Confirming known planetary trends using a photometrically selected
Kepler sample

Jonah T. Hansen,1⋆ Luca Casagrande,1 Michael J. Ireland1 and Jane Lin1
1Research School of Astronomy and Astrophysics, The Australian National University, Canberra, ACT 2611, Australia

Accepted XXX. Received YYY; in original form ZZZ.

ABSTRACT
Statistical studies of exoplanets and the properties of their host stars have been critical to informing models of planet formation. Numerous trends have arisen in particular from the rich Kepler dataset, including that exoplanets are more likely to be found around stars with a high metallicity and the presence of a “gap” in the distribution of planetary radii at 1.9 $R_\oplus$. Here we present a new analysis on the Kepler field, using the APOGEE spectroscopic survey to build a metallicity calibration based on Gaia, 2MASS and Strömgren photometry. This calibration, along with masses and radii derived from a Bayesian isochrone fitting algorithm, is used to test a number of these trends with unbiased, photometrically derived parameters, albeit with a smaller sample size in comparison to recent studies. We confirm that planets are indeed more frequently found around higher metallicity stars: planetary frequencies are 0.88 ± 0.12 percent for [Fe/H] < 0 and 1.37 ± 0.16 percent for [Fe/H] ≥ 0. We also recover the planet radius gap, along with a slight positive correlation with stellar mass. We conclude that this method shows promise to derive robust statistics of exoplanets. We also remark that spectrophotometry from Gaia DR3 will have an effective resolution similar to narrow band filters and allow to overcome the small sample size inherent in this study.

Key words: planets and satellites: fundamental parameters – catalogues – stars: planetary systems – stars: fundamental parameters

1 INTRODUCTION
Since the release of Kepler’s (Borucki et al. 2010) rich collection of over 4000 exoplanet transits, the study of exoplanet statistics has blossomed into a thriving field. Exoplanet demographic studies have led to many interesting results, including the preference of large exoplanets ($R_p > 4 R_\oplus$) to form around high metallicity stars (Santos et al. 2004; Fischer & Valenti 2005; Zhu 2019) and that close, multiple planet systems form preferentially around metal poor stars (Brewer et al. 2018). However, perhaps one of the most important results to have come out of these studies is that of the planet radius “gap” - a decrease in the number of planets with radii around 1.5-2.0 $R_\oplus$ (e.g., Owen & Wu 2013; Fulton et al. 2017). This particular radius is significant, as it separates the classification of “super-Earths” from “sub-Neptunes”. Reasons for the presence of this gap are numerous, ranging from UV photoevaporation of a planet’s atmosphere (Owen & Wu 2013, 2017; Lopez & Rice 2018) to core-powered mass loss (Ginzburg et al. 2018; Gupta & Schlichting 2020).

Recently, more separate studies have found evidence for this gap, using both the Kepler (Van Eylen et al. 2018; Berger et al. 2020b) and K2 surveys (Hardegree-Ullman et al. 2020; Cloutier & Menou 2020) and a handful have also identified the gap follows a trend with stellar mass: the drop in occurrence is found at smaller radii around less massive stars (Fulton & Petigura 2018; Berger et al. 2020b). The radius gap still has ambiguity in the strength of this deficiency, with studies such as the seminal Fulton et al. (2017) revealing a shallow gap, whereas Van Eylen et al. (2018), with a much smaller sample but with very precise parameters from asteroseismology, find a gap nearly devoid of planets. There is hence a question whether the strength of this gap is due to more precise parameters, small number statistics, selection effects or a combination of the three.

Part of the ambiguity around these planetary trends, including the radius gap, stem from the imprecise nature of the Kepler input catalogue (KIC). The KIC was a compilation of 13 million stars with optical photometry and stellar parameters created for the purpose of choosing Kepler targets, of which approximately 200,000 were chosen (Batalha et al. 2010; Brown et al. 2011). However, the parameters for these stars were lacking in precision and some critical parameters, such as the age and mass of these stars, were missing entirely. To investigate exoplanet demographics, precise stellar parameters are required which has led to many follow-up studies of the stars in the KIC (e.g., Bruntt et al. 2012; Molenda-Żakowicz et al. 2013; Huber et al. 2014; Petigura et al. 2017; Furlan et al. 2018; Berger et al. 2020a). Many of these studies rely on spec-
2 Hansen et al.

troscopy, for which sample selection effects are usually quite strong and observations are biased towards the brightest targets.

In this paper, we derive stellar parameters for a photometrically unbiased sample of confirmed Kepler transiting planets to study some of the known trends concerning exoplanetary demographics. We assemble our sample starting from the Strömgren survey for Astroseismology and Galactic Archaeology (SAGA, Casagrande et al. 2014) and complementing it with photometry from Gaia DR2 (Gaia Collaboration et al. 2018) and 2MASS (Skrutskie et al. 2006). Metallicity of the Kepler host stars is derived through a photometric calibration based on the APOGEE survey (Majewski et al. 2017), and effective temperatures are calculated through the Infra-Red Flux Method (IFRM; e.g., Blackwell & Shallis 1977; Casagrande et al. 2010) We then calculate the masses and radii of Kepler host stars through Gaia parallaxes and isochrone fitting by using the Bayesian isochrone fitting algorithm ELLI (Lin et al. 2018). This results in us obtaining a similar planet radius-mass trend to that of Fulton & Petigura (2018) and Berger et al. (2020b), as well as finding large planets preferentially form around metal rich stars and a slight trend that multiple exoplanet systems form around metal poor stars.

2 CATALOGUE COMPILATION

Multiple stellar catalogues were combined to leverage finding an appropriate metallicity index for our planet host star sample. Foremost of these was the aforementioned Kepler Input Catalogue (KIC), a catalogue of stars that lie in the Kepler field (Brown et al. 2011). Not all the stars present in the KIC contain useful data on their properties and so a subset of the catalogue was used: all KIC objects viewed in Quarter 15 of the Kepler mission that had long cadence data. This is a sample collected before Kepler’s reaction wheel had failed, and after stars with faulty data had been removed.

The KIC catalogue was matched with the Gaia DR2 catalogue (Gaia Collaboration et al. 2018) to obtain Gaia photometry and parallaxes for these stars. We remark that for the isochrone fitting described in Section 5 we use distances from Bailey-Jones et al. (2018). The Gaia data for these KIC stars was combined with other photometric catalogues, including the 2MASS catalogue’s $J, H$ and $K$ band photometry and the Strömgren catalogue’s $u, v, b$ and $y$ band photometry produced by Casagrande et al. (2014). These catalogues were cross-matched so that all stars contained the photometry from each survey; resulting in our catalogue of multi-band photometry encompassing around 30,000 stars. We note here that this is a small fraction of the KIC, primarily due to the small fraction of stars in the Kepler field currently with Strömgren photometry.

The photometry was corrected for reddening using the Schlegel et al. (1998) reddening map. This map is known to overestimate reddening along the galactic plane (see e.g., Arce & Goodman 1999; Schlafly & Finkbeiner 2011). Hence, it was re-scaled by the following formula where $b$ is the galactic radius:

$$E(B - V)_{\text{res}} = E(B - V) + 0.1 \log(|b| - 3) - 0.16$$

which is appropriate for the range $5 \leq b \leq 20$ encompassed by the Kepler field, and whose derivation is explained in Kunder et al. (2017) and Casagrande et al. (2019). Magnitudes were de-reddened using extinction coefficients from Casagrande & VandenBerg (2014, 2018).

The photometric stellar catalogue was finally combined with the list of Kepler Objects of Interest (KOI). This is a list of all candidate exoplanets found in the Kepler field, provided by the NASA Exoplanet Archive (https://exoplanetarchive.ipac.caltech.edu/). Objects with a KOI disposition flag of false positive were removed and the remaining entries were paired with the photometric data from their host star, producing a separate KOI catalogue of about 800 exoplanets and their host stars.

3 METALLICITY CALIBRATION

In order to derive homogeneous metallicities for all stars in our catalogue, we devise a metallicity calibration using the photometry assembled in Section 2. The largest sample of stars with spectroscopic metallicities in the Kepler field is from APOGEE (APO Galactic Evolution Experiment; e.g., Majewski et al. 2017), an infrared spectroscopic mission that was designed to measure the radial velocities and, more importantly, chemical abundances of over 100,000 red giants withing the Milky Way. We used the $[\text{M/H}]$ metallicity from APOGEE DR14 (Abolfathi et al. 2018) to calibrate our photometry; removing stars with the STAR_BAD label and combining this data with the photometric catalogue compiled above to obtain a total of 2415 stars in the KIC with known metallicities. These stars were then used to derive a metallicity relation for the KIC as a whole.

To derive the best relation a Principle Component Analysis (hereafter referred to as PCA) decomposition was performed over 84 linear combinations of colour indices. PCA can greatly reduce the dimensionality of the given data; in this case, reducing the dimensionality of the 84 candidate colour indices and their combinations up to second order. The colour indices chosen were those suggested to be sensitive to metallicity, including the well established $m_1 = (v - b) - (b - y)$ from the Strömgren photometric system, as well as $u - b, G - K, v - G, R_p - K,$ and $(B_p - R_p) - (R_p - K)$.

A singular value decomposition was performed on the collection of 84 colour indices for our APOGEE cross-matched photometric catalogue; outputting linear combinations of the input dimensions called “Principle Components”. The first of these vectors describes the linear combination of variables that produces the most variance in the data, the second giving the combination that produces the second most variance and so forth. Through an analysis of these principle components, we identified that the first principle component corresponded to the effective temperature of the star. The second component was the desired parameter - the metallicity. The correspondence between the second principle component and the APOGEE metallicity can be seen in Fig. 1.

With the metallicity principle component identified, we ran an iterative process to reduce the number of input parameters so that the resulting calibration was not over determined. The process is...
as follows: the decomposition was completed with \( n \) colour index dimensions and the parameter with the weakest contribution to the second principle component was removed. The decomposition was then performed again with \( n - 1 \) dimensions, repeating the process until a few input indices remained.

At the end of this procedure, we found that the best colour indices to use were the \( a - b \) index from the Strömgren photometric system and the \( G - K \) index combining the \( G \) band from Gaia’s photometry and the \( K \) band from 2MASS. A second round of PCA was conducted with these two indices, as well as the APOGEE metallicity itself, resulting in a linear calibration of the form

\[
[Fe/H]_{\text{cal}} = a_0(G - K)^2 + a_1(u - b) + a_2
\]

This calibration still had some residual trends, particularly in the \( G - K \) colour index. To correct this, we further fitted a 5th order polynomial to the residuals and subtracted this from the metallicity. Hence, our final calibration was of the form:

\[
[Fe/H]_{\text{cal}} = c_1(G - K)^5 + c_2(G - K)^4 + c_3(G - K)^3 + c_4(G - K)^2 + c_5(G - K) + c_6(u - b) + c_7
\]

with calibration parameters:

\[
c_1 = -0.346793 \quad c_2 = 3.108458
\]

\[
c_3 = -10.38798 \quad c_4 = 15.30547
\]

\[
c_5 = -10.86391 \quad c_6 = 1.729865
\]

\[
c_7 = 0.875089
\]

Our calibration into the \( G - K \) vs \( u - b \) plane is shown in the upper panel of Fig. 2, where stars are colour coded by their APOGEE metallicity. Also shown in the bottom panels is the residual of our photometric metallicity calibration as function of spectroscopic [Fe/H] and \( G - K \). The standard deviation of the residuals (shown in red) is 0.18 dex. It should be noted that for [Fe/H] \( \lesssim -1 \), our calibration systematically overestimates the true metallicity. However, this is of little concern for the sake of our study, since the bulk of planets lie well above this limit.

Especially when looking at the \( G - K \) residual plot, we can see by eye that this calibration does not hold well everywhere. As we will describe in the following section, we performed multiple colour and magnitude cuts to determine a selection for which the sample of KOI host stars is representative of the larger KIC population, as well as such that the metallicity calibration is well behaved. We determined that this range is:

\[
1.4 \leq u - b \leq 2.8
\]

\[
1.2 \leq G - K \leq 2.0
\]

The residuals for the colour cut are shown in blue in Fig. 2, with a smaller standard deviation of 0.16 dex.

To test the validity of the calibration, our photometric metallicities were tested against the spectroscopic metallicities measured by Petigura et al. (2017). It should be noted that this sample comprises of mostly main-sequence stars; the regime where most planet host stars reside. We also compare our metallicities against the SpecMatch pipeline metallicities from Furlan et al. (2018); another spectroscopic Kepler follow-up survey. The comparisons are shown in Fig. 3, with a mean difference (Our metallicity - Petigura et al. (2017)) of \(-0.05 \pm 0.14\) dex, and a difference (Our metallicity - Furlan et al. (2018)) of \(0.00 \pm 0.10\) dex. Both of these are well within the quoted uncertainty of our metallicity calibration.

### 4 DETERMINING A REPRESENTATIVE SAMPLE

In comparing the metallicity distribution of the KOI sample to the KIC sample, it is important to understand the extend of any differences in brightness or colour distributions, which could bias conclusions about planetary demographics. These differences could be potentially caused by a strong stellar-mass dependent planetary frequency (e.g., as is known for rare Jovian planets Bowler et al. 2010), or a strong effect of instrumental noise in planet detectability at fainter included KIC stars. However, if the samples are very similar in apparent magnitude and colour distributions, then we can compare metallicities of the KOI and KIC samples without needing to re-sample a subset of the KIC stars.

Since our sample is drawn from photometric catalogues, we can perform well defined magnitude and colour cuts and ensure the KOI sample represents the underlying sample of stars found in the KIC catalogue. This is very different from spectroscopically selected samples, where stars and KOI might be preferentially picked for their properties. We took our collection of KOIs and analysed their distribution of colours and magnitudes between all KOIs and the full photometric sample when the cuts discussed in Section 4 are applied. The residuals for the colour cut are shown in blue in Fig. 2, with a standard deviation of 0.18 dex.

We determined that this range is:

\[
1.4 \leq u - b \leq 2.8
\]

\[
1.2 \leq G - K \leq 2.0
\]

14.1 \( \leq G \leq 16.0 \)

In the following of the paper the full photometric catalogue will be referred to as the parent population when restricted to the above colour and magnitude ranges. The KS statistic and resultant p-value for the relevant parameters can be found in Table 1. A Wilcoxon rank sum test was also computed and the results can also be found in Table 1. Finally, both tests were completed for the subset of KOIs that have a confirmed disposition according to the NASA Exoplanet database. The results are in Table 2.

From this, it is evident that all three parameters for both samples do not have statistically significant low p-values. Hence we cannot reject the null hypothesis that the samples are drawn from the same parent population and we can therefore use the photometrically calibrated list of KOI host stars as a representative population of the full photometric catalogue. An HR diagram of the colour-cut photometric sample is shown in Fig. 4.

### 5 OBTAINING RADII AND MASSES

One trend we aimed to investigate was the “Planet-Radius gap”, a feature where there is a relative absence of planets with radii around 1.9 \( R_\oplus \) (e.g., Owen & Wu 2013; Fulton et al. 2017), with some studies showing that the depression follows a slight dependence on
Figure 2. Comparison of our photometric metallicity calibration against that of APOGEE. a) The $G - K$, $u - b$ colour plane coloured by the APOGEE [M/H] metallicity index. Our photometric metallicity calibration is shown by the lines for a given metallicity. b) Metallicity residuals (our metallicity - APOGEE) against APOGEE metallicity. Red plots the full catalogue of stars, whereas blue only plots stars that fall within the colour cut discussed in the text. c) Same as b, but against the $G - K$ colour index.

We derived stellar masses and radii using the Bayesian isochrone fitting algorithm \( \text{Elli} \) (Lin et al. 2018), which is built upon the MIST isochrones (Choi et al. 2016). The input parameters used by \( \text{Elli} \) are effective temperatures ($T_{\text{eff}}$), 2MASS $K$ magnitudes, reddening, \textit{Gaia} DR2 parallaxes, surface gravities $\log(g)$ and our photometric metallicities. In the following, we describe in detail our procedure.

To obtain effective temperatures we run the InfraRed Flux Method (IRFM) for all our KOIs. The IRFM is an almost model independent photometric technique originally devised to obtain angular diameters to a precision of a few per cent, and capable of competing against intensity interferometry should a good flux cal-

| Parameter | KS Statistic $D$ | KS $p$ | Wilcoxon Statistic $U$ | Wilcoxon $p$ |
|-----------|-----------------|--------|------------------------|-------------|
| $G - K$   | 0.061           | 0.714  | 0.233                  | 0.671       |
| $u - b$   | 0.091           | 0.233  | -1.575                 | 0.135       |
| $G$       | 0.062           | 0.709  | 0.425                  | 0.816       |
The calibration be achieved (e.g., Blackwell & Shallis 1977; Blackwell et al. 1980). We used the implementation described in Casagrande et al. (2010) which employs Gaia and 2MASS photometry to derive effective temperatures and angular diameters for stars of known metallicity and surface gravity. We adopted our photometric metallicities, and log(g) from the KOI catalogue. Effective temperatures derived from the IRFM were then fed into ELLI along the other parameters needed to derive stellar radii and masses. A new estimate of log(g) was computed, iterating between the IRFM and ELLI. Because of the mild dependence of the IRFM on the adopted metallicity and surface gravity (see e.g., Alonso et al. 1995; Casagrande et al. 2006) only a couple of iterations were necessary to converge on a final mass and radius for each star.

First, we compared our $T_{\text{eff}}$ against those published in Petigura et al. (2017) and Furlan et al. (2018), showing excellent agreement, with a mean difference of 30 K and a standard deviation of 90 K (Fig. 5). Since $T_{\text{eff}}$ from the IRFM are sensitive to reddening (where a change of ±0.01 in $E(B-V)$ has an impact of ±50 K), this comparison suggests that reddening is well under control.

In addition to stellar radii obtained from isochrone fitting, the availability of angular diameters and Gaia distances (Bailer-Jones et al. 2018) for all our targets allowed us to derive radii independently of stellar isochrones. We dub these “empirical radii” since they are virtually free from any stellar modelling assumption. Finally, the planet radius was determined by applying the planet to star radius ratio provided in the KOI catalogue; a parameter estimated from the transit depth.

Fig. 6 compares the stellar radii derived by ELLI against the empirical ones, showing relative residuals (ELLi radius - empirical radius)/ELLi radius with a mean of 0.00 ± 0.03 solar radii and a fractional error of 2 percent. This gives confidence that the radii and other stellar quantities derived from ELLi are robust. It should be noted that an outlier was removed, due to having abnormally large uncertainties on its parallax. From this point on, we adopt the empirical radius as our accepted stellar radius.

The empirical radii also have good relative uncertainties, with a mean of 3.4 percent, which is on par with that of Berger et al. (2020a) and Fulton & Petigura (2018). When multiplied by the Kepler planet to star radius ratio, we find our planet radii have uncertainties with a mean of 6.2 percent, highlighting that uncertainties in the Kepler radius ratio carry a significant contribution to the uncertainties in the planetary radius.

Finally, we tested our stellar parameters against those derived by Berger et al. (2020a), in particular the stellar luminosity (Fig. 7) and stellar radius (Fig. 8). Both of these parameters had extremely good agreement, with luminosity residuals of $-0.02 \pm 0.10 L_\odot$ and radius residuals of $0.01 \pm 0.03 R_\odot$. We also tested our masses against their catalogue, finding a mean difference and standard deviation of

---

**Figure 3.** Comparison of our photometric metallicity against that of Petigura et al. (2017) and Furlan et al. (2018), with residuals plotted against the literature values of metallicity, effective temperature and surface gravity.

**Figure 4.** HR Diagram of the representative colour-cut photometric sample. The colours represent the different subsets of the full photometric catalogue.
6 Hansen et al.

$-6 \pm 7$ percent, which again shows good agreement albeit with our mass tending to be lower than that of Berger et al. (2020a).

6 RESULTS AND ANALYSIS

6.1 Metallicity trends

We first investigated trends concerning metallicity between the KOI host stars and the parent population (ref. Section 4). The subset of KOIs which have been labelled as confirmed by the NASA Exoplanet Database was also extracted and compared, before finally splitting the confirmed KOIs between large ($R_p \geq 4R_\oplus$) and small ($R_p < 4R_\oplus$) planetary radii. The use of the confirmed sub-sample was to ensure that we were not affected by non-planetary companions, with the trade off of a smaller sample size. If a star had multiple planets, then it was classified according to the radius of the largest one. Histograms and cumulative distribution functions (CDFs) of these populations are shown in Fig. 9.

The mean metallicity of the KOIs ([Fe/H] = -0.01) and especially that of the confirmed KOIs ([Fe/H] = 0.02) is different from that of the parent population of stars as a whole ([Fe/H] = -0.03). This suggests that the exoplanet host stars are more metal rich than the rest of the candidate KOIs. To confirm this deviation, KS and Wilcoxon rank-sum tests were undertaken using the full metallicity distribution function (MDF), with each subset being tested against the parent population. The results are shown in Table 3.

With a $p \approx 9$ percent, the null hypothesis that the MDF of all KOI is drawn from that of the parent population cannot be rejected.
Figure 9. Histograms and cumulative distribution functions of the metallicity distribution of various exoplanet subsets. Confirmed KOI are those with a NASA Exoplanet Database designated disposition of confirmed, whereas “All KOI” refer to those with a disposition of confirmed or candidate. However, when restricting ourselves to the sample of confirmed KOIs, the significance drops to a mere 0.4 percent thus rejecting the null hypothesis. This is also the case for the two sub-samples with small and large planetary radii, although significance levels increase to around 3 percent. In particular, when looking at the histograms in Fig. 9, we see that the confirmed exoplanet host stars seem to favour higher metallicities than their non-exoplanet hosting counterparts, supporting earlier results (e.g., Gonzalez 1997; Santos et al. 2004; Fischer & Valenti 2005; Zhu 2019). The larger p-value from the sample of all KOIs (including those with a disposition of candidate), may be due to some of the candidate KOIs not being planetary companions.

To investigate further the trend with metallicity, a plot of the percentage of stars with a confirmed KOI for a given metallicity bin was created. This particular diagram was based on the work of Zhu (2019), and can be seen in Fig. 10. Also plotted in this figure is the percentage of stellar systems with multiple exoplanets, with the aim to compare to the findings by Brewer et al. (2018) as to whether multiple exoplanet systems were preferentially around lower metallicity host stars.

As predicted from Santos et al. (2004), Fischer & Valenti (2005) and Zhu (2019) among others, and as inferred from the cumulative distribution function, exoplanets (and especially those that have a radius greater than 4 Earth radii) appear to form preferentially around higher metallicity stars. Smaller exoplanets also appear to favour higher metallicities, peaking above solar metallicity, although then declining at very high metallicities. This furthers the work of Buchhave et al. (2012), who claims that while smaller exoplanets form around stars with a wide range of metallicities, large exoplanets form around primarily metal rich stars. We show that while the smaller exoplanets have a weaker trend than their larger counterparts, they still have a bias towards metallicities around and above solar metallicity.
Uncertainties are drawn from the Poisson statistic. We chose not to plotted the density distribution as contours in Fig. 11. We also plot 10,000 samples assigning each time normal errors in stellar mass 2D density plot through a Monte Carlo (MC) simulation. We drew planet-radiusgap. With the stellarmass and planetradius were cov-

As mentioned previously, one aim of this work was to study the metallicities. This indicates that the multiple exoplanet systems de-
solar luminosity as seen in Fig. 10, has an upwards trend at low
strength of this trend.

Wethusfindthatallexoplanetsaremorelikelytobeobserved around metal rich stars, while smaller exoplanets are 1.5 times as likely. We thus find that all exoplanets are more likely to be observed at higher metallicities, with the size of the exoplanet influencing the strength of this trend.

Furthermore, multiple planet systems follow the trend supported by Brewer et al. (2018), which while also having a peak at solar luminosity as seen in Fig. 10, has an upwards trend at low metallicities. This indicates that the multiple exoplanet systems detected by Kepler are likely to be compact, small exoplanet systems of the type described by Brewer et al. (2018), which favour lower metallicity stars.

### 6.2 The planet-radius gap

As mentioned previously, one aim of this work was to study the planet-radius gap. With the stellar mass and planet radius we recovered through the processes outlined in Section 5, we generated a 2D density plot through a Monte Carlo (MC) simulation. We drew 10,000 samples assigning each time normal errors in stellar mass and planet radius for each of our confirmed KOI data points, and plotted the density distribution as contours in Fig. 11. We also plot 1000 random samples from the MC simulation. We chose not to include the KOIs with a candidate disposition due to the potential presence of false positives in this sample (examined in Section 6.1).

What is particularly clear is the presence of a “gap” with a positive slope around 1.8-2.0 $R_\oplus$. This highlights a weak trend with stellar mass, similar to what was found by Fulton & Petigura (2018) and Berger et al. (2020b). To confirm this, we overplotted the best fit line to the radius gap from Berger et al. (2020b) with a slope of $d \log R_p/d \log M_{\star} = 0.26$ in Fig. 11. This line fits the data well, again supporting the conclusions of previous studies.

To view the gap more clearly, we contracted this plot over mass and normalised the data, creating a simple histogram of planet radii. This is shown in solid red in Fig. 12. We also show in solid blue the histogram of planet radii when KOI with a disposition of candidate are included. The most notable feature is a very clear bimodal distribution, with a gap again at 1.9 $R_\oplus$, supporting the conclusions of e.g. Fulton & Petigura (2018) and Berger et al. (2020b). The restriction of only including confirmed KOI influences the distribution by increasing the size of the second peak at ~ 2.5 $R_\oplus$ and decreasing the width of the first at ~ 1.6 $R_\oplus$, but the location of the gap does not change.

Following the work of Berger et al. (2020b), we also looked at how the radius histogram was affected by the incident stellar flux falling on the planet. We chose the separating flux value of 150 $F_\oplus$ from Berger et al. (2020b), to test to see whether we identified a similar trend; our sample had 114 planets designated as cool. This is plotted in Fig. 13, where again we have chosen to only plot confirmed KOI. We recover that planets with higher incident flux exhibit smaller radii than their cooler counterparts, possibly due to evaporation of the atmospheres of these planets. As with Berger et al. (2020b), we caution that these results may be a result of small number statistics and are likely fraught with Kepler selection effects.

### 7 CONCLUSIONS

We have compiled a photometric catalogue of stars in the Kepler field utilising photometry from Gaia, Strömgren and 2MASS catalogues. We created a metallicity calibration based on APOGEE spectroscopy to obtain a metallicity for our photometric sample, and then performed well defined colour and magnitude cuts to ensure our dataset was a representative sample of the underlying population of stars. We then derived temperatures and angular diameters using the IFRM, which were then used to derive stellar radii and stellar mass through Bayesian isochrone fitting. The resultant parameters were compared favourably with previous results from the literature, especially giving stellar radii with relative uncertainties around 3.4 percent. Planetary radii uncertainties of 6.2 percent hence indicate a major uncertainty contribution from the Kepler planet to star ratio.

The main findings from our analysis are as follows:

- We find that the stars hosting confirmed KOIs have a statistically different metallicity distribution than the parent population of stars in the Kepler field. We also find that this statistical claim is not valid for the sample of KOIs that include those with a disposition of candidate, consequences of the undetected false positives in the list of KOIs.
- We quantify the metallicity distribution differences between KOI and the larger sample of KIC stars, finding that KOI hosts tend to be more metal rich than their non-planet hosting counterparts. While holding especially true for large exoplanets larger than 4 $R_\oplus$, which has been known about in literature for some time (e.g., Buchhave et al., 2012), we also find this holds for smaller exoplanets albeit to a weaker extent. This follows the conclusions of (e.g., Zha 2019). Finally, we also find that Kepler stars with more than one exoplanet favour metal rich planets to a lesser degree than those with a single detection, supporting the conclusions of Brewer et al. (2018). In particular, we find an increase in the number of multiple exoplanet systems at the lowest metallicity bin in our sample.
- In a two dimensional histogram of planetary radius and stellar mass, we recover the planet radius gap at ~ 1.9 $R_\oplus$. We find that the gap exhibits a weak positive trend with stellar mass, corroborating the findings of Fulton & Petigura (2018) and Berger et al. (2020b).

We also tentatively find that there is a trend that planets with a high incident flux ($>150 F_\oplus$) tend to have smaller radii.

We note that our sample size is relatively small, especially compared to recent studies such as that of Fulton & Petigura (2018) and Berger et al. (2020b). The major reason for this is the limited number of stars in the Kepler field for which we had Strömgren photometry for. However, we anticipate that with the upcoming Gaia DR3 release we can obtain a larger sample of e.g., Strömgren photometry (or other suitable metallicity sensitivity indices) directly from the $BP$ and $RP$ spectra. With this larger sample, we believe that the methods described in this paper will be limited solely by Kepler uncertainties, thus allowing for more robust statistics and deeper insight into the demographics of Kepler’s exoplanet population.
Figure 11. 2D Monte Carlo distribution of the stellar mass and planetary radius of KOI with a confirmed disposition. Density contours are shown behind a random selection of 1000 samples. Note the presence of a gap around $R_p = 1.9 \, R_\oplus$. Also plotted in red is the best fit line to the radius gap from Berger et al. (2020b), with their slope of $d \log R_p / d \log M_{\text{star}} = 0.26$. It can be seen that our data also fits this line well, showing that the radius gap has a slight positive correlation with stellar mass.

Figure 12. Normalised histograms of the planetary radius of all KOI (blue) and only the confirmed KOI (red).

Figure 13. Normalised histograms of the planetary radius of confirmed KOI, separated by host star luminosities. All confirmed KOI are shown in black, whereas the red and blue histograms represent the subset with a planetary flux of less than and greater than 150 $F_\oplus$ respectively.

ACKNOWLEDGEMENTS

L.C. is the recipient of the ARC Future Fellowship FT160100402. M.I. acknowledges support from the ARC Discovery Scheme (DP170102233). This research has also 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. 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.
DATA AVAILABILITY

The data underlying this article are available in the article and in its online supplementary material.

REFERENCES

Abolfathi B., et al., 2018, ApJS, 235, 42
Alonso A., Arribas S., Martinez-Roger C., 1995, A&A, 297, 197
Arce H. G., Goodman A. A., 1999, ApJ, 512, L135
Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Mantelet G., Andrae R., 2018, AJ, 156, 58
Batalha N. M., et al., 2010, ApJ, 713, L109
Berger T. A., Huber D., van Saders J. L., Gaidos E., Tayar J., Kraus A. L., 2020a, AJ, 159, 280
Berger T. A., Huber D., Gaidos E., van Saders J. L., Weiss L. M., 2020b, AJ, 160, 108
Blackwell D. E., Shallis M. J., 1977, MNRAS, 180, 177
Blackwell D. E., Petford A. D., Shallis M. J., 1980, A&A, 82, 249
Borucki W. J., et al., 2010, Science, 327, 977
Bowler B. P., et al., 2010, ApJ, 709, 396
Brewer J. M., Wang S., Fischer D. A., Foreman-Mackey D., 2018, ApJ, 867, L3
Brown T. M., Latham D. W., Everett M. E., Esquerdo G. A., 2011, AJ, 142, 112
Brunth., et al., 2012, MNRAS, 423, 122
Buchhave L. A., et al., 2012, Nature, 486, 375
Casagrande L., VandenBerg D. A., 2014, MNRAS, 444, 392
Casagrande L., VandenBerg D. A., 2018, MNRAS, 479, L102
Casagrande L., Portinari L., Flynn C., 2006, MNRAS, 373, L3
Casagrande L., Ramírez I., Meléndez J., Bessell M., Asplund M., 2010, A&A, 512, A54
Casagrande L., et al., 2014, ApJ, 787, 110
Casagrande L., Wolf C., Mackey A. D., Nordlander T., Yong D., Bessell M., 2019, MNRAS, 482, 2770
Choi J., Dotter A., Conroy C., Cassiello M., Paxton B., Johnson B. D., 2016, ApJ, 823, 102
Cloutier R., Menou K., 2020, AJ, 159, 211
Fischer D. A., Valenti J., 2005, ApJ, 622, 1102
Fulton B. J., Petigura E. A., 2018, AJ, 156, 264
Fulton B. J., et al., 2017, AJ, 154, 109
Furlan E., et al., 2018, ApJ, 861, 149
Gaia Collaboration et al., 2018, A&A, 616, A1
Ginzburg S., Schlichting H. E., Sari R., 2018, MNRAS, 476, 759
Gonzalez G., 1997, MNRAS, 285, 403
Gupta A., Schlichting H. E., 2020, MNRAS, 493, 792
Hardegree-Ullman K. K., Zink J. K., Christiansen J. L., Dressing C. D., Ciardi D. R., Schlieder J. E., 2020, ApJS, 247, 28
Huber D., et al., 2014, ApJS, 211, 2
Kunder A., et al., 2017, AJ, 153, 75
Lin J., Dotter A., Ting Y.-S., Asplund M., 2018, MNRAS, 477, 2966
Lopez E. D., Rice K., 2018, MNRAS, 479, 5303
Majewski S. R., et al., 2017, AJ, 154, 94
Molenda-Zakowicz J., et al., 2013, MNRAS, 434, 1422
Owen J. E., Wu Y., 2013, ApJ, 775, 105
Owen J. E., Wu Y., 2017, ApJ, 847, 29
Petigura E. A., et al., 2017, AJ, 154, 107
Santos N. C., Israelian G., Mayor M., 2004, A&A, 415, 1153
Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Skrutskie M. F., et al., 2006, AJ, 131, 1163
Van Eylen V., Agentoft C., Lundkvist M. S., Kjeldsen H., Owen J. E., Fulton B. J., Petigura E., Snellen I., 2018, MNRAS, 479, 4786
Zhu W., 2019, ApJ, 873, 8

This paper has been typeset from a TEX/LATEX file prepared by the author.