MNRAS 000, 1–?? (2017) Preprint 23rd January 2019 Compiled using MNRAS LATEX style file v3.0

An asteroseismic view of the radius valley: stripped cores, not born rocky

V. Van Eylen[1]★, Camilla Agentoft[2], M. S. Lundkvist[2,3], H. Kjeldsen[2], J. E. Owen[4], B. J. Fulton[5], E. Petigura[5], I. Snellen[1]

[1] *Leiden Observatory, Leiden University, postbus 9513, 2300RA Leiden, The Netherlands*
[2] *Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark*
[3] *Zentrum für Astronomie der Universität Heidelberg, Landessternwarte, Königstuhl 12, 69117 Heidelberg, Germany*
[4] *Astrophysics Group, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK*
[5] *California Institute of Technology, Pasadena, California, USA*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Various theoretical models treating the effect of stellar irradiation on planetary envelopes predict the presence of a radius valley: i.e. a bimodal distribution of planet radii, with super-Earths and sub-Neptune planets separated by a valley at around $\approx 2\ R_\oplus$. Such a valley was observed recently, owing to an improvement in the precision of stellar, and therefore planetary radii. Here we investigate the presence, location and shape of such a valley using a small sample with highly accurate stellar parameters determined from asteroseismology, which includes 117 planets with a median uncertainty on the radius of 3.3%. We detect a clear bimodal distribution, with super-Earths ($\approx 1.5\ R_\oplus$) and sub-Neptunes ($\approx 2.5\ R_\oplus$) separated by a deficiency around $2\ R_\oplus$. We furthermore characterize the slope of the valley as a power law $R \propto P^\gamma$ with $\gamma = -0.09^{+0.02}_{-0.04}$. A negative slope is consistent with models of photo-evaporation, but not with the late formation of rocky planets in a gas-poor environment, which would lead to a slope of opposite sign. The exact location of the gap further points to planet cores consisting of a significant fraction of rocky material.

Key words: planets and satellites: physical evolution – planets and satellites: composition – planets and satellites: formation – planets and satellites: fundamental parameters

1 INTRODUCTION

Various theoretical models predict that planets at short orbital periods are strongly influenced by the radiation of their host stars. For example, at the shortest orbital period a "photoevaporation desert", i.e. an absence of sub-Neptune-size planets ($1.8 - 4.0\,R_\oplus$) and an increase in rocky planets (R<$1.8\,R_\oplus$) has been predicted (Lopez & Fortney 2013) and observed with increasing clarity as the precision of stellar parameters increased (Borucki et al. 2011; Lundkvist et al. 2016).

Furthermore, formation models predict that atmospheric erosion of short-period planets results in the presence of a "photoevaporation valley", i.e. a gap in the radius distribution of planets around $1.75{-}2\ R_\oplus$ (Owen & Wu 2013; Jin et al. 2014; Lopez & Fortney 2014; Chen & Rogers 2016; Lopez & Rice 2016; Owen & Wu 2017). This valley defines the boundary between planets with a mass large enough to hold on to their gas envelope, and planets which have been stripped of their atmospheres and consist of the remnant core. The specific shape and slope of the valley depends on the details of planet formation, the composititon of the formed planets and the physics of evaporation (e.g. Lopez & Rice 2016; Owen & Wu 2017).

Observing this valley is not straightforward and is complicated by the relatively high uncertainty in observed planet radii, a result of uncertain stellar radii (Owen & Wu 2013). Recently, Fulton et al. (2017) provided clear evidence of the valley by using a spectroscopic sample from the California-*Kepler* Survey (CKS), with better-constrained stellar parameters (Petigura et al. 2017; Johnson et al. 2017). Despite the clear detection of the bimodal radius distribution and a radius gap, Fulton et al. (2017) did not attempt to constrain the slope of this gap as a function of orbital period.

Here, we investigate the radius gap using a sample with homogeneously determined stellar parameters from astero-

★ E-mail: vaneylen@strw.leidenuniv.nl

seismology (Silva Aguirre et al. 2015; Lundkvist et al. 2016). This sample is smaller than the CKS sample, but has better constrained stellar parameters, which translate into more accurate planet parameters.

In Section 2 we describe our sample and parameter determination. In Section 3 we show the modeling of the radius valley. Finally, in Section 4 we compare our findings with theoretical predictions, and we draw conclusions in Section 5.

2 METHODS

In this work, we combine accurate stellar parameters from asteroseismology (Silva Aguirre et al. 2015; Lundkvist et al. 2016) with carefully modeled planet transits, to investigate the location, size, and shape of the so-called 'radius gap'. We first detail how we determine planet parameters, and then describe the properties of our sample.

2.1 Parameter Determination

To determine accurate planet parameters from transit surveys, accurate stellar parameters are required, because the transit depth only constrains $R_\mathrm{p}/R_\star$, where R_p and $R_\star$ are the planetary and stellar radius, respectively. We therefore start from a sample of exoplanet host stars with parameters homogeneously measured from asteroseismology, which can provide highly precise masses and radii for a sample of bright stars. For systems with multiple transiting planets, we use the planet modeling by Van Eylen & Albrecht (2015), which uses stellar parameters taken from the asteroseismic modeling by Silva Aguirre et al. (2015). For systems with a single transiting planet, planet modeling was similarly done by Van Eylen et al. (2017), which uses the slightly more complete asteroseismic catalogue by Lundkvist et al. (2016). Both asteroseismic catalogues are fully consistent (Lundkvist et al. 2016).

We summarize the planet modeling approach here. We start from the Presearch Data Conditioning (PDC) data (Smith et al. 2012). Using an iterative approach, the times of individual transits are determined using the transit model parameters. The individual transit times are then used to determine the best orbital period, and determine if any transit timing variations (TTVs) are present. The systems for which a sinusoidal TTV model is included are detailed in Van Eylen & Albrecht (2015) and Van Eylen et al. (2017). Planet transits are modeled with analytical transit equations (Mandel & Agol 2002). Our fitting procedure uses a Markov Chain Monte Carlo (MCMC) aproach using the *emcee* code (Foreman-Mackey et al. 2013), a Python implementation of the Affine-Invariant Ensemble Sampler (Goodman & Weare 2010). Eight planet parameters are sampled, namely the ratio of planet to star radius ($R_\mathrm{p}/R_\star$), the impact parameter (b), two combinations of eccentricity and angle of periastron e and ω ($\sqrt{e}\cos\omega$ and $\sqrt{e}\sin\omega$), the time of mid-transit (T_0), an offset in flux (F), and two stellar limb darkening parameters following a quadratic limb darkening law (u_1 and u_2). A flat prior is used for all parameters except limb darkening, for which a Gaussian prior was used, with the mean value predicted from a Kurucz atmosphere table (Claret & Bloemen 2011) and a standard deviation of 0.1. The stars are cross-checked for contamination from nearby stars from high-resolution imaging (e.g. Furlan et al. 2017). We refer to Van Eylen & Albrecht (2015) and Van Eylen et al. (2017) for a more detailed description of the transit analysis method. The stellar mass and radius, and the planet radius and orbital period are listed in Table 1 for all systems in our sample.

2.2 Sample Properties

As a starting point, we use the sample of planet host stars with homogeneously-determined asteroseismic parameters (Silva Aguirre et al. 2015; Lundkvist et al. 2016). As detailed in Van Eylen & Albrecht (2015) and Van Eylen et al. (2017), a few systems were removed from the initial sample, e.g. because they have subsequently been identified as false positives or likely false positives, or because they have not been observed in *Kepler*'s one minute short cadence sampling, which decreases the precision of the derived stellar and planetary parameters. Most of the planets have been confirmed or validated, while 17 are unconfirmed planet candidates that are likely to be bona fide planets (Morton et al. 2016; Van Eylen et al. 2017). The final sample contains 75 stars and 117 planets, which are listed in Table 1.

The requirement of measureable p-mode oscillations implies that our sample contains primarily bright stars (with mean *Kepler* magnitude 11.3). Their stellar types are centered around main sequence F and G stars, and a few more evolved stars. A histogram of the *Kepler* magnitude, stellar temperature and stellar radius is shown in Figure 1.

The stellar properties of our sample are broadly similar to those investigated by Fulton et al. (2017), which contains main sequence stars with a temperature range of 4700-6500 K. Our sample spans only the bright end of the Fulton et al. (2017) stars and is significantly smaller – 117 planets, compared to 900 in the adopted Fulton et al. (2017) sample. However, the parameters are determined to significantly greater precision, e.g. the median fractional uncertainty on the stellar radius is 2.2%, or 0.03 $R_\odot$, which can be compared to an 11% uncertainty in the CKS sample (Fulton et al. 2017) and a 25% uncertainty in the more general *Kepler* catalogue (Huber et al. 2014). This, in turn, leads to a median fractional uncertainty on the planet radius of 3.3% (or 0.068 $R_\oplus$), compared to 12% in the CKS analysis (Fulton et al. 2017).

3 RADIUS-PERIOD GAP

The planets in our sample are plotted in a period-radius plane in Figure 2, and compared to the sample by Fulton et al. (2017) which is larger but has higher uncertainties. We also plot the sample as a function of incident flux in Figure 3.

We now limit our sample to planets smaller than 4 $R_\oplus$. Even by eye, an absence of planets around $R \approx 2\ R_\oplus$ can be seen. In Figure 4, we show a histogram of the planet radius, which similarly shows a bimodal distribution with peaks roughly at $\approx 1.5\ R_\oplus$ and $\approx 2.5\ R_\oplus$, and a dip in between these peaks.

Figure 4 has not been corrected for transit probability, which is slightly lower for the planets above the gap, which

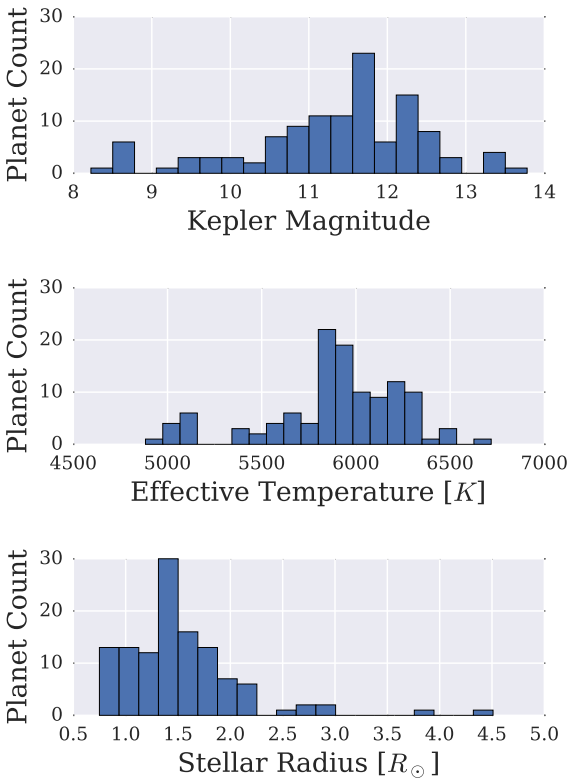

Figure 1. Histograms showing the basic properties of our sample: *Kepler* magnitude (top), stellar effective temperature (middle), and stellar radius (bottom). Stars with multiple planets are counted multiple times, but the shape of the histogram is not fundamentally changed if each star is only counted once.

occur at longer average periods, and has furthermore not been corrected for detection probability, which is lower at the smallest planets which are more likely to be missed. These corrections would be important to calculate absolute planet occurrence, but the sparseness of our sample makes it poorly suited for this purpose. However, any such correction would not affect the bimodal shape of the histogram.

We now constrain the shape and slope of the gap as a function of radius and orbital period. We first attempt to directly fit the absence of data points itself, using a linear model $\log R_{\rm mod} = m \log P_{\rm mod} + a$, where $R_{\rm mod}$ and $P_{\rm mod}$ are the modeled radius and period, and m and a are the slope and offset we set out to determine. To fit an absence of data points (the 'gap'), we invert the likelihood function, i.e.

$$\log L = -0.5 \sum_i \frac{(\log R_{\rm obs} - \log R_{\rm mod})^{-2}}{\sigma^2_{\log R}} - \log \frac{1}{\sigma^2_R}, \qquad (1)$$

where $R_{\rm obs}$ and $R_{\rm mod}$ are the observed and modeled planet radii, and σ_R is the uncertainty on the observed radius. Here, the power -2 ensures that the fit maximizes the distance to observations, fitting an absence of data, rather than the usual factor 2, when attempting to make a best fit through the observed data points. We then optimize the likelihood with an MCMC algorithm (*emcee,* Foreman-Mackey et al. 2013), using uninformative flat priors on the slope m and offset a, while limiting their range to $-0.5 \le m \le 0.5$ and $\log 1 \le a \le \log 4$, to ensure that the fit remains within our range of observations. We fit all data with $R \le 4\ R_\oplus$, and $1 \le P \le 100$ days. We sample with 10 walkers, taking 4000 steps each, after a burn-in phase of 2000 steps.

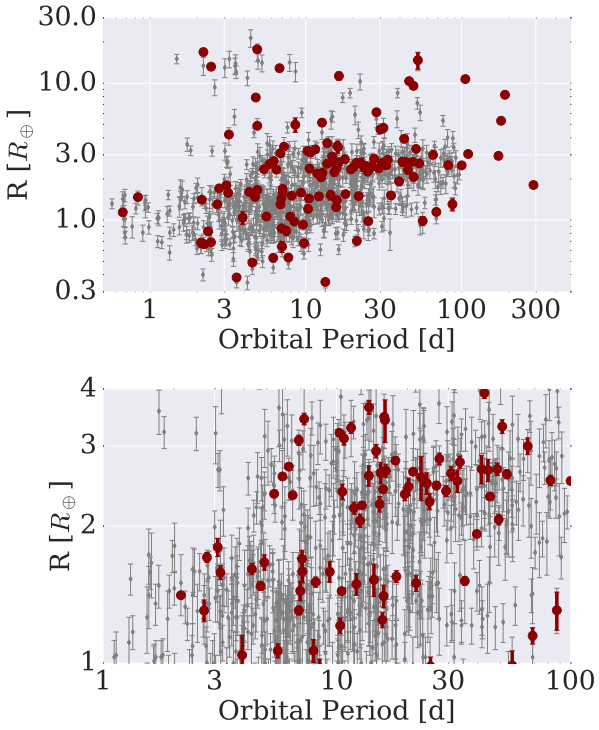

Figure 2. The planet radius as a function of orbital period. In grey, data points and uncertainties by Fulton et al. (2017) are shown, while the sample described here is shown in red. In many cases, the uncertainties are smaller than the symbol size. The bottom plot highlights the part of the sample where the radius gap occurs, at $R \approx 2\ R_\oplus$.

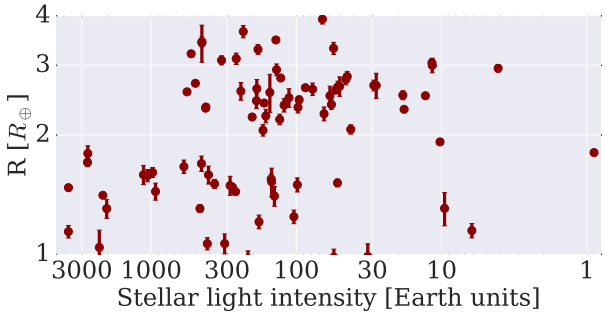

Figure 3. Similar to Figure 2, but with the planet radius as a function of incident flux rather than orbital period. In many cases, the uncertainties are smaller than the symbol size. The x-axis has been inverted, so that high incident flux (short orbital periods) are on the left. As before, only planets smaller than 4 $R_\oplus$ are shown.

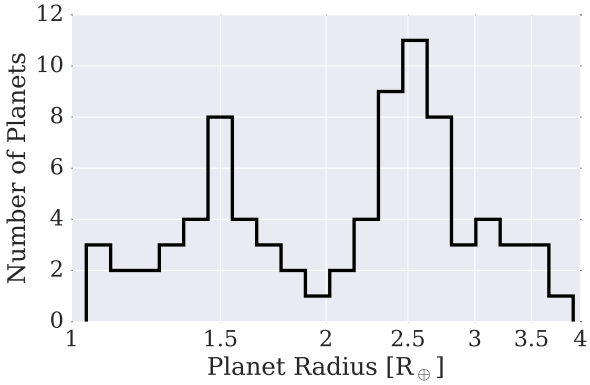

Figure 4. A histogram of the number of planets in the sample as a function of planet radius, with 1 $R_\oplus \leq R \leq 4\ R_\oplus$, using 20 logarithmic radius bins. Two peaks can be observed at $\approx 1.5\ R_\oplus$ and $\approx 2.5\ R_\oplus$, with a low density of planets in between.

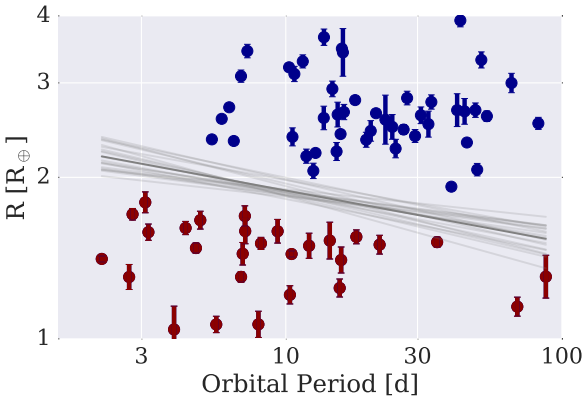

Figure 5. The grey lines show the best fits to the bootstrapped samples, 20 fits out of 1000 bootstrapped samples are shown, with the thicker line showing their average. We find a slope $m = 0.10 \pm 0.03$ and offset $a = 0.38 \pm 0.03$. We use these fits to separate our sample into planets below the gap (red) and planets above (blue).

We find median values $m = -0.08$ and $a = 0.34$. If we further limit our sample to $P \leq 25$ days, we find $m = -0.10$ and $a = 0.35$, showing that within the limitations of our sample, the measurement of the slope is largely independent of the precise period cut. A downside of this approach is that this likelihood function leads to unrealistically small uncertainties which depend heavily on the uncertainty of the observed radii. However, the true uncertainty of the slope of the radius valley is a result of the sparseness of the sampling, rather than the precision with which individual radii are measured.

To calculate the uncertainty due to our sampling, we make bootstrap versions of our sample, by generating new samples with the same size from our observed sample, allowing replacement. In these new, bootstrapped samples, some planets will be counted multiple times, while others may not be counted at all. In this way, we generate 1000 new samples, and apply the MCMC algorithm to each of these, as described above. We then take the 50% quantile for all samples of m and a, and use the 16% and 84% quantiles for the uncertainties. We find $m = -0.10 \pm 0.03$ and $a = 0.38 \pm 0.03$, which as expected results in similar values, but with higher, more realistic uncertainties. In Figure 5, we show 20 randomly drawn linear fits. We again check whether our result depends on orbital period by limiting our sample to $P < 25$ days, and find $m = -0.13^{+0.04}_{-0.05}$ and $a = 0.41 \pm 0.05$, which is a slightly steeper slope, albeit consistent at 1σ with the values above.

We can now use these fits to the gap to separate our sample into planets below and above the gap. Subsequently, we can look at the planets below the gap to directly estimate the slope of the gap, by looking at the maximum radius of these planets as a function of orbital period. We create four logarithmic bins as a function of period, and calculate the maximum radius in each bin, repeating the procedure by resampling our data, again allowing repetition of indivdual observations. We then apply a linear regression to each of the bootstrapped samples, and again calculate 16%, 50% and 84% quantiles. We find that the slope of the maximum of the lower part of the radius valley, i.e. the super-Earths, is $m = -0.05^{+0.01}_{-0.03}$ and $b = 0.26 \pm 0.02$. The result is shown in Figure 6.

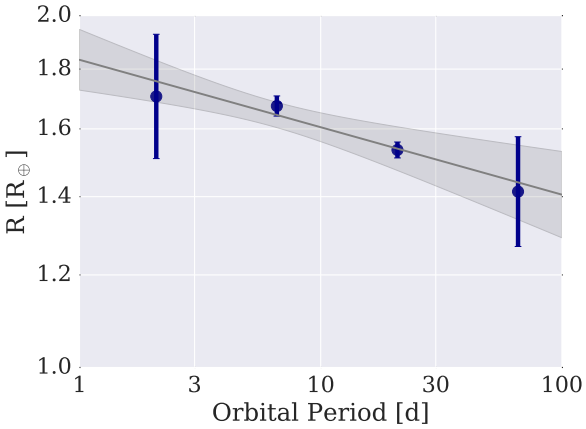

Figure 6. The data points show the maximum radius of planets below the gap as a function of orbital period, with the uncertainty derived from 1000 bootstrap iterations on the initial sample. The grey shows the best fit, together with a 68% confidence interval, again derived from the bootstrap iterations. We find that the slope of the gap is $m = -0.05^{+0.01}_{-0.03}$ and $b = 0.26 \pm 0.02$.

The downside of this approach is that it uses only a few observations (i.e. none of the sub-Neptunes were included) and is potentially sensitive to binning. A more robust approach makes use of support vector machines to determine the hyperplane of maximum separation between the planets above and below the valley. This line of separation maximizes the distance to points of the different classes of data (in this case, the super-Earths below the valley, and the sub-Neptunes above). Here, we use the Python implementation of support vector classification, *SVC*, in the scikit machine learning package *sklearn*. To determine the hyperplane

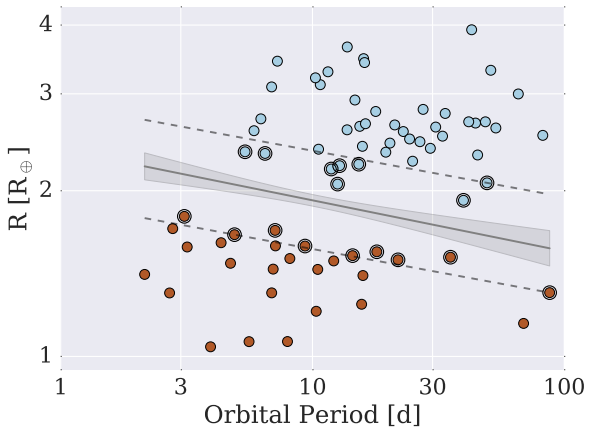

Figure 7. The slope of the radius valley as determined by support vector machines. The grey line represents the hyperplane of maximum separation, together with a 68% confidence interval derived from bootstrapping the original sample. The super-Earths below the radius valley are shown in red, while the sub-Neptunes above the valley are plotted in blue. The encircled data points are the support vectors, which determine the slope of the radius valley. The parallel dotted lines go through the support vectors, and are determined by offsets $a_{\rm low} = 0.29^{+0.04}_{-0.03}$ and $a_{\rm upp} = 0.44^{0.04}_{-0.03}$ respectively.

a penalty parameter C has to be chosen. This parameter determines the trade-off between maximizing the margin of separation and the tolerance for misclassification of observations, with high values of C allowing the lowest amount of misclassification.

This suggests that in this case, we want to use a high value of C, because the data points in our sample are well-separated into super-Earths and sub-Neptunes, and we only want to use the data points close to the gap to determine its shape. Indeed, if we pick a low value of C (e.g. $C = 1$), almost all data points are used to separate the sample, and we find that this no longer fits the radius valley (a high degree of misclassification) and leads to a steep (negative) slope which no longer visually matches the observed valley. By contrast, picking a very high value for C implies the hyperplane is determined by only very few support vectors (i.e. data points nearest to the valley). For example, for $C = 100$, the hyperplane is determined by only four support vectors, i.e. two super-Earths and two sub-Neptunes, leading to $m = -0.08^{+0.02}_{-0.01}$, where the uncertainties were calculated using 1000 bootstrapped samples as before. We finally calculate the hyperplane of maximum separation using a compromise between these extremes, i.e. $C = 10$. As can be seen in Figure 7, using this value, the slope of the valley is determined by about 15 support vectors, i.e. 15 planets closest to it. This provides results consistent with the lower C value above, but with more support vectors this leads to a larger uncertainty. Again following our bootstrapping approach, we find $m = -0.09^{+0.02}_{-0.04}$ and $a = 0.37^{+0.04}_{-0.02}$.

In summary, in this section we have used different methods to determine the slope of the observed radius valley. All these methods find consistent and distinctly negative slopes. Because support vector machines provided the most standardized way of separating samples, we use these as our preferred parameters, although some readers may prefer to use one of the other methods, or calculate their own slope based on the parameters listed in Table 1.

4 DISCUSSION

We observe a bimodal distribution of planet radius, broadly peaking at $\approx$ 1.5 $R_\oplus$ and $\approx$ 2.5 $R_\oplus$, with a valley at around $1.7 - 2$ $R_\oplus$ in between. The radius valley has also been observed recently by Fulton et al. (2017). The feature we observe here is broadly similar, although the valley is more pronounced in our sample, presumably because the stellar and planetary radii are determined more accurately in this work. The Fulton et al. (2017) sample is significantly larger, enabling a determination of occurrence rates of planets for different radii and periods, which is beyond the scope of this work. By contrast, owing to a highly precise asteroseismic sample of stellar parameters, we were able to measure the slope of the radius valley as a function of orbital period for the first time, and find $m = -0.09^{+0.02}_{-0.04}$.

A large body of theoretical work predicts and interprets the existence of a planet occurrence valley as a function of planet radius and orbital period or incident flux. Even when planets form with a continuous distribution of initial radii, photoevaporation can produce a deficit of planets around 2 $R_\oplus$ (Owen & Wu 2013). In such a model, planets can either maintain hydrogen envelopes, or not, depending on their XUV exposure, creating a bimodal distribution in planet sizes. Similarly, Lopez & Fortney (2013) predict an occurrence valley with a width of roughly 0.5 $R_\oplus$, occurring at larger radii for closer-in planets.

The physical reason for a deficit or gap is that planets around this radius would have a very small envelope, which is stripped off easily, even at low mass-loss rates. The mass-loss timescale peaks when the envelope approximately doubles the radius of the planet. Planets with a smaller envelope are unstable to complete evaporation, because the mass-loss timescale decreases during evaporation. On the other hands, planets with a larger envelope see their mass-loss timescale increase as mass is removed, which stabilises when they are double the core radius. As a result, planets that resisted full photo-evaporation end up with substantial envelopes, which contribute significantly to the planet radius, and make up $\approx 1-10$% of their mass (Lopez & Fortney 2014). Meanwhile, other planets end up with virtually no envelopes at all and remain as the stripped cores.

An alternative physical process to strip the atmosphere of some planets comes from the luminosity of the cooling rocky core itself (Ginzburg et al. 2017), and would similarly produce a radius valley. Another mechanism that may explain the large diversity in mean density of short-period planets, is late giant impacts which lead to atmospheric erosion (e.g. Liu et al. 2015; Inamdar & Schlichting 2016). However, while this mechanism would influence the mass distribution of these planets, it is unclear how it could lead to a clear period valley.

Lopez & Rice (2016) investigate the possibility that the short period super-Earths are a separate population of rocky planets which never had significant envelopes, rather than stripped cores of planets that lost their envelopes. This could

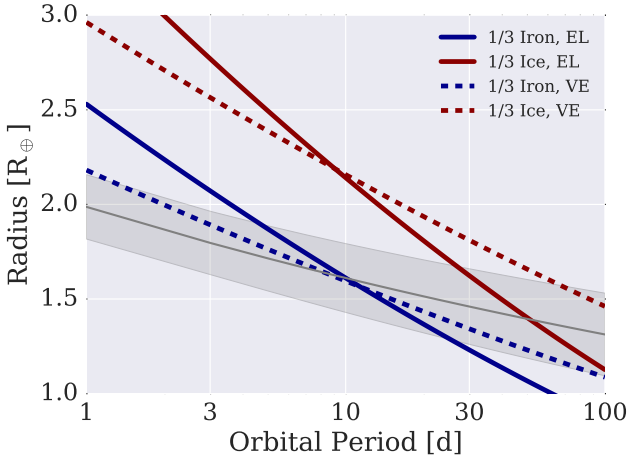

Figure 8. We compare the observed slope of the radius gap to theoretical models with different planet core compositions from Owen & Wu (2017), showing the position of the bottom of the evaporation valley, which is the largest super-Earth at a given orbital period. In grey, we show the best value and 1σ confidence interval from the support vector machine determination of the period valley, using the lower parallel line, shown in Figure 7. Different models for the bottom of the evaporation gap are shown, with solid lines showing a constant efficiency energy-limited (EL) models while dashed lines show evaporation models with variable efficiency (VE, see e.g. Owen & Jackson 2012). The thick lines show Earth-like composition cores (1/3 Iron). The thin lines show planets which consist of 1/3 ice and 2/3 silicates. We find that our observations provide the best match with Earth-like cores and a variable efficiency. We refer the reader to Owen & Wu (2017) for details about the models.

occur if these planets formed after the proto-stellar disks had already evaporated, in a similar way as to how the Earth has likely formed. Understanding whether the short-period super-Earths are the result of photo-evaporation or are primordial rocky planets is therefore important to constrain the frequency of planets like Earth in the habitable zone (Lopez & Rice 2016).

In the case of this late, gas-poor formation, the transition radius would be a function of the available solid material that a planet core can accrete due to collisions. This would result in a transit radius dependence on orbital period between $P^{0.07}$ and $P^{0.10}$, i.e. the radius valley increases with orbital period (Lopez & Rice 2016). This is in clear contrast with the photo-evaporation scenario. In that case, planets with the largest core mass are the most resistant to photo-evaporation, so that at short orbital periods, the transition radius is larger, and may scale with orbital period as $P^{-0.15}$ (Lopez & Rice 2016). Similarly, Owen & Wu (2017) find that the radius of the bottom of the valley depends on orbital period as $P^{-0.25}$ to $P^{-0.16}$, depending on the photo-evaporation model, and where the location of the valley depends on the properties of the remnant cores. Numerical models empirically give shallower slopes than analytic models for the same evaporation models, e.g. a slope of $P^{-0.12}$ is found from numerical models, for an analytical slope of $P^{-0.16}$ (Owen & Wu 2013).

The negative slope we observe here is consistent with physical models of photo-evaporation, but not with late formation, in a gas-poor environment after the disc has dissipated, which would predict a slope with a positive sign instead. The precise slope depends on the model of planet formation and the composition of the planets. In Figure 8, we compare the observed slope with different models by Owen & Wu (2017). Because the models use the maximum radius at the bottom of the valley, we compare them to the lower parallel support vector of Figure 7. We find that our slope is consistent at 2σ with the more complex models, including recombination and X-ray evaporation, and inconsistent with the steeper slope predicted for pure energy-limited evaporation (Owen & Wu 2017). Finally, it is clear from Figure 8 that the location of the photo-evaporation valley is more consistent with iron-rich cores than with icy cores. This was previously pointed out by Owen & Wu (2017) and Jin & Mordasini (2017), on the condition that the observed valley is indeed primarily caused by photo-evaporation – as the measurement of the valley's slope in this work appears to confirm.

We finally note that the presence of a clear gap in radius is evidence of largely homogeneous cores, with compositions similar to that of Earth, as a wide range of different compositions would smear out the radius gap (Owen & Wu 2017). Indeed, if sub-Neptune planets formed beyond the snowline, they would have large amounts of water and volatile ices (Rogers et al. 2011), which may completely eliminate the presence of any radius gap (Lopez & Fortney 2013). The presence of a clear gap can therefore be taken as evidence that the observed planets formed in-situ (e.g. Chiang & Laughlin 2013), which is also consistent with the observation of a desert of planets larger than 1.5 $R_\oplus$ at ultra-short periods (Lundkvist et al. 2016; Lopez 2017). The gap observed here is inconsistent with late time migration and suggests a core mass function peaking around 3 $M_\oplus$ (Owen & Wu 2017).

If photo-evaporation is indeed responsible for the observed super-Earths at short periods, this has implications for measuring the frequency of habitable zone Earth-like planets as well. Because such efforts often include planets slightly larger than Earth, or planets around later stellar types, they may include planets that are rocky only as a result of photo-evaporation, or are not rocky at all. This would result in an overestimate of the occurrence of true Earth analogs (Lopez & Rice 2016), indicating that great care must be taken when extrapolating findings of small planets at short orbital periods to more temperate Earth-sized planets.

5 CONCLUSIONS

Using a sample of planet host stars characterized with asteroseismology, we derive accurate stellar and planetary radii to investigate the presence, location and shape of a radius valley of planet occurrence. Within our sample of 117 planets, we detected a clear bimodal distribution, with super-Earth planets with radii of $\approx$ 1.5 $R_\oplus$ and sub-Neptune planets with radii of $\approx$ 2.5 $R_\oplus$, separated by a clear valley around 2 $R_\oplus$ where very few planets are observed.

- The location of the valley has a decreasing radius as a function of orbital period (see Figure 5). This negative slope is consistent with predictions for photo-evaporation, while

it is inconsistent with the exclusive late formation of gas-poor rocky planets, which would result in a slope with the opposite sign. Taking into account photo-evaporation will also be important when inferring the occurrence of Earth-like planets in the habitable zone (Lopez & Rice 2016).

- The presence of a clear valley implies a homogeneous core composition of the planets in our sample. Planets with a wide range of core compositions, or planets which have formed beyond the snow line, would wash out the valley (Owen & Wu 2017).
- When comparing the location of the valley with theoretical models, we find it to be broadly consistent with cores consisting of a significant fraction of iron, while inconsistent with mostly icy cores (Owen & Wu 2017; Jin & Mordasini 2017).

Determining the radii of planets and their host star is crucial for determining the location and shape of the radius valley. Here, asteroseismology achieves this precision (Silva Aguirre et al. 2015; Lundkvist et al. 2016). An important caveat for this approach is the limited sample size. Future transit surveys such as TESS (Ricker et al. 2014) and PLATO (Rauer et al. 2014) will lead to a larger sample with accurate parameters, and may allow to further refine the properties of the radius valley. Such a larger sample may also allow a detailed inference of the underlying occurrence rate of planets, which for now remains limited to larger but less accurately determined samples (Fulton et al. 2017).

Finally, because of the relative faintness of most stars observed by *Kepler*, no homogeneous inference of the mass of the planets in our sample is available. Future samples may allow the determination of planet mass and mean density, which would provide further tests for photo-evaporation models.

ACKNOWLEDGEMENTS

We thank Alan Heavens for discussions on support vector machines. Funding for the Stellar Astrophysics Centre is provided by The Danish National Research Foundation (Grant DNRF106). The research was supported by the ASTERISK project (ASTERoseismic Investigations with SONG and Kepler) funded by the European Research Council (Grant agreement no.: 267864). M.S.L. is supported by The Independent Research Fund Denmark's Sapere Aude program (Grant agreement no.: DFF–5051-00130). This research was made with the use of NASA's Astrophysics Data System and 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.

References

Borucki W. J., et al., 2011, ApJ, 736, 19
Chen H., Rogers L. A., 2016, ApJ, 831, 180
Chiang E., Laughlin G., 2013, MNRAS, 431, 3444
Claret A., Bloemen S., 2011, A&A, 529, A75
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Fulton B. J., et al., 2017, AJ, 154, 109
Furlan E., et al., 2017, AJ, 153, 71
Ginzburg S., Schlichting H. E., Sari R., 2017, preprint, (arXiv:1708.01621)
Goodman J., Weare J., 2010, Commun. Appl. Math. Comput. Sci., 5, 65
Huber D., et al., 2014, ApJS, 211, 2
Inamdar N. K., Schlichting H. E., 2016, ApJ, 817, L13
Jin S., Mordasini C., 2017, preprint, (arXiv:1706.00251)
Jin S., Mordasini C., Parmentier V., van Boekel R., Henning T., Ji J., 2014, ApJ, 795, 65
Johnson J. A., et al., 2017, AJ, 154, 108
Liu S.-F., Hori Y., Lin D. N. C., Asphaug E., 2015, ApJ, 812, 164
Lopez E. D., 2017, MNRAS, 472, 245
Lopez E. D., Fortney J. J., 2013, ApJ, 776, 2
Lopez E. D., Fortney J. J., 2014, ApJ, 792, 1
Lopez E. D., Rice K., 2016, preprint, (arXiv:1610.09390)
Lundkvist M. S., et al., 2016, Nature Communications, 7, 11201
Mandel K., Agol E., 2002, ApJ, 580, L171
Morton T. D., Bryson S. T., Coughlin J. L., Rowe J. F., Ravichandran G., Petigura E. A., Haas M. R., Batalha N. M., 2016, ApJ, 822, 86
Owen J. E., Jackson A. P., 2012, MNRAS, 425, 2931
Owen J. E., Wu Y., 2013, ApJ, 775, 105
Owen J. E., Wu Y., 2017, preprint, (arXiv:1705.10810)
Petigura E. A., et al., 2017, AJ, 154, 107
Rauer H., et al., 2014, Experimental Astronomy,
Ricker G. R., et al., 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. p. 20 (arXiv:1406.0151), doi:10.1117/12.2063489
Rogers L. A., Bodenheimer P., Lissauer J. J., Seager S., 2011, ApJ, 738, 59
Silva Aguirre V., et al., 2015, MNRAS, 452, 2127
Smith J. C., et al., 2012, PASP, 124, 1000
Van Eylen V., Albrecht S., 2015, ApJ, 808, 126
Van Eylen V., Albrecht S., et al. 2017, in prep.

Planet	$R_\star$ [$R_\odot$]	$M_\star$ [$M_\odot$]	$T_{\rm eff}$ [K]	Period [d]	R_p [$R_\oplus$]
Kepler-10b	1.0662 ± -0.0075	0.900 ± 0.030	5678 ± -50	0.83749026(29)	1.473 ± 0.026
Kepler-10c	1.0662 ± -0.0075	0.900 ± 0.030	5678 ± -50	45.294292(97)	2.323 ± 0.028
Kepler-23b	1.548 ± -0.048	1.00 ± 0.15	5828 ± -100	7.106995(73)	1.694 ± 0.076
Kepler-23c	1.548 ± -0.048	1.00 ± 0.15	5828 ± -100	10.742435(39)	3.12 ± 0.10
Kepler-23d	1.548 ± -0.048	1.00 ± 0.15	5828 ± -100	15.27430(16)	2.235 ± 0.088
Kepler-25b	1.299 ± -0.016	1.110 ± 0.090	6315 ± -70	6.2385369(33)	2.702 ± 0.037
Kepler-25c	1.299 ± -0.016	1.110 ± 0.090	6315 ± -70	12.7203678(35)	5.154 ± 0.060
Kepler-37b	0.7725 ± -0.0063	0.800 ± 0.030	5430 ± -50	13.36805(38)	0.354 ± 0.014
Kepler-37c	0.7725 ± -0.0063	0.800 ± 0.030	5430 ± -50	21.30207(92)	0.705 ± 0.012
Kepler-37d	0.7725 ± -0.0063	0.800 ± 0.030	5430 ± -50	39.792231(15)	1.922 ± 0.024
Kepler-65b	1.401 ± -0.014	1.169 ± 0.060	6193 ± -50	2.1549156(25)	1.409 ± 0.017
Kepler-65c	1.401 ± -0.014	1.169 ± 0.060	6193 ± -50	5.8599408(23)	2.571 ± 0.033
Kepler-65d	1.401 ± -0.014	1.169 ± 0.060	6193 ± -50	8.131231(21)	1.506 ± 0.040

Table 1. Stellar and planetary parameters of the objects in our sample. Stellar parameters were taken from Silva Aguirre et al. (2015) and Lundkvist et al. (2016), and planet parameters from Van Eylen & Albrecht (2015) and Van Eylen et al. (2017). A full version of this table is available online.