 planets). (b) Taking into account only those planets (1526 planets) having radii with errors less than 8%. (c) Same as panel (b) but also discarding planets with $P > 100$ days and planetary systems with $b > 0.7$ (this defines the “clean” sample having 965 planets). (d) Completeness-corrected histogram of planetary radii for our “clean” sample (dashed line) in comparison with the distribution of planet radii uncorrected for completeness (solid line). (e) Completeness-corrected planetary radii for our “clean” sample (red line) but compared with the one from Fulton & Petigura (2018; gray line). We confirm the presence of a gap in the occurrence distribution of close-in planets with orbital periods less than 100 days at $\sim 1.9 R_{\oplus}$.

(0.39 ± 0.01) , with R_{pl} in units of Earth radii and P in days, or R_{pl} scaling as $P^{-0.11}$; the slope, along with the corresponding prediction interval, is shown in Figure 12(b)). A similar fit was done for the valley observed between the planetary radii and the incident fluxes (Figure 13(b)), finding a linear relation defined as $\log(R_{\text{pl}}) = (+0.12 \pm 0.02) \log(F) + (0.04 \pm 0.03)$, with F in units of incident flux at the Earth, or R_{pl} scaling as $F^{+0.12}$.

Van Eylen et al. (2018) also detected a slope in the planetary radius valley as a function of the planetary orbital period from a small sample of 75 stars, well characterized by asteroseismology and having precise radii for their 117 associated planets. Their derived slope of $P^{-0.09^{+0.02}_{-0.04}}$ for the small-planet radius gap, although from a much smaller sample, is in good agreement with ours within the uncertainties. They also model the valley slope after restricting their sample to periods < 25 days and find $P^{-0.10}$; if we follow the same cut for the much larger CKS sample, we obtain a similar slope of $P^{-0.11 \pm 0.03}$.

The shape and, in particular, the value of the slope in the evaporation valley can constrain planetary formation models. Lopez & Rice (2016) used different models to point out that the transition radius between rocky super-Earths and sub-Neptunes with volatile envelopes scales differently with the orbital period depending on the planet formation scenario; the transition radius should decrease for longer orbital periods in the case of a photoevaporation scenario (R_{pl} scaling as $P^{-0.15}$). Contrarily, it should increase if the primordial rocky planets formed after the protoplanetary disks dissipated in a gas-poor formation model with a positive slope for the radius gap (R_{pl} scaling as $P^{+0.07 \pm 0.10}$).

According to models by Owen & Wu (2017), the slope of the transition radius (or the upper envelope of the super-Earth radius) with period derived through evaporation models can change when considering different evaporation efficiencies in these models, ranging from $P^{-0.25}$, when a constant evaporation efficiency is considered for all of the planets, to $P^{-0.16}$, when evaporation efficiency depends on the planet density. Owen & Wu (2017), as well as Jin & Mordasini (2018), also investigate how the bulk composition affects the planetary radii of super-Earths and sub-Neptunes as a function of period (or orbital semimajor axis). Depending on the mixture of iron, silicates, or ices, for example, the maximum radius of super-Earths changes as a function of orbital period. Using Figure 10 from Owen & Wu (2017) as an example, their model, which mimics the composition of the Earth (with \sim one-third iron) and considers a variable efficiency for the evaporation, provides the closest match to the upper envelope defining super-Earth radii versus period for periods less than ~ 8 days. The model then declines more steeply than the observed distribution toward increasing periods, with the result that the observed maximum radius for super-Earths falls well above the Owen & Wu (2017) model core with one-third iron. Since their “icy” models with one-third ice and two-thirds silicates have larger radii than the iron cores, the observed super-Earths with larger radii at longer periods may signal a shift in the overall compositions of super-Earths with short periods compared to those with long periods.

The location of the radius gap in Figure 12 and its negative slope agree with what is expected from photoevaporation models, as well as core-powered mass-loss models, which predict a slope of -0.11 (Gupta & Schlichting 2018). The change in the maximum radius versus orbital period observed for super-Earths in Figure 12 agrees qualitatively with model cores from Owen & Wu (2017); a quantitative comparison with model predictions can constrain the core compositions and may even be able to map compositions at different orbital periods around different types of host stars.

The shape and location of the planet radius versus orbital period valley is also in qualitative agreement with the core-powered mass-loss mechanism from Ginzburg et al. (2018). In this model, atmospheric mass loss is more effective if the equilibrium temperature of the planet, T_{eq} , is high, which is, in general, the case for more closely orbiting planets, or more massive main-sequence stars, which have larger luminosities. Ginzburg et al. (2018) evolved a model planetary distribution and found that the position of the valley minimum shifts from $\sim 2 R_{\oplus}$ for $P < 10$ days to $\sim 1.5 R_{\oplus}$ for longer orbital periods. Recently, Fulton & Petigura (2018) established a relation between the cumulative distribution of planets versus the incident flux as a function of stellar mass, with the distribution shifted to higher incident fluxes for larger stellar masses.

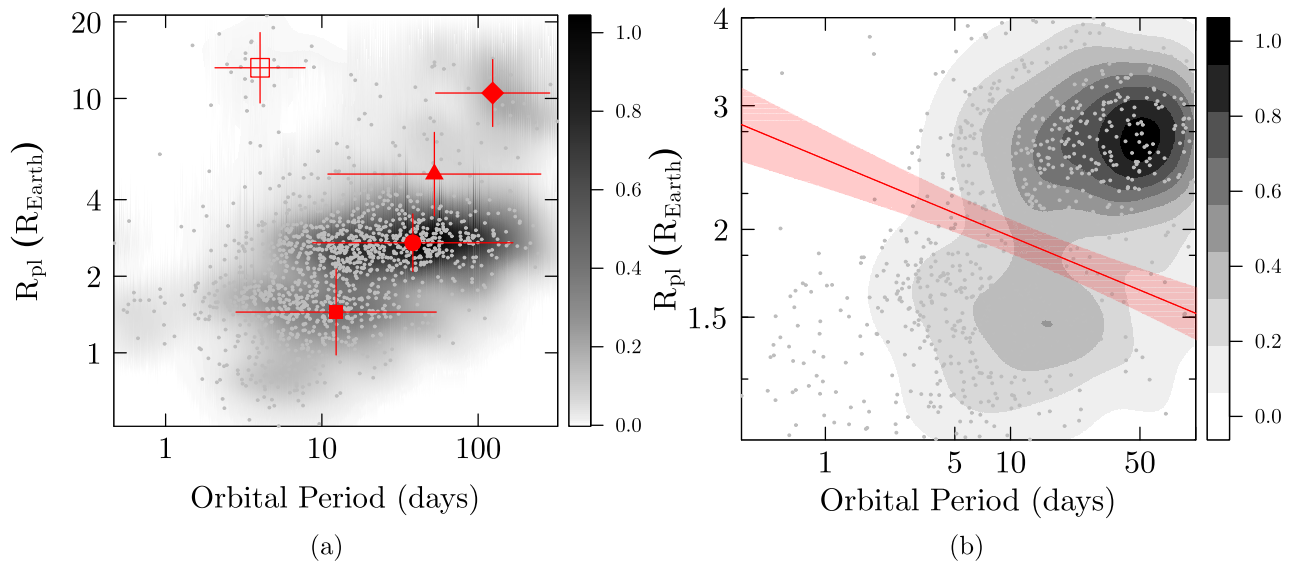


Figure 12. (a) Planetary radius as a function of planetary period for the entire sample. Red symbols and bars indicate the weighted median values and their weighted uncertainties, respectively, for planetary radii and orbital periods for each planet size domain: super-Earths (filled square), sub-Neptunes (circle), sub-Saturns (triangle), hot Jupiters (open square), and warm Jupiters (diamond). (b) Same as panel (a) but only considering the small-sized planet regime (radii less than $\sim 4 R_{\oplus}$) of our “clean” planet sample. The best-fit slope (red line) and prediction interval (shaded region) to the radius gap is shown.

Figure 13, where planetary radii are plotted versus incident flux, tells effectively the same story as Figure 12. The incident flux shown is the current flux, although within the photoevaporation model, Owen & Wu (2013, 2017) pointed out that it is the flux from the young, presumably active host star that is most important in sculpting the distributions of planetary radii in Figures 12 and 13. Photoevaporation is most effectively driven by the stellar integrated X-ray and extreme UV (EUV) fluxes during the first 100 Myr of the life of the star (Owen & Wu 2013). The planetary composition also plays a role, as H and He are most affected by the EUV flux, while the metals are more easily evaporated by the X-ray flux (Owen & Wu 2017). The planetary radii distributions as functions of orbital period and incident flux shown in Figures 12 and 13 are thus the result of their early X-ray and EUV radiation environments and distance from their young host star.

5.1.4. A Possible Correlation between Planetary Radii and Orbital Periods?

The CKS planetary sample studied here can be divided into four exoplanet size regimes as follows.

1. Jupiters with $8 R_{\oplus} < R_{\text{pl}} \leq 20 R_{\oplus}$.
2. Sub-Saturns with $4 R_{\oplus} < R_{\text{pl}} \leq 8 R_{\oplus}$.
3. Sub-Neptunes with $2 R_{\oplus} < R_{\text{pl}} \leq 4 R_{\oplus}$.
4. Super-Earths with $R_{\text{pl}} \leq 2 R_{\oplus}$.

The sample (in this case, we are considering the full sample of planets without any cuts) has a majority of exoplanets with small sizes, split roughly in equal numbers between the classes of super-Earths (736 exoplanets) and sub-Neptunes (706 exoplanets). It has a much smaller number of sub-Saturns (96 exoplanets) and 93 exoplanetary systems containing at least one Jupiter-like planet. About 20% of these systems have a single hot Jupiter without any detectable inner or outer companions, while almost half of the warm Jupiter sample (~ 30 out of ~ 70) is found around a variety of exoplanetary system architectures. We define hot Jupiters as those with

periods less than 10 days, while warm Jupiters have periods larger than 10 days.

Figure 12(a) also shows, as red symbols, the weighted median values and rms scatter of the planetary radii and orbital period distributions for each planet size domain: super-Earths (filled square), sub-Neptunes (circle), sub-Saturns (triangle), hot Jupiters (open square), and warm Jupiters (diamond). It is clear that the super-Earths have a median value for planetary radii and orbital periods that is lower than that of the sub-Neptunes, suggesting a possible correlation in the sense that the sizes of the exoplanets generally increase with orbital periods; some of this correlation is due to incompleteness in transit measurements given, for example, the difficulty in detecting small planets at larger distances from the parent star (with larger orbital periods). The correlation between planetary radii and orbital periods extends, however, toward the larger exoplanet groups containing the sub-Saturns and warm Jupiters. For these larger planets, the observational biases should not be significant, in particular for systems with orbital periods less than ~ 500 days (taking completeness values from Silburt et al. 2015).

The hot Jupiters do not fit into the trend of increasing orbital period with increasing exoplanet size. The median value of $R_{\text{pl}}-P$ for the warm Jupiters (diamond) generally follows the trend delineated by the smaller planets, while the hot Jupiters occupy a different locus in the R_{pl} -orbital period plane (median represented by the open square), having much shorter orbital periods, on average, and being much closer to their parent stars. Hot Jupiters are believed to have formed several astronomical units away from their parent stars and undergone extreme migration into the exoplanetary system’s inner region, destabilizing any small exoplanets, scattering them out of the system, or destroying them (Latham et al. 2011; Morbidelli 2014; Mustill et al. 2015). We also note that Huang et al. (2016) analyzed a sample of 45 hot Jupiters and 27 warm Jupiters from the *Kepler* catalog and proposed that warm Jupiters with low-mass companions would be formed in situ, not affecting their small neighbors, and that warm

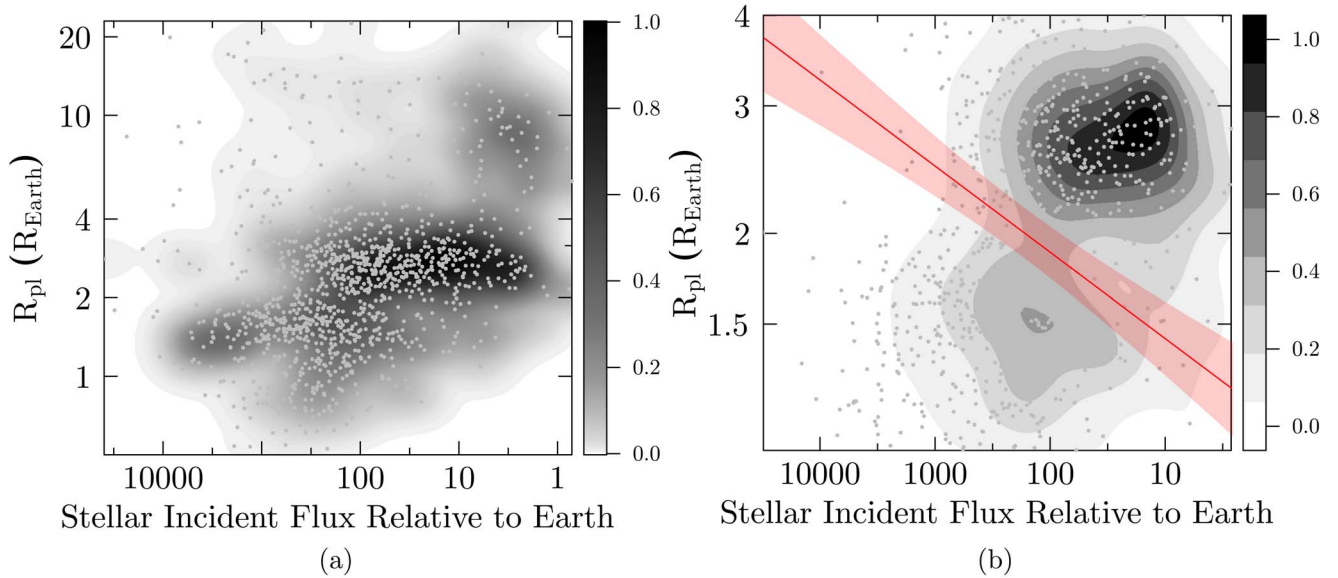


Figure 13. Panel (a): planetary radius as a function of stellar incident flux (relative to Earth) for the entire planet sample. Panel (b): same as panel (a) but only considering the small-sized planet regime (radii less than $\sim 4 R_{\oplus}$) of our “clean” planet sample. The best-fit slope (red line) and prediction interval (shaded region) to the photoevaporation valley are also shown.

Jupiters with no detectable companions may be a distinct population.

With the position of the hot Jupiters in the $R_{\text{pl}}-P$ plane dominated by migration, the other size groups (super-Earths, sub-Neptunes, sub-Saturns, and warm Jupiters) may define a seemingly tight correlation of median radius with median orbital period. This correlation is influenced by the dearth of sub-Neptunes with short orbital periods.

To evaluate how much the correlation depends on the limits adopted for cutting the planet sample in terms of radii and orbital periods, we analyze three different samples that will correspond to different levels of completeness. (1) We consider the entire sample of warm Jupiters, sub-Saturns, sub-Neptunes, and super-Earths at any orbital period, without any considerations about observational biases. A fit to the trend in log–log space results in a well-defined power-law relation, such that $R_{\text{pl}} \sim P^{0.84 \pm 0.11}$, with R_{pl} in Earth radii and orbital period in days. (2) If we limit the sample to include only those planets with $R_{\text{pl}} \geq 2 R_{\oplus}$ (equivalent to assuming that roughly all sub-Neptunes and larger planets can be detected at any orbital period), we obtain a steeper power law but with smaller errors in the fit: $R_{\text{pl}} \sim P^{1.1 \pm 0.3}$. (3) If we now add a cut in orbital period, considering only those planets with $R_{\text{pl}} \geq 2 R_{\oplus}$ and $P < 500$ days, we obtain the same power law as before but with an increase in the uncertainties in the fit ($R_{\text{pl}} \sim P^{1.09 \pm 0.13}$). We expect that the *Kepler* completion levels would be significant for this regime and verify from these tests that the correlation of planetary radii with orbital period does not change significantly in the three samples analyzed.

Helled et al. (2016) also suggested the existence of a correlation between planetary radius and orbital period for exoplanets with radii smaller than $4 R_{\oplus}$ using *Kepler* data. They did a statistical analysis that took into consideration completeness values for the detection of planets by *Kepler* (Silburt et al. 2015) to conclude that this correlation was not the result of a selection bias. Helled et al. (2016) obtained a power-law relation between planet radius and orbital period of $R_{\text{pl}} \sim P^{0.5-0.6}$, which is similar, within the uncertainties, to

the value of ~ 0.8 obtained here. If true, the correlation between radii and orbital periods found for the smaller planets in the CKS sample may imply that larger planets would also more likely form at larger distances from the host star.

6. Conclusions

1. We have conducted a homogeneous, quantitative spectroscopic analysis of 1232 exoplanet host stars using the high-resolution Keck/HIRES spectra made publicly available by the CKS team (Johnson et al. 2017; Petigura et al. 2017). Stellar parameters (T_{eff} and $\log g$) were derived from an EW measurement analysis of a sample of 158 Fe I and 18 Fe II lines and using the automated pipeline described in Ghezzi et al. (2010, 2018).
2. *Gaia* DR2 parallaxes (Gaia Collaboration et al. 2018) for the CKS stars were used to determine precise distances in Bailer-Jones et al. (2018), which were then used as the foundation for determining stellar luminosities in this study. With tightly constrained effective temperatures and luminosities, stellar radii were computed for the entire sample, with a median internal uncertainty of 2.8%.
3. Our derived stellar radii from precise parallaxes and spectroscopic determinations of the effective temperatures achieve a comparable precision against stellar radii obtained from asteroseismology, with no significant systematic offsets. Precise stellar radii are important to constrain the planetary radii—a crucial parameter necessary to unveil planetary composition.
4. Considering the sample of those stars with R_{\star} uncertainties less than 10% and planetary transit depths from Thompson et al. (2018), we derive planetary radii for 1633 planets, with a median uncertainty of 3.7%. Comparisons of our derived planetary radii with those from Fulton & Petigura (2018) indicate that our planetary radii are systematically larger than theirs by $\sim 4.3\%$.
5. The derived planetary radii clearly exhibit two peaks in the completeness-corrected planetary radii distributions. In particular, for our “clean” planet sample (planets with

radii errors $< 8\%$, $P < 100$ days, and $b < 0.7$), we obtain peaks in the radius distributions corresponding to $1.47 R_{\oplus}$ (super-Earths) and $2.72 R_{\oplus}$ (sub-Neptunes), with a clean minimum (the “gap”) at $1.9 R_{\oplus}$.

6. Given the good internal precision in the derived radii, it was possible to evaluate not only the location but also the shape of the radius gap. Our results indicate that the radius gap for the CKS sample does not fall at a constant value of radius but rather changes as functions of both planetary orbital period and incident stellar flux. The position of the radius gap decreases with orbital period, and this decrease can be fit by a power law of the form $R_{\text{pl}} \propto P^{-0.11}$; this agrees well with the recent value of -0.09 from Van Eylen et al. (2018) for a much-reduced sample of 117 planets having precise radii from asteroseismology.
7. According to Owen & Wu (2017), the value of the slope in the evaporation valley can constrain the planet core compositions at different orbital periods around different types of host stars. In our case, $R_{\text{pl}} \propto P^{-0.11}$ matches a terrestrial-like composition model for the planets in the transition radius. In addition, core-powered mass-loss models also predict a slope of -0.11 (Gupta & Schlichting 2018).
8. The value of the radius gap increases with increasing incident stellar flux, such that $R_{\text{pl}} \propto F^{+0.12}$ provides an excellent fit to our results. The trend of radius gap position with incident flux (and orbital period) agrees with models of photoevaporation (e.g., Owen & Wu 2013, 2017; Jin & Mordasini 2018).
9. If we divide our planetary sample into warm Jupiters (Jupiters with orbital periods larger than 10 days), sub-Saturns, sub-Neptunes, and super-Earths, we find that the mean values for planetary radii and orbital periods seem to suggest a possible correlation: larger planets seem to form more distantly from their parent stars. Considering only those planets with radii $\geq 2 R_{\oplus}$ and orbital periods < 500 days (for which *Kepler* completeness levels should be high), we obtain $R_{\text{pl}} \propto P^{0.8}$; this slope is similar to what has been found previously for small planets by Helled et al. (2016).

We thank the referee for giving suggestions that improved the paper. We warmly thank the California-*Kepler* Survey team for making their reduced data publicly available. We thank the European Space Agency (ESA) mission *Gaia* and NASA Exoplanet Archive for the data used in this work. C.M. acknowledges the financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). C.M. also thanks D.F. Morell for his kindly assistance in the R language. K.C. and V.S. acknowledge that their work is supported in part by the National Aeronautics and Space Administration under grant 16-XRP16 2-0004, issued through the Astrophysics Division of the Science Mission Directorate. L.G. is thankful for the financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Software: Python packages: *astropy* (Astropy Collaboration et al. 2013), *isoclassify* (Huber et al. 2017); and R packages: *astro* (Lee 2014), *mclust* (Scrucca et al. 2016), *MASS* (Venables & Ripley 2002).

ORCID iDs

Cintia F. Martinez  <https://orcid.org/0000-0003-2745-8241>
Luan Ghezzi  <https://orcid.org/0000-0002-0464-2655>

References

- Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, **47**, 481
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **588**, A33
Bailer-Jones, C. A. L., Rybizki, J., Fousneau, M., Mantelet, G., & Andrae, R. 2018, *AJ*, **156**, 58
Berger, T. A., Huber, D., Gaidos, E., & van Saders, J. L. 2018, *ApJ*, **866**, 99
Bilir, S., Ak, S., Karaali, S., et al. 2008, *MNRAS*, **384**, 1178
Borucki, W. J. 2016, *RPPH*, **79**, 036901
Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Sci*, **327**, 977
Brewer, J. M., & Fischer, D. A. 2018, *ApJS*, **237**, 38
Brewer, J. M., Fischer, D. A., Valenti, J. A., & Piskunov, N. 2016, *ApJS*, **225**, 32
Buchhave, L. A., Latham, D. W., Johansen, A., et al. 2012, *Natur*, **486**, 375
Castelli, F., & Kurucz, R. L. 2003, in IAU Symp. 210, Modelling of Stellar Atmospheres, ed. N. Piskunov, W. W. Weiss, & D. F. Gray (San Francisco, CA: ASP), A20
Castelli, F., & Kurucz, R. L. 2004, *A&A*, **419**, 725
Chaplin, W. J., Basu, S., Huber, D., et al. 2014, *ApJS*, **210**, 1
Chen, H., & Rogers, L. A. 2016, *ApJ*, **831**, 180
Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, **823**, 102
Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2015, *ApJ*, **810**, 95
Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2016, *ApJ*, **828**, 99
Ciardi, D. R., Fabrycky, D. C., Ford, E. B., et al. 2013, *ApJ*, **763**, 41
Coelho, P., Barbuy, B., Meléndez, J., et al. 2005, *A&A*, **443**, 735
Dorn, C., Noack, L., & Rozel, A. B. 2018, *A&A*, **614**, A18
Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, *ApJS*, **178**, 89
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, *AJ*, **156**, 264
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, *A&A*, **616**, A1
García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, *AJ*, **151**, 144
Ghezzi, L., Cunha, K., Schuler, S. C., & Smith, V. V. 2010, *ApJ*, **725**, 721
Ghezzi, L., Montet, B. T., & Johnson, J. A. 2018, *ApJ*, **860**, 109
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
Gupta, A., & Schlichting, H. E. 2018, arXiv:1811.03202
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, *A&A*, **486**, 951
Helled, R., Lozovsky, M., & Zucker, S. 2016, *MNRAS*, **455**, L96
Holtzman, J. A., Hasselquist, S., Shetrone, M., et al. 2018, *AJ*, **156**, 125
Huang, C., Wu, Y., & Triaud, A. H. M. J. 2016, *ApJ*, **825**, 98
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
Ikoma, M., & Hori, Y. 2012, *ApJ*, **753**, 66
Inamdar, N. K., & Schlichting, H. E. 2016, *ApJL*, **817**, L13
Jin, S., & Mordasini, C. 2018, *ApJ*, **853**, 163
Jin, S., Mordasini, C., Parmentier, V., et al. 2014, *ApJ*, **795**, 65
Johnson, J. A., Petigura, E. A., Fulton, B. J., et al. 2017, *AJ*, **154**, 108
Jönsson, H., Allende Prieto, C., Holtzman, J. A., et al. 2018, *AJ*, **156**, 126
Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, *ApJL*, **713**, L79
Kurucz, R. L. 1992, in IAU Symp. 149, The Stellar Populations of Galaxies, ed. B. Barbuy & A. Renzini (Dordrecht: Kluwer), 225
Latham, D. W., Rowe, J. F., Quinn, S. N., et al. 2011, *ApJL*, **732**, L24
Lee, K. 2014, astro: Astronomy Functions, Tools and Routines, R package version 1.2, <https://CRAN.R-project.org/package=astro>
Lindgren, L., Hernández, J., Bombrun, A., et al. 2018, *A&A*, **616**, A2
Lopez, E. D., & Fortney, J. J. 2014, *ApJ*, **792**, 1
Lopez, E. D., Fortney, J. J., & Miller, N. 2012, *ApJ*, **761**, 59
Lopez, E. D., & Rice, K. 2016, arXiv:1610.09390
Lundkvist, M. S., Huber, D., Silva Aguirre, V., & Chaplin, W. J. 2018, arXiv:1804.02214
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, **154**, 94
Mamajek, E. E., Torres, G., Prsa, A., et al. 2015, arXiv:1510.06262
Morbideilli, A. 2014, *RSPTA*, **372**, 20130072
Morton, T. D. 2015, isochrones: Stellar model grid package, Astrophysics Source Code Library, ascl:1503.010
Mulders, G. D., Pascucci, I., Apai, D., & Ciesla, F. J. 2018, *AJ*, **156**, 24

- Mulders, G. D., Pascucci, I., Apai, D., Frasca, A., & Molenda-Żakowicz, J. 2016, *AJ*, **152**, 187
- Mustill, A. J., Davies, M. B., & Johansen, A. 2015, *ApJ*, **808**, 14
- Owen, J. E., & Wu, Y. 2013, *ApJ*, **775**, 105
- Owen, J. E., & Wu, Y. 2017, *ApJ*, **847**, 29
- Petigura, E. A., Howard, A. W., Marcy, G. W., et al. 2017, *AJ*, **154**, 107
- Pinsonneault, M. H., Elsworth, Y. P., Tayar, J., et al. 2018, *ApJS*, **239**, 32
- Rogers, L. A. 2015, *ApJ*, **801**, 41
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. 2016, *The R Journal*, **8**, 1
- Seager, S., & Mallén-Ornelas, G. 2003, *ApJ*, **585**, 1038
- Serenelli, A., Johnson, J., Huber, D., et al. 2017, *ApJS*, **233**, 23
- Silburt, A., Gaidos, E., & Wu, Y. 2015, *ApJ*, **799**, 180
- Silva Aguirre, V., Davies, G. R., Basu, S., et al. 2015, *MNRAS*, **452**, 2127
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, **131**, 1163
- Snedden, C. A. 1973, PhD thesis, Univ. Texas at Austin
- Sousa, S. G., Santos, N. C., Adibekyan, V., Delgado-Mena, E., & Israelian, G. 2015, *A&A*, **577**, A67
- Stassun, K. G., Collins, K. A., & Gaudi, B. S. 2017, *AJ*, **153**, 136
- Thompson, S. E., Coughlin, J. L., Hoffman, K., et al. 2018, *ApJS*, **235**, 38
- Valenti, J. A., & Piskunov, N. 1996, *A&AS*, **118**, 595
- Van Eylen, V., Agentoft, C., Lundkvist, M. S., et al. 2018, *MNRAS*, **479**, 4786
- Venables, W. N., & Ripley, B. D. 2002, *Modern Applied Statistics with S* (New York: Springer)
- Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, *Proc. SPIE*, **2198**, 362
- Weiss, L. M., & Marcy, G. W. 2014, *ApJL*, **783**, L6
- West, R. G., Gillen, E., Bayliss, D., et al. 2018, arXiv:1809.00678
- Wu, Y., & Lithwick, Y. 2013, *ApJ*, **772**, 74
- Zinn, J. C., Pinsonneault, M. H., Huber, D., & Stello, D. 2018, arXiv:1805.02650