Studying The Factors That Lead to Intra-System Diversity in Kepler Exoplanet Sizes

Mason Takeuchi

Received April 11, 2025
Accepted September 05, 2025
Electronic access October 31, 2025

Sub-Neptune-sized planets are a key population to study to better understand and refine theories on planet formation. Previous research has shown that these planets typically reside in multi-planet systems, and planets within the same system show remarkable uniformity in size, spacing, eccentricity, and inclination. In this work, we explore how the properties of planets in multi-planet systems depend on those of their host star, which is one of the best indicators we have for formation conditions. We survey the Kepler planets to assess correlations between the intra-system masses and radii, and properties of their host star. We measure tentative linear trends; for example, the fitted slope for the dispersion mass to stellar mass relationship was -0.1471. Though tentative, most of these trends agree with previous results. Our conclusions were limited by the precision of the data for the masses and radii of planets, therefore there is a need for more precise data as well as a more comprehensive sample in order to establish clear links between stellar properties and planetary formation processes.

Introduction

Exoplanets, or extra-solar planets, are planets that exist outside of the Solar System. Before 1992, scientists had not confirmed the presence of any planets outside of the Solar System[1]. Now scientists know there are thousands of these planets[2,3], largely due to efforts like the Kepler mission. The Kepler mission was a space telescope launched in 2009 that was designed to point at a specific region of the sky to detect Earth-sized planets[4]. The mission revealed more than 2,700 planets around the surveyed stars alone, revealing that there could be billions of planets in the entire Milky Way galaxy[2]. The telescope used a popular planet detection method called the transit method to find such planets, which involves looking at a star until a planet transits, or crosses, between it and the telescope. This leads to a certain amount of stellar light being blocked, which, when coupled with the properties of the star, can be analyzed to determine the relative size of the planet transiting[3]. Most interestingly, the mission discovered that more than half of sun-like stars have planets with tight orbits that are in between the size of Earth and Neptune[5], known as the "small- planet population." It is also important to note that there may be bias for intra-system comparison among the Kepler dataset due to the methods the telescope uses to make such detections. For example, because Kepler uses the transit method, which requires a planet to pass in between the star and the viewer, this could introduce bias towards systems with planets sharing mutual high inclinations.

Among other methods to characterize exoplanets is the radial velocity method, in which astronomers look for the wobble of the host star due to the gravitational pull of its planets, appearing as the star moving towards and away from the observer[6]. Due to the Doppler effect, the stellar spectrum shifts during the course of one orbit of the planet, tracing a sinusoidal curve with an amplitude corresponding to parameters such as planetary mass. The resulting movement of the star is given by

$$V\ [\mathrm{m\,s^{-1}}] = 28.4 \left(\frac{P}{1\ \mathrm{yr}}\right)^{-1/3} \left(\frac{M_p \sin i}{M_{\mathrm{Jup}}}\right) \left(\frac{M_*}{M_\odot}\right)^{-2/3} \quad (1)$$

The observed stellar spectra correspond to a lower uncertainty in derived parameters. Additionally, this method is biased towards systems with higher inclinations.

Another method to measure planet masses is the transit timing variation (TTV) technique[7]. Observers look for the slight change in the exact mid-transit time within photometric light curves because of the presence of another planet in the system. This method can only be used to determine the mass of the perturbing planet, making multi-planet systems the only subject of this method.

Small planets are a class of planets characterized by radii smaller than that of Neptune and larger than that of Earth[8]. Many small planets exist in multi-planet systems with high uniformity of the intra-system masses, radii, eccentricity, and spacing[9].

There is a particularly compelling science case for studying sub-Neptune planets (planets with radii between 2 and 4 $M_\oplus$) because none exist within our own Solar System. These types of planets are also the most common types of planets detected within 1 Astronomical Unit (AU) of their host stars[10]; it is imperative that we study these samples to refine formation theories. Notably, many multi-planet systems with several sub-Neptune

planets demonstrate uniformity in terms of their masses, radii[11], and eccentricities[12]. Determining the cause of this uniformity could reveal more about the formation of our own Solar System and the factors that affect the formation process during the early stages of a stellar system's life.

Another salient feature of the small planet class is the bimodality of the radius distribution of the Kepler planets, with a lack of planets between 1.5 and $2R_\oplus$. This splits the small planet category in two - the smaller are known as super-Earths ($R_p < 1.5\,R_\oplus$), while the larger planets ($R_p > 2\,R_\oplus$) are known as sub-Neptunes[9]. Planets within the range of $1.5\,R_\oplus < R_p < 2\,R_\oplus$ are relatively rare, which is the reason for the existence of the radius gap. Several theories suggest that the radius bimodality forms because of atmospheric mass loss[13]. One theory proposes that planets in the early stages of their lives can be subject to photoevaporation depending on the activity of their star - if the planet does not have sufficient mass, it is more likely to lose a substantial part of its atmosphere to stellar radiation[14]. Another theory puts forward core-powered mass loss, in which the residual heat from formation coming from a planet itself drives atmospheric escape as the thermal energy of the atmospheric particles exceeds the gravitational potential energy[15]. However, core-powered heating is theorized to act on longer time scales compared to photoevaporation wherein the stellar activity and XUV output is concentrated in the first hundred Myr[14].

A wide range of types of planets have been documented. From large, Jupiter-sized planets to small, terrestrial planets, there are many archetypes that are represented across available data sets. Many discovered planets fall into the super-Earth to sub-Neptune range and as mentioned above, many sub-Neptune planets exist in multi-planet systems[9]. In the standard core accretion model of planetary formation dust particles collide and form planetesimals within the protoplanetary disk that surrounds the star. Over time, planetesimals also collide and form planetary cores. Once they reach this stage, they have enough gravity of their own to attract gas to create an envelope around the core, resulting in gas planets[16]. The use of categorization is helpful when thinking about how different planets form. For example, they can help answer how giant planets close to the star form compared to earth-sized planets that form much further away. I plan to use established datasets to compare properties of planets in multi-planet systems, such as mass and radius, while looking for similarities and trends between them. The Kepler mission data is well suited to this type of research because it used long observation times and thus detected many multi-planet systems. I plan to study multi-planet systems because comparisons of planet masses and radii within the same system and formation conditions provide a useful control variable. Another benefit to using data from the Kepler mission is that because the Kepler mission produced a large sample from exoplanet detections, and because the mission was well characterized, it is possible to account for detection biases[9]. Several key studies have investigated the relationships of certain parameters among planets of the same system, such as Weiss et al. 2022 and Millholland et al. 2017. Weiss et al. primarily study the uniform architecture of sub-Neptune systems, and how changes from this uniformity might reveal trends in planetary formation processes. They conclude that while compact multi-planet systems form with a high degree of similarity, more observational data must be collected, and a more predictive model of planet formation must be found to gain more insight into why planets occur so uniformly. Millholland et al. investigate trends among sub-Neptune planets as well, but dive more specifically into mass and radius relations, coming to the conclusion that planets within a system have more similar masses and radii than a random selection of planets would if they were placed in the same system.

Planet parameters compiled in the NASA Exoplanet Archive contain summaries from different independent research teams using different analysis methods. Most have different evaluations of the size of error bars. Therefore, using consistent methods to measure parameters is important when evaluating and comparing error bars and trends.

Since the compilation of some of these papers, hundreds more exoplanets have been discovered, characterized, and added to databases available online. Alongside the main goal of searching for intra-system trends among small planets regarding radii, masses, and eccentricity when larger planets are present, this research also seeks to use updated data and revises trends of papers published more than six to seven years ago. By using similar graphs and metrics, I hope to either support their conclusions or show that some conclusions are less robust.

Data Summary & Methodology

The NASA Exoplanet Archive, which is regularly updated to contain all confirmed exoplanets, includes discoveries by missions like TESS, Kepler, KELT, and more. This research project uses the Kepler data because the Kepler mission observed stars for extended periods of time, allowing more planets with longer orbital periods to be detected and allowing the detection of multiple planets from the same system. However, measurements are most often from a collection of different data and analyses. We sample the PSCompPars[?] table [one composite row per planet with a combination of calculated values from different studies] of the exoplanet archive using astroquery[17], which provides a row of information for each planet. Other packages used in this project were pandas[18], matplotlib[19], NumPy[20], and SciPy[21]. Hadden & Lithwick 2017 (henceforth HL2017) used TTVs to measure the masses of planets within multi-planet systems and includes radius, mass, and period information, as well as respective error bounds for each property. Their dataset contains information from 145 planets in 55 systems detected by Kepler[22].

An issue we faced by using the NASA Exoplanet Archive

was that each planet's data (specifically the mass and radius) was often measured by different teams, using different methods, which were then selected randomly for use from the PSCompPars table. To partially combat this, and because all HL2017 values are already included in the Exoplanet Archive, we manually select many of the mass values of planets from HL2017 for this project, since it offers uniform calculations of mass values.

Data Filtering

During import using the astroquery package, mass values are filtered to only include planets $< 100 M_{\oplus}$, and to only include those planets in Kepler systems. For evaluations of the metrics described below, we use a sub-sample of planets with mass and radii constrained to better than 50%, since values with high uncertainty make it difficult to discern clear trends; we name this the 'constrained sample' (shown in blue, Fig. 1). The number of planets in the constrained sample changes when different constraints are placed on the uncertainty threshold. For example, when the sample is limited to planets with mass and radius uncertainties lower than 30%, the number of planets in the sample drops from 199 to 47. At 70%, the number of planets in the sample is 72 and at 50% there are 66 planets. A threshold of 50% allows us to obtain a moderately constrained threshold without excessively reducing the number of usable planets (see e.g., Otegi et al., 2019)[23]. For completeness, we also quote values for the whole sample of planets, the 'unconstrained sample'. Part of the filtering process also involves removing single planets from the data because the main aim of the project is to look at relationships among multi-planet systems. While this shifts the project away from detailed investigations of single planets and their properties, analysis of multi-planet systems can shed light on formation processes and trends and reveal more about general trends between systems.

Graphing data

In this project, we only use the "default" mass from the HL2017 data when manually selecting for the masses of the exoplanet data. HL2017 tested two different priors when measuring the mass (a 'default' logarithmic prior and a 'high mass' uniform prior) and associated eccentricity, to test for the sensitivity of the measurement to the prior. HL2017 showed that in some cases the use of a lower versus a higher mass prior led to significant differences in the measured mass. Therefore we only manually select for the mass in our table using the default mass measurements from HL2017 if their two measurements agree within 2 sigma. Further, we only use the mass from HL2017 in place of the NASA data if it has a more constrained mass measurement than the data the exoplanet archive randomly selects. Mean error bars for both masses were then calculated from HL2017 data for low and high error bounds.

An important aspect of the study was defining an intra-system dispersion, Δ, in the masses and separately the radii of planets within a given system. In this study, we use the first equation shown in Millholland et al. 2017 and modify it slightly to calculate Δ for each system, as given by

$$\Delta_M = \frac{1}{N_{\rm pl} - 1} \sum_{i=1}^{L-N_{\rm pl}} \log \frac{M_{i+1}}{M_i}, \tag{2}$$

$$\Delta_R = \frac{1}{N_{\rm pl} - 1} \sum_{i=1}^{L-N_{\rm pl}} \log \frac{R_{i+1}}{R_i}. \tag{3}$$

where Δ_M is the mass dispersion, Δ_R is the radius dispersion, M_p is the planetary mass, R_p the planetary radius, and $N_{\rm pl}$ is the number of planets in the given system. We use a Monte Carlo propagation to calculate the error propagation for both Δ_M and Δ_R using a distribution of $N = 10000$. For each system's Δ calculation, planets are ordered by orbital period, following the methodology of Millholland et al. (2017)[11]. We evaluate the metrics given in Equation 2 and 3 for our sample of systems, which are shown in Fig. 1.

Results & Discussion

We compute statistical metrics to confirm whether these apparent correlations are numerically supported through the use of the Spearman's rank correlation coefficient (ρ). A positive coefficient represents positive correlation, and a negative coefficient represents a negative correlation. Correlation strength scales from 0 to 1, with numbers close to 0 representing virtually no correlation, and numbers closer to 1 representing extremely strong correlations. Additionally, each coefficient value is accompanied by a p-value which represents the probability that a trend appearing in the data is due to chance. A small p-value means that any resulting trend would be unlikely if the sample had a random distribution. The spearmanr function from the scipy.stats package is used to calculate both coefficient and p-values[21]. In Fig. 2 we show the relationship between Δ_M and Δ_R among systems in the data. Otegi et al. 2022 sample the NASA Exoplanet Archive for planets with mass uncertainty smaller than 50% and radius uncertainty smaller than 16%, concluding that these planets have a radius similarity three times stronger than that of mass. Otegi et al. 2022 finds that there is a "weak correlation between the similarities in mass and radii" for mass versus radius correlations. Through analysis of the constrained planets data set, we find a moderate positive correlation between Δ_M and Δ_R, with a Spearman ρ coefficient of 0.4717 (p-value = 0.0001, see Table 1), which somewhat agrees with the conclusion of Otegi et al. 2022. They also demonstrate that limiting the sample to lower mass planets yields an increase in correlation larger than that expected due to the smaller sample size; as

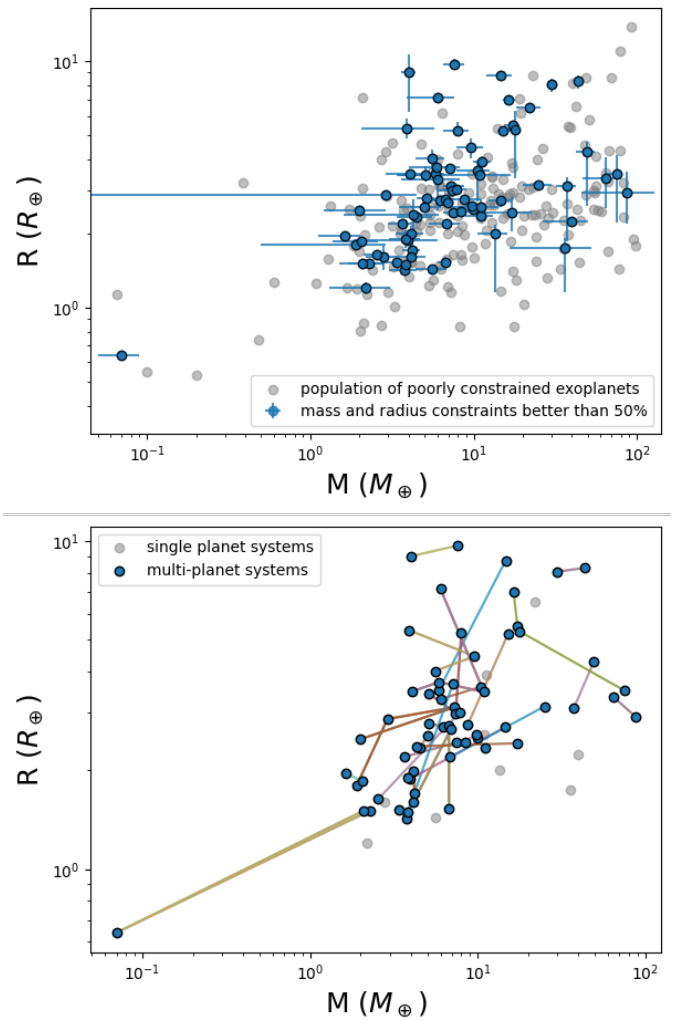

Fig. 1 A visualization of the exoplanet samples analyzed in this work. Top: Blue points are planets with masses and radii uncertainties constrained to better than 50%, which form the primary sample of this study. The points in gray are planets that have a greater than 50% uncertainty. Bottom: Blue points are planets in systems with more than one planet from the constrained sample. Lines of the same color connect planets in the same system. Gray points are single planet systems and are dropped from the data. Note the use of log scaling for both axes.

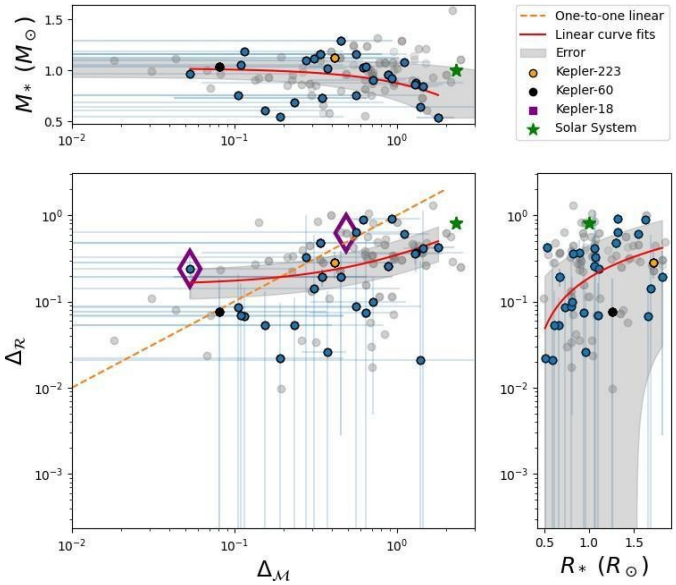

Fig. 2 Δ_M versus the Δ_R of all the systems. Gray points are systems which include planets from the unconstrained sample, included for reference, while blue and colored points are systems with planets from the constrained sample. The solar system is indicated with a green star for reference, while the black and orange points represent Kepler-60 and Kepler-223, respectively. The orange dashed line is a one-to-one linear relationship between Δ_M and Δ_R. The solid red lines show linear fits for the constrained sample (the linear fits appear curved due to the log-scaling of the axes).

the number of pairs they use drops from 58 to 46, the p-value increases from 4×10^{-10} to 3×10^{-4}.

We are also interested in investigating whether stellar properties like stellar density (ρ_*) or metallicity have an impact on planet formation. The constrained sample seems to support the possibility that the Δ_M of a system may be correlated with the parent star's mass, density, and possibly the metallicity. Δ_M has a tentative correlation (Spearman $\rho = -0.2502$, p-value $= 0.0427$) when compared to stellar mass (M_*), while Δ_R supposedly has a moderate positive correlation (Spearman $\rho = 0.4111$ p-value = 0.0006) when compared to stellar radius (R_*). Error bars shown in Fig. 2 are calculated by using the Monte-Carlo error propagation introduced after Equation 2 and 3.

Whole Sample Analysis

System-System Relations

We fit linear relationships between x and y variables using the scipy curve fit functionality[21]. In Fig. 1, the lines in red represent optimized curve fits for each relationship.

The slope of the line in Δ_M vs. Δ_R is 0.1918 ± 0.0566. The slope of the line in Δ_M vs. M_* is -0.1471 ± 0.0512. The slope of the line in Δ_R vs. R_* is is 0.5930 ± 0.1948. If the correlation were strong, Δ_M could be predicted from Δ_R and vice versa. A large Δ_M in a system would imply a strong Δ_R. But in reality, the correlation is not particularly strong. In fact, the general shift below the one-to-one line (dotted orange line, Fig. 1) implies that on average, the Δ_M of a system is higher than the Δ_R of a system. Studies find that planets in the same system have more similar radii than would be expected by chance, and this effect is stronger for radius than it is for mass, where masses in the same system tend to vary more widely[11]. This suggests that

Table 1 Spearman ρ correlation coefficients between intra-system Δ's and host star properties, for the constrained, unconstrained, and sub-Neptune-only target samples, with p-values and 95% confidence intervals for each respective relationship. Δ_M is intra-system mass dispersion, Δ_R is intra-system radius dispersion, and M_*, R_*, ρ_*, and [Fe/H] are the stellar mass, stellar radius, stellar density, and metallicity respectively. The sample size of the constrained sample is 199, the sample size of the constrained sample is 66, and the sample size of the sub-Neptune sample is 35.

Exoplanet sample	Independent variable	Dependent variable	Spearman ρ	p-value	95% CI
Constrained	Δ_M	Δ_R	0.4717	0.0001	[0.2493, 0.6519]
Constrained	Δ_M	$M_*(M_\odot)$	-0.2502	0.0427	[-0.4821,0.00002]
Constrained	Δ_R	$M_*(M_\odot)$	0.2731	0.0265	[0.0123, 0.5018]
Constrained	Δ_R	$R_*(R_\odot)$	0.4111	0.0006	[0.1696 0.6141]
Constrained	Δ_M	ρ_* $(\mathrm{g\,cm^{-3}})$	0.2405	0.0518	[-0.0130 0.4694]
Constrained	Δ_M	[Fe/H](dex)	0.0899	0.4727	[-0.2085 0.3932]
Unconstrained	Δ_M	Δ_R	0.4115	0.0001	[0.2840, 0.5255]
Unconstrained	Δ_M	$M_*(M_\odot)$	-0.0053	0.9409	[-0.1609, 0.1302]
Unconstrained	Δ_R	$M_*(M_\odot)$	0.1558	0.028	[-0.0260, 0.2577]
Unconstrained	Δ_R	$R_*(R_\odot)$	0.1422	0.0451	[-0.0543, 0.2335]
Unconstrained	Δ_M	ρ_* $(\mathrm{g\,cm^{-3}})$	0.0905	0.2038	[-0.0458, 0.2319]
Unconstrained	Δ_M	[Fe/H](dex)	0.1465	0.0389	[0.0003, 0.2853]
Sub-Neptune	Δ_M	Δ_R	0.3776	0.0253	[-0.0058, 0.7047]
Sub-Neptune	Δ_M	$M_*(M_\odot)$	-0.0355	0.8394	[-0.4090, 0.2957]
Sub-Neptune	Δ_R	$M_*(M_\odot)$	0.6642	0.0001	[0.3943, 0.8310]
Sub-Neptune	Δ_R	$R_*(R_\odot)$	0.7023	0.0001	[0.4508, 0.8723]
Sub-Neptune	Δ_M	ρ_* $(\mathrm{g\,cm^{-3}})$	0.058	0.7407	[-0.2904, 0.4202]
Sub-Neptune	Δ_M	[Fe/H](dex)	0.389	0.0209	[-0.0091, 0.7144]

planets can form in many different ways where the mass and radius variability among planets in a system are not necessarily tied together.

Table 2 Linear coefficients between intra-system $\Delta's$ and host star properties for the constrained target sample, with uncertainties for each respective relationship. Δ_M is intra-system mass dispersion, Δ_R is intra-system radius dispersion, and M_* and R_* are the stellar mass and radius respectively.

Exoplanet sample	Independent variable	Dependent variable	Linear Fit Slope
Constrained	Δ_M	Δ_R	0.1918 $\pm$ 0.0566
Constrained	Δ_M	M_*	-0.1471 $\pm$ 0.0512
Constrained	Δ_R	R_*	0.5930 $\pm$ 0.1948

Planet-Star Relations

Interestingly, the relationship between Δ_M and M_* has a weak correlation coefficient of -0.2502 among constrained planets with a low error (p-value of 0.0427, see Table 1). Additionally, Δ_R for the constrained sample has a strong correlation when compared to R_* (Spearman ρ = 0.4111, p-value = 0.0006) and a weaker, but still statistically significant correlation when compared to M_* (Spearman ρ = 0.2731, p-value = 0.0265) (see Table 1). The presence of a correlation in the constrained sample between M_* and Δ_M is relatively unexpected, as stellar properties are not thought to have a large impact on the characteristics of orbiting small planets, including Δ_M [9].

Analysis of specific planetary systems

The one-to-one line represents Δ_M's and Δ_R's that are the same. If density is uniform, systems are expected to lie on a $\Delta_M = 3 \times \Delta_R$ line [24]. The system with the smallest Δ_M is Kepler-18 (indicated in a purple box, Fig. 2, $N_{\mathrm{pl}} = 3$), a system of three planets orbiting a G-type star of $R_* = 0.923 R_\odot$ [25]. The Kepler-18 system is notable because its Δ_R value of $\Delta_R = 0.2401$ is significantly different from its Δ_M value of $\Delta_M = 0.0534$. This difference is especially apparent when comparing this system to the system with the closest Δ_M value to Kepler-18 which lies on the 1-to-1 line, Kepler-60 (black point in Fig. 2, $N_{\mathrm{pl}} = 3$). Both Kepler-60 and Kepler-18 are G-type stars, and while Kepler-60 has 3 planets, Kepler-18 has 2 (see 3.3). While these

systems are similar in terms of Δ_M, something during their respective formation processes has caused them to form with a lower Δ_R than Kepler-18. We discuss possible reasons for the Δ discrepancy in the case of Kepler-18:

1 Stellar Density (ρ_*) is$1.0100 g/cm^3$ for Kepler-18, but is $0.3400 g/cm^3$ for Kepler-60. We can see a visual positive correlation and a weaker numerical Spearman ρ correlation coefficient of 0.2405 when Δ_M is compared to ρ_* in the constrained sample (p-value = 0.0518, Table 1).

   Essentially this means that as ρ_* increases, the Δ_M of planets in that system typically increases as well. While ρ_* typically increases with Δ_M, stellar metallicity may not be similarly linked. Testing stellar metallicity from the NASA Exoplanet Archive versus Δ_M yields a Spearman ρ value of 0.0899 with a p-value of 0.4727 for the constrained sample (Table 1). Owen & Murray-Clay 2018 found that stars with a higher metallicity create systems containing planets with larger rocky cores[26]. This could broaden the range of potential planetary masses and potentially lead to an increase the Δ in a system. However, we find little correlation, which may mean stellar metallicity does not have an impact on the mass differences of surrounding planets, or alternatively the data are insufficient to test this hypothesis.

2 The difference between Kepler-18 and its neighbor on the graph could also be attributed to the period ratio, which is calculated by taking the difference between the outer planet's period and the inner planet's period, divided by the inner planet's period. The result is then divided by R^* to normalize it, as given by:

$$\delta P = \frac{|P_{i+1} - P_i|}{R_*} \tag{4}$$

   where P is the period of each planet and R^* is stellar radius. Kepler-18 has a relative δ_P of 0.8524 while Kepler-60 has a relative δ_P of 0.2658. However, this difference is smaller than the differences between the aforementioned ρ_* values, and plotting Δ_M against δ_P results in no statistically significant correlations, assessed with the Spearman ρ coefficient.

In either case, something different has happened in this system that is deviating from formation theories. Weiss et al. 2022 discuss how uniform planetary radii in a system can indicate uniform planetary mass, and that mass uniformity is a fundamental artifact of the planet formation process. These planets can form in different ways, most likely due to variations in the formation conditions and the properties of their host star. The study of these stellar parameters is important because they can be indicative of the formation conditions of the system early in its lifetime. While we have data from protoplanetary disks from programs such as ALMA, which exceeds at detecting protoplanetary disks around young stars, for the majority of exoplanets we study, protoplanetary disks in the system have long since dispersed. Therefore, the study of planet atmospheres as well as the star itself is crucial for inferring information about the conditions and environment of planet formation around the star.

We also look deeper at trends in Δ_M compared to stellar composition (metallicity, ρ_*, and M_*) for sub-Neptune planets in the Kepler dataset, by looking only at planets in the data between 2 and 4 $R_\oplus$. This sample is constructed of any system which contains a sub-Neptune planet, regardless of the other types of planets in that system. Major differences from the entire population include a weak and statistically insignificant correlation between Δ_R and M_* of -0.0355 (p-value = 0.8394) and a stronger correlation between and Δ_R and R_* of 0.7023 (p-value = 0.0001) for sub-Neptune planets between 2 and 4 $R_\oplus$ (see Table 1). This is relatively unexpected as previous studies show that there is little to no correlation between variation of planet sizes within a planetary system and host star mass or radius[9]. Additionally, we note that because we narrow down the sample of sub-Neptune planets to 35, this study lacks the power to confirm weak correlations; results are exploratory[27].

We analyze the systems that fall into the super-Earth category (between 1 and 1.5 $R_\oplus$); only 2 planets in the constrained sample fall into this range: Kepler-128 b and Kepler-36 b. Kepler-128 b has a Δ_M and Δ_R of 0.1145 and 0.0680. The Kepler-128 system also contains a planet c which, with a adius of $R_p = 1.521 R_\oplus$, sits on the border between being a super-Earth and sub-Neptune. Kepler-36 b has a Δ_M and Δ_R of 0.6214 and 0.8985. The Kepler-36 system has another planet that falls into the sub-Neptune category ($R_p = 3.679 R_\oplus$). Despite these differences, neither system is extremely different from other systems in the Δ_M-Δ_R parameter space, and they remain consistent with the overall trends in Fig. 2

Assumptions

If we were to use planets with higher than a 50% error (the unconstrained version of the data), the inferred trend has the potential to change dramatically (see Table 1). As an example of this, the Kepler-18 system has an inner planet, Kepler-18 b, dropped because its uncertainty exceeds 50%. With the mass and radius reincorporated into the Δ calculations, the position of the point in the main panel of Fig. 2 moves from the rectangular purple point on the left in the constrained version of the data to the rectangular purple point in the upper right (in grey since this is in the unconstrained sample). As shown here, the tight constraints used for the filtering of the main sample can result in a constrained sample with distinct evaluations of the metrics Equation 2 and Equation 3, thereby changing the position of the systems in the parameter space of Fig. 2 and the overall trend. On the other hand, there are systems such as Kepler-223, which

also has a dropped planet. In this case, however, the planet has a very similar mass and radius to others in the constrained sample, so its presence or absence does not change the position of the system in the graph much. In summary, we assume that systems will move within the Δ_M-Δ_R parameter space depending on the constraints imposed on the sample selection.

However, we accept this because dropping high uncertainty planets from certain systems improves the robustness of our measurements overall.

Conclusion

In this study, we attempted to find whether stellar parameters affect the masses and radii of planets around those stars. We used the Δ parameter to study the distribution of planets in multi-planet systems in the Kepler dataset. We found correlations in agreement with previous studies. However, it is still uncertain how exactly the stellar parameters (M_*, R_*, ρ_*) contribute to the large difference between Δ_M and Δ_R for planets in Kepler systems.

Further study of systems that have large Δ values in mass and radius, as well as more accurate measurements for both could shed light on the planet formation process and what leads to systems forming in a "peas-in-a-pod" structure versus a less uniform pattern with large differences in planet sizes. As demonstrated by Weiss et al. 2022, it is possible that a difference in mass-determining mechanisms within a system could be the cause behind the differences apparent in planets in the system. As it is difficult to show these formation mechanisms through mass and radius data alone, it is also difficult to demonstrate which mechanisms specifically can lead to variation in planet sizes. We acknowledge that the use of the Kepler dataset in this project could lead to bias towards planets that orbit nearer to their host star or have high inclinations. Therefore our conclusions are only valid for the types of small planets discussed within this paper. Further, advances in the field to achieve better observational data such as gaining mass measurements for planets missing them, as well as improving uncertainties for mass and radius values- would tighten the constraints on overall trends, helping to better understand the formation and evolution of planetary systems.

Based on our current data, we can conclude that stellar properties do have an effect on surrounding planets. Using the Δ metric we found weak but statistically significant evidence supporting previous conclusions that the properties of stars seem to have an affect on the properties of their planets[24]. Trends with Δ_R were largely the same between our paper and others, finding a weak to moderate positive correlation. Specifically, Otegi et al. (2022)[24] concluded that higher mass stars tended to have a higher Δ_R. Most trends with Δ_M were similar to previous studies as well. However, our correlation between Δ_M and M_* was negative, in contrast to what other studies have found. In Otegi et al. (2022), a sample size of 144 planets and 48 systems is used. We use a smaller sample size of 66 planets and 28 systems. In their methodology, Otegi et al. (2022) use a moving sample analysis, which is not used in this paper. Both of these could be a source of differing results between our conclusions and those of Otegi et al. (2022)[24]. Differences in stellar properties like metallicity may have an effect on the resulting mass and Δ_R's of planets in a system. However, more data points and more accurate measurements of stellar properties are needed since we rely on both for actual findings. To reiterate once again, the trends we explore can vary significantly depending on the mass and radius measurements used, highlighting the importance of well-constrained parameters. Further study of stellar properties such as well-constrained M_*, R_*, and ρ_*-through stellar spectra are imperative for us to fully understand the nature of the surrounding planets.

References

1 D. Wolszczan and D. Frail, *A planetary system around the millisecond pulsar PSR1257+12*, 1992.

2 Caltech/IPAC, *NASA Exoplanet Archive*, `https://exoplanetarchive.ipac.caltech.edu`, 2025, Accessed 2025.

3 H. J. Deeg and R. Alonso, *Transit Photometry as an Exoplanet Discovery Method*, 2018.

4 W. J. Borucki, D. Koch, G. Basri, N. Batalha, T. Brown, D. Caldwell, J. Caldwell, J. Christensen-Dalsgaard, W. D. Cochran, E. Devore, E. W. Dunham, A. K. Dupree, T. N. G. III, J. C. Geary, R. Gilliland, A. Gould, S. B. Howell, J. M. Jenkins, Y. Kondo, D. W. Latham, G. W. Marcy, S. Meibom, H. Kjeldsen, J. J. Lissauer, D. G. Monet, D. Morrison, D. Sasselov, J. Tarter, A. Boss, D. Brownlee, T. Owen, D. Buzasi, D. Charbonneau, L. Doyle, J. Fortney, E. B. Ford, M. J. Holman, S. Seager, J. H. Steffen, W. F. Welsh, J. Rowe, H. Anderson, L. Buchhave, D. Ciardi, L. Walkowicz, W. Sherry, E. Horch, H. Isaacson, M. E. Everett, D. Fischer, G. Torres, J. A. Johnson, M. Endl, P. Macqueen, S. T. Bryson, J. Dotson, M. Haas, J. Kolodziejczak, J. V. Cleve, H. Chandrasekaran, J. D. Twicken, E. V. Quintana, B. D. Clarke, C. Allen, J. Li, H. Wu, P. Tenenbaum, E. Verner, F. Bruhweiler, J. Barnes and A. Prsa, *Kepler Planet-Detection Mission: Introduction and First Results*, 2010.

5 J. Fulton, E. A. Petigura, A. W. Howard, H. Isaacson, G. W. Marcy, P. A. Cargile, L. Hebb, L. M. Weiss, J. A. Johnson, T. D. Morton, E. Sinukoff, I. J. M. Crossfield and L. A. Hirsch, *The California-Kepler Survey. III. A Gap in the Radius Distribution of Small Planets*, 2017.

6 V. Bozza, L. Mancini and A. Sozzetti, *Methods of Detecting Exoplanets: 1st Advanced School on Exoplanetary Science*, 2016.

7 J. Miralda-Escudé, *Orbital Perturbations of Transiting Planets: A Possible Method to Measure Stellar Quadrupoles and to Detect Earth-Mass Planets*, 2002.

8 L. Zeng, S. B. Jacobsen, D. D. Sasselov, M. I. Petaev, A. Vanderburg, M. Lopez-Morales, J. Perez-Mercader, T. R. Mattsson, G. Li, M. Z. Heising, A. S. Bonomo, M. Damasso, T. A. Berger, H. Cao, A. Levi and R. D. Wordsworth, *Growth model interpretation of planet size distribution*, 2019.

9 L. M. Weiss, S. C. Millholland, E. A. Petigura, F. C. Adams, K. Batygin, A. M. Bloch and C. Mordasini, *Protostars and Planets VII*, 2022.

10 E. A. Petigura, G. W. Marcy, J. N. Winn, L. M. Weiss, B. J. Fulton, A. W. Howard, E. Sinukoff, H. Isaacson, T. D. Morton and J. A. Johnson, *The California-Kepler Survey. IV. Metal-rich Stars Host a Greater Diversity of Planets*, 2018.

11 S. Millholland, S. Wang and G. Laughlin, *Kepler Multi-planet Systems Exhibit Unexpected Intra-system Uniformity in Mass and Radius*, 2017.

12 V. V. Eylen and S. Albrecht, *Eccentricity From Transit Photometry: Small Planets in Kepler Multi-Planet Systems Have Low Eccentricities*, 2015.

13 J. L. Bean, S. N. Raymond and J. E. Owen, *The nature and origins of sub-Neptune size planets*, 2020.

14 J. E. Owen and Y. Wu, *Kepler Planets: A Tale of Evaporation*, 2013.

15 S. Ginzburg, H. E. Schlichting and R. Sari, *Core-powered mass-loss and the radius distribution of small exoplanets*, 2018.

16 J. B. Pollack, O. Hubickyj, P. Bodenheimer, J. J. Lissauer, M. Podolak and Y. Greenzweig, *Formation of the Giant Planets by Concurrent Accretion of Solids and Gas*, 1996.

17 Astroquery Project, *Astroquery Documentation*, `https://astroquery.readthedocs.io/en/latest/`, 2024.

18 Pandas Development Team, *Pandas Library*, `https://pandas.pydata.org`, 2024.

19 Matplotlib Development Team, *Matplotlib Library*, `https://matplotlib.org`, 2024.

20 NumPy Developers, *NumPy Library*, `https://numpy.org`, 2024.

21 SciPy Community, *SciPy Library*, `https://scipy.org`, 2024.

22 S. Hadden and Y. Lithwick, *Kepler Planet Masses and Eccentricities from TTV Analysis*, 2017.

23 J. F. Otegi, F. Bouchy and R. Helled, *Revisited Mass-Radius relations for exoplanets below 120 Earth masses*, 2019.

24 J. F. Otegi, R. Helled and F. Bouchy, *The similarity of multi-planet systems*, 2022.

25 T. A. Berger, D. Huber, E. Gaidos and J. L. van Saders, *Revised Radii of Kepler Stars and Planets Using Gaia Data Release 2*, 2018.

26 J. E. Owen and R. Murray-Clay, *Metallicity-dependent signatures in the Kepler planets*, 2018.

27 R. Luque and E. Palle, *Density, not radius, separates rocky and water-rich small planets orbiting M dwarf stars*, 2022.