THE UNIVERSITY OF CHICAGO

CONSTRAINING EXOPLANET COMPOSITION DEMOGRAPHICS WITH LARGE-SCALE SURVEYS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ASTRONOMY AND ASTROPHYSICS

BY

ANDREW RUSSELL NEIL

CHICAGO, ILLINOIS

DECEMBER 2021

Dedicated to all the teachers and mentors who have nurtured my growth as a person

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I acknowledge that the University of Chicago and its surrounding neighborhood of Hyde Park, where I have lived and worked for the past six years, are located on the unceded land of the Peoria, Miami, Kickapoo, and Potawatomi Nations. I also recognize that the University of Chicago has benefited from centuries of forced labor and economic extraction from enslaved African Americans and their descendants.

First and foremost, I'd like to thank my advisor, Leslie Rogers. Thank you for taking me on as a student when I knew next to nothing about exoplanets and molding me into the scientist I am today. You have supported me every step of the way and have continually pushed me out of my comfort zone, from applying to fellowships and conferences, joining you as a TA for the Fermi's paradox course, or taking on a mentorship role for younger students. You have made me into a more confident person and it has been a pleasure working with you over the past five years.

I'd also like to thank my committee: Daniel Fabrycky, Jacob Bean, and Daniel Holz. Thank you all for your support and for providing valuable perspectives on my work.

Much of this thesis wouldn't have been possible without two of my collaborators: Jessica Liston and Isaac Malsky. Thank you Jessica for being willing to have me as a mentor and for sticking with our project for two years. I hope the experience was valuable; I know it was immensely helpful and rewarding for me. It was a pleasure to work with you and I look forward to seeing where your future takes you. Thank you Isaac for teaching me about MESA and for readily helping out whenever it was needed.

Thank you to all of the exoplanet scientists I have had the pleasure to interact with at UChicago, especially the past and present members of the Rogers exoplanet group: Nadejda, Jisheng, Isaac, Quang, Jessica, Bridget, Fei, Andrea, Anora, Greg, and others. It was helpful to learn from all of you and I'm grateful for our group and one-on-one discussions. Thank you to my office mates, including Emily, Cory, Adina, Greg, and Megan for keeping me sane

while working in a windowless office.

Thank you to all of the support staff and workers in ERC and the University of Chicago, whose labor is essential to the work that we all do. Thank you to Laticia Rebeles and Jennifer Smith for the assistance they provide to graduate students in the department.

Grad school wouldn't have been the same without the friends that I've made in the department. Thank you to Amy, Gourav, Nora, Jason, Mandy, Phil, Meng Xiang, Cory, Celeste, Emily, Jisheng, James, Dylan, Ross, Michael, Fei, Huanqing, Maya, John, Rostom, Rebecca, Adina and Abby. I will cherish the memories that we all made together and I hope you all find success beyond grad school.

Lastly, I'd like to thank my family for their support of me not only during the past six years, but in getting me this far to begin with: my mother, my father, my two brothers and their respective partners. Thank you to my two cats, Sampson and Audrey, for their crucial emotional support. Thank you to my partner's parents, for welcoming me to Canada during a pandemic and for always looking out for my partner and me above themselves. Finally, thank you to my partner Qing Yang. We made it through grad school and experienced all of its associated ups and downs together. Thank you for all that you do for me and for our family. I couldn't have done it without you.

ABSTRACT

Large-scale surveys of transiting exoplanets such as *Kepler* have revolutionized the study of exoplanet demographics. This dissertation focuses on the insights into planetary composition that can be gained by combining transit surveys with constraints on those planets' masses from radial velocity follow-up and transit timing variations. We first investigate second-order effects on the empirical planet mass-radius relation, finding the sample of *Kepler* and other transiting planets from small surveys with mass measurements to be consistent with no dependence on host star mass. We then show how the joint mass-radius-period distribution of planets can be constrained using a mixture model to include several compositional subpopulations. We create a suite of models and employ model selection techniques to show that the inclusion of at least three subpopulations (planets with gaseous envelopes, evaporated rocky cores, and intrinsically rocky planets) is supported by the data. We find similar support for models that include or exclude photoevaporation, as well as models that include or exclude water worlds, highlighting the degeneracies inherent to the planet population in the mass-radius-period plane. We use our models to calculate $\eta_\oplus$, finding our estimate to be highly model dependent, obtaining a significantly lower $\eta_\oplus$ when photoevaporation is included in the model. Finally, we evolve a dense grid of planet evolution models that can enable future population studies to translate between the fundamental properties of planets (mass, envelope mass fraction, incident flux, age) and the observable plane (mass, radius, period). This thesis builds towards the robust characterization of the exoplanet composition distribution which will provide insights into competing theories of planet formation and the occurrence rate of habitable planets in the galaxy.

CHAPTER 1

INTRODUCTION

The advent of large-scale surveys of transiting exoplanets has revolutionized the study of exoplanet demographics. The *Kepler* space telescope, launched in 2009, has since discovered roughly 2400 confirmed transiting exoplanets and characterized their radii and periods (Borucki et al., 2011a). *Kepler*'s sensitivity to planets down to radii of $1\,R_\oplus$ and periods up to 1 year have enabled the robust characterization of the radius-period distribution of small, close-in planets. This has revealed the ubiquity of planets between the size of the Earth and Neptune, deemed "Super-Earths" or "Mini-Neptunes" depending on their composition (Howard et al., 2012). Additionally, high-precision spectroscopy of host stars, enabling the tight constraint of *Kepler* planet radii, has revealed the existence of a gap in the radius distribution of small exoplanets between 1.5 and $2\,R_\oplus$, deemed the "Fulton gap" (Fulton et al., 2017). This gap has several proposed explanations, from envelope mass loss due to photoevaporation or core-powered mechanisms, to compositional diversity of planetary cores (e.g. Ginzburg et al., 2018; Fulton & Petigura, 2018; Zeng et al., 2019).

This dissertation focuses on the insights into planetary composition that can be gained by combining large-scale transit surveys like *Kepler* with constraints on those planets' masses from radial velocity follow-up and transit timing variations. An exoplanet's bulk density can be constrained through combined measurements of radius and mass, offering a window into the space of possible compositions. In Chapter 2, we use the available sample of exoplanets with mass and radius measurements to investigate whether host star properties affect exoplanet composition through the mass-radius relationship. In Chapters 3 and 4, we move beyond the mass-radius relationship to explore how multiple compositional subpopulations can imprint themselves in the mass-radius-period plane of exoplanets. In Chapter 3, we develop a technique to model the joint mass-radius-period distribution of exoplanets, using a mixture model to include multiple subpopulations of planets. We extend our mass-radius-

period distribution model in Chapter 4 to include exoplanets with icy core compositions, allowing us to construct a suite of models to compare and contrast. Throughout these three chapters, we employ several model selection techniques to assess the validity of our models and their support from the data. In Chapter 5, we evolve a large, dense grid of planet models to provide planetary radii as a function of planet mass, envelope mass fraction, incident flux and age, enabling future population studies to robustly incorporate mass-radius relations from evolution models. We review future exoplanet surveys, extensions of this work and conclude in Chapter 6.

1.1 Exoplanet Observables

Our ability to constrain the properties of the exoplanet population that exists in the galaxy is limited by the quantity and quality of the data available to us. In this section we will review the observables of exoplanets as measured by the transit and radial velocity techniques, as well as their typical properties.

1.1.1 Transits

The transit technique relies on measuring the brightness of a star in order to detect periodic dimming that occurs due to the transit of an exoplanet orbiting that star. The fundamental quantity measured by the transit technique is the transit depth, or the fraction of light blocked by the planet:

$$\delta = \left(\frac{R_p}{R_*}\right)^2 \tag{1.1}$$

where δ is the transit depth, R_p is the radius of the planet, and R_* is the radius of the star. A consequence of this equation is that the precision of the measured planet radius is dependent on the precision of the measured stellar radius. More precisely measured stellar

radii translates into lower uncertainties in the observed planet radii. The Fulton gap in the planet radius distribution was only unveiled through the use of a subset of *Kepler* planets that had high-resolution spectroscopic follow-up of the host stars, resulting in low planet radius uncertainties (Fulton et al., 2017). The period of the planet's orbit is the second main observable obtained by the transit technique, measured as the time between successive transits. Other observables directly measured from the transit technique include impact parameter and transit duration, and possible inferred parameters include the inclination and eccentricity of the planet's orbit.

Kepler is the largest transit survey to date, and the source of the datasets used in Chapters 2, 3 and 4 (Borucki et al., 2011a). There are currently ≈ 2400 confirmed *Kepler* planets, with a similar number of planet candidates yet to be confirmed. Confirmed *Kepler* exoplanets have a median radius uncertainty of 18%. The California-*Kepler* Survey (CKS), a subset of *Kepler* that have lower radius uncertainties due to high-resolution spectroscopic measurements of the host stars, as well as cross-matching with *GAIA* parallaxes, is used in Chapters 3 and 4 rather than the full *Kepler* sample (Petigura et al., 2017; Johnson et al., 2017). The CKS sample (after cuts) has 1130 total planets and a median radius uncertainty of 4.8%, a significant improvement over the full sample. Periods are measured so precisely with *Kepler* that we treat the period of planets as having no uncertainty.

Another important aspect of an observational survey is the detection efficiency, or the ability of the survey to detect planets across different ranges of the observable properties. For the transit technique, smaller planets are harder to detect due to the smaller transit depth, and longer period orbits are harder to detect due to the limited observing time of *Kepler* and the need to stack multiple transits to build a signal. The detection efficiency of *Kepler* is well-studied through injection-recovery tests, where transits are injected at the pixel-level and run through the *Kepler* pipeline to see which transits are recovered and which are missed (Christiansen et al., 2020). Through this process, we can create a mapping of radius and

period to the fraction of transiting planets detected by *Kepler*. In addition to planets missed by the *Kepler* pipeline, we also need to account for the planets missed because their orbits do not transit the star with respect to the telescope. We describe the steps to calculate the detection efficiency in more detail in Chapters 2 and 3, with the combined detection efficiency and transit probability as a function of radius and period shown in Figure 3.1. The detection efficiency is necessary ingredient that enables us to properly constrain the true underlying planet population as opposed to the observed population.

TESS is an ongoing all-sky transit survey that has been operational since 2018 (Ricker et al., 2015). *TESS* has greater sensitivity to planets orbiting M-dwarf stars compared to *Kepler*, and currently has a similar quantity of planet candidates compared to *Kepler*, although with much fewer confirmed. Due to timing issues, *TESS* data is not used in this thesis. However, a simulated *TESS* catalog is used in Chapter 2 in relation to the exoplanet mass-radius relation, as the sensitivity of TESS to planets around lower mass stars is conducive to the detection of host star mass dependence in the mass-radius relation.

1.1.2 Radial Velocity

The radial velocity technique measures the Doppler shift in the absorption lines of a star's spectrum induced by the motion of the star around the star-planet center of mass. The fundamental observable with the radial velocity technique is the radial velocity semi-amplitude K:

$$K = 0.6387 \left(\frac{P}{\mathrm{day}}\right)^{-1/3} \left(\frac{M}{M_\oplus}\right) \frac{\sin i}{\sqrt{1-e^2}} \times \left(\left(\frac{M_{*,}}{M_\odot}\right) + 3 \times 10^{-6} \left(\frac{M}{M_\oplus}\right)\right)^{-2/3} \tag{1.2}$$

where P is the planet's orbital period, M is the planet's mass, i is the inclination of the planet's orbit, e is the eccentricity of the planet's orbit, M_* is the mass of the host star,

and K is the semi-amplitude given in meters per second. Similarly to the transit technique, the radial velocity semi-amplitude is dependent on the mass of the host star, so a higher precision measurement of the host star mass will lead to higher precision on the planet's mass. The semi-amplitude also depends on the inclination of the planet's orbit. For planets with RV measurements that also transit their star, which is the subset we are concerned with in this thesis, this inclination is constrained to be nearly 90° due to the requirements of transit geometry. The period is also measured through the repeating sinusoidal nature of the radial velocity semi-amplitude curve, again with high precision compared to the precision on the mass.

As opposed to the *Kepler* transiting planet sample used in this thesis, radial velocity measurements of transiting planets are taken with a variety of different spectrographs (such as MAROON-X (Seifahrt et al., 2018)) and are each subject to their own instrumental precision. Mass uncertainties are generally higher by several factors than radius uncertainties. For the subset of CKS planets with radial velocity mass measurements used in Chapters 3 and 4, the average mass uncertainty is 22% with a total sample size of 68 planets. Due to the nature of the radial velocity technique requiring high-resolution spectroscopy of host stars, taking RV measurements of planets is expensive and astronomers need to pick and choose which planets to observe. This selection process is often not as well-documented and poorly characterized, making it difficult to properly model in statistical applications.

1.1.3 Transit Timing Variations

Another technique that can characterize the masses of transiting planets is transit timing variations (TTV). As the name suggests, this involves precisely measuring the transit times of a planet in order to capture variations induced by the presence of other exoplanets in the system. As such, TTVs can only be used to measure the masses of transiting planets in multi-planet systems. Additionally, TTVs that are strong enough to be measured typically

occur in exoplanet systems that have specific configurations where planets are in orbital mean motion resonance, having orbital periods in integer ratios with each other. The equations for TTVs are more complicated than the equations governing the transit depth and semi-amplitude, and not convenient to write down here. Importantly, unlike radial velocities and transits, the masses derived from TTVs are independent of the measurement of the host star mass (Holman et al., 2010).

In Chapter 2, we use a sample of transiting planets with TTV mass measurements to assess the host star mass dependence of the exoplanet mass-radius relation. Our sample had a comparable number of planets with TTV masses compared to planets with RV masses (63 vs 88). However, in Chapters 3 and 4, we solely use *Kepler* planets with RV mass measurements, excluding the TTV sample. Planets with TTV masses have been shown to have systematically lower densities than planets with RV masses, possibly due to observational biases between the two techniques, or potential physical differences between the two populations (Jontof-Hutter et al., 2014; Mills & Mazeh, 2017). For this reason, it is often simpler to stick with RV masses, although both samples are valid to use.

1.2 Data Analysis Techniques

In this section I will highlight three main classes of data analysis techniques that I used throughout my thesis: hierarchical Bayesian modeling, model selection, and simulations. While these three topics are by no means exhaustive of the techniques I utilized, they occur repeatedly in the following chapters and deserve some brief introduction.

1.2.1 Hierarchical Bayesian Modeling

Hierarchical Bayesian modeling serves as the foundation for much of the research presented in this thesis. Specifically, Chapters 2, 3, and 4 all follow the same basic structure: construct a hierarchical Bayesian model, fit it to the appropriate data, and analyze the resulting fit in

order to answer a research question.

The Bayesian aspect of a hierarchical Bayesian model arises from Bayes theorem:

$$\mathrm{p}(\theta|y) = \frac{\mathrm{p}(y|\theta)\mathrm{p}(\theta)}{\mathrm{p}(y)} \tag{1.3}$$

where θ represents a set of model parameters and y represents the data. The term on the left-hand side represents the posterior probability, what we hope to constrain: the probability of a set of model parameters given the data available. The first term in the numerator on the right-hand side is the likelihood function, which tells us how likely the data that we have is given a particular set of parameters. The second term in the numerator is the prior, where we distill our beliefs about what we expect about the distribution of model parameters. The term in the denominator is called the evidence, sometimes used in model selection, but otherwise mostly ignored, thus turning this equation into a proportionality.

In practice, to construct a Bayesian model we first choose a parameterization for the distribution we are attempting to constrain. For example, if we're constraining the mass distribution of exoplanets, we might choose a log-normal distribution. In the simplest case, this sets the likelihood - the probability density function of a log-normal distribution. We then must assume a prior for the parameters of the distribution. A log-normal has a mean μ and a scatter σ as parameters. We might choose a normal distribution centered on 0 for the prior of μ, and a half-normal distribution (to restrict to positive values) for the prior on σ. In Chapters 3 and 4, our likelihood is more complex owing to the fact that we have to account for the detection aspect of our planet sample - not every planet in the underlying population is detected, and the probability of detection is a function of the planet's radius and period. The likelihood we use in Chapters 3 and 4 is called an inhomogeneous Poisson process likelihood and contains an integral over the three dimensions that we constrain that equates to the expected number of planets detected.

After deciding on the parametrization, the likelihood, and the priors, we can then fit the

model to the data. A widespread technique for the fitting of Bayesian models is Markov chain Monte Carlo (MCMC), which involves using randomized means to build a Markov chain of parameter values whose distribution is sampled from the posterior distribution. In Chapters 2, 3, and 4, we use the Stan software package to perform MCMC (Carpenter et al., 2017). Stan uses a particular implementation of MCMC called the No-U-Turn Sampler (NUTS), itself an extension of another implementation called Hamiltonian Monte Carlo (HMC) (Hoffman & Gelman, 2014). These samplers are particularly suited for the models we build in Chapters 2, 3, and 4, due to the size of the datasets involved and the hierarchical nature of the models.

The hierarchical aspect of hierarchical Bayesian modeling arises from the fact that we are given imperfect data that has some observational uncertainty associated with it. As such, we can separate out the *observed* properties of the planets from the *true* properties of the planets. The true planet properties are unknowable - the observed properties give us some idea of what they might be, but we must incorporate the true properties as parameters to be modeled. This separates our model parameters into two layers: we now have a set of parameters that govern the distributions of the true planet properties, called the hyperparameters, and a set of parameters that directly model the true properties of each individual planet. This layering makes the model hierarchical. At the top, we have the hyperparameters - the mean and spread of a log-normal distribution in the previous example - that inform the distribution of the true planet properties. The true planet properties, themselves parameters in the hierarchical model, then inform the observed properties of the planets that end up in the catalog. A graphical example of a hierarchical model is shown in Figure 2.2 in Chapter 2.

The inclusion of true planet properties as parameters in the model drastically increases the scale of the model, as you require as many additional parameters as planets in your sample. In Chapters 3 and 4, the number of hyperparameters is on the order of tens, but the number of true planet parameters is over one thousand. This necessitates an algorithm like

NUTS or HMC that can handle such a large number of model parameters. An additional benefit of hierarchical Bayesian modeling is that the model returns a posterior distribution for the true properties of each planet alongside the hyperparameters. Within the context of a specific model, you are given a better picture of the underlying planet sample used in your fit.

1.2.2 Model Selection

Throughout this thesis, we construct hierarchical Bayesian models and fit them to data to answer scientific questions, as discussed in Section 1.2.1. Fitting a single model in isolation to test a hypothesis can be done, but without additional models to compare it to, the interpretation of the model loses context. For this reason, we often want to construct multiple models to fit to a given dataset, with the differences between model fits offering key insights. For example, in Chapter 2, we want to test whether or not the planetary mass-radius relation has any dependence on host star mass. To do so, we construct a model that includes this dependence as well as a model that has no dependence. By comparing the resulting model fits we can begin to answer the question of whether or not this dependence is justified.

Once we have multiple models, how do we know which one to "trust" more? Which model should we accept as the model that provides the best fit to the data? How can we quantify this? This is the field of model selection. In the previous example, performing model selection on the model that includes dependence and the model that has no dependence can give us a quantitative answer to the question of whether or not this dependence is supported by the data.

The hallmark of a good model is its predictive power. If a model provides a good fit to the data, it should be able to predict the properties of data that it has not been exposed to. For example, say we discover a new planet and have characterized its radius. The mass-radius relations that we have modeled to both include and not include dependence on host star

mass can offer separate predictions for the new planet's mass. We can then compare these predictions to RV or TTV measurements of the planet's mass and assess quantitatively how well the predictions did. Since a single planet is far too small a sample size, ideally we'd want a sizeable sample of planets for the models to predict and compare to observations. The "best fit" model would then be the model that better predicts planet masses on average compared to the other model.

An issue arises because naturally, we'd want to use all of the data available to us to fit the model. Where then can we obtain the separate sample of planets for prediction purposes? Do we have to wait around for new planets to be discovered?

One way around this problem is to see how well the models perform on the data that has already been used to fit the model. We can use the likelihood, or log probability, of each individual planet as a measure of how well that planet is fit by the data. Using data that has been used to fit the model to assess the model itself gives rise to a serious problem. Of course a model is going to provide a good fit to the data that was used to fit it; that's the whole point of fitting the model! Furthermore, the more flexibility we give a model, for instance by giving it additional parameters, the better the model can fit each individual data point. The best models would simply be the models with the most flexibility, because they would be able to contort themselves to whatever shape the data demands. If we expose such a hyper-flexible model to new data, it would likely provide catastrophically wrong predictions, because it fit the specifics of the initial dataset without being generalizable to new data. This problem is called overfitting.

A simple solution is to include a penalty that scales with the number of parameters in the model. This gives rise to a class of model selection metrics called information criteria. The best models are then the models that provide the best balance between how well they fit the data and minimize the number of parameters. Examples of information criteria include Aikake information criterion (AIC), Bayesian information criterion (BIC), Deviance

information criterion (DIC) and the Widely Applicable Information Criterion (WAIC). AIC is defined as:

$$\text{AIC} = -2 \log p(y|\hat{\theta}_{\text{MLE}}) + 2k \tag{1.4}$$

where the first term is the likelihood of the data given by the maximum likelihood estimate of the parameters of the model, and the second term is a penalty for the number of parameters k. BIC takes a similar form to AIC, except the penalty scaling changes to $k \ln n$, where n is the number of data points. DIC replaces the explicit penalty on the number of parameters with a penalty on the deviance of the total log-likelihood over the posterior distribution, intending to provide the same effect of penalizing flexibility. DIC also utilizes the posterior mean to calculate the log-likelihood, rather than the maximum likelihood estimate.

None of the three aforementioned are fully Bayesian (despite BIC having "Bayesian" in the name), as they use pointwise estimates of the posterior rather than using the full information of the posterior distribution (Gelman et al., 2014). WAIC aims to correct this. Its penalty (the equivalent of the second term in Equation 1.4) takes the form:

$$p_{\text{WAIC}} = 2 \sum_{i=1}^{n} V_{s=1}^{S} (\log p(y_i|\theta^s)) \tag{1.5}$$

where we compute the sample variance of each data point i's log likelihood over posterior samples of the hyperparameters given by θ^s, and then take the sum over n data. WAIC averages over the posterior distribution rather than using a pointwise estimate as in the AIC, BIC, and DIC. In Chapter 2, we use WAIC to compare between models of the mass-radius relation. These information criteria are not always appropriate, however, and can fail in the regime of high model complexity with multivariate data or with large numbers of parameters. For this reason, we move beyond WAIC and apply other model selection methods in Chapters 3 and 4.

Another solution to the issue of assessing out-of-sample predictive accuracy is to set aside a portion of our sample for the purpose of performing model selection. In machine learning language, the subset that we set aside is typically called the "validation set", whereas the subset that we use to fit the data is called the "training set". We'd fit the model to the training set, and then made predictions on the validation set to assess the model's performance.

How do we decide how to break up the data into these two subsets? Ideally, to maximize our sample size, we'd be able to use the entire sample to perform predictions. We can achieve this through repeatedly breaking down the dataset into different training sets and validation sets. We'd divide the data once, fitting to the training set and performing predictions on the validation set. We'd then divide the original dataset again to obtain a new validation set and repeat the process until we have performed predictions on our entire sample. This is called cross-validation (Vehtari et al., 2017).

How many times should we break up the data like this? To provide the best possible prediction for an individual data point, we'd ideally choose that single data point as the validation set with the entirety of the remaining data as the training set. If we repeat this for each data point, with n total data points, this would necessitate n total model fits. This is called leave-one-out cross-validation. Naturally, this can be computationally expensive, especially with large datasets. Instead of this, we often simply choose a number of times k to divide the data, called the number of folds. This technique is called k-fold cross-validation. In k-fold cross-validation, we first divide the dataset into k different subsets. We then fit the model k separate times, each time using a separate subset as the validation set to perform predictions, and using the remainder of the data as the training set. As k becomes smaller, the predictions become less accurate because you are left with less and less data in your training set. $k = 10$, i.e. 10-fold cross-validation, is a common choice to balance computational resources and predictive accuracy. Each training set in 10-fold

cross-validation contains 90% of the data in the original sample.

Cross-validation should naturally deal with overfitting, as the process of splitting the dataset into training sets and validation sets means that we are assessing the model by its prediction on unseen data. As such, there is no need for an explicit penalty for the number of parameters and we can simply compute the log likelihood of each data point. In Chapters 3 and 4, we use k-fold cross validation to perform model selection on our suite of joint mass-radius-period distribution models. In Chapter 4, we explore the effectiveness of k-fold cross validation by running a set of simulations to see if the preferred model is the model that was used to generate the data. We find that in the context of the mixture models we constructed, k-fold cross validation still had issues in preferring models with more complexity. Unfortunately, with high-complexity models that include large numbers of parameters, multivariate data, or multiple mixtures, model selection is not a straightforward task. Further research is necessary into model selection techniques that function well in these circumstances.

1.2.3 Simulations

Simulations are an extremely common tool in astronomy and the word can convey a variety of different meanings. In the context of this thesis, we mean the creation of simulated planet catalogs in order to do analysis. This can have several purposes. A common purpose is to demonstrate the validity of a technique. With a catalog of real planets, we may not know what to expect when fitting a model; with a simulated catalog, we know exactly what to expect and can confirm the technique is working properly. Another purpose is to simulate the conditions of a future dataset in order to compare to the capabilities of the current dataset. Finally, simulations can be used as a computational convenience to approximate complicated analytic distributions; for example, we can plot the mass distribution constrained by a model fit by simulating a large number of planets from that mass distribution.

In generating a simulated planet catalog, the first step is to set the values of the hyperparameters. This can be done in several ways. We can simply take the median values of a set of posteriors, if we are trying to simulate planets from a given model fit. We can also sample over the posteriors instead of taking the median values, in order to more robustly capture the variation within the posterior distribution. This can be done by drawing a random posterior sample, simulating a number of planets, and then repeating by drawing more random posterior samples. If we are not interested in simulating planets from a model fit, we can instead handpick the parameter values to whatever is desired. This is useful if exploring how planet catalogs might differ under parameters far removed from the actual model fit. If we aim to simulate a catalog of observed planets, rather than planets belonging to the underlying population, we also need to choose a method to apply observational errors as well as apply the effects of detection. For example, if we are simulating a *Kepler*-like catalog, we might draw relative errors on radius from the true *Kepler* sample to apply to our simulated catalog, and decide whether planets are detected or not using the *Kepler* detection efficiency as a function of radius and period. Another common choice is to simply apply flat errors of a fixed percentage to all planets, or to do the same but to adjust percentages as a function of radius or mass.

One major aspect of simulated catalogs that makes them useful is that the parameters that generated them are known. We can use this to validate that analysis techniques are working properly. For example, we can generate a simulated catalog with a known set of hyperparameters θ, fit the model to the simulated catalog, and verify that we retrieve the same model hyperparameter values. We can also generate multiple catalogs to study the effects of sample size, observational errors, and detection bias on the ability of the model to retrieve the correct input values. Another example of applying simulated catalogs to verify a technique is the study on the effectiveness of k-fold cross-validation in Chapter 4. We generate a simulated catalog with one specific model, fit a suite of models to that catalog,

and then perform model selection on each of those models. Since one model was used to generate the data, that generative model should be the model preferred by the model selection process. Unfortunately, these simulations revealed that this was not always the case.

Simulations can also allow you to explore the possibilities that a future dataset might bring. For example, in Chapter 2 we assess whether or not there is evidence for host star mass dependence in the mass-radius relation in the available dataset at the time. We find there to be no dependence, but we are also limited by the number of planets in the dataset orbiting low-mass stars. Perhaps a dependence exists, but we cannot yet detect it due to the limitations of the dataset. We perform a simulation of an expanded mass-radius dataset that might arise from *TESS*, a transit survey that launched in 2018, by estimating the expected size of the dataset and level of observational error. We find that with this expanded dataset, this dependence can be detected if it exists and is strong enough.

Lastly, simulations provide a convenient tool to approximate distributions that are otherwise analytically intractable. In Chapters 3 and 4, we develop mixture models of the exoplanet mass-radius-period distribution. The radius distribution in this model is conditioned on mass, and differs depending on the mixture in question. In order to plot the retrieved analytic 1D radius distribution of the underlying planet population, we'd need to integrate out these dependencies on the mixture and mass, as well as integrate over the posterior distribution. Instead, it is much simpler to generate simulated planets and take their radius distribution as the radius distribution of the model. Care must be taken to ensure that a large enough number of planets are simulated so that statistical fluctuations do not alter the shape of the distribution. There are many examples of figures in this thesis which were generated in this manner, including Figure 3.3 in Chapter 3 and Figure 4.2 in Chapter 4.

CHAPTER 2

HOST STAR DEPENDENCE OF SMALL PLANET MASS-RADIUS DISTRIBUTIONS

Kepler's sensitivity to planets orbiting not only FGK dwarfs, but also M dwarfs, allowed for comparative studies of occurrence rates of planets orbiting stars of different spectral types. These studies found that small planets have higher occurrence rates around M dwarf stars than FGK stars (Dressing & Charbonneau, 2015; Gaidos et al., 2016). Further, this higher occurrence rate directly translates into higher amounts of planetary heavy-element mass in short orbits around M dwarfs (Mulders et al., 2015a), rather than planets around M dwarfs simply being smaller but more numerous on average than planets around FGK dwarfs. This finding is contrary to expectations from planetary formation, as the protoplanetary disk mass scales with the host star mass - the opposite trend that we see with the planetary mass.

This finding from Mulders et al. (2015a) was enabled by using a mass-radius relationship to translate radii of *Kepler* planets orbiting M and FGK dwarfs into masses. However, a key assumption in the use of this mass-radius relationship was that the mass-radius relation is independent of the host star - planets orbiting M dwarfs and planets orbiting FGK dwarfs follow the same mass-radius relationship. If planets orbiting M dwarfs are on average less dense than planets orbiting FGK dwarfs, this would impact the mass-radius relation and subsequently alter the trend of planetary heavy-element mass with host star type. Conflicting physical processes make it unclear whether planets around M dwarf stars would be more or less dense than their counterparts around FGK dwarfs - but planets around M dwarfs certainly form in a different environment. M dwarfs have a longer pre-main sequence phase, with a larger difference in luminosity from formation to the main sequence stage. Additionally, young M dwarfs undergo periods of intense activity, with elevated XUV radiation that could blow away the envelopes of forming exoplanets.

In this chapter, we adapt the probabilistic mass-radius relation developed by Wolfgang

et al. (2016) to include dependence on both host star mass and incident flux. We then fit this host star mass (and incident flux) dependent mass-radius relation to the sample of transiting planets with either RV or TTV mass measurements. Our fits are consistent with no host star mass dependence. We then apply our model fits to the trend of planetary heavy-element mass in short orbits as a function of host star mass, confirming the results of Mulders et al. (2015a) and showing that even when you allow for the mass-radius relation to depend on host star mass, planetary heavy-element mass in short orbits is higher around M dwarfs than FGK dwarfs. We further test several other assumptions and find this trend to hold up.

A key limiting factor in detecting a dependence of the mass-radius relation on host star mass is the number of planets with radius and mass measurements orbiting M dwarfs. At the time this work was published, there were only six of these planets. This number has since increased dramatically with the advent of *TESS*. As part of this chapter, we perform simulations to see whether we would be able to detect a difference in the mass-radius relation with host star mass using a hypothetical TESS dataset and find the answer to be yes, under certain conditions. The time is right for this topic to be revisited.

This chapter was completed in collaboration with Leslie Rogers and published in the Astrophysical Journal (Neil & Rogers, 2018). We thank our anonymous referee for providing valuable feedback and suggesting additions that rounded out the paper. We also thank Angie Wolfgang, Daniel Fabrycky, Emily Gilbert and Gregory Gilbert for useful discussion. LAR gratefully acknowledges support from NASA Exoplanet Research Program grant NNX15AE21G and from NSF FY2016 AAG Solicitation 12-589 award number 1615089.

2.1 Abstract

The planet formation environment around M dwarf stars is different than around G dwarf stars. The longer hot protostellar phase, activity levels and lower protoplanetary disk mass of M dwarfs all may leave imprints on the composition distribution of planets. We use

hierarchical Bayesian modeling conditioned on the sample of transiting planets with radial velocity mass measurements to explore small planet mass-radius distributions that depend on host star mass. We find that the current mass-radius dataset is consistent with no host star mass dependence. These models are then applied to the *Kepler* planet radius distribution to calculate the mass distribution of close-orbiting planets and how it varies with host star mass. We find that the average heavy-element mass per star at short orbits is higher for M dwarfs compared to FGK dwarfs, in agreement with previous studies. This work will facilitate comparisons between microlensing planet surveys and *Kepler*, and will provide an analysis framework that can readily be updated as more M dwarf planets are discovered by ongoing and future surveys such as *K2* and *TESS*.

2.2 Introduction

The *Kepler* survey (Borucki et al., 2011b) has discovered thousands of transiting exoplanets, leading to the robust characterization of the radius distribution of small planets (Burke et al., 2015; Fulton et al., 2017). Subsequent radial velocity follow-up of transiting planets from *Kepler* and other surveys, as well as transit timing variations, have constrained planet masses and allowed exploration of the composition distribution of planets, often in the form of a mass-radius relationship. Constraining the composition distribution of planets provides insights into different formation pathways and the prevalence of Earth-like rocky planets.

A mass-radius relationship (hereafter M-R relation) is a key ingredient necessary for the comparison of different exoplanet populations. While *Kepler*, and in the near future *TESS* (Ricker et al., 2015), are able to characterize the radii of planets, *WFIRST*, a future microlensing survey, will only be able to characterize masses. If we wish to combine transit and microlensing surveys to constrain planet occurrence rates, an M-R relation is essential for translating masses into radii, and vice versa. For example, Suzuki et al. (2016) showed that the break in the mass-ratio function of microlensing planets from the *MOA-II* survey

occurs at a higher mass than the peak in the *Kepler* mass distribution by converting *Kepler* radii to masses with an M-R relation. However, *WFIRST* is much more sensitive to planets around M dwarf hosts than *Kepler*, due to the innate characteristics of the microlensing method. Given these disparate stellar distributions, if planet formation differs around M dwarfs compared to FGK dwarfs, this may manifest itself in the M-R relation. In this paper, we consider a mass-radius relationship that depends on host star mass.

2.2.1 Planet formation around M dwarfs

Important differences during the protostellar phase between M and FGK dwarfs could impact planet formation. Before reaching the main sequence, stars undergo a period of contraction whereby the luminosity declines several orders of magnitude. For low mass stars, this pre-main sequence phase lasts longer and the luminosity difference is more pronounced, leading to an initially distant snow line that migrates far inward while the star is contracting (Kennedy & Kenyon, 2008). This has important consequences for planet composition. Raymond et al. (2007) and Lissauer (2007) found that planets formed in situ around low mass stars may be deficient in volatiles, as water-bearing planetesimals would not be scattered as often from beyond the snow line. However, if the surface density of water content in planetesimals is much higher than in our own solar system, volatile delivery to planets inside the snow line would increase (Ciesla et al., 2015) Alternatively, if the disk is long-lived, planets that form beyond the snow line would have more time to migrate inward, leading to volatile-rich planets at short orbits (Alibert & Benz, 2017).

Another characteristic feature of M dwarfs is their intense activity at young stellar ages, during which flares can output bursts of XUV radiation and relativistic particles. For planets formed in situ, this high energy radiation could provide enough heating to evaporate the atmosphere and any water content on the surface (Scalo et al., 2007). Alternatively, for gaseous planets that form outside the snow line and migrate inward, XUV radiation from

the star may be enough to fully evaporate their envelopes, resulting in "habitable evaporated cores" (Luger et al., 2015). This process is more likely to occur for planets less massive than 2 $M_{\oplus}$, which could result in a population of volatile-rich rocky planets. If the mechanisms described can efficiently erode planet atmospheres, we expect close-in planets around M dwarfs to experience more mass loss and to be more dense than their counterparts around FGK stars.

Kepler has shown that the radius distribution of planets around M dwarfs differs from that around FGK dwarfs, with higher occurrence rates of small planets around M dwarfs (Dressing & Charbonneau, 2015; Gaidos et al., 2016). Similarly, radial velocity surveys have shown differences in the the mass distribution, the most notable being a comparative lack of giant planets around M dwarfs (Bonfils et al., 2013). Occurrence rate studies with *Kepler* have also suggested a trend of increasing planetary heavy-element mass in short orbits for decreasing stellar mass, seemingly at odds with the protoplanetary disk mass scaling with stellar mass (Mulders et al., 2015a). However, it is not yet clear as to how the composition distribution differs, or how this may be impacted by the stellar environment. If planets around M dwarfs have lower volatile content or experience more envelope mass loss, then we would expect planets to be more dense around low mass stars. However, this could be counteracted by a higher surface density of water content in the protoplanetary disk or longer disk lifetimes. While differences in the stellar environment exist for planets around different types of stars, there is currently no predictive model for how these differences may imprint themselves in the observed population of planets. Answering this question motivates characterizing more planets around M dwarf stars.

2.2.2 M-R relations

Empirically derived planet M-R relations have traditionally been cast as simple power laws. Lissauer et al. (2011) found $M/M_{\oplus} = (R/R_{\oplus})^{2.6}$ by fitting a power law to Earth and Saturn.

Utilizing a sample of 22 planet pairs with TTV measured masses, Wu & Lithwick (2013) found a linear relation with $M/M_\oplus = 3(R/R_\oplus)$. Weiss et al. (2013) introduced incident flux dependence into the M-R relation, using a sample of 35 planets with mass and radius measurements to find $(R/R_\oplus) = 1.78(M/M_\oplus)^{0.53}(F/\text{erg s}^{-1}\text{ cm}^{-2})^{-0.03}$ for $M < 150M_\oplus$. Weiss & Marcy (2014) used a sample of 42 *Kepler* planets with RV-measured masses to fit a broken power law with a linear density relation $\rho_P = 2.43 + 3.39(R/R_\oplus)$ g cm^{-3} for $R/R_\oplus < 1.5$ and $M/M_\oplus = 2.69(R/R_\oplus)^{0.93}$ for $1.5 < R/R_\oplus < 4.0$.

More recently, Wolfgang et al. (2016) used hierarchical Bayesian modeling to derive a more statistically robust, probabilistic M-R relation. They include intrinsic scatter in planet composition by modeling the M-R relation as a power-law with dispersion that is normally distributed. They find $M/M_\oplus = 2.7(R/R_\oplus)^{1.3}$ for $R/R_\oplus < 4.0$ with an intrinsic scatter of $1.9M/M_\oplus$, constrained to physically plausible densities, as the best-fit relation. The benefit of the Bayesian approach is that the uncertainties in the power law parameters and intrinsic scatter are fully quantified. Chen & Kipping (2017) used a similar approach to fit broken probabilistic power laws to a much wider range of planet (and stellar) masses, including the transition points as additional parameters in the model. In this paper, we extend the approach by Wolfgang et al. (2016) to include dependence on host star mass.

This paper is organized as follows. In section 2.3 we outline the steps taken to obtain our planet sample, detail our model and how we fit it to the data. The results of our model fitting are described in section 3.6, along with applications to planet heavy element mass and planet mass distributions. We discuss model selection, limitations and future extensions of this work in section 2.5, and conclude in section 2.6.

2.3 Methods

2.3.1 Data

In order to constrain the M-R relation, we need a sample of planets with well-characterized mass and radius measurements and uncertainties. For homogeneity, we choose to use the sample of transiting planets with radial velocity mass measurements. There is a comparable number of transiting planets with TTV mass measurements, but those planets have been shown to have systematically lower densities than planets with RV mass measurements (Weiss & Marcy, 2014; Jontof-Hutter et al., 2014), likely due to different observational biases in the two techniques (Steffen, 2016; Mills & Mazeh, 2017). We use only planets with radii below $8R_\oplus$, as we are most interested in the M-R relation of small planets (rocky planets and mini-Neptunes / super-Earths). There is evidence for a transition in the M-R relation from Neptunian to Jovian planets at around $\sim 11R_\oplus$ (Chen & Kipping, 2017), so adopting a cutoff of $8R_\oplus$ is a conservative measure to ensure we are safely outside of the giant planet regime. Physically, as a planet becomes more massive, the core becomes more dense and degeneracy pressure plays an increasingly important role, causing the M-R relation to flatten at around a Jupiter radius (Chabrier & Baraffe, 2007).

The planet catalog for this work was downloaded from the NASA Exoplanet Archive[1] on June 13, 2017. Our first cuts are made to exclude planets with $R > 8R_\oplus$, circumbinary planets, and those without either transit or RV measurements. As our model incorporates host star uncertainties, we require that a planet has both a well-defined radial velocity semi-amplitude measurement (K) and a planet radius to host star radius ratio ($r = R_p/R_*$) measurement, with uncertainties. While transit depth may be seen as a more fundamental parameter derived from a transit light curve, stellar limb-darkening causes the transit depth to deviate from the simple $\delta = (\frac{R_p}{R_*})^2$ relation (Mandel & Agol, 2002), with the limb-

1. https://exoplanetarchive.ipac.caltech.edu/

darkening law and coefficients varying from star to star. To avoid having to treat limb-darkening for each individual star, we use the radius ratio instead of the transit depth as the primary transit parameter. For each planet, we verify the source paper for the RV measurement to ensure each value reported in the Exoplanet Archive is correct. We remove any planets without reported K or r values, or those with only upper limits. In the case of planets with transit depth measurements but no r (HD 219134 b and c), we calculate r from the transit depth assuming no limb darkening. For those planets without reported eccentricity or inclination measurements, we assume $e = 0$ and $i = 90^{\circ}$.

The final planet catalog includes 88 planets. The distribution of host star mass for these planets is shown in Figure 2.1. Given that we are most interested in how the M-R relation differs around stars of low mass, it is unfortunate that we only have six planets orbiting stars with $M < 0.7 M_{\odot}$; the majority of planets in the sample orbit G type stars. This is largely due to most radial velocity follow-up campaigns targeting GK dwarfs as solar analogues. However, there are many ongoing and future radial velocity observing programs that are specifically designed to characterize planets around M dwarfs (e.g.*MAROON-X* (Seifahrt et al., 2016), *SPIRou* (Moutou et al., 2015), *CARMENES* (Quirrenbach et al., 2014), *ESPRESSO* (González Hernández et al., 2017)) as a follow-up to transit observations (e.g. *TESS*, *MEarth* (Berta et al., 2012), *ExTra* (Bonfils et al., 2015), *SPECULOOS* (Burdanov et al., 2017), *APACHE* (Christille et al., 2013)), so this number is expected to increase in the near future (Kains et al., 2016). Despite the current lack of planets with well-characterized radii and masses around M dwarfs, our analysis will provide a framework that can be revisited when the sample size increases drastically in the near future. A host star mass dependent M-R relation is also necessary to study trends in exoplanet populations with host star mass, and to allow comparison between *Kepler* planets and planets discovered with the microlensing method.

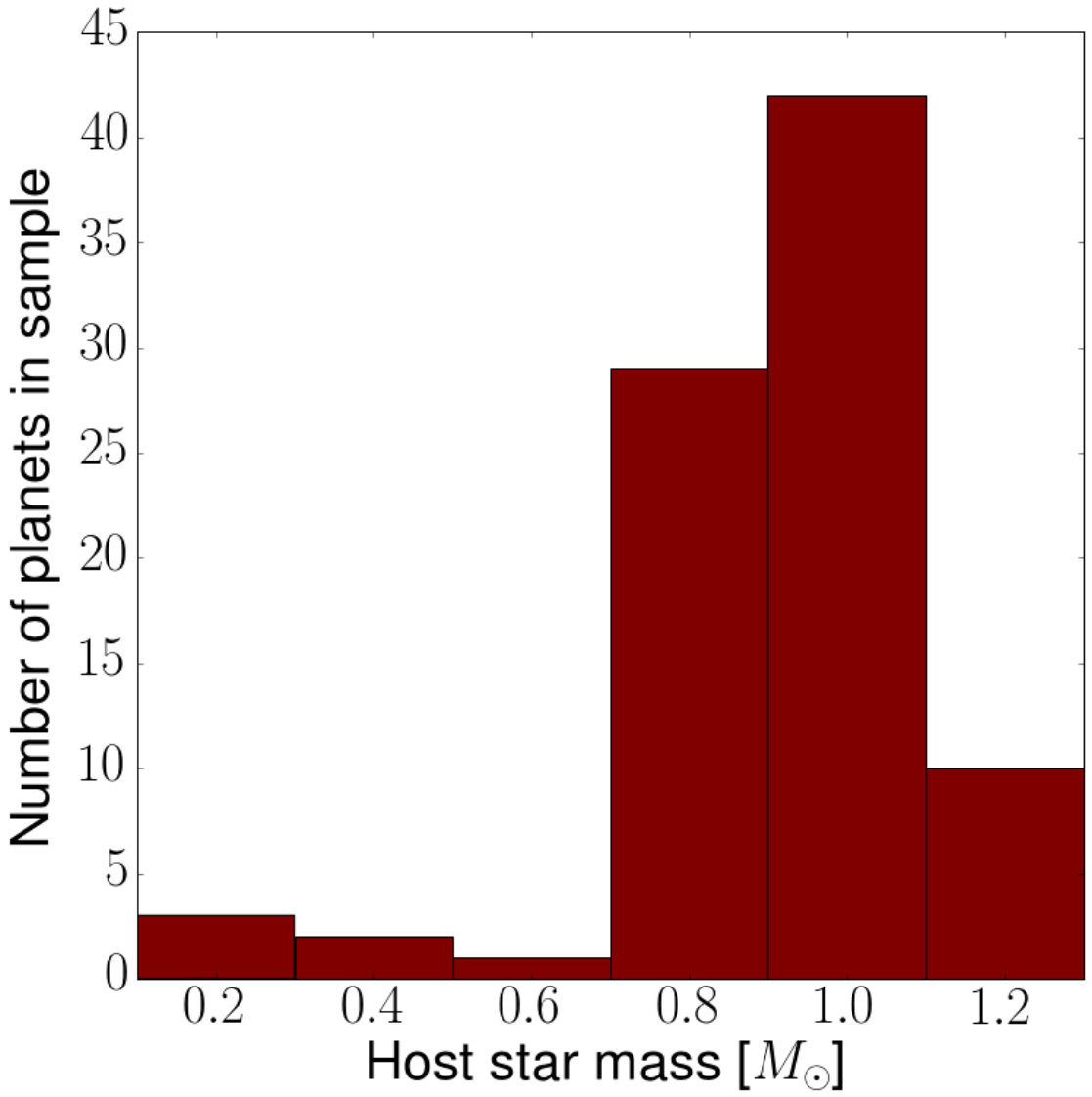

Figure 2.1: A histogram of planet host star mass in our sample. We use a sample of 87 transiting planets with radius $< 8R_\oplus$ and RV-measured masses. Of those 87, only six orbit low-mass ($< 0.7M_\oplus$) stars: GJ 436, GJ 1132, GJ 1214, GJ 3470, LHS 1140 and K2-18. The median host star mass of the sample is $0.95M_\odot$.

2.3.2 Model

Since we are investigating the dependence of the M-R relation on host star mass, we choose to start with the same framework as Wolfgang et al. (2016) in order to isolate this dependence. At a given radius, a planet's mass is drawn from a Gaussian distribution where the mean of the Gaussian distribution is a power-law:

$$\frac{M}{M_\oplus} \sim \text{Normal}\left(\mu = C\left(\frac{R}{R_\oplus}\right)^\gamma, \sigma = \sigma_M\right) \tag{2.1}$$

and the standard deviation parametrizes the intrinsic scatter. The fact that the mass is drawn from a distribution makes this a probabilistic, rather than deterministic, relation.

In Wolfgang et al. (2016), C, γ and σ_M are the three parameters to be fit to the data. In our case, we expand each of these to include host star mass dependence:

$$\gamma = \gamma_0 + \ln\left(\frac{M_*}{M_\odot}\right)\gamma_s \tag{2.2}$$

$$C = C_0 + \ln\left(\frac{M_*}{M_\odot}\right)C_s \tag{2.3}$$

$$\sigma_M = \sqrt{\sigma_0^2 + \ln\left(\frac{M_*}{M_\odot}\right)\sigma_s} \tag{2.4}$$

where we have introduced three new parameters that parameterize the host star dependence: γ_s, C_s, σ_s. In the case where these three parameters are all zero, then we obtain the original M-R relation in Wolfgang et al. (2016), which is independent of host star mass. In Equations (2.2 - 2.4), M_* is the mass of the planet's host star, such that at a solar mass $M_\odot$, the host star mass dependent terms drop out, and the M-R relation is given solely by γ_0, C_0 and σ_0. Parameterizing the M-R relation in this manner allows both the mean planet mass and intrinsic scatter to smoothly vary as a function of host star mass. This specific choice of

parametrization is chosen for simplicity, similar to the choice of parameterizing the M-R relation as a power-law. Our aim is to allow the M-R relation to change flexibly with host star mass, while still minimizing the number of parameters.

Wolfgang et al. (2016) considered an additional model, where the intrinsic scatter in mass is allowed to vary as a function of planet radius. While this may be physically reasonable, they find that this parameter is consistent with zero and there is no strong evidence for requiring its inclusion. For this reason and for the sake of simplicity, we do not include any dependence of the intrinsic scatter on radius. However, following Wolfgang et al. (2016), we do include a maximum density constraint, based on the mass of a planet composed of a pure iron core. At a given planet radius, the maximum physically plausible planet mass is given by:

$$\log_{10}\left(M_{\text{pureFe}}\right) = \frac{-b + \sqrt{b^2 - 4a(c - R)}}{2a} \tag{2.5}$$

where $a = 0.0975, b = 0.4938$ and $c = 0.7932$ (Fortney et al., 2007a). Imposing this constraint has the effect of truncating the normal distribution of masses at a given radius. This is more constraining for small planet radii and severely limits the range of masses that these small planets can have. At around $1.5R_{\oplus}$, this maximum mass limit no longer constrains the M-R relation significantly. In this way, although we have not included a parameter that allows the intrinsic scatter to vary with radius, the range of masses is being constrained at small radii by this maximum mass limit.

A graphical representation of our model is shown in Figure 2.2. As input for our model, we take the following for each planet: observed radial velocity semi-amplitude K_{obs} with uncertainty σ_K, observed planet radius to host star radius ratio r_{obs} with uncertainty σ_r, period P, eccentricity e, and inclination i. We also take as input a list of observed host star masses $M_{*,\text{obs}}$ with uncertainty σ_{M_*}, as well as observed stellar radii $R_{*,\text{obs}}$ with uncertainty σ_{R_*}. We only include each host star once such that multiple planets around the same star will

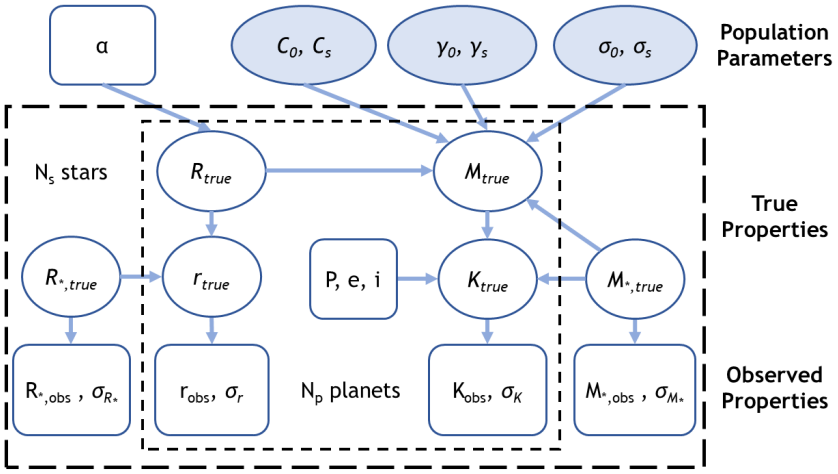

Figure 2.2: A graphical model of our host star mass dependent M-R relation as given by Equations (2.1 - 2.7). Rectangles represent input or assumed data; circles are parameters that are fitted for in our model. The shaded circles are the population level parameters of interest, whereas unshaded circles are inferred latent parameters.

be drawing from the same host star properties during each step of the fitting. If asymmetric error bars were reported for a given measurement, we use the average of the two as the uncertainty.

For each planet, we introduce the parameters $R_{\rm true}$, $M_{\rm true}$ which represent the true radius and mass of that planet. The true planet radii are assumed to follow some underlying distribution given by α (a uniform distribution for the purposes of this paper), while the true planet masses are given by our host star mass dependent M-R relation. The true planet masses and radii are both restricted to be above zero, and the true planet masses are restricted to be below the iron density constraint given above. We also introduce parameters corresponding to the true values of our observables: $K_{\rm true}$, $r_{\rm true}$, $M_{*,\rm true}$, $R_{*,\rm true}$. $K_{\rm true}$ is calculated from the planet's period, inclination, eccentricity, true mass and host star mass while $r_{\rm true}$ is calculated from the planet's true radius and host star radius, as shown below.

$$
\begin{aligned}
K_{\rm true} &= 0.6387 \left(\frac{P}{\rm day}\right)^{-1/3} \left(\frac{M_{\rm true}}{M_\oplus}\right) \frac{\sin i}{\sqrt{1-e^2}} \\
&\times \left(\left(\frac{M_{*,\rm true}}{M_\odot}\right) + 3\times 10^{-6}\left(\frac{M_{\rm true}}{M_\oplus}\right)\right)^{-2/3} \quad [m/s] \\
r_{\rm true} &= 0.009168 \left(\frac{R_{\rm true}/R_\oplus}{R_{*,\rm true}/R_\odot}\right)
\end{aligned}
\tag{2.6}
$$

Our full hierarchical model, including our list of priors for these parameters, is shown in Equations (2.7 - 2.8). U represents a uniform distribution, N represents a normal distribution, T represents truncation bounds, and $\sim$ represents "is distributed as".

$$
\begin{aligned}
K_{\rm obs} &\sim {\rm N}(K_{\rm true}, \sigma_K) \\
r_{\rm obs} &\sim {\rm N}(r_{\rm true}, \sigma_r) \\
M_{*,\rm obs} &\sim {\rm N}(M_{*,\rm true}, \sigma_{M_*}) \\
R_{*,\rm obs} &\sim {\rm N}(R_{*,\rm true}, \sigma_{R_*}) \\
M_{\rm true} &\sim {\rm N}(\mu, \sigma_M)\ T(0, M_{\rm true,Fe})
\end{aligned}
\tag{2.7}
$$

$$
\begin{aligned}
C_0 &\sim \mathrm{N}(5, 10)\ T(0,) \\
C_s &\sim \mathrm{N}(0, 10) \\
\gamma_0 &\sim \mathrm{N}(1, 1) \\
\gamma_s &\sim \mathrm{N}(0, 1) \\
\sigma_0 &\sim \mathrm{N}(3, 10)\ T(0,) \\
\sigma_s &\sim \mathrm{N}(0, 20) \\
R_{\mathrm{true}} &\sim \mathrm{U}(0.2, 20) \\
M_{*,\mathrm{true}} &\sim \mathrm{U}(0.07, 2) \\
R_{*,\mathrm{true}} &\sim \mathrm{U}(0.05, 3)
\end{aligned}
\tag{2.8}
$$

2.3.3 Fitting the Model

To fit our model to the data, we use the Python implementation of the Stan statistical software package (Carpenter et al., 2017)[2]. Stan uses the No-U-Turn sampler (NUTS) MCMC algorithm, an efficient method of numerically evaluating hierarchical Bayesian models. For our host star mass dependent model, we ran 8 chains each with 30,000 iterations. The first 5000 iterations of each chain are thrown out as "burn-in" to allow the chain to reach its equilibrium distribution, such that we are left with 200,000 posterior samples of each parameter. To assess the convergence and independence of each chain, we look at the output effective sample size (ESS) and Gelman-Rubin convergence diagnostic, $\hat{R}$. For each parameter, we ensure that $\hat{R} < 1.01$ and that the ESS is large. For our six parameters of interest, the ESS are all $> 20,000$ while for our true planet and host star parameters, the ESS are all $> 40,000$. Given these diagnostics, we assume the posteriors have converged to an acceptable degree.

2. http://mc-stan.org

Parameter	15.9%	Median	84.1%
C_0	2.45	2.97	3.52
C_s	-1.77	-0.26	0.67
γ_0	1.16	1.29	1.43
γ_s	-0.09	0.22	0.61
σ_0	2.59	3.35	4.31
σ_s	-12.06	-4.18	1.39

Table 2.1: Summary statistics of the six population level parameters from our model posteriors.

2.4 Results

2.4.1 Model fit

Table 2.1 summarizes the posteriors for our host star mass dependent M-R relation, and they are visualized in Figure 2.3. We find that the three parameters encoding host star mass dependence $\{C_s, \gamma_s, \sigma_s\}$ are each consistent with zero. This can be seen in the set of 2D contour plots shown in Figure 2.3, where the zero points for these three parameters (shown in blue) all lie within the 1σ contours. Further, the median values of C_0, γ_0 and σ_0 all closely match those found in Wolfgang et al. (2016) and are much less sensitive to the priors than the three host star mass parameters, suggesting that the M-R relation does not have a strong dependence on host star mass.

Though the host star mass parameters are all consistent with zero, it is possible that the current mass-radius dataset is insufficient to robustly reveal any true host-star mass dependencies in the planet M-R distribution. Our analysis constrains the extent to which the host star mass parameters can deviate from zero; all the posterior distributions in Figure 3 are more strongly peaked than the assumed priors. The distribution of γ_s peaks at $\sim$0.25, indicating slight preference for a steeper slope in the M-R relation towards higher host star masses, but is only discrepant from 0 at a 0.43σ level. The host star mass dependent scatter, σ_s, shows the most evidence for being discrepant from 0, with a median of -4.18 and 75%

of the posterior samples being below zero. This would suggest a larger scatter in the M-R relation toward lower host star masses. However, given that all three parameters do not rule out no host star mass dependence, we stress that this is only what the current dataset suggests with this model, and may not reflect underlying trends in the planet population.

Figure 2.4 shows our host star mass dependent M-R relation marginalized over the posterior distribution for two different host star masses, representing M stars (in red) and F stars (in blue). As previously stated, the model prefers a shallower M-R relation with higher intrinsic scatter towards lower host star mass. This difference is most apparent at higher planet radii, with the spread of planet masses for rocky planets ($< 1.5R_\oplus$) being negligibly different between M and F host stars. Two of the most dominant limiting factors for this model are the small number of transiting planets with RV measured masses with $2.5 < R/R_\oplus < 8.0$, as well as the small number of those planets orbiting M dwarfs. This can be seen in Figure 2.4, where the majority of planets fall between $1.5R_\oplus$ and $2.5R_\oplus$, and orbit solar-mass stars. These two limitations are largely due to the selection criteria of RV followup, where brighter (up to $T_{\rm eff} \sim 6200$K) stars and lower radii planets are preferred, in the quest to find and characterize small, potentially rocky planets.

2.4.2 Incident Flux Dependence

As M dwarfs can have luminosities hundreds to thousands of times fainter than their GK counterparts, planets with the same period will have vastly different incident fluxes from their host star depending on the type of star. Overall, given similar period distributions, we expect planets around M dwarfs to have much lower incident fluxes. Incident flux has clear physical connections to planet composition. For example, there is a lack of ultra-short period planets with radius $< 2R_\oplus$, thought to arise from photoevaporation (Sanchis-Ojeda et al., 2014) due to the extreme incident fluxes of these close-in planets. Evidence for photoevaporation can also be found in *Kepler* radius distribution, which has been shown to have a gap at

$\sim 1.8R_\oplus$ (Fulton et al., 2017). Therefore, there is the possibility that any differences in the M-R relation due to host star mass can be attributed to different distributions of incident flux. We briefly explore incident flux dependence in the M-R relation and examine its effects on the host star mass dependence.

We model incident flux dependence into the M-R relation in an analogous fashion to the host star mass dependence in our primary model. We introduce three new parameters $\{C_f, \gamma_f, \sigma_f\}$ that modify the power-law slope, normalization and intrinsic scatter and are scaled by the natural log of the incident flux. Equations (2-4) then become:

$$C = C_0 + \ln\left(\frac{M_*}{M_\odot}\right)C_s + \ln\left(\frac{S}{100S_\oplus}\right)C_f \tag{2.9}$$

$$\gamma = \gamma_0 + \ln\left(\frac{M_*}{M_\odot}\right)\gamma_s + \ln\left(\frac{S}{100S_\oplus}\right)\gamma_f \tag{2.10}$$

$$\sigma_M = \sqrt{\sigma_0^2 + \ln\left(\frac{M_*}{M_\odot}\right)\sigma_s + \ln\left(\frac{S}{100S_\oplus}\right)\sigma_f} \tag{2.11}$$

where the incident flux S for a given planet is calculated with the following equation:

$$\frac{S}{S_\oplus} = \left(\frac{R_*}{R_\odot}\right)^2 \left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right)^4 \left(\frac{M_*}{M_\odot}\right)^{-2/3} \left(\frac{P}{P_\oplus}\right)^{-4/3} \tag{2.12}$$

Our incident flux dependent model is then a nine parameter model, where each different combination of incident flux and host star mass gives a distinct M-R relation.

The posterior distributions for $\{C_s, \gamma_s, \sigma_s\}$ for our standard six parameter model compared to the nine parameter incident flux dependent model are shown in Figure 2.5. We find that the posterior distributions for C_s and γ_s widen when incident flux dependence is introduced. We also note that while the original model prefers shallower slopes toward lower mass stars, with incident flux dependence this trend is reversed. With the current dataset, it is difficult to disentangle the effects of incident flux and host star mass on the M-R rela-

tion. Both may affect planet composition, but currently there is no empirical proof for either affecting the M-R relation and no evidence to prefer one dependence over the other.

While this model does include both host star mass and incident flux dependence, this extra set of parameters is not justified by the limited dataset available. In section 2.5.1 we demonstrate by information criterion considerations that this model is not strongly preferred over the six parameter model. For this reason and for simplicity, we stick to the six parameter host star mass dependent M-R relation as the primary relation in the following sections. We have tested the conclusions of the following sections with the incident flux dependent model and have found the conclusions to be unchanged.

2.4.3 Exploring Other Datasets

TTV Dataset

As a consistency check, we repeat the modeling above using the TTV dataset instead of the RV dataset, as well as the combined RV + TTV datasets. We compiled the TTV dataset from the NASA Exoplanet Archive, with an identical cut on planet radius of $R < 8R_{\oplus}$ as well as a cut on planet mass of $M < 25M_{\oplus}$ to remove planets with physically unlikely masses. For TTV planets, we replace the radial velocity semi-amplitude in our model with the planet to host star mass ratio, M/M_{*}. We derive these values from the reported masses and host star masses, which is valid because host star uncertainties do not factor into TTV modeling (Holman et al., 2010). For the combined RV and TTV dataset, if a planet has both RV and TTV mass measurements, we assume the RV measurement. The TTV dataset contains 63 planets compared to 88 RV planets, but has the advantage of more planets orbiting low-mass stars (13 planets with host star masses below $0.7M_{\odot}$, compared to 6 planets with RV), although they come from fewer distinct systems. Due to overlap between the two datasets, the combined RV + TTV dataset has 143 planets.

The posterior distributions of the three host star mass dependent parameters for our

standard model fitted to the RV, TTV and RV + TTV datasets are shown in Figure 2.6. With the TTV dataset, we find that there is no evidence for host star mass dependence, as the posteriors for the three host star mass dependent parameters are consistent with zero. The median value of γ_s using the TTV dataset is -0.08, which translates to a slightly shallower power law slope towards higher host star masses, a trend opposite to what we found with the RV dataset. Rather than revealing anything insightful about the underlying population, this is likely a result of our limited sample and further evidence for no host star mass dependence evident in the current sample of planets. The median value of σ_s is -1.88 using the TTV dataset, in the same direction as the RV dataset. With the RV+TTV dataset, the posteriors tighten, but are still centered at 0.

Potential TESS Dataset

The lack of evidence for host star mass dependence in the M-R relation from fitting our model to the RV, TTV and RV + TTV datasets begs the question of how much data would be needed to demonstrate a dependence if such a dependence were to exist. The biggest hope for the near future is the *TESS* survey and subsequent radial velocity follow-up, given the increased number of M dwarf planets *TESS* expects to find compared to *Kepler*. Using a simulated catalog of *TESS* planets from Sullivan et al. (2015), we construct a hypothetical future dataset of transiting planets with RV measured masses. We assume that each planet orbiting an M dwarf with a V magnitude above some cut will have a radial velocity mass measurement. We use a cut of $V_{\rm mag} < 14$ for potentially rocky planets ($\frac{R_p}{R_\oplus} < 2$) and a cut of $V_{\rm mag} < 13$ for those with gaseous envelopes ($2 < \frac{R_p}{R_\oplus} < 8$) to simulate the preference of many radial velocity follow-up programs to characterize rocky planets. We perform similar cuts on the FGK sample, using a cut of $K_{s,\rm mag} < 10$ for rocky planets and $K_{s,\rm mag} < 7$ for gaseous planets. We then generate radial velocity semi-amplitudes for each planet using two different models: the first using our six-parameter model

with $\{C_0 = 3.0, C_s = 0.5, \gamma_0 = 1.3, \gamma_s = -0.5, \sigma_0 = 2.0, \sigma_s = -3.0\}$, and the second using a three-parameter model with no host star mass dependence $\{C = 3.0, \gamma = 1.3, \sigma = 2.0\}$. We make a further cut on the sample to ensure each planet has $K_{\rm expected} > 1$m/s (taken from Sullivan et al. (2015)), in line with the sensitivity of radial velocity spectrographs such as *MAROON-X* (Seifahrt et al., 2016). Finally, we generate normally distributed fractional errors for K, r, R_*, M_* using values similar to those found in the current sample of small planets with mass and radius measurements. These mean fractional errors are 0.2, 0.015, 0.03, and 0.1, respectively. The true values of these parameters are perturbed with these errors to simulate our observables.

We fit both datasets to each of the two models used to generate them: the six-parameter model with host star mass dependence and the three-parameter model without. We do this for 10 realizations of each dataset. We then calculate the difference in WAIC (further discussed in section 2.5.1) for each dataset between the two models, as a means to test whether we can successfully distinguish which is the correct model. We find that this information criterion correctly prefers the model with no host star mass dependence for the dataset generated without host star mass dependence in 9 of 10 realizations. Further, compared to the smaller, current RV dataset, the 1σ uncertainties for the three host star mass dependent parameters decrease by a factor of ~ 3. For the dataset generated with host star mass dependence, we similarly find that 10 out of 10 realizations correctly preferred the model with host star mass dependence by at least 2σ. We also note that 7 out of 10 realizations excluded 0 for C_s at a 1σ level, with 10 out of 10 and 4 out of 10 realizations excluding zero for γ_s and σ_s, respectively.

These results suggest that, while variance is still a factor, for this set of model parameters we would likely be able to find evidence for host star mass dependence with a future *TESS* dataset. However, these results are contingent upon a few factors. While we generated 10 realizations of the radial velocity measurements, each realization used the same set of radii

taken from the simulated catalog of Sullivan et al. (2015). The true variance between datasets should be higher, if radii were generated alongside masses. We used a simplistic treatment of both the errors on the masses and radii, as well as the radial velocity follow-up strategy. The real catalog may look much different from the ones studied here. Further, we only tested one set of model parameters, and only one model parametrization with host star mass dependence: there are many plausible combinations of these parameters that may describe the underlying M-R relation. Despite these limitations, we present this as one method of determining whether or not *TESS* may help us distinguish between M-R relations with and without host star mass dependence.

2.4.4 Planet Mass Distributions

The *Kepler* survey has provided a wealth of potential planets, with over 4500 planet candidates discovered around a wide variety of host stars. This large sample of planet candidates has enabled many exoplanet population studies, including how exoplanet systems may differ around various types of stars. A pair of papers by Mulders et al. (2015b,a) (hereafter MPA15a,b) has provided evidence for increasing heavy-element mass in planetary systems in short orbits (period between 2-150 days) around lower-mass stars. They find that the average heavy-element mass in short orbits scales roughly inversely with host star mass, from 3.6 $M_\oplus$ around F stars to 7.3 $M_\oplus$ around M stars. This trend is at apparent odds with the observed trend of protoplanetary disk masses increasing with host star mass, derived from millimeter-wave observations (Andrews et al., 2013).

As with any analysis of the *Kepler* sample involving planet masses, the results in MPA15b are heavily reliant on the assumed M-R relation. They derived their results using both the deterministic M-R relation in Weiss & Marcy (2014) as well as the probabilistic M-R relation in Wolfgang et al. (2016), and found the trend to be robust to the assumed M-R relation. However, they use the best-fit deterministic relation from Wolfgang et al. (2016) rather than

using the truly probabilistic formulation. More specifically, they did not take into account the intrinsic scatter of the M-R relation or the maximum density constraint derived from a pure iron core. This would have the effect of overestimating planet masses at small radii, which could skew the observed trend given that smaller planets have higher occurrence rates around lower mass stars. Additionally, uncertainties in the adopted M-R relation parameters are not taken into account, which would cause the errors reported to be underestimated. In this section, we seek to test the robustness of this trend, and to see to what extent it depends on the assumed planet M-R relation. We will use the posteriors from our host star mass dependent M-R relation for this purpose. We largely follow the methodology of MPA15b in deriving planet occurrence rates, which we briefly describe below.

For the purposes of this calculation, we use the Q1-16 KOI (Mullally et al., 2015) catalog along with the corresponding list of target stars observed in these quarters. In order to isolate the effect of using our own M-R relation posteriors, we ensure that our KOI catalog completely matches that used in MPA15b. Giant stars are removed from the stellar catalog given their position in $\log g$-$T_{\rm eff}$ space according to the prescription in Ciardi et al. (2011). Stellar noise during a transit is derived from the Combined Differential Photometric Precision (CDPP) metric (Christiansen et al., 2012) at 3, 6 and 12 hour timescales for each observing quarter, which were downloaded from the MAST archive[3]. For each star, we take the median of the CDPP of all quarters at each timescale and fit a power law. The CDPP for a given star and a given transit duration is then calculated using the star's CDPP power law fit.

The occurrence rate $f_{\rm occ}$ of a given planet with radius and orbital period R_p, P in a bin of stellar effective temperature $T_{\rm eff}$ is defined as the inverse of the number of stars in that bin for which the planet would be detectable N_*, multiplied by a factor $f_{\rm geo}$ to account for the geometric probability for the planet to transit:

3. http://archive.stsci.edu/kepler/data_search/search.php

$$f_{\rm occ}(\{T_{\rm eff}\}, R_p, P) = \frac{1}{f_{\rm geo} N_*(\{T_{\rm eff}\}, R_p, P)} \tag{2.13}$$

We use $T_{\rm eff}$ bins that correspond to M, K, G and F stars based on the recommendation of the Exoplanet Study Analysis Group (SAG) 13[4], with upper bin edges of 3900, 5300, 6000, and 7300 K. For each star, the S/N of the planet is calculated given the transit depth, the number of transits and the stellar noise that the system would have with the planet orbiting that star. The detection efficiency given that S/N is then calculated based on the number of transits and a cumulative gamma function that is empirically derived from planet transit injection and recovery tests (Christiansen et al., 2015). For detailed calculations, see equations (2-9) in MPA15a, as well as MPA15b.

In order to calculate planetary heavy-element mass as a function of host star mass, we first sample from our host star mass dependent M-R relation to obtain a set of the six parameters $\{C_0, C_s, \gamma_0, \gamma_s, \sigma_0, \sigma_s\}$. At the same time, we resample the KOI population by bootstrapping (drawing N_p planets from the sample with replacement). Then, for each planet we sample N_m masses using the set of host star mass dependent M-R relation parameters, as well as sampling the planet's radius and host star mass from normal distributions. To retrieve the heavy-element mass instead of the total mass, we cap the mean of the probabilistic M-R relation at $22 M_\oplus$ in order to replicate the measured median core masses of giant planets (Miller & Fortney, 2011). Each mass is sampled from a truncated normal distribution where the mean and standard deviation are given by our M-R relation, and the truncation is due to the maximum density of a planet with a pure iron core. For each sample of a planet's mass, we calculate the planet's contribution to the average heavy-element mass by multiplying its mass by the occurrence rate of the planet in its $T_{\rm eff}$ bin. We repeat the outer level sampling of the M-R relation parameters and bootstrap resampling of the population N_b times.

Our results are shown in Figure 2.7. For these results, $N_m = 25$ and $N_b = 500$. The

4. https://exoplanets.nasa.gov/exep/exopag/sag/

Assumptions	M	K	G	F
1) MPA15b	$11.2\ ^{+4.9}_{-3.6}$	$7.4\ ^{+1.6}_{-1.3}$	$7.0\ ^{+1.2}_{-1.0}$	$5.4\ ^{+1.1}_{-0.9}$
2) $M_Z \propto \sqrt{M_p}$	$10.6\ ^{+4.2}_{-3.1}$	$6.7\ ^{+1.5}_{-1.3}$	$6.2\ ^{+1.2}_{-1.0}$	$4.9\ ^{+1.1}_{-0.9}$
3) Snowline scaling + 2	$8.4\ ^{+3.1}_{-2.3}$	$5.9\ ^{+1.3}_{-1.1}$	$6.1\ ^{+1.1}_{-1.0}$	$5.0\ ^{+1.1}_{-0.9}$
4) Incident flux dep. + 3	$8.4\ ^{+3.2}_{-2.4}$	$5.7\ ^{+1.6}_{-1.3}$	$5.6\ ^{+1.7}_{-1.2}$	$4.7\ ^{+1.6}_{-1.1}$
5) TTV + 3	$5.9\ ^{+3.2}_{-1.9}$	$3.9\ ^{+1.1}_{-0.9}$	$3.9\ ^{+0.9}_{-0.8}$	$3.1\ ^{+0.9}_{-0.7}$
6) TTV + 4	$7.0\ ^{+3.5}_{-2.5}$	$4.4\ ^{+1.6}_{-1.2}$	$4.4\ ^{+1.5}_{-1.1}$	$3.8\ ^{+1.7}_{-1.2}$
7) RV+TTV + 3	$7.6\ ^{+3.1}_{-2.2}$	$5.2\ ^{+1.4}_{-1.1}$	$5.2\ ^{+1.5}_{-1.1}$	$4.4\ ^{+1.3}_{-1.1}$
8) RV+TTV + 4	$7.6\ ^{+3.0}_{-2.2}$	$5.4\ ^{+1.2}_{-1.0}$	$5.5\ ^{+1.1}_{-0.9}$	$4.6\ ^{+1.0}_{-0.9}$

Table 2.2: Average planet heavy-element mass per star, under different assumptions. Assumptions are described more explicitly in the text.

data shown is the median over N_b and N_m samples, with error bars representing the 15.9 and 84.1 percentiles. For each $T_{\rm eff}$ bin, we plot the corresponding host star mass by taking the median of the masses of each star in the bin, with error bars again representing the 15.9 and 84.1 percentiles. We find that our results are consistent with the findings in MPA15b. The average planetary heavy-element mass increases from $5.4^{+1.1}_{-0.9} M_\odot$ around F stars to $11.7^{+4.9}_{-3.6} M_\odot$ around M stars. With respect to MPA15b, our error bars are roughly twice as large and our average masses slightly higher, but the overall trend is consistent.

With the assumed model of the host star mass dependent M-R relation, our results are consistent within error with neither a scaling similar to the scaling of protoplanetary disk masses with host star mass, nor a flat trend. We fit a power-law to the results and find a power-law index of -0.9. We also fit a power-law to each of the N_b individual samples of the planet population and M-R relation parameters. These are shown as faded lines in Figure 2.7 and give a rough visualization of the spread of this trend with host star mass.

We have tested several combinations of parameters to see what would be consistent with a flat trend of planetary heavy-element mass with host star mass. Keeping $\{C_0, \gamma_0, \sigma_0\}$ fixed to their median values as listed in Table 2.1, we find that the following combination of $\{C_s = 1, \gamma_s = 1.6, \sigma_s = -5\}$ is sufficient to produce a flat trend. This results in a flat M-R relation for planets with $R > 1.3 R_\oplus$ orbiting M dwarf ($M \sim 0.42 M_\odot$) stars. As seen in our

model posteriors in Figure 2.3, this set of parameters is well outside the 2σ contours and is heavily disfavored by the current dataset.

In order to test the robustness of this trend in heavy-element mass, we would like to know to what extent the increase at lower host star mass is dependent on our assumptions. First, we test the assumption that heavy-element mass is equivalent to planet mass up to a limiting mass, chosen to be $22M_\oplus$ in MPA15b due to the median core mass of giant planets. Recent evidence shows that the correlation between heavy-element mass and total mass for giant planets roughly scales as $M_Z \propto \sqrt{M}$, based on thermal and structural evolution models (Thorngren et al., 2016). This correlation is fit to masses above $\sim 20M_\oplus$. We adopt this scaling for planets with $R > 4R_\oplus$, which roughly corresponds to a mass of $20M_\oplus$ in our model, with $M_Z \propto M$ below $4R_\oplus$. To smoothly transition between these two functions, we adopt a logistic function with a transition point at $4R_\oplus$ and a scale parameter of 1. For giant planets with $R > 10R_\oplus$, the M-R relation changes to become approximately flat (Chen & Kipping, 2017), and so we assume the heavy-element mass is flat past this radius.

The second assumption we test is the period cut we apply on the planet sample. A planet with a period of 150 days has a significantly different irradiation depending on what type of star it is orbiting. By applying a cut of $2 - 150$ days regardless of host star mass, we are applying an uneven cut in planet irradiation. To account for this, we use the location of the snowline as a proxy for planet irradiation and use the scaling found in Ida & Lin (2005), where $a_{\rm snowline} = 2.7{\rm AU}(M_*/M_\odot)$. This is equivalent to a linear scaling of the period with host star mass, by Kepler's third law. We take $2 - 150$ days to be the period bounds for a host star with $1M_\odot$, and scale both the inner and outer bound by $M_*/M_\odot$. For example, for a planet with a host star mass of $0.5M_\odot$, it is only included in the sample if its period falls between $1 - 75$ days.

For these two cases (where we also include the M_Z scaling in the second case), along with our initial assumptions, we calculate the average heavy element mass per star in bins

of heavy element mass, for each of our four host star mass bins. We then integrate this curve along the heavy element mass axis to obtain the total average heavy element mass per star. The results are shown in Table 2.2. We find that scaling the heavy element mass as $\sqrt{M}$ decreases the total heavy element mass across all four host star mass bins. Scaling the period bounds with the location of the snowline decreases the heavy element mass for M and K stars, making the slope shallower. Despite changing our assumptions for the heavy element mass and period scaling, the trend of increasing heavy element mass towards lower host star mass is still present. Figure 2.8 shows the contribution of different heavy element mass regimes to the total heavy element mass per star; the integral of each curve in Figure 2.8 gives the total heavy element mass shown in Figure 2.7. Both the snowline and heavy element mass scalings are included in Figure 2.8. We can see that the increase in the average heavy element mass for M dwarfs is due to planets with heavy element masses up to $\sim 16 M_\oplus$ contributing more mass than those around F, G and K stars. Only at masses higher than $16 M_\oplus$ do planets contribute more heavy element mass for F, G and K stars compared to M stars.

Using similar methodology to calculating the planet heavy-element mass, we also calculate mass distributions for *Kepler* planets around M stars as well as FGK stars, shown in Figure 2.9. For this calculation, we use the period range scaling with the snowline, and we bin by planet mass as well as stellar $T_{\rm eff}$. We find that planet occurrence is higher for planets with $3 < M/M_\oplus < 32$ around M dwarfs than FGK dwarfs. Beyond $32 M_\oplus$, we are limited by the small number of planets discovered around M dwarfs, as well as the limited radius range to which we fit the M-R relation. Below $3 M_\oplus$, the *Kepler* sample is thought to be incomplete for the period range under consideration (Christiansen et al., 2015).

We find an occurrence rate of $1.03^{+0.24}_{-0.20}$ planets per M dwarf with $M < 10 M_\oplus$ and $1 < P < 100$ d. Compared to the *HARPS* M dwarf sample for the same mass and period bounds which found 0.88 planets per star (Bonfils et al., 2013), our derived result is a factor

of 1.17 higher but consistent within 1 sigma. We find $0.30^{+0.19}_{-0.12}$ planets per M dwarf with $10 < M/M_{\oplus} < 100$ and $1 < P < 100$ d, which is a factor of 6 higher than the *HARPS* result of 0.05. We attribute this higher frequency of massive planets to our sampling of the masses of each planet, which has the effect of flattening the distribution. Since our masses are more uncertain than directly measured radial velocity masses, planets with smaller radii at higher occurrence rates will have some significant posterior probability at higher masses.

2.4.5 Minimum-mass extrasolar nebula

The minimum-mass solar nebula (MMSN) is an estimate of how much mass must have been in the solar protoplanetary disk to form the planets had they formed in situ. Kuchner (2004) first showed how one could extend the concept of the MMSN to exoplanets, using a sample of radial velocity planets in order to calculate the minimum-mass extrasolar nebula (MMEN). More recently, Chiang & Laughlin (2013) calculated the surface density profile of the MMEN using the *Kepler* sample of transiting planets. They find a similar power-law slope compared to the MMSN (-1.6 and -1.5, respectively) and an overdensity compared to the MMSN of about a factor of five. However, their primary result has several limitations. The M-R relation they use is the one derived in Lissauer et al. (2011): a power-law fitted to six solar system planets, which has since been greatly improved upon by utilizing the sample of transiting exoplanets with mass measurements. The host star mass for each individual Kepler planet is not taken into account, neither for the calculation for orbital radius nor for the M-R relation. Finally, the occurrence rate of each individual planet is not factored in. Here, we redo the MMEN calculation, making several improvements over the initial work.

We use the *Kepler* occurrence rates, planet heavy-element mass and host star mass samples as described in section 2.4.4 for the following calculation. For each sample, the solid surface density of the planet is given by the following equation:

$$\Sigma_{\rm solid} = \frac{M_Z}{2\pi a^2} \tag{2.14}$$

In Figure 2.10, we plot a 2D weighted histogram of the MMEN surface density of solids and semimajor axis for M, K, G, and F dwarfs separately, where the weights are the occurrence rates of the KOIs. We find that the surface density of solids at short orbits is higher for M dwarfs than it is for FGK dwarfs. Between F, G and K dwarfs there is little difference, mirroring Figure 2.8. Compared to Chiang & Laughlin (2013), we find that the surface density profile of the MMEN for FGK dwarfs exhibits a shallower slope with a power-law index ranging from -0.9 to -1.2, whereas the power-law index for M dwarfs, -1.8, more closely matches their result of -1.6. We find a similar normalization for the surface density of solids, which is several factors higher than that of the MMSN. While the MMEN suggests a higher surface density of solids at short orbits for disks around M dwarfs, migration likely plays a pivotal role in this trend and thus the MMEN may not reflect the true initial protoplanetary disk surface density profile (Mulders et al., 2015a). This is supported by recent work from Raymond & Cossou (2014), which showed that the surface density profiles of multiple planet systems have a wide range of power-law slopes.

2.5 Discussion

2.5.1 Model selection

In order to test whether we are justified in adding host star mass dependence to the M-R relation, we would like to estimate the predictive accuracy of models with and without host star mass dependence. Cross-validation is a robust method of evaluating the predictive accuracy of a model, but requires multiple model fits and is computationally expensive. Luckily, several alternatives exist which approximate cross-validation and are simpler to compute. For the purposes of this paper, we choose Watanabe-Akaike information criterion (WAIC)

Model	RV	TTV	RV+TTV
No M_* dependence	-14.8 ± 4.6	-8.4 ± 2.4	-17.3 ± 4.6
C_s	-14.8 ± 2.9	-5.2 ± 2.1	-17.0 ± 2.7
γ_s	-11.7 ± 2.1	-7.4 ± 2.2	-11.9 ± 2.6
σ_s	-7.5 ± 4.0	-2.8 ± 1.6	-9.4 ± 3.7
C_f, γ_f, σ_f	-5.9 ± 3.8	2.5 ± 2.2	-8.0 ± 3.3
$C_s, C_f, \gamma_s, \gamma_f, \sigma_s, \sigma_f$	4.3 ± 3.2	40.1 ± 6.3	-1.7 ± 4.7

Table 2.3: Difference in Watanabe-Akaike information criteria (WAIC) compared to the standard six-parameter host star mass dependent model, along with the error in the difference. Results are shown for three datasets and six different models. Negative numbers favor the model in question over the six-parameter model. We find that models with fewer parameters are favored.

to be our predictive measure of choice (Watanabe, 2012). Similar to other information criterion, WAIC relies on computing the log predictive density for the data sample used to fit the model, and applying a bias correction to estimate the log predictive density for a hypothetical new data sample. Unlike alternatives such as AIC and DIC, WAIC is fully Bayesian in the sense that it averages over the posterior distributions, rather than relying on point estimates (Gelman et al., 2014). Given that the posteriors for our six parameters are not all normally distributed (σ_s in particular), this makes WAIC an appealing choice.

We compute WAIC for seven different models: our six-parameter host star mass dependent M-R relation, a three-parameter model with no host star mass dependence, three four-parameter models that correspond to adding one of C_s, γ_s, and σ_s to the three-parameter model, our six-parameter incident flux dependent model, and our nine-parameter model with both incident flux and host star mass dependence. We then calculate the difference in WAIC between the six-parameter model and each other model, along with the error in the difference. We repeat this for the TTV and RV+TTV datasets discussed in section 2.4.3. Our results are shown in Table 2.3. Lower values for a given model indicate better predicted out-of-sample fit compared to the six-parameter model. We find that generally, models with fewer parameters have a higher predicted out-of-sample fit compared to our standard six-parameter model. Our results for the incident flux dependent model vary between datasets,

with incident flux dependence being most strongly favored over host star mass dependence for the combined RV+TTV dataset, but slightly disfavored for the TTV only dataset. For the RV and TTV datasets, the nine-parameter model with both host star mass and incident flux dependence is disfavored, whereas for the combined RV+TTV dataset, the nine-parameter model is slightly favored but with a large error in the difference. Given that the error of these estimates is often comparable to the calculated differences, nothing definitive can be said about which model should be preferred. This strengthens our conclusion that there is no evidence in the current M-R dataset for host star mass dependence in the M-R relation.

2.5.2 Limitations

Data set

The limited number of transiting planets with RV mass measurements and their uneven occupation of parameter space is perhaps the most significant factor limiting this work. 48% of planets in our sample have a host star mass between $0.9 - 1.1 M_{\odot}$, and only six planets have a host star mass below $0.7 M_{\odot}$. Given our model parametrization, where host star dependence scales as $\ln M_*$, the difference between the coefficient of the host star mass dependent parameters is $\sim$0.2 between $0.9 - 1.1 M_{\odot}$, but $\sim$0.6 between $0.5 - 0.9 M_{\odot}$. Under this scaling, the M-R relation is most significantly different for planets around M dwarf stars, but we only have six such planets. Furthermore, 58% of planets in our sample have radii between $1.5 - 4.0 R_{\oplus}$. More planets with radii between $4.0 - 8.0 R_{\oplus}$ would further constrain the slope of the power-law and allow a more substantive investigation into whether the scatter of the M-R relation changes with radius. These problems are twofold: first, there is the detection of transiting planets within our parameters of interest, and secondly, the radial velocity follow-up of these planets. Both issues need to be addressed, although mass-radius studies are more sensitive to the number of planets subjected to RV follow-up given the wealth of data provided by *Kepler*.

Fortunately, there are several surveys and experiments scheduled to become operational in the near future that will alleviate these issues. The Transiting Exoplanet Survey Satellite (*TESS*), scheduled to launch in 2018, is an all-sky NASA-sponsored mission designed to monitor $\sim$200,000 of the brightest nearby stars. *TESS* is expected to find $\sim$400 planets with $R < 2R_{\oplus}$ hosted by M dwarfs (Sullivan et al., 2015), compared to $\sim$130 found by *Kepler*. On the RV follow-up side, several high-resolution spectrographs (e.g. *MAROON-X*, *SPIRou*, *ESPRESSO*) are under development and planned to coincide with *TESS*. There are also efforts currently underway to search for planets around nearby M dwarfs (e.g. *CARMENES*, *MEarth*, *IR RV* spectrograph). Since *TESS* is scheduled to observe the brightest nearby M dwarfs, RV follow-up of planets around M dwarfs in the era of TESS will be more feasible. As shown in section 2.4.3, a future dataset of planets from *TESS* with radial velocity mass measurements may be able to distinguish between M-R relations with and without host star mass dependence.

Physical Basis

Given that we are empirically fitting the M-R relation, there is no shortage of parametrizations we could have considered. Much like the decision to characterize the M-R relation as a power law, the decision to scale the M-R relation parameters by the natural log of host star mass was based on simplicity and intuitive understanding, rather than any physical basis. There is no reason to think that this scaling should be physically preferred over a power law scaling, for instance. Our motivation for this paper was to allow for host star mass dependence and see how much information is in the current dataset. For this reason, we do not consider alternate parametrizations.

Ultimately, M-R relations should move away from strictly empirical relations and toward a physically motivated distribution. One possible step in this direction is to use a mixture model. Results from the *Kepler* survey show a gap in the radius distribution of small planets

at short orbital periods (Fulton et al., 2017). This bimodal distribution is thought to arise from two separate planet populations: those with significant H/He envelopes and those without. The gap is consistent with evidence that planets below $\sim 1.6 M_{\oplus}$ having densities consistent with a purely rocky composition (Rogers, 2015; Weiss & Marcy, 2014; Dressing et al., 2015). Modeling these two planet populations separately using a mixture model would give each planet a probability of falling into either population, depending on its radius. This would be an improvement over current efforts that model a break in the power-law around $1.6 M_{\oplus}$, as it would account for the overlap between these two populations.

Mass Conditioned on Radius

In this paper we parameterize the M-R relation in terms of mass as a function of radius, in order to directly apply the relation to the *Kepler* sample of planets, which generally lack mass measurements. Framing the M-R relation as *M(R)* allows mass estimates of *Kepler* planets to be readily obtained. Furthermore, the radius measurements for the planets in our sample are generally much more precise than the mass measurements. However, one could argue that mass is the more fundamental physical quantity and the relation should be cast as radius as a function of mass. A *R(M)* relation would also be applicable to the sample of microlensing planets, which have mass constraints but no radius measurements. If desired, one could easily obtain radius as a function of mass by switching mass and radius in Equation (2.1) and fitting for a new set of parameters. Ultimately, a combined joint mass-radius distribution would solve this issue, which would allow one to obtain either relation.

Selection Effects

In using the sample of transiting planets with radial velocity mass measurements, we are subject to a host of poorly characterized selection effects. For example, the choice to follow up a *Kepler* planet is not completely transparent, and while guidelines exist, it is often

ultimately a human decision. Planets orbiting M dwarfs are typically not favored to be chosen for RV follow-up, due to their intrinsic faintness and the desire to characterize planets around solar analogs. Furthermore, upper limits when a planet is not detected are not always published, which could lead to a bias towards more massive planets as they are more likely to be detected. The heterogeneity of the data set also poses a problem: some planets in the sample were discovered by both radial velocity and transit methods independently, and we do not restrict our catalog to one specific survey or program such as *Kepler*. Finding a way to model these selection effects is a difficult task, but necessary to remove any potential biases.

Incident Flux Dependence

In section 2.4.2 we explored adding incident flux dependence to the M-R relation. We found that with the current dataset, we are unable to distinguish between the need for host star mass dependence and the need for incident flux dependence. This is in agreement with previous work by Weiss et al. (2013) which found weak dependence of planet radius on incident flux for small planets. Despite this, incident flux is a key parameter to take into account when constraining planet composition distributions. The incident flux on a planet can affect whether or not a planet retains its atmosphere and the thermal evolution of a planet, particularly for those in close-in orbits (Scalo et al., 2007). An increased sample of planets with mass and radius measurements will warrant revisiting this incident flux dependence. Additionally, the sample of transiting planets with TTV mass measurements will provide another avenue to explore this dependence, given that TTV techniques prefer planets on longer periods than RV methods (Mills & Mazeh, 2017).

2.6 Conclusion

We have modeled host star mass dependence in the planet M-R relation by introducing three new parameters to the probabilistic M-R relation first established in Wolfgang et al. (2016). We fit the model to the current sample of transiting planets with RV measured masses and find that the host star mass dependent parameters are consistent with zero and there is no strong evidence for host star mass dependence in the M-R relation. We have tested the observed trend in Mulders et al. (2015a) of increasing planetary heavy element mass towards lower mass stars and have found this trend to be robust against many of their assumptions. This trend also manifests itself in the minimum-mass extrasolar nebula, with the surface density profile for M dwarfs exhibiting a steeper slope than that of FGK dwarfs. This work provides a framework for including host star mass dependence in the M-R relation, and can be revisited when more planets around M dwarfs have their masses and radii characterized by upcoming surveys such as *TESS* and subsequent radial velocity follow-up.

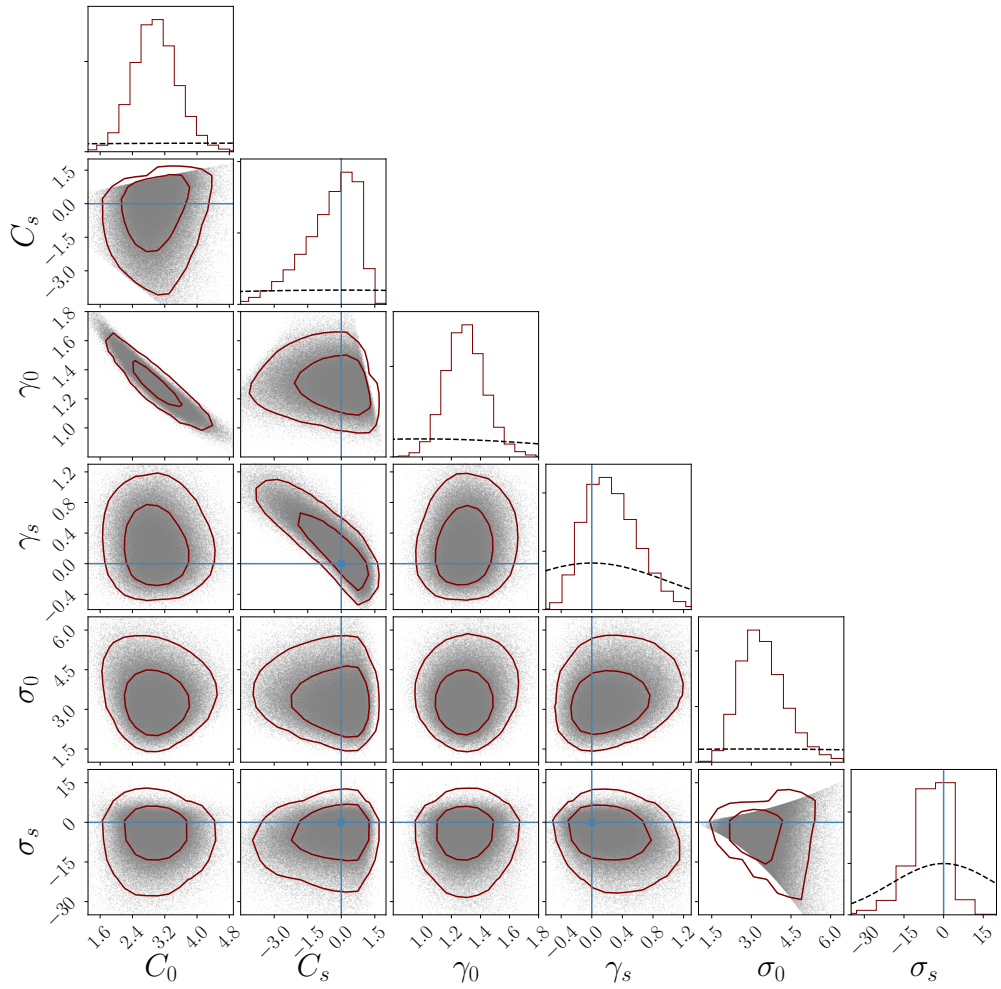

Figure 2.3: Posterior distributions from our six-parameter host star mass dependent M-R relation. In addition to power law slope, normalization, and intrinsic scatter, we introduce three parameters to allow the possibility of host star mass dependence. Contours show 68% and 95% confidence levels. All host star mass dependent parameters are consistent with zero (shown by the blue points), indicating that there is no robust evidence for host mass dependence in the current planet mass-radius dataset. The host star mass dependent intrinsic scatter, σ_s, with a median of -4.18 and a long negative tail, has the most support far from zero. Sharp ridges in the posterior distributions of σ_s and C_s arise from the constraint for σ_M and C to be positive. The black dotted lines in the 1D distributions indicate the priors, generally chosen to be weakly informative. Generated with corner.py (Foreman-Mackey, 2016).

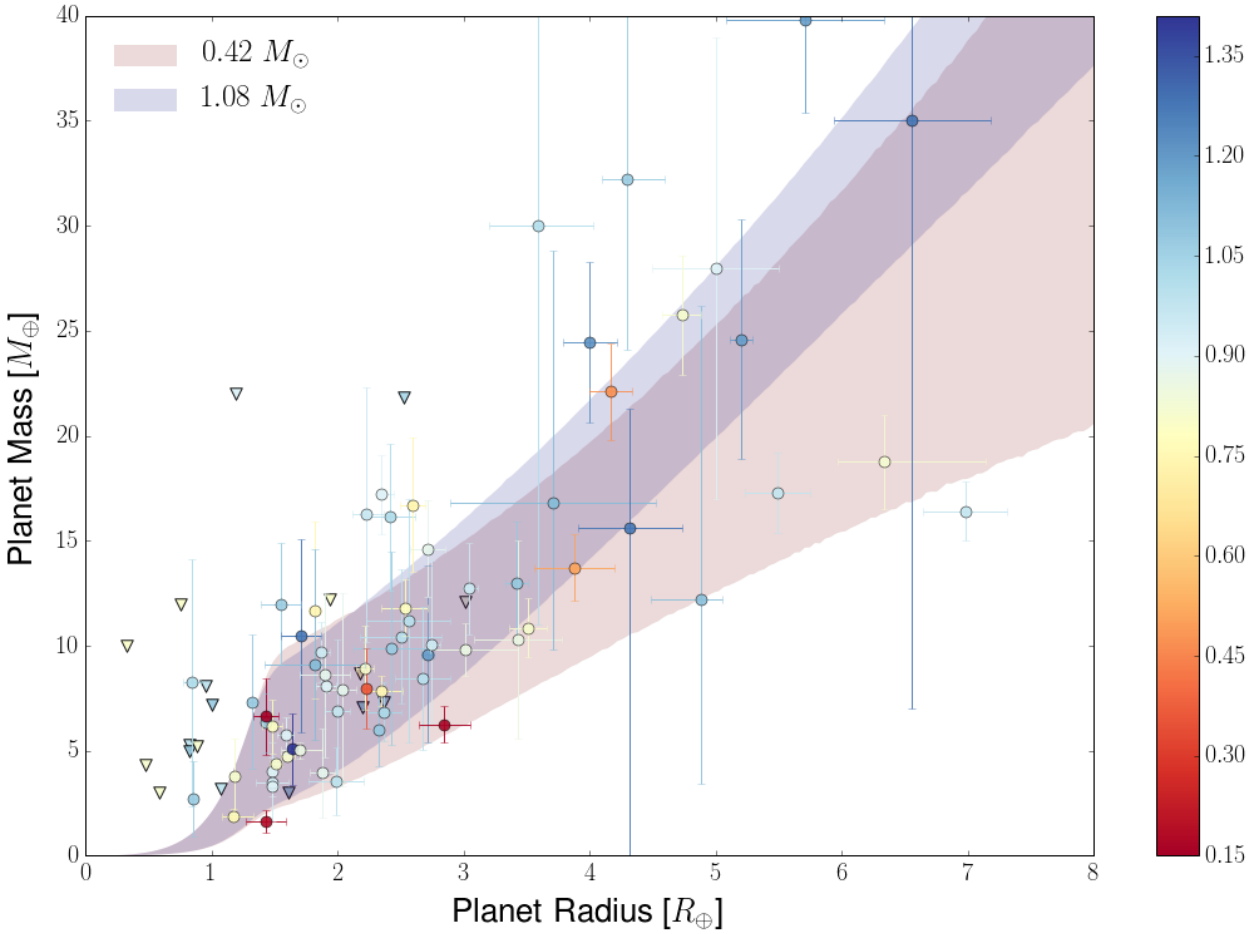

Figure 2.4: The host star mass dependent M-R relation for two characteristic stellar masses: 0.42 $M_\odot$, a typical M dwarf (red) and 1.2 $M_\odot$, a typical F dwarf (blue). The shaded region corresponds to the central 68% of masses drawn at a given radius with the M-R relation parameters marginalized over their posterior distributions. The narrow feature below 1.5 $M_\odot$ is due to the pure iron maximum density restriction. The colored points represent the observed masses and radii of planets in our sample along with their reported uncertainties. The color of the points represents the host star mass as given by the colorbar, with redder points having lower host star mass and bluer points having higher host star mass. Triangles are upper limits in mass as reported by the original authors, although we use more complete information in our modeling. Compared to the representative M-R relation for F stars, the M-R relation for M dwarfs has a shallower slope and higher intrinsic scatter. However, the posterior distributions for the host star mass dependent parameters are wide enough to be consistent with no host star mass dependence.

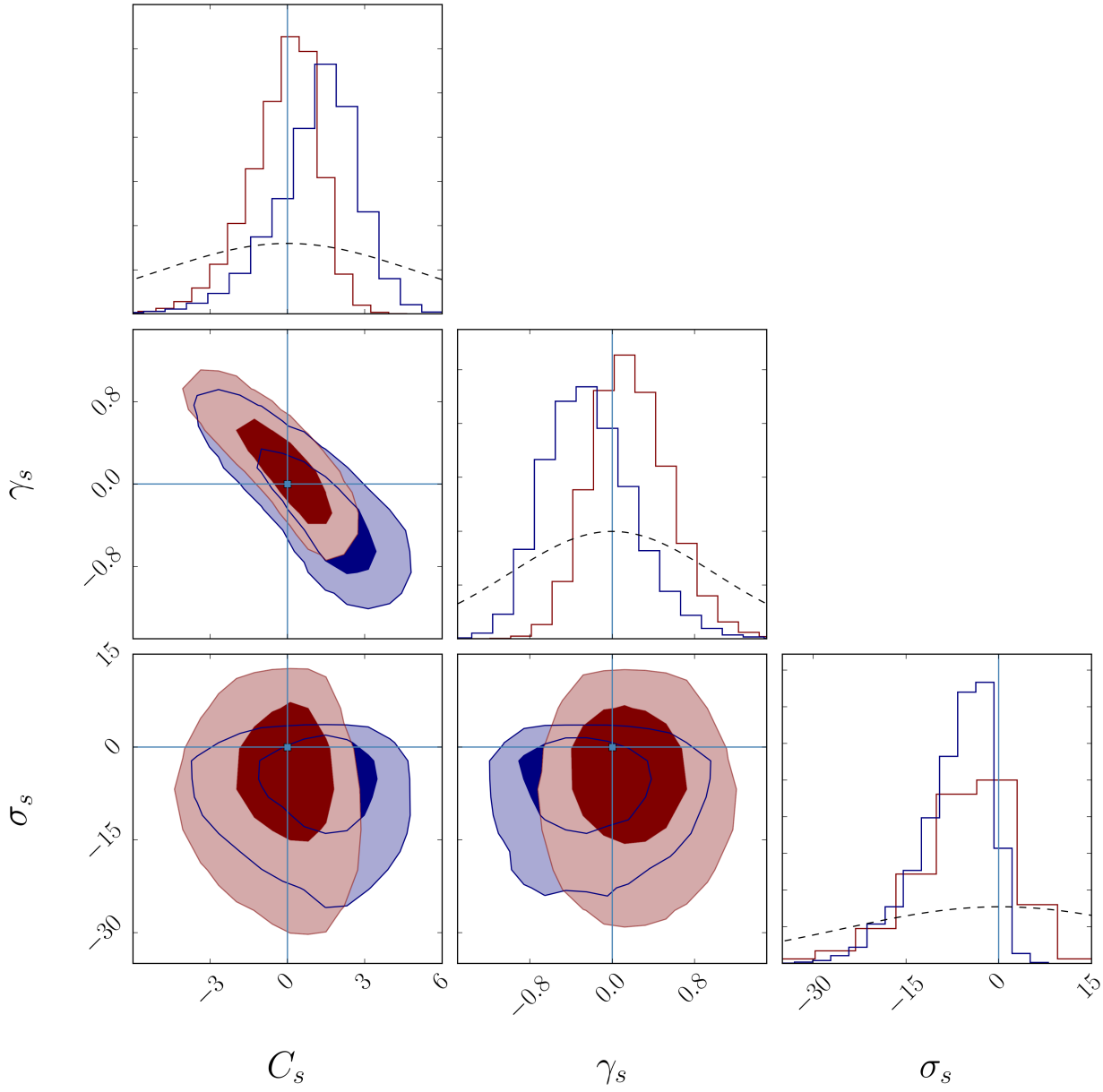

Figure 2.5: Posterior distributions from our nine parameter model that includes both incident flux and host star mass dependence in the M-R relation. Here we show only the host star mass dependent parameters, $\{C_s, \gamma_s, \sigma_s\}$. The posteriors from the six parameter model with only host star mass dependence are shown in red, while those from the nine parameter model with incident flux dependence are shown in blue, with the 1σ and 2σ contours shown. We find that the posteriors for C_s and γ_s widen when incident flux dependence is included, but the posterior distribution for σ_s is slightly tighter. Additionally, the peak of the γ_s distribution shifts from positive to negative, which corresponds to preferring a steeper slope in the M-R relation for low mass stars when incident flux dependence is included.

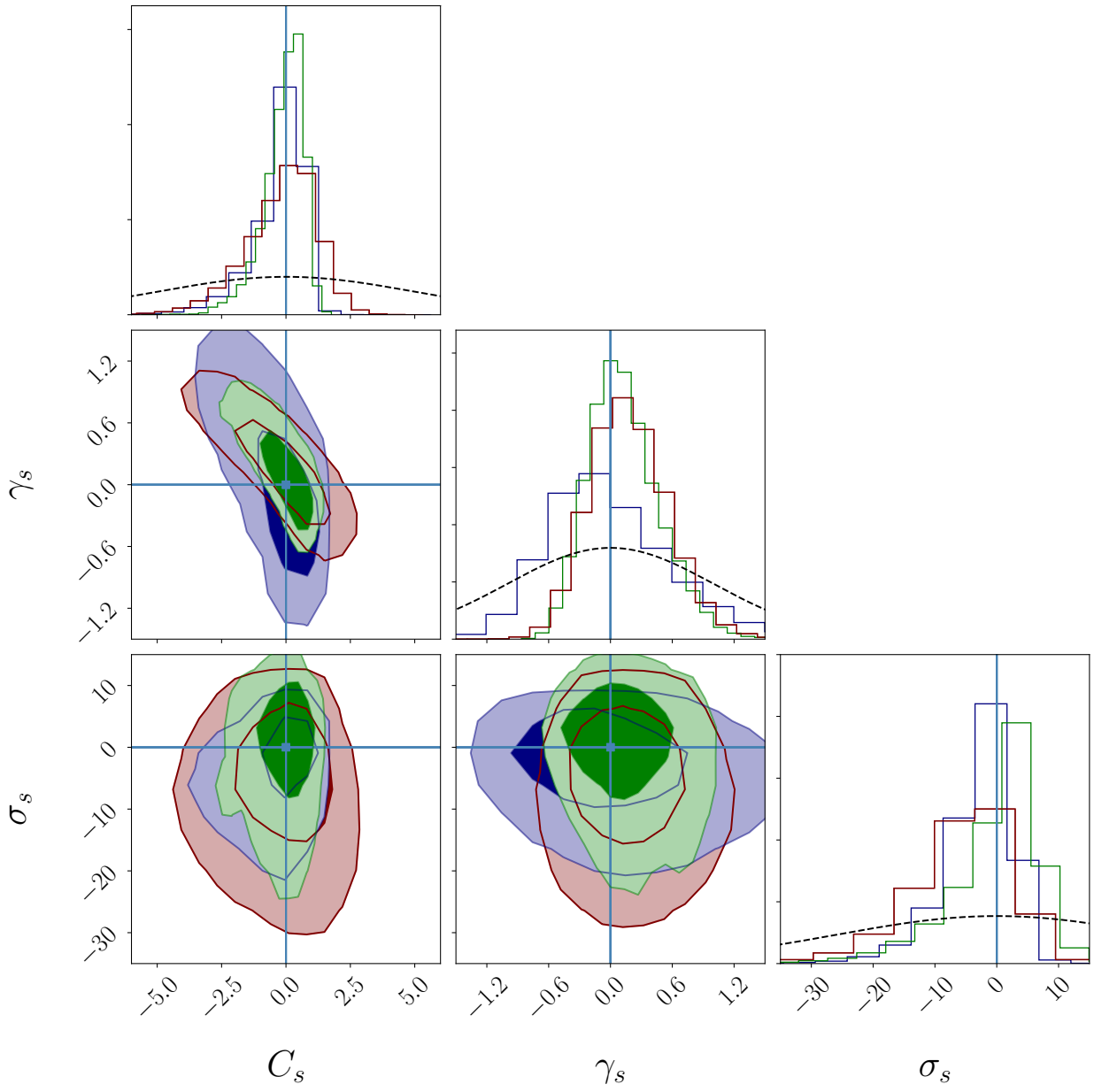

Figure 2.6: Posterior distributions from our six parameter model fit with three different datasets: RV (red), TTV (blue), and RV + TTV (green). As in Figure 2.5, we show only the host star mass dependent parameters, $\{C_s, \gamma_s, \sigma_s\}$. We find that the posterior distributions agree between the three datasets, with the TTV data also showing no evidence for host star mass dependence.

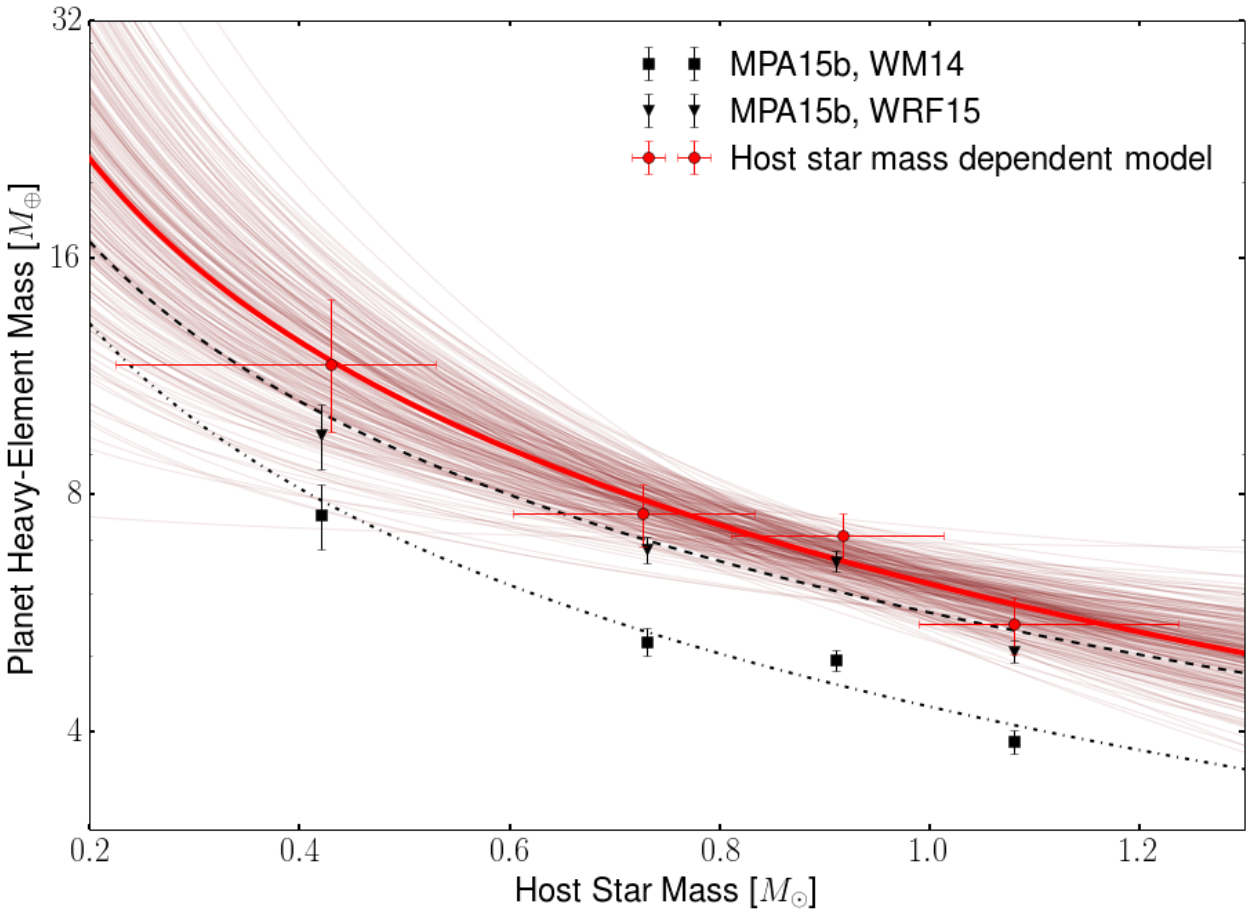

Figure 2.7: The total planetary heavy-element mass per star vs host star mass. Our current work is in red; black points are obtained from the best-fit WRF15 M-R relation (square) and from the best-fit WM14 relation (triangle). The points show the median heavy-element mass in that bin, with the error bars covering the central 68%; the lines are fitted power laws to these four points. We fully incorporate M-R relation uncertainties as well as Poisson errors in obtaining our result: faded lines show different power law fits for each sampling of the M-R relation posteriors. Compared to MPA15b, we find a shallower slope but the overall trend of increasing planet heavy-element mass towards lower mass stars remains.

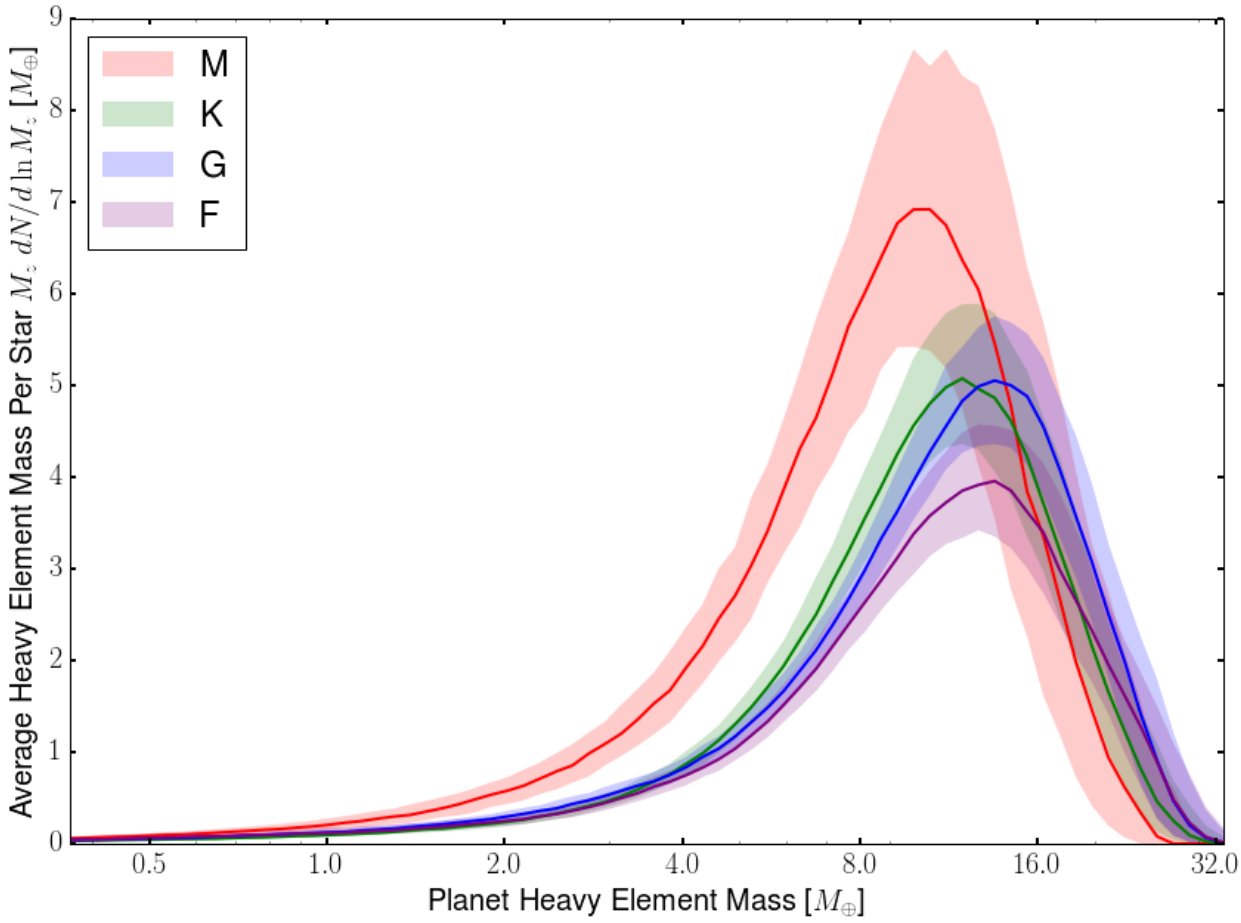

Figure 2.8: The average heavy element mass per star, binned and scaled by heavy element mass. The red curve is for planets around M dwarf stars, with the green, blue and purple curves representing K, G and F stars respectively. The shaded region shows the 68% region, obtained through bootstrap resampling and sampling of the M-R relation parameters. These results are for the third line in Table 2.2, where we scale the heavy element mass as $\sqrt{M_p}$ and scale the period bounds with the location of the snowline. The integral under each curve gives the total heavy-element mass reported in Figure 2.7. We find that the increased planetary heavy element mass around M dwarfs at short periods is due to higher occurrence of planets with $M_Z < 20M_\oplus$.

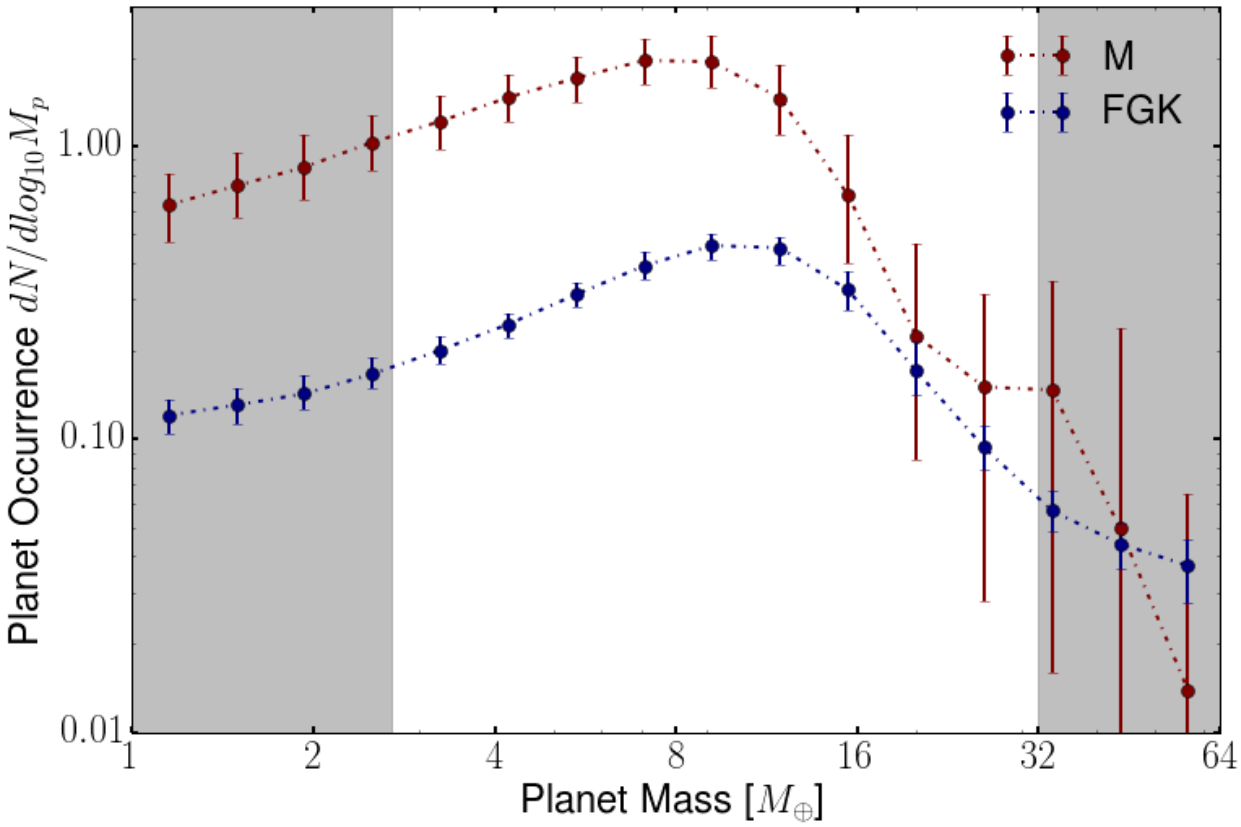

Figure 2.9: Mass distributions for *Kepler* planets with short period orbits around M dwarfs (Red) and FGK dwarfs (Blue). Distributions were obtained by sampling from the probabilistic host star mass dependent M-R relation posteriors. Error bars correspond to the central 68% of occurrence rates drawn at a given mass bin. The grey shaded regions indicate incompleteness of the sample: below 3 $M_\oplus$, the *Kepler* sample is incomplete, and above 32 $M_\oplus$, the M-R relation we fit has to be extrapolated. The mass distribution for planets around M dwarfs peaks at a lower mass and has higher occurrence for less massive planets than the distribution for planets around FGK dwarfs.

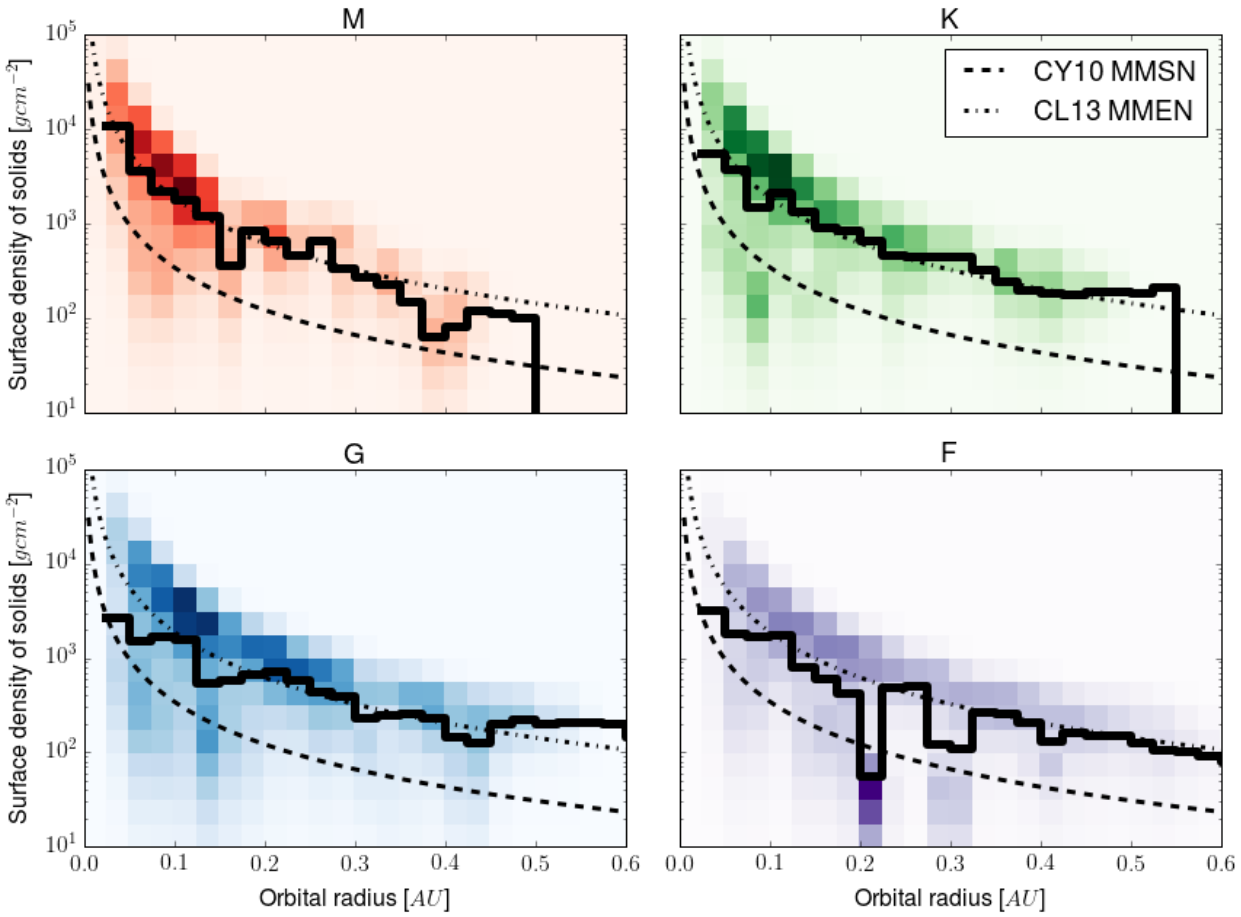

Figure 2.10: A 2D weighted histogram of the MMEN for different types of stars, using the *Kepler* sample. Samples are drawn for each KOI using our host star mass dependent M-R relation posteriors, and weighted by the occurrence rate of the KOI. The solid black line indicates the cumulative center of each orbital radius bin, and the dashed and dot-dashed lines show the results from the Chiang & Youdin (2010) MMSN and the Chiang & Laughlin (2013) MMEN. We find that the slope of the surface density profile of the MMEN is steeper for M dwarfs, resulting in more mass at short orbits around M dwarfs.

CHAPTER 3

A JOINT MASS-RADIUS-PERIOD DISTRIBUTION OF EXOPLANETS

Prior to this work, the mass-radius-period plane was primarily studied in three separate forms. Transit surveys give us the radius-period distribution, which has been well studied and resulted in such discoveries as the "Fulton gap" (Fulton et al., 2017), a valley in the occurrence rate of planets between 1.5 and 2 $R_\oplus$. Radial velocity surveys give us the equivalent mass-period distribution, which has primarily been used to study giant planets, as radial velocity surveys aren't well-suited to study the demographics of small planets. Finally, combining transit surveys with radial velocity follow-up and transit timing variations gives a sample of planets with both radius and mass measurements, allowing us to study the mass-radius relationship. The previous chapter explored the possibility of host star mass dependence in the mass-radius relationship.

While they are typically studied separately, the ingredients exist that enable us to constrain the joint mass-radius-period distribution of exoplanets. We simply require a sample of transiting exoplanets, a subset of which have mass measurements. Constraining the 3D joint distribution allows us to perform the same analyses that make use of the radius-period distribution, mass-period distribution, or mass-radius relation, while providing a more complete picture. The previous paper made use of a mass-radius relation in order to translate the radius-period distribution of *Kepler* exoplanets into a mass-period distribution. The joint mass-radius-period distribution allows you to perform that analysis self-consistently, without relying on two separate fits.

In this chapter, we utilize a subset of the *Kepler* sample that have high-quality radius measurements, a subset of which have radial velocity mass measurements, in order to constrain the joint mass-radius-period distribution of exoplanets. Furthermore, we explore the existence of multiple subpopulations of exoplanets - planets with gaseous envelopes, evapo-

rated rocky cores, and intrinsically rocky planets - through the use of mixture models. The Fulton gap suggests the existence of a subpopulation of exoplanets with gaseous envelopes and a subpopulation without. We construct four separate models that include different combinations of these subpopulations in order to assess the support for this hypothesis. Using model selection, we find that models that include all three subpopulations of planet are preferred over simpler models that only include one or two subpopulations. Models that include these three subpopulations are also better able to match the observed radius distribution of *Kepler* exoplanets and reproduce the Fulton radius gap. We apply our models to calculate planet occurrence rates, finding the answers to be model dependent. This is especially true for the case of $\eta_{\oplus}$, the occurrence rate at Earth's radius and period, due to the extrapolation involved.

These models are extended in Chapter 4, where we include planets with icy compositions alongside the three subpopulations included in this chapter, expanding the number of models from four to ten. Additionally, we perform more robust model selection that has consequences for the outcome of this chapter.

This chapter was completed in collaboration with Leslie Rogers and published in the Astrophysical Journal (Neil & Rogers, 2020). We thank our anonymous referee for providing valuable feedback and suggestions that improved the paper. LAR gratefully acknowledges support from NASA Exoplanet Research Program grant NNX15AE21G and from NSF FY2016 AAG Solicitation 12-589 award number 1615089.

3.1 Abstract

The radius-period distribution of exoplanets has been characterized by the *Kepler* survey, and the empirical mass-radius relation by the subset of *Kepler* planets with mass measurements. We combine the two in order to constrain the joint mass-radius-period distribution of *Kepler* transiting planets. We employ hierarchical Bayesian modeling and mixture models

to formulate four models with varying complexity and fit these models to the data. We find that the most complex models that treat planets with significant gaseous envelopes, evaporated core planets, and intrinsically rocky planets as three separate populations are preferred by the data and provide the best fit to the observed distribution of *Kepler* planets. We use these models to calculate occurrence rates of planets in different regimes and to predict masses of *Kepler* planets, revealing the model dependent nature of both. When using models with envelope mass loss to calculate $\eta_{\oplus}$, we find nearly an order of magnitude drop, indicating that many Earth-like planets discovered with *Kepler* may be evaporated cores which do not extrapolate out to higher orbital periods. This work provides a framework for higher-dimensional studies of planet occurrence and for using mixture models to incorporate different theoretical populations of planets.

3.2 Introduction

With the advent of large-scale exoplanet surveys, exemplified by NASA's *Kepler* mission (e.g. Borucki et al., 2011a), we have arrived at a much more detailed picture of the demographics of exoplanets. The *Kepler* survey has discovered thousands of transiting exoplanets, and with its well-studied detection efficiency, has characterized the radius-period distribution of small planets to high precision (Burke et al., 2015; Thompson et al., 2018). Early studies were able to reveal that planets smaller than Neptune are common, with an occurrence rate (number of planets per star) on order unity (Howard et al., 2012; Fressin et al., 2013). Later studies improved methodology by employing hierarchical Bayesian modeling (Foreman-Mackey et al., 2014; Burke et al., 2015) as well as approximate Bayesian computation (Hsu et al., 2018), avoiding a bias in earlier papers that underestimated the occurrence rate of planets at the threshold of detection. Later studies also benefited from a better characterized detection efficiency (Christiansen et al., 2015; Christiansen, 2017), as well as a stellar sample with high-precision radii and masses (Petigura et al., 2017), the latter allowing for the discovery

of a gap in the radius distribution of small planets between $1.5-2.0R_{\oplus}$ (Fulton et al., 2017). Finally, with the release of Data Release 25, the final data release of *Kepler*, we have the best characterization of the radius-period distribution of transiting planets to date (Fulton & Petigura, 2018; Mulders et al., 2018; Hsu et al., 2019).

While the radius-period plane is the most natural plane to constrain planet occurrence rates for transit surveys, it is but a projection of a high dimensional space where the occurrence of planets can depend on many factors, both those intrinsic to the planet and those relating to the host star that the planet orbits. Planet mass is one of the most fundamental properties of a planet and has been studied statistically through radial velocity (RV) and microlensing surveys (Howard et al., 2010; Mayor et al., 2011; Suzuki et al., 2016). With RV follow-up of transiting planets, the mass-radius distribution of planets, commonly characterized by a mass-radius relationship, has been empirically constrained. Since the data is limited by the rate of RV and transit-timing variation (TTV) characterization of small planets, most studies have parametrized the mass-radius relation as a power-law with multiple breaks (Wu & Lithwick, 2013; Weiss et al., 2013; Weiss & Marcy, 2014). Later studies have improved upon the statistical methodology and have accounted for the intrinsic scatter in the mass-radius relation (Wolfgang et al., 2016; Chen & Kipping, 2017; Neil & Rogers, 2018; Ning et al., 2018). While the theory of planet interior structure and its manifestation in the mass-radius relation has been studied for over a decade (Valencia et al., 2006; Fortney et al., 2007b; Seager et al., 2007; Rogers et al., 2011; Lopez & Fortney, 2013; Howe & Burrows, 2015; Zeng et al., 2015; Chen & Rogers, 2016), these efforts have largely been separate from empirically derived mass-radius relations.

The 2D distributions of radius-period, mass-period, and mass-radius have all been characterized using various datasets and statistical techniques. To date, no effort has been made to constrain all three distributions simultaneously and self-consistently. In this paper, we constrain the 3D joint mass-radius-period distribution of exoplanets using the *Kepler* survey,

a subset of which have mass measurements. In doing so, we can combine the state-of-the-art data and techniques used to constrain each 2D distribution, and expand the analysis into 3D space. This 3D joint distribution can then be marginalized to obtain each individual 2D distribution, which can be used for such purposes as calculating occurrence rates or predicting individual masses or radii of planets.

This paper is organized as follows. In section 3.3 we outline how we arrived at our planet sample and detection efficiency function. Four joint mass-radius-period distribution models with increasing complexity are presented in section 3.4. The fitting process is described in section 3.5. The results of our model fitting are described in section 3.6, along with model selection, occurrence rate calculations and individual planet mass and radius predictions. We discuss caveats, how to use these results and future extensions of this work in section 3.7, and conclude in section 3.8.

3.3 Data

In order to constrain the joint mass-radius-period distribution, we use the California-*Kepler* Survey (CKS), a subset of transiting planets from *Kepler* with high-resolution spectroscopic follow-up of their host stars (Petigura et al., 2017; Johnson et al., 2017), cross-matched with *Gaia* data. This sample has several significant advantages. First, it is by far the largest sample of planets with high-precision radius measurements. The CKS sample was the first sample to uncover the gap in the *Kepler* radius distribution, due to the low median uncertainty in planet radii of 13% (Fulton et al., 2017). Cross-matching the sample with *Gaia* parallaxes achieves even lower radius uncertainties (median of 5%) and allowed a stellar mass dependence in the radius gap to be revealed (Fulton & Petigura, 2018). Second, there is a subset of planets in the sample that have mass measurements, either through RV or TTV, allowing us to constrain the mass-radius relation. Finally, the *Kepler* survey has a well-defined detection efficiency as a function of period and radius, essential for correcting

for the detection bias in our estimate of the true occurrence rate density of planets.

We follow Fulton & Petigura (2018) and apply cuts to the CKS sample to ensure high quality data. In brief, we use the magnitude-limited sample (brighter than 14.2 *Kepler* magnitude), take out evolved stars with a $T_{\rm eff}$-dependent cut on R_*, and further limit the sample to G and K dwarfs (4700 K $< T_{\rm eff} <$ 6500 K). Additionally, we remove stars that have high dilution based on *Gaia* and imaging data. We remove any planet candidates that have been designated as false-positives by the *Kepler* team, and any planets with grazing transits ($b > 0.9$) due to uncertainty in planet radii at high impact parameters. Lastly, we remove any planets with properties that fall outside our orbital period and radius bounds, which are $0.3 - 100$ days and $0.4 - 30 R_\oplus$, respectively (see section 3.4 for why these bounds were chosen). The final planet radii we use are taken from Fulton & Petigura (2018), which were calculated using stellar radii derived from both CKS spectroscopy and *Gaia* parallaxes. Our final sample has 1130 planets, with a median radius uncertainty of 4.8%.

On the mass side of things, we use mass measurements where available for planets in the CKS-*Gaia* sample. Given the systematic differences in the density of planets measured with RV vs TTV (Jontof-Hutter et al., 2014; Steffen, 2016; Mills & Mazeh, 2017), we limit our mass sample to RV-measured masses only, leaving the analysis including TTV masses to a future work. We discuss the impact of this choice in section 3.7.2. We take RV mass measurements from the NASA Exoplanet Archive[1], downloaded on September 25, 2018. The CKS-*Gaia* sample we use in this paper has 53 planets with RV mass measurements.

In measuring the detection efficiency of the *Kepler* survey, we follow the procedure for *Kepler* Data Release 25 as outlined in Christiansen (2017), Burke & Catanzarite (2017), and Thompson et al. (2018). We describe the steps briefly here, and refer the reader to the aforementioned resources for details. We first apply the same cuts to the *Kepler* Q1-Q17 DR25 stellar target sample as we applied to the CKS planet candidate sample, leaving

1. https://exoplanetarchive.ipac.caltech.edu/

us with 30,070 stars. Following Christiansen (2017), we use the pixel-level injected light curves (using only injections that match our CKS cuts) to calculate the fraction of simulated transits recovered by the pipeline as a function of the expected multiple event statistic (MES), a measure of the strength of the transit signal relative to the noise calculated by the Transiting Planet Search algorithm. We then calculate the fraction of transits correctly dispositioned as planet candidates by the Robovetter, a tool used by the *Kepler* team to disposition transits as candidates or false positives, as a function of MES, as in Thompson et al. (2018). This average detection efficiency of the combined *Kepler* Pipeline and the Robovetter is then fit by a gamma CDF function:

$$\mathrm{p} = F(x|a,b,c) = \frac{c}{b^a \Gamma(a)} \int_0^x t^{a-1} e^{-t/b} dt \tag{3.1}$$

where x is the MES, p is the probability that a planet at a given MES is detected, and a, b, and c are coefficients that are retrieved to be $a = 29.41$, $b = 0.284$, $c = 0.891$. We then use the KeplerPORTS[2] Python package to calculate the detection efficiency individually for each target star in a grid of planet radius and period, using our fitted gamma CDF function to convert from estimated MES to fraction of planets detected. We then multiply this completeness of the *Kepler* pipeline by the transit probability p_{tra}. Our final detection efficiency is then taken to be the average over our stellar sample, and is shown in Figure 3.1.

3.4 Models

We constrain the planet occurrence rate density in radius, mass and period, which is defined as the number of planets per star per interval in radius, mass, and period. We denote the planet occurrence rate density by $\Gamma(P, M, R)$:

2. https://github.com/nasa/KeplerPORTs

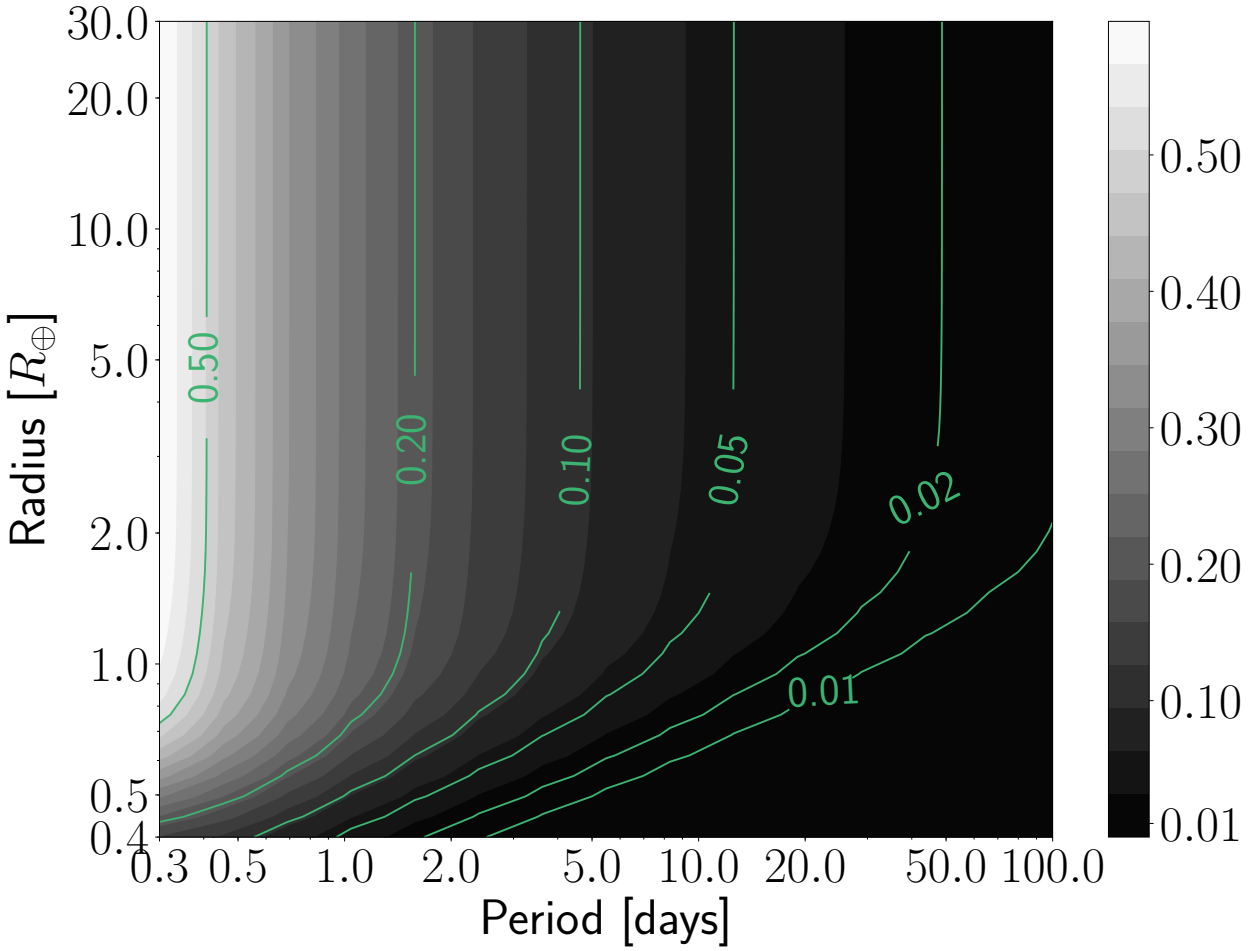

Figure 3.1: The average detection efficiency of the combined *Kepler* Pipeline and Robovetter for our stellar sample of interest as a function of planet period and radius. The shading of the grid corresponds to the fraction of planets detected at a given radius and period. Transit probability is factored in. The colored lines show contours ranging from 1% to 50% of planets detected.

$$\Gamma(P, M, R) = \frac{dN}{dP\ dM\ dR} \tag{3.2}$$

Note that the standard set by other occurrence rate papers is to define the occurrence rate density using log intervals, but it is trivial to convert between the two by either dividing or multiplying by radius, period, and mass. We choose this definition because it simplifies the notation for the normal, log-normal and power-law parametrizations that we use herein.

In constraining the planet occurrence rate density in radius, mass and period, we must assume some parametrization. In the following sections, we present several such parametrizations that start simple and gradually add further complications. To start, we consider a model with a single population of planets. We later incorporate mixture models to account for multiple, distinct, physically motivated populations of planets.

For each parametrization, we start with the following breakdown of the occurrence rate density, $\Gamma(P, M, R)$:

$$\Gamma(P, M, R) = \Gamma_0\, \mathrm{p}(P)\, \mathrm{p}(M)\, \mathrm{p}(R|M) \tag{3.3}$$

where we have separated the occurrence rate density into four components: a period distribution $p(P)$ that is independent of mass and radius, a mass distribution $p(M)$, a radius conditioned on mass distribution $p(R|M)$ given by a probabilistic mass-radius relationship, and an overall normalization term Γ_0. For all of our models, the period P is assumed to be in units of days, and the planet radius R and planet mass M are in Earth units ($R_\oplus$ and $M_\oplus$, respectively). All period, mass and radius distributions are normalized to 1 over the boundaries we set and define later in the text.

With the use of a mixture model, Equation (3.3) becomes:

$$\Gamma(P, M, R) = \sum_{q=0}^{N-1} \Gamma_0\, \mathrm{p}(P|q)\, \mathrm{p}(M|q)\, \mathrm{p}(R|M, q)\, \mathrm{p}(q|M, P) \tag{3.4}$$

where $0 \leq q < N$ is the index of the mixture component with N total mixture components. The mixture probability $p(q)$ may depend on mass and period, as we show in the case of envelope mass loss in section 3.4.2. In the case of $N = 1$, Equation (3.4) reduces to Equation (3.3) above. In the following subsections, we describe four models corresponding to $N = 1, 2, 3, 4$.

For all models, we choose to parametrize the period distribution with a broken power-law, as has been done in previous studies of *Kepler* occurrence rates (e.g. Mulders et al., 2018) in order to account for the flattening of planet occurrence rates at around 10 days:

$$\begin{aligned} \mathrm{p}(P) &= AP^{\beta_1}, \ P < P_{\mathrm{break}} \\ \mathrm{p}(P) &= AP_{\mathrm{break}}^{\beta_1-\beta_2} P^{\beta_2}, \ P > P_{\mathrm{break}} \end{aligned} \tag{3.5}$$

where A is a normalization factor (dependent on β_1, β_2, P_{break}) such that the distribution is normalized to 1 between the lower and upper period bounds of 0.3 and 100 days, in line with *Kepler*'s sensitivity limits. We have also modeled the period distribution as a single power-law, but found it to be a poor fit as the occurrence rate planets with short and long periods (close to the boundaries of 0.3 and 100 days, respectively) was being significantly overestimated, and the occurrence rate of planets with periods in between was consequently underestimated.

Since mass is a fundamental quantity for the structure of a planet and is one of the factors that determines a planet's radius, we choose to parametrize our joint distribution in terms of an independent mass distribution and a radius conditioned on mass distribution, instead of the converse. Despite this, the mass distribution of *Kepler* planets is more uncertain than the radius or period distributions given that previous studies have relied on empirical mass-radius relations to translate *Kepler* planet radii into masses (Mulders et al., 2015a; Neil & Rogers, 2018; Pascucci et al., 2018). By fitting the mass-radius relation simultaneously

with the mass distribution, we aim to infer a more statistically robust mass distribution of *Kepler* planets. The mass distribution has been most commonly parametrized as a broken power-law, with evidence for a break at around $\sim 20M_{\oplus}$ for microlensing planets, which mostly orbit M dwarfs (Suzuki et al., 2016), and $\sim 10M_{\oplus}$ for *Kepler* planets, which mostly orbit G dwarfs (Pascucci et al., 2018; Neil & Rogers, 2018). For all models, we choose to use a log-normal parametrization, which can take into account the peak-like nature of the mass distributions inferred in previous studies:

$$\mathrm{p}(M) = \mathrm{lnN}(M|\mu_M, \sigma_M) \tag{3.6}$$

with a lower bound of $0.1M_{\oplus}$, roughly corresponding to the predicted mass of the smallest *Kepler* planet, and an upper bound of $10,000M_{\oplus}$, well below the hydrogen burning limit. In model 4 we explore a more flexible parametrization that can reveal subtler features of the *Kepler* mass distribution.

We parametrize our radius conditioned on mass distribution with a probabilistic mass-radius relationship, as first done in Wolfgang et al. (2016). At a given planet mass, a planet's radius is drawn from a normal distribution with a mean given by a power-law and an intrinsic scatter:

$$\mathrm{p}(R|M) = \mathrm{N}(R|\mu(M), \sigma\mu(M)) \tag{3.7}$$

where we use a fractional scatter given by $\mu\sigma$ rather than the flat scatter in Wolfgang et al. (2016), to account for the wide range of masses we are modeling. Following Chen & Kipping (2017), we separate the mass-radius relation into three regimes, corresponding to different mass ranges. Each regime has its own power-law slope as well as fractional scatter:

$$
\begin{aligned}
\mu_0 &= C M^{\gamma_0} \\
\mu_1 &= C M_{\mathrm{break},1}^{\gamma_0-\gamma_1} M^{\gamma_1} \\
\mu_2 &= C M_{\mathrm{break},1}^{\gamma_0-\gamma_1} M_{\mathrm{break},2}^{\gamma_1-\gamma_2} M^{\gamma_2}
\end{aligned}
\tag{3.8}
$$

with the subscripts 0, 1, 2 indicating the low mass range, intermediate mass range, and high mass range, respectively. Instead of abrupt transitions at the break points $M_{\mathrm{break},1}, M_{\mathrm{break},2}$, we use a logistic function to smooth between the different power-laws:

$$
\begin{aligned}
S_1 &= \frac{1}{1+\exp(-5(\log M - \log M_{\mathrm{break},1})} \\
S_2 &= \frac{1}{1+\exp(-5(\log M - \log M_{\mathrm{break},2})}
\end{aligned}
\tag{3.9}
$$

$$
\begin{aligned}
\mu &= (1-S_1)\mu_0 + S_1(1-S_2)\mu_1 + S_1 S_2 \mu_2 \\
\sigma &= (1-S_1)\sigma_0 + S_1(1-S_2)\sigma_1 + S_1 S_2 \sigma_2
\end{aligned}
\tag{3.10}
$$

with μ and σ serving as inputs to Equation (3.7). Finally, we introduce a lower-bound truncation to Equation (3.7) as in Wolfgang et al. (2016), taken as the radius of a pure-iron planet of a given mass. However, rather than use the analytic function fits in Fortney et al. (2007b), which were only fit up to $100 M_\oplus$, we use the analytic scaling relation in Seager et al. (2007) (hereafter S07), which was fit up to a larger mass range appropriate for this work:

$$
\begin{aligned}
R &= R_1 \cdot 10^{k_1 + \frac{1}{3}\log_{10}(M/M_1) - k_2\ (M/M_1)^{k_3}} \\
R_1 &= 2.52; M_1 = 5.80 \\
k_1 &= -0.20949; k_2 = 0.0804; k_3 = 0.394
\end{aligned}
\tag{3.11}
$$

For a $0.1 M_\oplus$ planet, the corresponding radius of a pure-iron planet is $\approx 0.4 R_\oplus$, which we set as a hard lower bound in radius. The upper bound is taken to be $30 R_\oplus$, a generously inclusive upper limit based on the largest *Kepler* planets.

For each model, we include a separate mass-radius relation for rocky planets, with differing implementations. Ideally, we would have a survey that is complete at the lowest planet radii of interest, and would have accurate RV masses of these small, rocky planets. While *Kepler* has discovered hundreds of potentially rocky planets, the subset of these planets with RV masses remains small, and the uncertainties on the available RV masses are large (with a median uncertainty of 22% in our sample). If we attempt to fit for the mass-radius relation of rocky planets (below $\sim 1.6 R_\oplus$ (Rogers, 2015)), the lack of mass measurements in our sample may be problematic and lead to large uncertainties. Fortunately, the mass-radius relation of rocky planets of a given composition can be constrained from interior structure models, as in S07. Objects outside our data sample, such as Solar System planets and moons, as well as exoplanet systems with TTV mass measurements, can also constrain the rocky mass-radius relation, as done in Chen & Kipping (2017). The current sample of potentially rocky planets with mass and radius information exhibit a low scatter about a theoretical pure-silicate mass-radius relation (Dressing et al., 2015; Zeng et al., 2015). The prior information from the Solar System planets and interior structure models will be more constraining than the CKS sample used in this work. Therefore, we fix the rocky mass-radius relation to the analytic mass-radius relation of a pure silicate planet from S07:

$$
\begin{aligned}
R &= R_1 \cdot 10^{k_1 + \frac{1}{3}\log_{10}(M/M_1) - k_2\ (M/M_1)^{k_3}} \\
R_1 &= 3.90; M_1 = 10.55 \\
k_1 &= -0.209594; k_2 = 0.0799; k_3 = 0.413
\end{aligned}
\tag{3.12}
$$

While each individual model has its own unique parametrization, with varying number of mixture populations, they all share the universal elements of the above distributions. Each model uses a broken power-law for period distributions, log-normals for mass distributions, and probabilistic broken power-laws for mass-radius relations.

3.4.1 Baseline Model with a Single Population of Planets

As our baseline model, we consider the exoplanet mass-radius-period distribution to be comprised of a single population of planets, that span a range of compositions from rocky to gas giants. This is consistent with most previous works that constrained either the mass-radius relation or radius-period distribution (Wolfgang et al., 2016; Chen & Kipping, 2017; Foreman-Mackey et al., 2014; Burke et al., 2015).

For this model, our single population of planets is described by a broken power-law period distribution, a log-normal mass distribution, and a double broken power-law mass-radius relation. We replace the first equation describing the low-mass end of the mass-radius relation in Equation (3.8) with the rocky mass-radius relation from S07 in Equation (3.12). We fix this relation for reasons described in the preceding section, but with varying intrinsic fractional scatter σ_0, as well as the low-to-intermediate-mass transition point $M_{\rm break,1}$. This correspondingly changes the normalization for the intermediate and high mass power-laws in Equation (3.8). We are left with a total of 13 parameters for our baseline model: $\{\Gamma_0, \beta_1, \beta_2, P_{\rm break}, \mu_M, \sigma_M, \gamma_1, \gamma_2, \sigma_0, \sigma_1, \sigma_2, M_{\rm break,1}, M_{\rm break,2}\}$. In the following sections, we will refer to this model as "model 1", since it is both the first model we present and only

has one mixture component.

3.4.2 Mixture Model with Envelope Mass Loss

A single population of planets, described by a mass-radius relation and a log-normal mass distribution, is insufficient to fully characterize the current sample of planets. Detailed analysis of the radius distribution of *Kepler* planets has revealed a bimodal distribution, with a gap between $1.5 - 2.0 R_\oplus$ (Fulton et al., 2017). This is strong evidence for two separate populations of planets (super-Earths and sub-Neptunes) that vary in envelope gas mass. This feature is not captured in the most frequently used empirical mass-radius relations (Wolfgang et al., 2016; Chen & Kipping, 2017; Ning et al., 2018), and thus will not be reflected in the radius or mass predictions using those relations.

We directly model the two hypothesized populations of planets using a mixture model, building upon our baseline model introduced in the previous section. To start, we consider two populations of planets that share a common formation history and form with a primordial hydrogen/helium envelope, but one population retains their envelopes (gaseous), and the other loses their envelopes due to the incident X-ray and extreme ultraviolet flux (hereafter XUV flux) on the planet (rocky evaporated/remnant cores). These two populations are then both drawn from the same mass and period distribution, but differ by the radius conditioned on mass distribution $\mathrm{p}(R|M, q)$. Note that what we are calling the gaseous population need not have a large H/He envelope mass fraction: a planet with 1% by mass H/He has a transit radius nearly double that of a planet with the same mass but without a H/He envelope (e.g. Lopez & Fortney, 2013; Chen & Rogers, 2016). In addition, for planets with massive enough envelopes, most of the H/He will not be in the gaseous phase.

For our envelope mass loss, we use the prescription for hydrodynamic, XUV-driven mass loss from Lopez et al. (2012). The mass-loss timescale is given by:

$$t_{\rm loss} = \frac{GM_{\rm env}^2}{\pi \epsilon R_{\rm prim}^3 F_{\rm XUV,E100}} \frac{F_\oplus}{F_p} \tag{3.13}$$

where $M_{\rm env}$ is the envelope mass of the planet, ϵ is the mass-loss efficiency, $F_{\rm XUV,E100}$ is the XUV flux at the Earth when it was 100 My old, and F_p is the bolometric incident flux on the planet. The incident flux on the planet is calculated using the planet's orbital period, as well as its host star mass, radius and temperature. The radius of the planet here, $R_{\rm prim}$ is the primordial radius at formation when it had an envelope and is not necessarily the current radius of the planet. We approximate this from the mass of the planet using the median mass-radius relation of the gaseous population (i.e., the mass-radius relation in Equation (3.8) substituting in the central values of our priors for each free parameter; see Table 3.1 for our priors). The mass of the envelope is calculated from mass of the planet, by smoothing between an envelope mass of $M_{\rm env} = 0.1 M_p$ and $M_{\rm env} = M_p - \sqrt{M_p}$, the latter taken from the scaling of planet heavy-element mass with total mass, calculated from thermal and structural evolution models of giant planets (Thorngren et al., 2016). The smoothing is done via a logistic function as in Equation (3.9), with a transition at $20 M_\oplus$, where the scaling from Thorngren et al. (2016) is valid.

Denoting the gaseous mixture as $q = 0$ and the rocky mixture as $q = 1$, the probability that a planet retains its envelope is then:

$$\begin{aligned} {\rm p}(q=0) &= {\rm p}_{\rm ret} = \min\left(\alpha \frac{t_{\rm loss}}{\tau}, 1\right), \\ {\rm p}(q=1) &= (1 - {\rm p}_{\rm ret}) \end{aligned} \tag{3.14}$$

where τ is the age of the star, and α is an additional free parameter in the model. The nominal values of ϵ, $F_{\rm XUV,E100}$ and τ we take to be 0.1, 504 erg s^{-1} cm^{-2} and 5 Gyr for each star. In order to account for the uncertainties in these three values, we include a

prefactor α to scale the retention probability either up or down:

$$\alpha = \left(\frac{0.1}{\epsilon}\right)\left(\frac{504 \text{ erg s}^{-1}}{F_{\text{XUV,E100}}}\right)\left(\frac{5 \text{ Gyr}}{\tau}\right) \tag{3.15}$$

This prescription for the envelope retention may seem fine-tuned, given the numerous assumptions we made. Our goal here is just to include a reasonable, simple model of mass loss, understanding that there may be complexity that we are not accounting for. We allow α to vary to partially account for this uncertainty. The photoevaporation that we model here has been shown to be able to reproduce the Fulton radius gap by itself (Van Eylen et al., 2018; Fulton & Petigura, 2018). However, there are other mass loss mechanisms that can reproduce the radius gap without the need for photoevaporation, such as core-powered mass loss (Ginzburg et al., 2018; Gupta & Schlichting, 2019). We do not include multiple mass loss mechanisms as it would unnecessarily complicate the model and leave exploring additional mass loss prescriptions to future work.

As previously mentioned, our two populations both formed with a gaseous envelope, but one population lost their envelopes through hydrodynamic mass loss and are currently rocky, and the other population retained their envelopes and are currently gaseous. For this model and subsequent models, we introduce the following subscripts: "fr" to represent formed rocky, "fg" to represent formed gaseous, "cr" to represent currently rocky, and "cg" to represent currently gaseous. Our two populations differ in the radius conditioned on mass distribution $\mathrm{p}(R|M,q)$ only. Thus, using Equation (3.4), our final mixture model rate density becomes:

$$\begin{aligned}
\mathrm{p}(R|M,q=0) &= \mathrm{p}(R|M)_{\text{cg}} \\
\mathrm{p}(R|M,q=1) &= \mathrm{p}(R|M)_{\text{cr}} \\
\Gamma(P,M,R) &= \Gamma_0\, \mathrm{p}(P)_{\text{fg}}\, \mathrm{p}(M)_{\text{fg}} \left[\mathrm{p}_{\text{ret}}\ \mathrm{p}(R|M)_{\text{cg}}\right. \\
&\qquad \left. + (1-\mathrm{p}_{\text{ret}})\mathrm{p}(R|M)_{\text{cr}}\right]
\end{aligned} \tag{3.16}$$

We use Equation (3.12) from S07 for the rocky component of the mass-radius relation $\mathrm{p}(R|M)_{\mathrm{cr}}$, with a fixed 5% scatter. The conditional distribution then becomes:

$$\mathrm{p}(R|M)_{\mathrm{cr}} = N(R_{\mathrm{S07}}(M), 0.05 \cdot R_{\mathrm{S07}}(M)) \tag{3.17}$$

For the gaseous mixture component mass-radius relation $\mathrm{p}(R|M)_{\mathrm{cg}}$, we use Equation (3.8), allowing C and γ_0 to vary along with the scatter σ_0. With the low amount of mass-radius information in the dataset at these low masses, we don't expect the model to tightly constrain these parameters, but we value flexibility by including them. In total, our mixture model has 16 free parameters: $\{\Gamma_0, \beta_1, \beta_2, P_{\mathrm{break}}, \mu_M, \sigma_M, C, \gamma_0, \gamma_1, \gamma_2, \sigma_0, \sigma_1, \sigma_2,$ $M_{\mathrm{break,1}}, M_{\mathrm{break,2}}, \alpha\}$. We will refer to this model with two mixture components as "model 2" in the following sections.

3.4.3 Two Populations of Rocky Planets

Our previous model assumed a shared formation history for the two populations of planets. Next, we allow for the possibility that the rocky and gaseous planets arise from different formation scenarios. This manifests as the sub populations having distinct mass and period distributions. We still maintain the envelope mass loss introduced in the previous model, leaving us with three populations of planets: an "intrinsically rocky" population, a population of evaporated cores (which lost their envelopes through hydrodynamic mass loss), and a population of planets with H/He envelopes. The intrinsically rocky and evaporated cores share a common radius conditioned on mass distribution, and the evaporated cores and gaseous envelope planets arise from a shared underlying mass and period distribution (though due to the dependence of envelope mass loss on mass and period, their mass and period distributions will be different). Following the notation from the preceding section, we denote planets belonging to the "intrinsically rocky" population as $q = 2$. Instead of modeling the mixture probability $\mathrm{p}(q = 2)$ using physics, we can simply model the prior $p(q)$

as a 2-simplex:

$$\begin{aligned} \mathrm{p}(0,1) &= 1 - Q_{\mathrm{fr}} \\ \mathrm{p}(2) &= Q_{\mathrm{fr}} \end{aligned} \tag{3.18}$$

where $0 < Q_{\mathrm{fr}} < 1$ is an additional parameter in our model parameterizing the total fraction of intrinsically rocky planets compared to "intrinsically gaseous" planets ($q = 0, 1$). Note that $\mathrm{p}(q = 0)$ and $\mathrm{p}(q = 1)$ are further modified by the envelope retention from the previous model, becoming:

$$\begin{aligned} \mathrm{p}(0) &= (1 - Q_{\mathrm{fr}})\ \mathrm{p}_{\mathrm{ret}} \\ \mathrm{p}(1) &= (1 - Q_{\mathrm{fr}})\ (1 - \mathrm{p}_{\mathrm{ret}}) \end{aligned} \tag{3.19}$$

Following a similar expansion as with the previous model, we obtain the final rate density:

$$\begin{aligned} \Gamma(P, M, R) = \Gamma_0 \Big[&(1 - Q_{\mathrm{fr}})\ \mathrm{p}(P)_{\mathrm{fg}}\, \mathrm{p}(M)_{\mathrm{fg}} \left[\mathrm{p}_{\mathrm{ret}}\ \mathrm{p}(R|M)_{\mathrm{cg}} \right. \\ &\left. + (1 - \mathrm{p}_{\mathrm{ret}}) \mathrm{p}(R|M)_{\mathrm{cr}} \right] \\ &+\ Q_{\mathrm{fr}}\ \mathrm{p}(P)_{\mathrm{fr}}\, \mathrm{p}(M)_{\mathrm{fr}}\, \mathrm{p}(R|M)_{\mathrm{cr}} \Big] \end{aligned} \tag{3.20}$$

For our intrinsically rocky population, we use the same parametrization of the period and mass distributions:

$$\begin{aligned} \mathrm{p}(P)_{\mathrm{fr}} &= A_{\mathrm{fr}} P^{\beta_{1,\mathrm{fr}}},\ P < P_{\mathrm{break,fr}} \\ \mathrm{p}(P)_{\mathrm{fr}} &= A_{\mathrm{fr}} P_{\mathrm{break,fr}}^{\beta_{1,\mathrm{fr}} - \beta_{2,\mathrm{fr}}} P^{\beta_{2,\mathrm{fr}}},\ P > P_{\mathrm{break,fr}} \\ \mathrm{p}(M)_{\mathrm{fr}} &= \mathrm{lnN}(M | \mu_{M,\mathrm{fr}}, \sigma_{M,\mathrm{fr}}) \end{aligned} \tag{3.21}$$

with $A_{\rm fr}$ as a normalization constant dependent on $\beta_{1,\rm fr}, \beta_{2,\rm fr}$ and $P_{\rm break,fr}$ as in Equation (3.5).

After adding the the three parameters from the period distribution, the two parameters from the mass distribution, and the fraction of intrinsically rocky planets $Q_{\rm fr}$, this model has a total of 22 free parameters. Following suit with our previous models, we will refer to this model with three populations of planets as "model 3".

3.4.4 Mixture of Log-Normals

Perhaps the biggest weakness of the previous models is the lack of flexibility in the mass distribution. By assuming a log-normal distribution, we are potentially missing features in the *Kepler* mass distribution that can't be captured with a two-parameter distribution.

We modify our mixture model as described in the previous section by using a mixture of log-normals as the mass distribution of the gaseous population. In practice, this is implemented by extending the previous model to include additional mixtures, modeling $\mathrm{p}(q)$ as an $(N-1)$-simplex:

$$\begin{aligned} \mathrm{p}(q) &= Q_q \\ \sum_{q=0}^{N-2} Q_q &= 1 \end{aligned} \tag{3.22}$$

where q is the qth mixture and Q_q is the fraction of planets belonging to that mixture. While technically this mixture of log-normals is implemented as additional mixtures in the model, qualitatively the multiple mixtures are still considered to be part of intrinsically gaseous population, so we still consider this to be a model with three populations of planets. In this paper we try $N = 4$, which corresponds to a intrinsically gaseous population of planets drawn from a mixture of two log-normals, but higher N can be easily modeled. Models 1, 2, and 3

described in the preceding sections correspond to $N = 1, 2, 3$ respectively. Our model with $N = 4$ requires three additional parameters: the mean and spread of the second component of the intrinsically gaseous mass distribution, $\mu_{M,2}$ and $\sigma_{M,2}$, as well as the overall fraction of planets belonging to this component, Q_2. This model has a total of 25 parameters and is the final model we present in this paper, labeled "model 4".

3.5 Fitting

3pt

We model our *Kepler* planet catalog as draws from an inhomogeneous Poisson process. This technique has been used previously to constrain the occurrence rate of *Kepler* planets in radius-period space (Foreman-Mackey et al., 2014). We use a corrected form of their likelihood:

$$\mathrm{p}(\{w_k\}|\theta) = \exp\left(-\int \hat{\Gamma}_\theta(w)\mathrm{d}w\right) \prod_{k=1}^{K} \Gamma_\theta(w_k)\mathrm{p_{tr}}(w_k) \tag{3.23}$$

where w represents the planet parameters of interest (in our case the true planet mass, M_{true}, the true planet radius R_{true}, and the planet period P), k is the index for a particular planet for a total of K planets, $\{w_k\}$ is the collection of all planet properties, θ represents our population-level parameters, p_{tr} is the a priori probability of a planet to transit (R_{star}/a) and $\hat{\Gamma}_\theta$ is the observable occurrence rate density:

$$\hat{\Gamma}_\theta(w) = \Gamma_\theta(w)\mathrm{p_{det}}(w)\mathrm{p_{tr}}(w) \tag{3.24}$$

with p_{det} representing the probability that a planet of a given radius and period would be detected by the *Kepler* pipeline and subsequently identified as a planet candidate by the Robovetter. The true planet mass and radius, M_{true} and R_{true} respectively, are considered to be free parameters in the model. As in Foreman-Mackey et al. (2014), we expand the

Parameter	Units	Model 1	Model 2	Model 3	Model 4	Prior
Γ_0	$N_{\rm pl}/N_{\rm s}$	$1.28^{+0.06}_{-0.06}$	$1.14^{+0.06}_{-0.05}$	$1.04^{+0.05}_{-0.05}$	$1.05^{+0.05}_{-0.05}$	lnN(0, 1)
C	$R_\oplus$	-	$2.37^{+0.20}_{-0.22}$	$2.08^{+0.28}_{-0.23}$	$2.32^{+0.31}_{-0.34}$	N(2.5, 1)
γ_0	-	-	$0.00^{+0.05}_{-0.05}$	$0.08^{+0.06}_{-0.07}$	$0.02^{+0.09}_{-0.07}$	N(0, 0.1)
γ_1	-	$0.42^{+0.01}_{-0.01}$	$0.74^{+0.05}_{-0.04}$	$0.75^{+0.05}_{-0.05}$	$0.61^{+0.04}_{-0.03}$	N(0.6, 0.1)
γ_2	-	$0.08^{+0.07}_{-0.07}$	$0.04^{+0.06}_{-0.06}$	$0.04^{+0.05}_{-0.06}$	$-0.01^{+0.05}_{-0.05}$	N(0, 0.1)
σ_0	-	$0.07^{+0.02}_{-0.01}$	$0.18^{+0.02}_{-0.02}$	$0.16^{+0.02}_{-0.02}$	$0.16^{+0.02}_{-0.02}$	lnN(-1.8, 0.25)*
σ_1	-	$0.27^{+0.03}_{-0.03}$	$0.34^{+0.06}_{-0.05}$	$0.33^{+0.06}_{-0.05}$	$0.23^{+0.05}_{-0.04}$	lnN(-1.3, 0.25)
σ_2	-	$0.11^{+0.03}_{-0.02}$	$0.10^{+0.02}_{-0.02}$	$0.10^{+0.02}_{-0.02}$	$0.11^{+0.02}_{-0.02}$	lnN(-2.3, 0.25)
$M_{\rm break,1}$	$M_\oplus$	$0.43^{+0.08}_{-0.08}$	$17.4^{+2.5}_{-2.3}$	$20.0^{+3.4}_{-3.0}$	$9.76^{+1.94}_{-1.43}$	lnN(2, 1)
$M_{\rm break,2}$	$M_\oplus$	$267.08^{+51.6}_{-38.7}$	$175.7^{+25.3}_{-20.5}$	$168.9^{+24.0}_{-19.0}$	$162.2^{+24.9}_{-20.1}$	lnN(5, 0.25)
μ_M	$\ln(\frac{M}{M_\oplus})$	$0.29^{+0.14}_{-0.16}$	$1.00^{+0.07}_{-0.08}$	$1.72^{+0.13}_{-0.14}$	$0.60^{+0.49}_{-0.47}$	N(1, 2)*
σ_M	$\ln(\frac{M}{M_\oplus})$	$1.72^{+0.09}_{-0.07}$	$1.65^{+0.06}_{-0.06}$	$1.38^{+0.08}_{-0.07}$	$2.39^{+0.29}_{-0.24}$	lnN(1, 0.25)*
$\mu_{M,2}$	$\ln(\frac{M}{M_\oplus})$	-	-	-	$1.72^{+0.08}_{-0.08}$	N(2, 0.5)
$\sigma_{M,2}$	$\ln(\frac{M}{M_\oplus})$	-	-	-	$0.63^{+0.07}_{-0.06}$	lnN(0, 0.25)
$\mu_{M,\rm fr}$	$\ln(\frac{M}{M_\oplus})$	-	-	$-0.15^{+0.25}_{-0.34}$	$-0.31^{+0.24}_{-0.31}$	N(0, 2)
$\sigma_{M,\rm fr}$	$\ln(\frac{M}{M_\oplus})$	-	-	$1.61^{+0.18}_{-0.14}$	$1.42^{+0.19}_{-0.15}$	lnN(0.5, 0.25)
β_1	-	$0.90^{+0.08}_{-0.07}$	$0.88^{+0.07}_{-0.07}$	$1.24^{+0.19}_{-0.17}$	$1.14^{+0.15}_{-0.15}$	N(0.5, 0.5)
β_2	-	$-0.69^{+0.06}_{-0.05}$	$-0.76^{+0.06}_{-0.06}$	$-0.71^{+0.07}_{-0.06}$	$-0.73^{+0.06}_{-0.07}$	N(-0.5, 0.5)
$P_{\rm break}$	days	$7.01^{+0.44}_{-0.49}$	$7.16^{+0.45}_{-0.44}$	$8.09^{+0.71}_{-0.57}$	$7.81^{+0.59}_{-0.58}$	lnN(2, 1)
$\beta_{1,\rm fr}$	-	-	-	$0.90^{+0.17}_{-0.15}$	$0.82^{+0.17}_{-0.18}$	N(0.5, 0.5)
$\beta_{2,\rm fr}$	-	-	-	$-1.25^{+0.19}_{-0.26}$	$-1.11^{+0.19}_{-0.24}$	N(-0.5, 0.5)
$P_{\rm break,fr}$	days	-	-	$4.38^{+0.60}_{-0.44}$	$4.58^{+1.06}_{-0.60}$	lnN(2, 1)
α	-	-	$7.98^{+1.40}_{-1.36}$	$7.29^{+1.84}_{-1.52}$	$8.39^{+1.53}_{-1.45}$	lnN(0, 1)
$Q_{\rm fr}$	-	-	-	$0.20^{+0.06}_{-0.05}$	$0.21^{+0.05}_{-0.05}$	U(0, 1)*
Q_2	-	-	-	-	$0.26^{+0.07}_{-0.06}$	D(1)

Table 3.1: The retrieved median and 1-σ intervals for each parameter of interest for each model, as well the units and the assumed prior. N represents a normal distribution, lnN represents a lognormal distribution, and D represents a Dirichlet distribution with parameter length equal to the number of mixture components. Descriptions of each parameter can be found in section 3.4. Parameters with units listed as '-' are dimensionless. The priors with an * differ depending on the model. In model 1, the σ_0 parameter takes a $\ln \mathrm{N}(-2.8, 0.25)$ prior to emulate the low scatter in the mass-radius relation for small, rocky planets. The μ_M and σ_M priors are changed to $\mathrm{N}(0, 0.5)$ and $\ln \mathrm{N}(0, 0.25)$ respectively in model 4 in order to impose identifiability in the mixture model by clearly separating the two mass distribution components. In addition for model 4, the parameter $Q_{\rm fr}$ is fit together with the other mixture components as a simplex and is assigned a Dirichlet prior that gives any combination of mixture probabilities equal weighting.

likelihood to account for the fact that our inputs are observed masses and radii, rather than the true values:

$$\begin{aligned} \mathrm{p}(\{x_k\}|\theta, \{w_k\}) = \exp\left(-\int \hat{\Gamma}_\theta(w)\mathrm{d}w\right) \\ \cdot \prod_{k=1}^{K} \Gamma_\theta(w_k)\mathrm{p}(x_k|w_k)\mathrm{p}_{\mathrm{tr}}(w_k) \end{aligned} \tag{3.25}$$

where x_k represents the measurements of the kth planet's mass and radius. The $\mathrm{p}(x_k|w_k)$ term is given by the following:

$$\begin{aligned} \mathrm{p}(x_k|w_k) = N(R_{\mathrm{true},k}|R_{\mathrm{obs},k}, \sigma_{R,\mathrm{obs},k}) \\ \cdot N(M_{\mathrm{true},k}|M_{\mathrm{obs},k}, \sigma_{M,\mathrm{obs},k}) \end{aligned} \tag{3.26}$$

where $R_{\mathrm{obs,k}}, \sigma_{\mathrm{R,obs,k}}$ are the measured radius and radius uncertainty for the kth planet, and similarly for the measured masses. In the case where a planet does not have a mass measurement, Equation (3.26) changes to only include the first term. Note that due to the high precision to which periods of *Kepler* planets are measured, we neglect observational uncertainty in the orbital period and take $P_{\mathrm{obs,k}} = P_{\mathrm{true,k}}$

The difference between the likelihood used here and that presented in Foreman-Mackey et al. (2014) is the absence of the p_{det} term in the multiplicative sum. Including the p_{det} term in both the integral and the individual planet likelihoods is incorrectly conditioning on the detection twice (Loredo, 2004; Mandel et al., 2019). The geometric transit probability, p_{tr}, appears in both because the subset of planets that are transiting is the population we are able to detect.

For our mixture models, we use a modified form of the inhomogeneous Poisson process likelihood:

$$\mathrm{p}(\{w_k\}|\theta) = \exp\left(-\int\left(\sum_{q=0}^{N-1}\mathrm{p}(q)\hat{\Gamma}_{\theta,q}(w)\right)\mathrm{d}w\right) \cdot \prod_{k=1}^{K}\left(\sum_{q=0}^{N-1}\mathrm{p}(q)\Gamma_{\theta,q}(w_k)\mathrm{p}_{\mathrm{tr}}(w_k)\right) \tag{3.27}$$

where $\mathrm{p}(q)$ are the mixture probabilities of the qth mixture, for a total of N mixtures.

Given that a minority of transiting planets in our sample have RV mass measurements, and given the complexity of some of the models we are attempting to fit, we are forced to adopt reasonably informative priors for our population level model parameters. We refer to the literature to determine our priors where previous work has been done (without explicitly using past results as our priors), striking a balance between flexibility and being informative. Our priors for each parameter in our models are listed in the last column of Table 3.1.

To fit our models to the data, we use the Python implementation of the Stan statistical software package (Carpenter et al., 2017)[3]. Stan uses the No-U-Turn Sampler (NUTS) MCMC algorithm, a method of numerically evaluating hierarchical Bayesian models. The ability of NUTS to efficiently handle large dimensional spaces makes it particularly suitable for this work. The CKS sample contains over a thousand planets, and each is modeled with a "true" mass and radius, totalling over two thousand parameters. For each model, we ran 8 chains each with 3,000 iterations. The first 2,000 iterations of each chain are used for warm up, where they are allowed to reach their equilibrium distributions, and the MCMC algorithm fine-tunes its internal parameters. We are left with 8,000 posterior samples of each parameter. To assess the convergence and independence of each chain, we look at the Gelman-Rubin convergence diagnostic, $\hat{R}$. For each parameter, we ensure that $\hat{R} < 1.01$, a reasonable benchmark for chains mixing well.

3. http://mc-stan.org

In order to assess the validity of our model fits, we perform tests on simulated data for each of our models. For each model, we generate 10 simulated catalogs per set of chosen model parameters. We ensure that the model retrieval for each parameter is centered on the true value, without being biased high or low. We repeat this for several sets of model parameters drawn from our prior distributions to ensure we can sufficiently distinguish between these different parameters using the same priors.

Since the inhomogeneous Poisson process likelihood in Eqs (3.23) and (3.27) contains a three-dimensional integral, the computational speed is severely bottlenecked by the grid of period, mass and radius points we use to compute the integral, especially in the mixture model case where the mass and radius distributions are no longer independent of period. We use a grid of size 11x41x41 in period-mass-radius space, with a detection efficiency grid of the same size to compute this integral. This grid size was chosen based on the condition of minimizing both the run time and error; we keep errors down to a few percent when compared to using a denser grid.

3.6 Results

3.6.1 Model Fits

Our results are summarized in Table 3.1, where we present the median and 1σ uncertainties (68% confidence interval) for each parameter, for each of our 4 models, along with the assumed prior. Of course, to understand the full picture, one must look at the full N-dimensional posterior to account for correlations between parameters rather than looking at each parameter independently. For all of the following results and plots, we sample from the posteriors of each of our models such that the uncertainties from the model fits are folded in.

For model 1, the best comparison for the mass-radius portion of the model is to Chen &

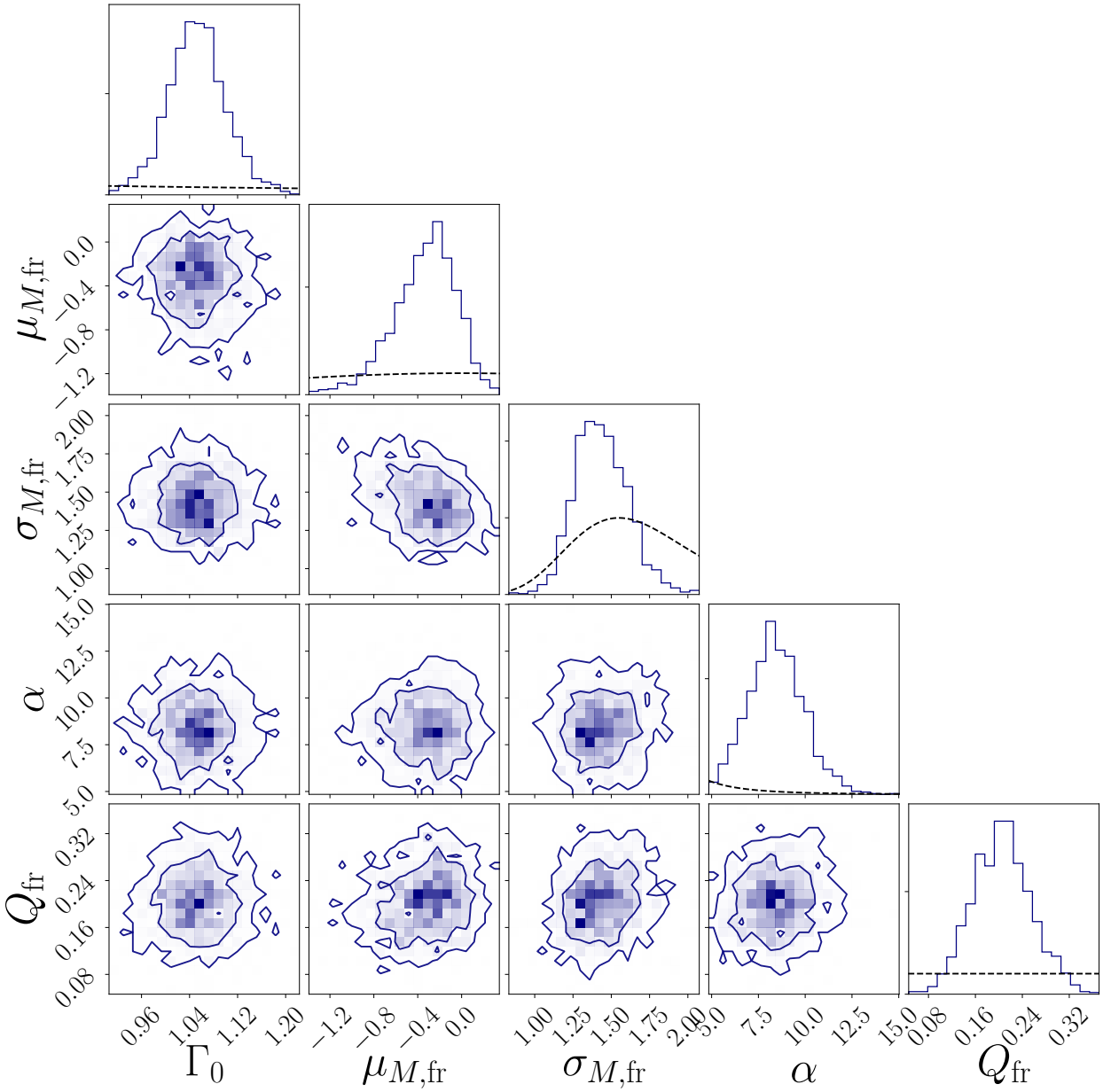

Figure 3.2: The 2D posteriors for a selection of parameters constraining the rocky mixture with model 4. The parameters are: Γ_0, the overall occurrence rate normalization (planets per star); $\mu_{M,\mathrm{fr}}$, the mean of the log-normal distribution for the intrinsically rocky population; $\sigma_{M,\mathrm{fr}}$, the spread of the log-normal distribution for the intrinsically rocky population; α, the scaling of the mass-loss equation; Q_{fr}, the total fraction of intrinsically rocky planets. The contours are at the 1σ and 2σ levels, corresponding to the 68% and 95% confidence intervals, with the shading indicating the density of posterior samples in a given bin. The black dotted lines in the 1D distributions show the assumed priors.

Kipping (2017) (hereafter CK17), as we have modeled our mass-radius relation after their work. We find a shallower mid-size slope γ_1 ($0.42^{+0.01}_{-0.01}$ in our work vs $0.59^{+0.04}_{-0.03}$ in CK17) with a slightly positive slope for γ_2 ($0.08^{+0.07}_{-0.07}$ vs $-0.04^{+0.02}_{-0.02}$), in the giant planet regime. Our breaks in the mass-radius relation are found at significantly lower ($0.43^{+0.08}_{-0.08}$ vs $2.04^{+0.66}_{-0.59}$ $M_\oplus$ for $M_{\rm break,1}$) and higher (267^{+52}_{-39} vs 132^{+18}_{-21} $M_\oplus$ for $M_{\rm break,2}$) masses. This low value for the low mass break is below most planets in the sample, perhaps indicating weak support for including this break based on the CKS sample. This is supported by the nonparametric mass-radius relation fit by Ning et al. (2018), which found no evidence for a low mass break in the mass-radius relation with the current mass-radius sample. The scatter parameters σ_0, σ_1, σ_2 that we retrieve are higher across the board ($0.07^{+0.02}_{-0.01}$ vs $0.04^{+0.01}_{-0.01}$, $0.27^{+0.03}_{-0.03}$ vs $0.15^{+0.02}_{-0.01}$, and $0.11^{+0.03}_{-0.02}$ vs $0.07^{+0.01}_{-0.01}$, respectively). For each of our parameters, the uncertainty is much higher than in CK17, likely due to the much more limited sample of planets with mass and radius measurements that we condition on. Whereas we only include masses for planets in the CKS sample with RV mass measurements (described in Section 2.3.1), CK17 includes many more giant planets and solar system planets, among others. The fact that we are modeling the full 3D distribution and are accounting for non-detections and selection effects while CK17 simply modeled the mass conditioned on radius distribution also likely widens the posteriors for these parameters. We note that the goal of this work to fit the mass-radius relation together with the mass and period distributions self-consistently using one dataset, and the mass-radius relation is just one aspect of this overall distribution.

The overall normalization (planets per star) within our range of parameter space ($0.4 < \frac{R}{R_\oplus} < 30$, $0.3 < \frac{P}{1\text{ day}} < 100, 0.1 < \frac{M}{M_\oplus} < 10,000$), Γ_0, is found to be $1.28^{+0.06}_{-0.06}$ for model 1 (for continued discussion of occurrence rates, see section 3.7.2). The log-normal mass distribution peaks at a value of $\mu_M = 0.29^{+0.14}_{-0.16}$ (corresponding to $1.34 M_\oplus$) with a spread in natural log space of $\sigma_M = 1.72^{+0.09}_{-0.07}$. Finally, the period distribution has a break at $7.0^{+0.44}_{-0.49}$ days with a low period power-law slope of $0.90^{+0.08}_{-0.07}$ and a high period power-law

slope of $-0.69^{+0.06}_{-0.05}$. Mulders et al. (2018), who also parametrized the period distribution of *Kepler* planets with a broken power-law, found a break at 12^{+3}_{-2} days and slopes of $1.5^{+0.5}_{-0.3}$ and $0.3^{+0.1}_{-0.2}$. However, they defined their occurrence rate density in terms of $\log P \log R$, which means their slopes correspond to slopes of 0.5 and -0.7 in our model. Our slopes are thus compatible within uncertainties, noting that they consider a period range of 2-400 days, compared to our range of 0.3-100 days, and parametrize the rest of the model differently, which could account for any differences.

Introducing XUV-driven hydrodynamic mass loss with model 2 introduces several new parameters, and also significantly affects several existing model parameters. Replacing the fixed low-mass mass-radius relation in model 1, we fit for the slope γ_0 and normalization C of this low-mass gaseous planet regime. We find a nearly flat slope of $0.04^{+0.06}_{-0.06}$, a normalization of $2.37^{+0.20}_{-0.22}$ $R_\oplus$, a first break in the mass-radius relation at $17.4^{+2.5}_{-2.3}$ $M_\oplus$, and a scatter of $0.18^{+0.02}_{-0.02}$. Given the flat slope, this normalization should correspond to the 2.4 $R_\oplus$ peak of the bimodal radius distribution first uncovered in Fulton et al. (2017); indeed, we find our value to be consistent with their reported location of the second peak. Allowing the possibility of envelope mass loss shifts the intermediate-mass slope γ_1 of the mass-radius relation to $0.74^{+0.05}_{-0.04}$, the peak of the log-normal mass distribution to higher mass ($1.00^{+0.07}_{-0.08}$, or 2.72 $M_\oplus$), and the second break in the mass-radius relation to lower mass ($175.7^{+25.3}_{-20.5}$ $M_\oplus$). Finally, we allowed the prefactor governing envelope mass loss to vary, and found that it favored easier envelope retention than our assumed values for stellar age, mass loss efficiency, and XUV flux from the star ($\alpha = 7.98^{+1.40}_{-1.36}$).

With model 3 we add a population of intrinsically rocky planets with separate mass and period distributions from the gaseous and evaporated core populations in model 2. Compared to the gaseous and evaporated core populations, the intrinsically rocky population has a log-normal mass distribution that peaks at a lower mass ($\mu_{M,\mathrm{fr}} = -0.15^{+0.25}_{-0.34}$ vs $\mu_M = 1.72^{+0.13}_{-0.14}$, or 0.86 $M_\oplus$ vs 5.58 $M_\oplus$), and a period distribution with a shallower slope at low periods

($\beta_{1,\mathrm{fr}} = 0.90^{+0.17}_{-0.15}$ vs $\beta_1 = 1.24^{+0.19}_{-0.17}$) and steeper slope at high periods ($\beta_{2,\mathrm{fr}} = -1.25^{+0.19}_{-0.26}$ vs $\beta_2 = -0.71^{+0.07}_{-0.06}$), leading to more planets at shorter periods. While the overall period distribution for the intrinsically rocky population is shifted towards shorter periods, there are also more intrinsically rocky planets at longer orbital periods compared to the evaporated core population, as most of these longer period intrinsically gaseous planets retain their envelopes instead of becoming evaporated cores. Adding this additional population does not significantly affect the mass-radius relation for the gaseous population nor the α factor for the envelope mass loss. We find the total fraction of intrinsically rocky planets to be $Q_{\mathrm{fr}} = 0.20^{+0.06}_{-0.05}$. While this intrinsically rocky population is in many ways degenerate with the evaporated core population, the fact that it is well constrained above zero perhaps indicates the need for rocky planets either at higher masses or longer periods than the mass loss prescription would allow. For further discussion of the need for the intrinsically rocky population and the meaning of these labels, see Section 3.7.1. We include a subset of the 2D posteriors for model 3 in Figure 3.2, for the parameters relevant to the intrinsically rocky distribution. We find that the overall normalization is not strongly correlated with any of these parameters. However, there exists a weaker correlation with the mass loss scaling factor α: higher values, which correspond to higher envelope retention probabilities, lead to a higher fraction of intrinsically rocky planets Q_{fr}, to balance with the fewer number of evaporated cores. This is evidence for the degeneracy between the evaporated core and intrinsically rocky populations

In model 4, we allow a more flexible mass distribution for the gaseous and evaporated core populations by using a mixture of two log-normal distributions. We find that the mean of the high mean mass log-normal in this model is close to that of model 3 ($\mu_{M,2} = 1.72^{+0.08}_{-0.08}$, or 5.58 $M_\oplus$) but with a much lower scatter ($\sigma_{M,2} = 0.63^{+0.07}_{-0.06}$). The low mean mass log-normal, by contrast, has a mean at lower masses ($\mu_{M,1} = 0.60^{+0.49}_{-0.47}$, or 1.82 $M_\oplus$) with a high scatter ($\sigma_{M,1} = 2.39^{+0.29}_{-0.24}$. The high mean mass component has about twice the weight of the

low mean mass component ($Q_3 = 0.51$ vs $Q_2 = 0.26$). This has the effect of concentrating most of the mass between $1 - 30 M_\oplus$, with long tails toward low and high masses. The high mean mass component accounts for the bulk of the super-Earths and mini-Neptunes found by the *Kepler* survey, while the low mean mass component includes planets of all sizes, from sub-Earth sizes to gas giants. Adding this extra component in the mass distribution also has some significant effects in the mass-radius relation, such as shifting the intermediate-mass segment towards lower masses ($\gamma_1 = 0.61^{+0.04}_{-0.03}$ vs $0.75^{+0.05}_{-0.05}$, $M_{\rm break,1} = 9.76^{+1.94}_{-1.43}$ vs $20.0^{+3.4}_{-3.0}$ $M_\oplus$), with lower scatter ($\sigma_1 = 0.23^{+0.05}_{-0.04}$ vs $0.33^{+0.06}_{-0.05}$).

We summarize the joint mass-radius-period distribution for model 4 in Figure 3.3. We include the resulting 1D mass, period and radius distributions for each mixture, compared to the observed distribution of the sample. We also include a representation of the model in mass-radius space, showing both the mass-radius relation and relative occurrence for all three mixtures, along with the mass-radius measurements included in the sample. We note that these 1D and 2D distributions in Figure 3.3 are all projections, and do not fully capture the 3D nature of the distribution. While there is no explicit radius or mass dependence in the period distribution, they are not independent in models 2-4 due to the irradiation flux dependence of the envelope mass loss.

We provide an enlarged version of the radius distribution for model 4 in Figure 3.4 for short period and long period planets, along with the fraction of planets belonging to each mixture as a function of radius in Figure 3.5. Figure 3.4 is best compared to Figure 7 in Fulton et al. (2017). We find similar locations for the two peaks, although the gap in between the two is not as deep and even seems to nearly disappear for planets with periods longer than 10 days. We note two possible explanations for this. First, this radius distribution is period dependent. Compared to the radius distribution of planets with $P > 10$ days, the radius distribution of planets with $P < 10$ days is shifted towards lower radii, with the 1.3 $R_\oplus$ peak exceeding the height of the 2.4 $R_\oplus$ peak. This is because there are fewer

gaseous planets and more evaporated cores and intrinsically rocky planets at short periods. Second, there may not be enough flexibility in our model to allow for a deep gap in the radius distribution, perhaps because of the nature of the mass loss and the smoothness of the log normal distributions. This is further evidenced in Figure 3.7, where we compare the observed radius distribution of the CKS sample with a simulated observed distribution with our model. The most discrepant bin is the location of the gap, at slightly below 2.0 $R_\oplus$, where our model predicts a significantly lower number of planets.

The transition between the gaseous planets and evaporated cores as shown in both Figures 3.4 and 3.5 is steep, with significant overlap between the two only in a small range in radius between 1.5 and 2.0 $R_\oplus$. This is likely mostly due to the envelope mass loss prescription used in the model. We find similar numbers of intrinsically rocky and evaporated core planets, with slightly more intrinsically rocky, especially at low radii ($R < 0.8R_\oplus$) and high radii ($R > 1.8R_\oplus$).

We note that the intrinsically rocky population extends to higher masses and radii than expected. Rogers (2015) found that most observed planets with $R > 1.6$ $R_\oplus$ are not consistent with rocky compositions, yet we find significant amounts of them. For example, as shown in Figure 3.5, at $R = 2.0R_\oplus$ about 5% of planets are intrinsically rocky, and there exists a small tail that goes up to $R = 3.0$ $R_\oplus$. This is most likely a shortcoming of the model and a data-driven approach: a log-normal mass distribution leads to these high-mass tails, and there is nothing in the model that precludes these planets despite physical expectations for why they should be rare. Simply put, these high-mass rocky planets are massive enough to have accreted significant amounts of gas in the disk during formation (Ikoma & Hori, 2012; Bodenheimer & Lissauer, 2014).

We plot the radius-period distributions of the three populations of planets for model 4 in Figure 3.6. The negative slope of the radius valley found by Van Eylen et al. (2018) can be seen by the division between the gaseous planets and the evaporated cores. However, the

intrinsically rocky population serves to wash out the radius gap, as the period distribution of the intrinsically rocky population is independent of its radius distribution. While the gap is washed out, adding the intrinsically rocky planets does not eliminate the negative slope of the radius gap, as the fraction of evaporated cores is constrained to be above zero. Therefore we find evidence that there is a population of planets subject to photoevaporation, consistent with Van Eylen et al. (2018), but the gap that we infer is not as clearly distinguished.

3.6.2 Occurrence Rate Comparisons

We compute occurrence rates using each of our four models. Occurrence rates are calculated analytically by integrating Equation 3.4 over various ranges in period, mass and radius. We sample over the population-level parameter posteriors in order to incorporate these uncertainties into our occurrence rate estimates. Our reported values and uncertainties are the median and 1-σ percentiles from analytically calculating these occurrence rates over $1,000$ posterior samples of the population-level parameters. Our uncertainties do not reflect the uncertainty in the choice of model or parametrization thereof. All reported values are conditional on the model chosen and the data sample used to fit that model.

We compare occurrence rates derived from our models with several values from recent works in Table 3.2. We find broad agreement with these previously calculated occurrence rates. Differences likely arise due to the dissimilarities in our data sample and, likely more importantly, model parametrization. Our four 3D joint distribution models are much more complex in their parametrization compared to the common method of constraining the heights of predefined bins. We also note that computing occurrence rates for periods extending past 100 days (the upper limit for our sample and model) involves an extrapolation of our results to these longer periods.

Overall, for the widest radius and period ranges, as in the first two rows of Table 3.2, we find that the occurrence rate generally decreases with the added complexities of models 2, 3

and 4. The occurrence rate estimates from each model presented here are consistent within error of the result from Mulders et al. (2018), who modeled the radius-period distribution as independent broken power-laws. The occurrence rates based on our model 1 (with no envelope mass loss) show the best agreement with Mulders et al. (2018), as expected given that model 1's low complexity most closely matches the independent broken power-laws assumed by Mulders et al. (2018). For the mini-Neptune occurrence rates measured by Fulton et al. (2017) (rows 3 and 4 in Table 3.2), we find that the more complex models predict a higher occurrence rate relative to model 1 and Fulton et al. (2017), despite a lower overall occurrence rate integrated over our full parameter space range. Interestingly, these models also predict a similar occurrence rate for planets between 1.4 and 2.8 $R_\oplus$, and planets between 2 and 4 $R_\oplus$, whereas Fulton et al. (2017) find the former to be slightly more common. Our estimates of the Hot Jupiter occurrence rate are consistent with RV measurements from Mayor et al. (2011), but are dependent on the model assumed. Our estimates of the Super-Earth occurrence rate are higher than the RV measurements from Howard et al. (2010), especially for model 4, which is higher by nearly a factor of 3. We find a rate consistent with previous *Kepler* measurements of the Hot Jupiter occurrence rates with model 1, and a higher rate with models 2 and 3 (Petigura et al., 2017). Interestingly, the more flexible mass distribution used in model 4 decreases this rate in line with model 1 and previous *Kepler* results.

Our estimates of $\eta_\oplus$, as defined by the occurrence rate of planets around GK stars within 20% of $R = 1R_\oplus$ and $P = 1$ year, range from 0.08 to 0.008, and are consistent with the wide range of uncertainty reported in Burke et al. (2015). However, as with any estimate of $\eta_\oplus$ based on current data, our estimate presented in Table 3.2 requires extrapolation, and thus is highly model dependent and subject to a large degree of uncertainty not taken into account in the error bars. While our models only constrained the period distribution out to 100 days, calculating $\eta_\oplus$ to within 20% of a year means that we extrapolate our period distribution

to over four times its stated boundary. This assumes the period distribution continues as a power-law from $P_{\rm break}$ to 438 days. This is a wide range to assume a single power-law; there could easily be another break in the period distribution that we are missing, or perhaps a power-law does not fit this range well at all. Secondly, in this wider period range, any dependence of the period distribution on radius or mass is likely to be more significant. While our modeling of three mixture populations allows for correlations between mass or radius and period (e.g. fewer low-density planets at short orbital periods due to hydrodynamic envelope mass loss), there can be additional correlations between these quantities within a mixture that we are not accounting for. Thus, care must be taken to avoid overinterpretation of these (and all other) extrapolated $\eta_\oplus$ estimates.

Despite these caveats, there are important lessons to take away in our extrapolation of $\eta_\oplus$. We find that our estimates vary by nearly an order of magnitude depending on the choice of model: $\eta_\oplus$ calculated with model 1 is consistent with Burke et al. (2015), whereas models 2-4 find a much lower $\eta_\oplus$. In model 1, the period distribution is completely independent of the radius and mass distributions, so rocky planets and gaseous planets extrapolate similarly to longer periods, and the ratio between the two remains the same at short orbits versus longer orbits. When we introduce envelope mass loss with model 2, as we go out to longer orbital periods the proportion of gaseous planets relative to rocky planets increases, as planets at these large distances from their star can more easily retain their envelopes. Accordingly, the $\eta_\oplus$ estimate drops by nearly an order of magnitude, as planets with H/He envelopes typically have radii larger than 1.2 $R_\oplus$.

Though the overall extrapolated $\eta_\oplus$ value is similar for models 2-4, the mass (and hence radius) distribution of the extrapolated planet populations at $\approx$ 1AU are very different. To illustrate this, we refer to Figure 3.6, where we highlight the bounds in which we calculate $\eta_\oplus$. For model 2, planets that fall in the $\eta_\oplus$ bin are split between the gaseous and evaporated core populations (see section 3.7.2 for further discussion). The inclusion of an intrinsically

rocky population in models 3-4 allows for more massive rocky planets at these longer orbital periods, compared to the evaporated core population. Accordingly, as shown in Figure 3.6, the intrinsically rocky planets are much more evenly distributed in radius inside the bin, whereas the evaporated core planets are preferentially lower radii.

In our calculation of $\eta_\oplus$, we used the same boundaries in radius and period as Burke et al. (2015), to make the comparison direct. However, we are neglecting the true power of the full mass-radius-period distribution in doing so. As mentioned above, in our calculation of $\eta_\oplus$ we were including planets of a wide range of densities in this bin. By adding constraints on the mass range in calculating $\eta_\oplus$, we can ensure the planets are more Earth-like. We add a bound of a factor of 2 on either side of an Earth mass (0.5 - 2 $M_\oplus$) and recalculate $\eta_\oplus$. We find that the numbers drop even more, with $0.055^{+0.011}_{-0.009}$ for model 1, and $0.005^{+0.003}_{-0.002}$, $0.006^{+0.004}_{-0.003}$, and $0.008^{+0.006}_{-0.004}$ for models 2, 3 and 4, respectively. We note that with these bounds, model 3 is higher than model 2, in contrast to when calculated with no mass bounds. Clearly, this wide range of answers should give even more pause when considering any extrapolated $\eta_\oplus$ calculations for planets around GK dwarfs at face value.

The comparisons in Table 3.2 were all derived using *Kepler* data, with the exception of the comparison to the RV-measured Hot Jupiter occurrence rate from Mayor et al. (2011) as well as the RV-measured Super-Earth occurrence rate from Howard et al. (2010). With the full mass-radius-period distribution modeled here, comparisons with surveys that measure planet mass, such as RV and microlensing surveys, are more direct and straightforward than if we were to simply model the radius-period distribution. However, differences in period and host star distributions, among other effects, still make direct comparisons with these surveys difficult. Nonetheless, our occurrence rates derived here are more generalizable to a wider range of parameter space, and our comparisons demonstrate their validity by showing consistency with prior results. The model dependence of our results, and to a further extent all occurrence rate calculations, must be kept in mind to avoid overinterpretation.

Variable Range	Model 1	Model 2	Model 3	Model 4	Literature Value	Reference
$0.4 < \frac{R}{R_\oplus} < 30$, $0.3 < \frac{P}{1\text{ day}} < 100$	$1.36^{+0.07}_{-0.06}$	$1.31^{+0.06}_{-0.06}$	$1.10^{+0.05}_{-0.05}$	$1.12^{+0.06}_{-0.05}$	-	-
$0.5 < \frac{R}{R_\oplus} < 6$, $2 < \frac{P}{1\text{ day}} < 400$	$2.28^{+0.22}_{-0.18}$	$1.89^{+0.18}_{-0.13}$	$1.70^{+0.18}_{-0.17}$	$1.70^{+0.21}_{-0.17}$	$2.4^{+0.5}_{-0.5}$	Mulders et al. (2018)
$1.4 < \frac{R}{R_\oplus} < 2.8$, $0.3 < \frac{P}{1\text{ day}} < 100$	$0.43^{+0.02}_{-0.02}$	$0.47^{+0.03}_{-0.03}$	$0.48^{+0.04}_{-0.04}$	$0.51^{+0.05}_{-0.05}$	$0.43^{+0.02}_{-0.02}$	Fulton et al. (2017)
$2 < \frac{R}{R_\oplus} < 4$, $0.3 < \frac{P}{1\text{ day}} < 100$	$0.30^{+0.01}_{-0.01}$	$0.45^{+0.02}_{-0.02}$	$0.46^{+0.03}_{-0.03}$	$0.49^{+0.04}_{-0.04}$	$0.37^{+0.02}_{-0.02}$	Fulton et al. (2017)
$50 < \frac{M}{M_\oplus} < 10000$, $0.3 < \frac{P}{1\text{ day}} < 11$	$0.005^{+0.001}_{-0.001}$	$0.010^{+0.002}_{-0.001}$	$0.009^{+0.001}_{-0.001}$	$0.005^{+0.001}_{-0.001}$	$0.009^{+0.004}_{-0.004}$	Mayor et al. (2011)
$3 < \frac{M}{M_\oplus} < 10$, $0.3 < \frac{P}{1\text{ day}} < 50$	$0.19^{+0.01}_{-0.01}$	$0.22^{+0.01}_{-0.01}$	$0.22^{+0.02}_{-0.02}$	$0.32^{+0.03}_{-0.03}$	$0.12^{+0.04}_{-0.04}$	Howard et al. (2010)
$0.8 < \frac{R}{R_\oplus} < 1.2$, $292 < \frac{P}{1\text{ day}} < 438$	$0.076^{+0.016}_{-0.011}$	$0.008^{+0.003}_{-0.002}$	$0.008^{+0.004}_{-0.003}$	$0.009^{+0.007}_{-0.004}$	$0.1^{+1.9}_{-0.1}$	Burke et al. (2015)
$0.8 < \frac{R}{R_\oplus} < 1.2$, $292 < \frac{P}{1\text{ day}} < 438$, $0.5 < \frac{M}{M_\oplus} < 2.0$	$0.055^{+0.011}_{-0.009}$	$0.005^{+0.003}_{-0.002}$	$0.006^{+0.004}_{-0.003}$	$0.008^{+0.006}_{-0.004}$	-	-

Table 3.2: The occurrence rate within different intervals in mass, period and radius for each of the four models, along with a comparison to previous works.

3.6.3 Model Selection

In the preceding sections we have presented four viable models for a joint mass-radius-period distribution, with increasing complexity. While the additional complexities of the latter models are motivated by physical processes and planet formation theory, we would like an objective measure of whether or not these additional complexities are justified, or even necessitated, by the current data. In order to assess this, we must measure each model's predictive accuracy, or its ability to generalize to an independent sample of planets, while accounting for problems such as overfitting.

Cross-validation is a class of model selection techniques which uses the sample of interest in order to estimate the out-of-sample predictive accuracy (Hastie et al., 2001). The most comprehensive cross-validation is leave-one-out cross-validation (LOO), where for each data point the model is rerun leaving out that point, and then the likelihood of the left-out point is calculated using that model fit. Given that our dataset contains 1130 planets, this would require 1130 model runs and thus is computationally prohibitive. There are several approximations to LOO that don't necessitate additional model runs, such as Watanabe-Akaike information criterion (Watanabe, 2010), used in previous analyses (Neil & Rogers, 2018), but diagnostics (e.g. the variance over the posterior samples of the log-likelihood of a given planet) indicate that these are invalid for our current models, possibly due to their three-dimensional nature in period, mass and radius (Vehtari et al., 2017). In the field

of exoplanets, cross-validation has been used for model selection in predicting additional transiting planets in *TESS* systems (Kipping & Lam, 2017) as well as improving transit classification (Ansdell et al., 2018). We turn to K-fold cross-validation as our model selection method of choice.

We choose 10 folds for our cross-validation as a compromise between computational time and robustness. We divide the planet sample into 10 subsets, separately dividing the planets with mass measurements and those without mass measurements to ensure the ratio between the two is consistent between subsets. We then create our 10 new planet samples by excluding the planets from one of the 10 subsets. We fit our four models on each of these 10 samples, and calculate the expected log predictive density (elpd) for each planet using the model fit in which the planet was excluded:

$$\begin{aligned} \widehat{\mathrm{elpd}}_k &= \log p(x_k | \left\{ x_{(-j)} \right\}) \\ \log p(x_k | \left\{ x_{(-j)} \right\}) &= \log \left(\frac{1}{S} \sum_{s=1}^{S} p(x_k | \theta^{j,s}) \right) \end{aligned} \tag{3.28}$$

where the data for a given planet k is summarized by x_k; for a given subset j of the planet sample, the sample excluding that subset is $x_{(-j)}$ and the resulting model fit is θ^j, and we can use a total of S draws from our model posteriors to summarize our posterior distribution (Vehtari et al., 2017). For our joint distribution, generalizing to mixture models, this probability turns into:

$$\begin{aligned} p(x_k|\theta) = & \sum_{q=0}^{N_{\mathrm{mix}}} p(q|M_{obs,k}, P_k, \theta)(P_k|\theta, q) \\ & \cdot p(R_{obs,k}|M_{obs,k}, \theta, q) p(M_{obs,k}|\theta, q) \end{aligned} \tag{3.29}$$

Model	$\widehat{\text{elpd}} - \widehat{\text{elpd}}_{\text{M1}}$
Model 2	-257 ± 18
Model 3	-380 ± 19
Model 4	-376 ± 20

Table 3.3: Difference in expected log predictive density for each model compared to model 1, along with the error in the difference. Negative numbers favor the model in question over model 1. We find that each model with envelope mass loss is strongly preferred over the model without. Models 3 and 4, which include multiple populations of rocky planets, are preferred over model 2, which only has evaporated cores. From these results, Models 3 and 4 are statistically indistinguishable from each other and one is not preferred over the other.

for planets with mass measurements, where q gives the index for a given mixture component. For planets without mass measurements we remove the last term and instead marginalize over the mass distribution for the radius term. Our total expected log predictive density (elpd) for the entire model is then:

$$\widehat{\text{elpd}} = \sum_{k=1}^{K} \log p(x_k | x_{(-j)}) \tag{3.30}$$

where we are summing over each planet's contribution from Equations (3.28), (3.29), with K representing the number of planets. We can then compare the expected log predictive density between two models directly, with the standard error (se) of the difference as follows:

$$\text{se}(\widehat{\text{elpd}}^A - \widehat{\text{elpd}}^B) = \sqrt{NV_{k=1}^K(\widehat{\text{elpd}}_k^A - \widehat{\text{elpd}}_k^B)} \tag{3.31}$$

where A and B represent two different models, and V represents the sample variance (in this case the variance of the differences).

We present our results in Table 3.3. The results presented are in terms of the difference in expected log predictive density between the given model and model 1, which has no envelope mass loss. Negative numbers favor the model in question. We find that each model with envelope mass loss (models 2, 3 and 4) is strongly preferred over the basic model with a single population of planets and no envelope mass loss (model 1). Furthermore, models 3

and 4, which include a second, independent population of rocky planets, are both preferred over model 2. The difference in the expected log predictive density between models 3 and 4 is only 4, with an uncertainty in the difference of 11; thus, these results do not strongly prefer either model 3 or 4 over the other.

Given these results, we can confidently suggest that we are justified in the additional complexities introduced in the latter models. Not only are these complexities easily physically motivated, but the cross-validation shows that these models better predict out-of-sample data. One of the main purposes of this cross-validation is to make sure we are not overfitting the data and fitting peculiarities in this particular sample that do not generalize to the exoplanet population as a whole. That seems to not be the case. While 25 population-level parameters for model 4 seems like a large number, our dataset contains 1130 planets, with each planet having a radius and period measurement, and 53 of those planets having a mass measurement, for a total of 2313 measurements. This is nearly 100 times the number of parameters. Further complexity than what we have modeled here may even yield further improvement in the expected log predictive density, especially in the radius and period distributions, for which we have plenty of data.

The cross-validation results do not strongly prefer model 3 over model 4, or vice versa. However, there are other pieces of evidence we can turn to in assessing these models. The difference between the observed radius distribution of the real dataset and a simulated dataset generated with the model fit, as shown in Figure 3.7 for models 3 and 4, shows that model 4 better predicts the radius distribution at the highest radii (above $4R_\oplus$), due to the longer tail towards high masses in the mass distribution. Model 4 also better predicts the radius distribution at small radii (below $1.5R_\oplus$). These features may not show up in the cross-validation results because the expected log predictive density is a sum of all the log likelihoods for each planet, and giant planets as well as very small planets represent a small fraction of the overall dataset. Model 4 also has a smaller tail towards high masses for planets of about

$2R_\oplus$. This is evident in Figure 3.8, where for Kepler-60d the mass prediction using model 4 falls off more steeply past $15M_\oplus$ than for either model 2 or 3. These effects are relatively small, and it is reasonable that the cross-validation can not distinguish between model 3 and 4.

3.6.4 Mass/Radius Predictions

With a joint mass-radius-period distribution, we can predict planet masses from radii and vice versa. Compared to predicting masses or radii using a mass-radius relationship alone, our models have several benefits. First, the fact that we are constraining the underlying distributions, rather than just the relationship between mass and radius, allows us to weight the observation accordingly. For example, if a planet has a measured mass of $20M_\oplus$ with a Gaussian error of $5M_\oplus$, the low end of the mass range should be more strongly weighted than the high end, since in this case planets in this less massive range are more common. This would lead to a smaller radius prediction than if one had used a mass-radius relation in isolation. Secondly, the additional complexities that we model here, including envelope photo-evaporation and mixtures of distinct sub-populations of planets, are physically motivated and would be difficult to model with a mass-radius relation alone. This allows for situations where the resulting radius or mass prediction can be bimodal, depending on which mixture the planet falls into. The predictions for two similar size planets can also differ, if their periods are dissimilar.

We illustrate the latter effect in Figure 3.8, where we show mass predictions using our models for two different planets, 55 Cnc e and Kepler-60d. These two planets have similar radii (1.91 vs 1.99 $R_\oplus$, respectively), but dissimilar periods (0.74 vs 11.9 days). For our models with envelope mass loss (models 2-4), predicting masses for these two planets leads to two substantially different predictions. 55 Cnc e is on a very short orbit, and thus it is likely that the planet has lost its envelope due to hydrodynamic mass loss, and should

have a rocky composition (in the context of our model since we are not including ices). The mass predictions for this planet using the models that take into account envelope mass loss consequently predict a peak mass of around $8M_{\oplus}$, in line with the RV observations, with a sharp cutoff below $4M_{\oplus}$. On the other hand, Kepler 60-d is on a longer orbit and may have retained a gaseous envelope. The mass predictions for this planet using the envelope mass loss models thus allow for lower mass predictions and peak at around $5M_{\oplus}$, again in line with the TTV mass measurement.

55 Cnc e and Kepler-60d are similar sizes, but their disparate periods allow for different mass predictions when envelope mass loss is taken into account. By contrast, estimates using the probabilistic broken powerlaw mass-radius relation of CK17 or our model 1, which don't take into account envelope mass loss or the existence of multiple exoplanet populations, predict nearly the same mass for both planets. Note that we are not assessing the generalized accuracy of our mass or radius predictions. Properly testing this generalized accuracy would require not just two test cases, but an independent sample of planets with mass and radius measurements that were not included in fitting the model. Our example in Figure 3.8 serves to illustrate one of the consequences of using our models for predicting planet masses.

3.7 Discussion

3.7.1 Mixture Labels

While we have given each mixture component physically intuitive labels, i.e. "gaseous", "evaporated cores" and "intrinsically rocky", we note that these labels do not necessarily reflect the formation history for each individual planet. These mixture components are an abstraction that we have used to approximate the various physical populations of planets. For instance, given the large amount of overlap between the evaporated core and intrinsically rocky populations, some planets that fall under one label could be the other, and the

intrinsically rocky population could be serving to give more flexibility to the period and mass distributions of the evaporated cores. As a further example, since the mass-radius relation is described by a normal distribution at a given mass, the low radius tail of gaseous planets at low masses would have compositions consistent with rocky, and would not be considered gaseous planets. These high density "gaseous" planets fall into the estimate for $\eta_{\oplus}$, as described in section 3.6.2.

In order to visualize the distribution of these labels across our planet sample, we present a ternary plot in Figure 3.9 which shows the membership probabilities (gaseous vs evaporated core vs intrinsically rocky), as well as incident flux, for every planet in our sample. We find that most planets in the sample fall along two axes: the bottom axis, which is split between intrinsically rocky and gaseous planets, and the left axis, which is split between evaporated core and intrinsically rocky planets. There is also a substantial number of planets with low (less than 10%) gaseous probability that fall somewhere between the evaporated core and intrinsically rocky axis. Notably, there is a significant lack of planets with roughly even probabilities between the three mixtures, and a significant lack of planets along the right axis, with close to zero probability of being intrinsically rocky and split between evaporated core and gaseous. There are also no planets we can definitively say are intrinsically rocky and not evaporated cores, based on our model. We take this as evidence that the intrinsically rocky population is adding flexibility to the mass and period distributions of the evaporated cores, and cannot necessarily prove the existence of this separate population based on our data sample.

As a further test of the validity of including this intrinsically rocky population, we tested two additional models, as modifications of our four models presented in this work. Our first alternate model adds another component to the intrinsically gaseous population in model 2, with a separate period and mass distribution. This can be seen as an alternative to model 3, where we add flexibility to the intrinsically gaseous population rather than

creating an intrinsically rocky population, so we label it "model 3b". The second modified model adds a separate period distribution to the second component of the mass distribution for the intrinsically gaseous population in model 4, so we label it "model 4b". The aim of these modified models is to test whether the additional flexibility to the intrinsically gaseous population eliminates the need for the intrinsically rocky population. We find this to not be the case. Using the 10-fold cross-validation in Section 3.6.3, we find that model 3 is significantly preferred over model 3b ($> 3\sigma$), indicating that the flexibility that the intrinsically rocky population offers is better supported than adding additional flexibility to the intrinsically gaseous population. While we find that model 4b is somewhat preferred over model 4 ($< 3\sigma$), we note that the fraction of intrinsically rocky planets $Q_{\rm fr}$ increases rather than decrease or let alone become consistent with zero. While we cannot definitively prove the existence of intrinsically rocky planets, we have shown that their additions to the mixture model in the paper are justified.

3.7.2 Caveats

While this work represents a major step forward in determining the occurrence rates of planets across a wide range of periods, masses and radii, and the interconnectedness of these properties, there exist several limiting factors that need to be taken into account when considering the results presented herein. We describe three broad caveats here, and continue the discussion in our mention of future work in Section 3.7.4.

High dimensional distribution

Traditional occurrence rates using *Kepler* have naturally limited their scope to the radius-period plane, as these are the two quantities directly measured with the transit technique. Here, we have included masses of *Kepler* planets measured with RV observations in order to expand this 2D distribution into a 3D distribution of mass, radius and period. In doing so,

we have created models with up to 25 parameters and included three populations of planets to represent this complex 3D distribution. Despite this, the distribution presented here is still a projection of an even higher dimensional distribution.

Mass, radius and period are the three main observables and some of the most fundamental quantities for exoplanets, but they just represent three of many possible factors that affect which planets are common, and how common those planets are. Together, planet radius and mass can be used as a proxy for composition, but it remains a proxy, and ultimately planet occurrence as a function of planet composition remains an important question.

While here we have treated planets as independent, many *Kepler* planets are found in multiple-planet systems, with these planets exhibiting uniformity in mass and radius, as well as regularly spaced orbits (Millholland et al., 2017; Weiss et al., 2018).

There are also a variety of factors associated with the host star that can affect planet occurrence. Host star metallicity has shown to have a significant effect not only on gas giant occurrence rates, but also sub-Neptune and terrestrial planets, with higher metallicities associated with higher planet occurrence (Wang & Fischer, 2015; Mulders et al., 2016; Dong et al., 2018; Petigura et al., 2018). The presence of stellar companions has been linked to lower occurrence rates (Wang et al., 2014; Kraus et al., 2016). Stellar obliquity (alignment of planet orbits with respect to the spin of the host star)) is correlated with both planet size and period (Muñoz & Perets, 2018). While we only consider planets around G and K dwarfs in this work, planets around M dwarfs have been shown to have a different radius and mass distribution from those around G and K dwarfs, with a significant lack of gas giants (Bonfils et al., 2013; Dressing & Charbonneau, 2015), and possible period-dependent differences (Mulders et al., 2015a; Neil & Rogers, 2018).

Most planets discovered via transits and RV are relatively close to the Sun (within ~ 1500 parsecs), but the galactic distribution of planets could be substantially different, which will be measured with the future *WFIRST* survey (Montet et al., 2017). These factors, among

others, could all affect planet occurrence, and the true distribution of planets has many dimensions.

Data Set

Kepler's discovery and characterization of several thousand exoplanets, and in particular the high-quality sample of 1130 planets used in this paper, has enabled the statistical study of the occurrence rate of planets in the 3D space of radius, period and mass. However, the statistical analysis will always be limited by the quality and quantity of the available data, and this paper is no exception. *Kepler*'s detection efficiency drops off towards lower radii and longer periods, and these regions of high incompleteness are of high interest and contain many mysteries. The differences between the two separate populations of rocky planets modeled in this paper would manifest in these regions of high incompleteness where we do not have many planets in the sample. Extending the sample to longer orbital periods would allow us to more easily study the interplay between the period and radius/mass distributions, as the differences would be more pronounced.

The availability of mass measurements is an additional major limitation. Our main sample contains only 53 planets with RV mass measurements, yet we constrain the mass-radius relation for gaseous planets based on these 53 planets. Furthermore, there is a significant lack of RV measurements for small, potentially rocky planets. As a result, our rocky planet mass-radius relationship is fixed and based on theoretical curves: our planet sample would not be able to constrain this well at all. In order to model a population of gaseous planets together with a population of rocky planets, we cannot solely rely on the data, but must include physics in the form of hydrodynamic mass loss. With more RV mass measurements of planets in the size regime where there is some overlap between rocky and gaseous compositions (broadly $1.4 < \frac{R}{R_\oplus} < 2.0$), we would be able to more strongly constrain the difference in occurrence rates between the two populations. With the ongoing *TESS* search for tran-

siting planets and subsequent RV follow-up, we can expect this dataset to continue to grow in the future.

When compiling our dataset we made the choice of using RV mass measurements only, thereby excluding a comparable number of TTV masses. Planets with masses measured with RV vs TTV have been shown to have systematic differences in their densities. This is likely due to observational biases, with the TTV technique having higher sensitivity than RV to planets with lower masses and longer periods (Jontof-Hutter et al., 2014; Steffen, 2016; Mills & Mazeh, 2017). We exclude TTV planets in order to keep the dataset as homogeneous as possible, although we acknowledge that modeling the RV and TTV masses together is a critical next step. Given the differences in densities, we expect one major effect of including TTV masses would be to increase the scatter in the mass-radius relation. Further, since the TTV technique is more sensitive to longer period planets, including them may require extending our model to longer orbital periods than our analysis here. This may also allow further exploration of the connection between the mass-radius plane and orbital period, as mass-radius occurrence may depend on orbital period beyond the mixture models and envelope mass loss utilized in this paper. We leave more detailed analysis of the TTV sample to future work.

In our analysis, we have taken careful steps to fully account for the incompleteness of the *Kepler* pipeline and subsequent dispositioning of planet candidates by the Robovetter software. While this should give us unbiased results as a function of radius and period, the incompleteness of RV followup of transiting planets has not been accounted for in our analysis and could bias our results. The RV followup process is less systematic and thus more difficult to properly account for. Burt et al. (2018) simulated several strategies for RV follow-up of *TESS* planets, and found that the practice of only reporting masses measured with a high statistical significance lead to biased mass-radius relations. This bias results in a mass-radius relation pushed towards higher densities below $10 M_{\oplus}$, and has been found to

affect previous empirical fits of the mass-radius relation. In our work, this could result in an underestimation of the number of low-mass gaseous planets, and a mass-radius relation that is shifted towards higher densities. In order to fully account for this RV selection effect, all non-detections regardless of statistical significance need to be reported (e.g. as done in Marcy et al. (2014)) and subsequently included in the data sample for papers doing statistical analysis.

Model Dependence

In this work we presented four models, of increasing complexity. As shown by our occurrence rate calculations in Section 3.6.2 and mass predictions in Section 3.6.4, our conclusions are highly dependent on the choice of model. For a planet of radius $1.9R_\oplus$ and period 0.74 days, using model 1, we predict a mass distribution that peaks at $2M_\oplus$, whereas we predict a mass distribution that peaks at $6 - 8M_\oplus$ for the same planet using models 2-4. While we have put a lot of consideration into our choice of parametrizations for the models presented here, constructing them using both physical reasoning and past precedence, there remain many areas of potential improvement. For instance, in our model there is no constraint restricting the density of gaseous planets to be lower than that of a rocky planet of the same mass. As a result, high-density "gaseous" planets are included in the calculation of $\eta_\oplus$. Changing the parametrization of one distribution (i.e. the period distribution) can have rippling effects throughout the entire 3D distribution. It is of the utmost importance to consider the results presented here in the context of the model that was used to generate them.

Just as the choice of parametrization significantly affects the results, the choice of prior can often have a large effect. For instance, adopting a wide prior on α, the parameter that governs the scaling for the envelope mass loss, can cause α to be retrieved several orders of magnitude higher than our default assumption of $\alpha = 1$, which is based on the default values for mass loss efficiency, stellar XUV flux, and age of the star assumed in Lopez & Fortney

(2013). This would lead to planets that are more easily able to retain their envelopes, greatly reducing the number of evaporated cores. However, such high or low values of α may be unphysical ($\alpha < 0.1$ or $\alpha > 10$) and would require the XUV flux of stars to strongly deviate from what we assumed. With this case and with other parameters, we have tried to strike a balance between restricting the prior to reasonable ranges of the parameter (based on physical models and past measurements), and allowing the model flexibility. Uncertainty ranges quoted herein represent credible intervals inferred in the context of a specified model. They do not include uncertainties in the choice of model parametrization or priors.

While our results are model dependent, our application of mixture models has shown to be justified by the available data, and we have shown that doing so requires the addition of the physics of envelope mass loss. We provide the results from multiple models, allowing us to assess how much this model choice impacts our conclusions. Finally, model dependence is not limited to our work and must be considered for any analysis of the mass, period, and/or radius distributions of exoplanets.

3.7.3 How to use these results

The main purpose of this paper is to provide a framework for constraining higher dimensional planet occurrence rate distributions and including mixture models in these distributions. While we present occurrence rate measurements, mass/radius predictions, and measurements of the amount of gaseous vs evaporated core vs intrinsically rocky planets, our goal is not to put forward our results as definitive. All results must be considered in context of the model. Instead of "we predict a mass of x", it is instead more prudent to say "we predict a mass of x, if we use this model that includes hydrodynamic envelope mass loss". Broadly, when using our population inferences, we recommend using model 4, as this model includes the three populations of planets that are shown to be justified by the data in Section 3.6.3, and provides the best fit to the observed distribution of Kepler radii and periods. By providing

fits to multiple models, we offer the exoplanet community the chance to test for robustness against the choice of model.

3.7.4 Composition Distribution

Previous efforts to empirically constrain the radius-period distribution and mass-radius relationship of planets have relied on simple functions and distributions such as power-laws, normals and log-normals. While the physics of planet formation sometimes entered the scene, as in Wolfgang et al. (2016) where planets are constrained to have a density lower than that of a pure iron core, these scenarios are few and far between and there remains a gulf between observations and theory. Here, we have included the physics of envelope mass loss as a necessary step to model rocky planets alongside gaseous planets and disfavor unphysical small planets with gaseous envelopes on short orbits. This can make the model somewhat incongruous and unsatisfying: for some aspects we rely on physics, and for some we trust the data to constrain the distributions for us.

One possible step towards a more physically grounded model would be to constrain a composition distribution. Rather than working with the observables mass and radius, we can choose mass and some proxy for composition, such as envelope mass fraction. Using input from theoretical models of planet interior structure, we can go from mass and envelope mass fraction to an output radius. Thus, we should be able to recreate the double broken power-law mass-radius relationship described here, but as a natural outcome of planet interior structure. As we obtain more and more planet mass measurements, and reveal areas in the planet mass-radius space where composition visibly differs, moving towards this composition distribution will become easier and more necessary. With such a distribution, we can close the gap between theory and observations in our modelling.

3.8 Conclusion

We have used hierarchical Bayesian modeling to constrain the joint mass-radius-period distribution of *Kepler* transiting exoplanets. We construct four models of increasing complexity, the latter three using mixture models and XUV-driven hydrodynamic mass loss to independently model populations of planets with gaseous envelopes, evaporated cores, and intrinsically rocky planets. We find through model selection techniques that these additional complexities are warranted, and better match the observed radius and period distributions of *Kepler*. We calculate occurrence rates in different regions of parameter space using these models, including extrapolation to $\eta_\oplus$, finding broad agreement with previous results but also finding the estimates of $\eta_\oplus$ to be significantly model dependent. We use these models to predict masses of *Kepler* planets, showing the benefit of employing mixture models by comparing the predictions for two planets of similar radius but dissimilar periods. Although our results are highly dependent on the parametrization and choice of priors for our models, we provide a framework for the modelling of higher-dimensional distributions of exoplanets together with the inclusion of mixture models. We find the use mixture models to be a valuable tool for capturing the complexity inherent to planet formation and evolution.

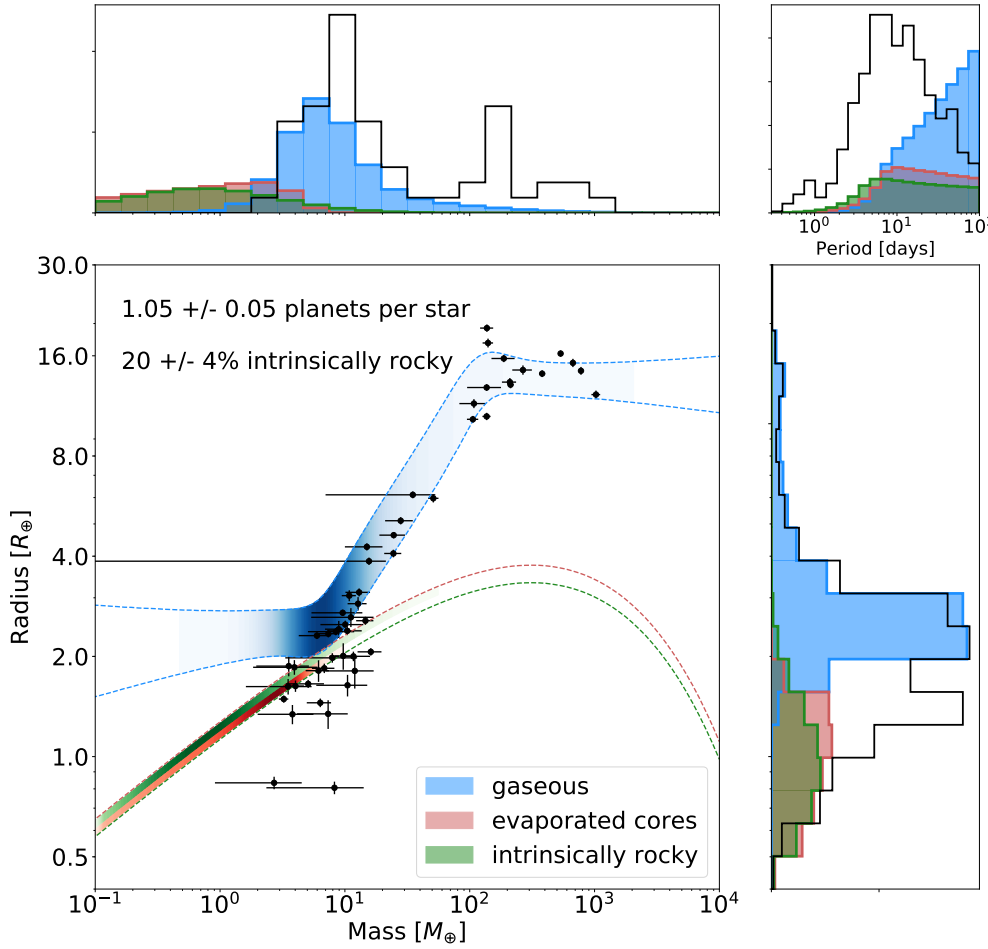

Figure 3.3: The joint Mass-Radius-Period distribution using our model 4 which has three mixture populations of planets. Our results for the gaseous population are shown in blue, evaporated cores in red, and intrinsically rocky in green. In the bottom left panel, we show the mass-radius relation. The dotted lines indicate the central 68% of planets drawn at a given mass, and the shading indicates the occurrence at that mass relative to the overall occurrence of that mixture. The evaporated core and intrinsically rocky mixtures are stacked on top of each other in the mass-radius plane for clarity; both follow the same mass-radius relation and the central 68% of planets in these two mixtures fall between the red and green dotted lines. The *Kepler* planets in our sample with RV mass measurements are shown in black with their respective measurement errors. At the top we show the underlying mass distributions of the three populations, as well as the distribution of mass measurements in our sample in black. Similarly, we show the underlying radius distributions on the right with the radius distribution of the *Kepler* sample in black, and we show the period distributions in the top right.

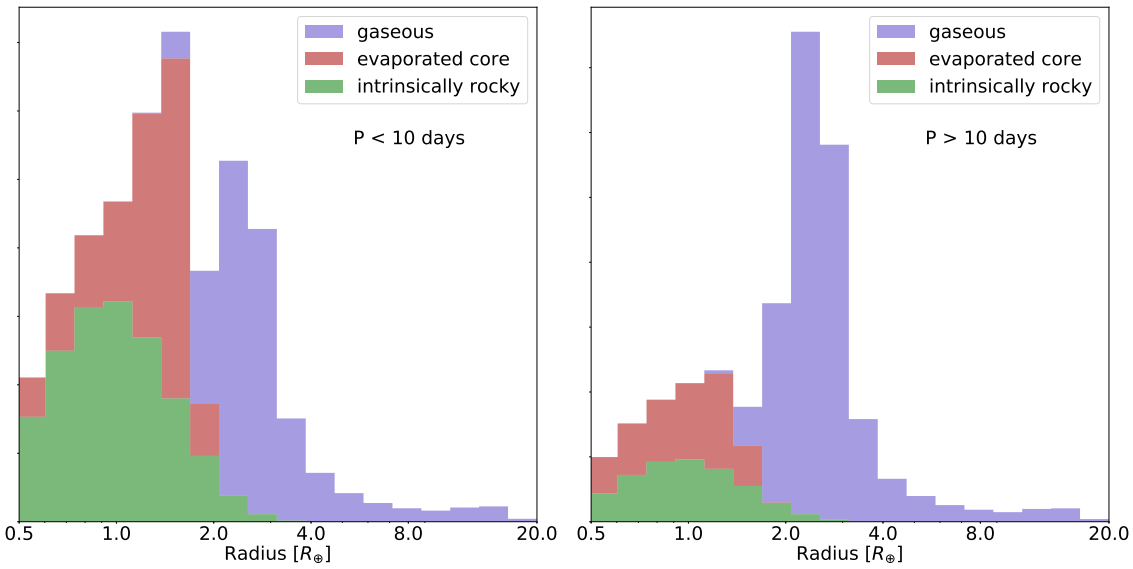

Figure 3.4: The underlying radius distribution of planets from our model 4, broken down into the three populations of planets: gaseous shown in blue, evaporated cores shown in red, and intrinsically rocky shown in green. On the left we show the radius distribution of planets with short periods ($P < 10$ days) and on the right we show the radius distribution of planet with long periods ($P > 10$ days). The three populations are stacked on top of one another, such that the outline shows the overall intrinsic radius distribution. The two radius distributions are normalized separately rather than to each other (there are more planets at $P > 10$ days than $P < 10$ days, which is not shown by these distributions)

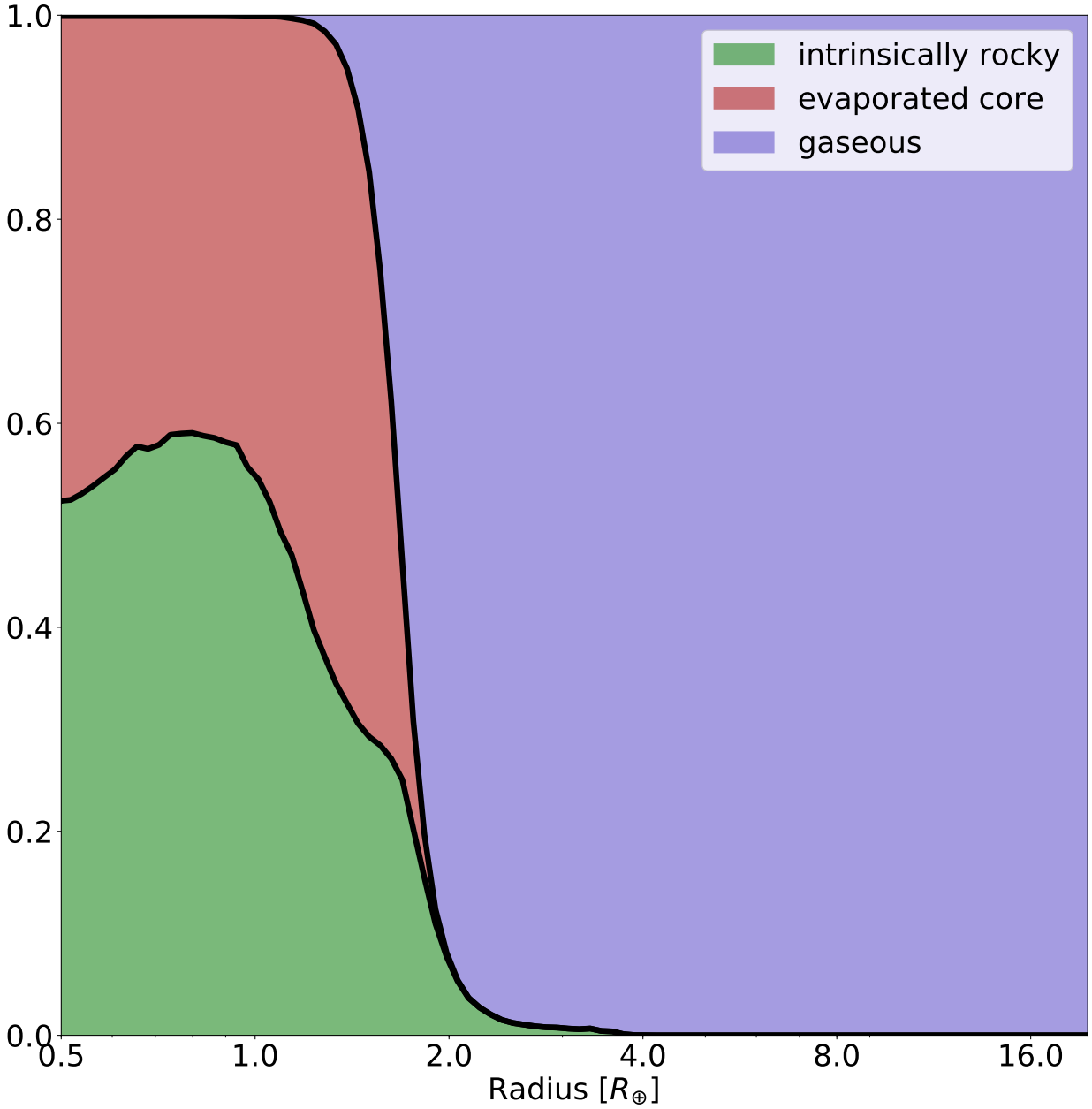

Figure 3.5: The fraction of planets belonging to each mixture component at a given radius, using our model 4 with three populations of planets: gaseous shown in blue, evaporated cores shown in red, and intrinsically rocky shown in green.

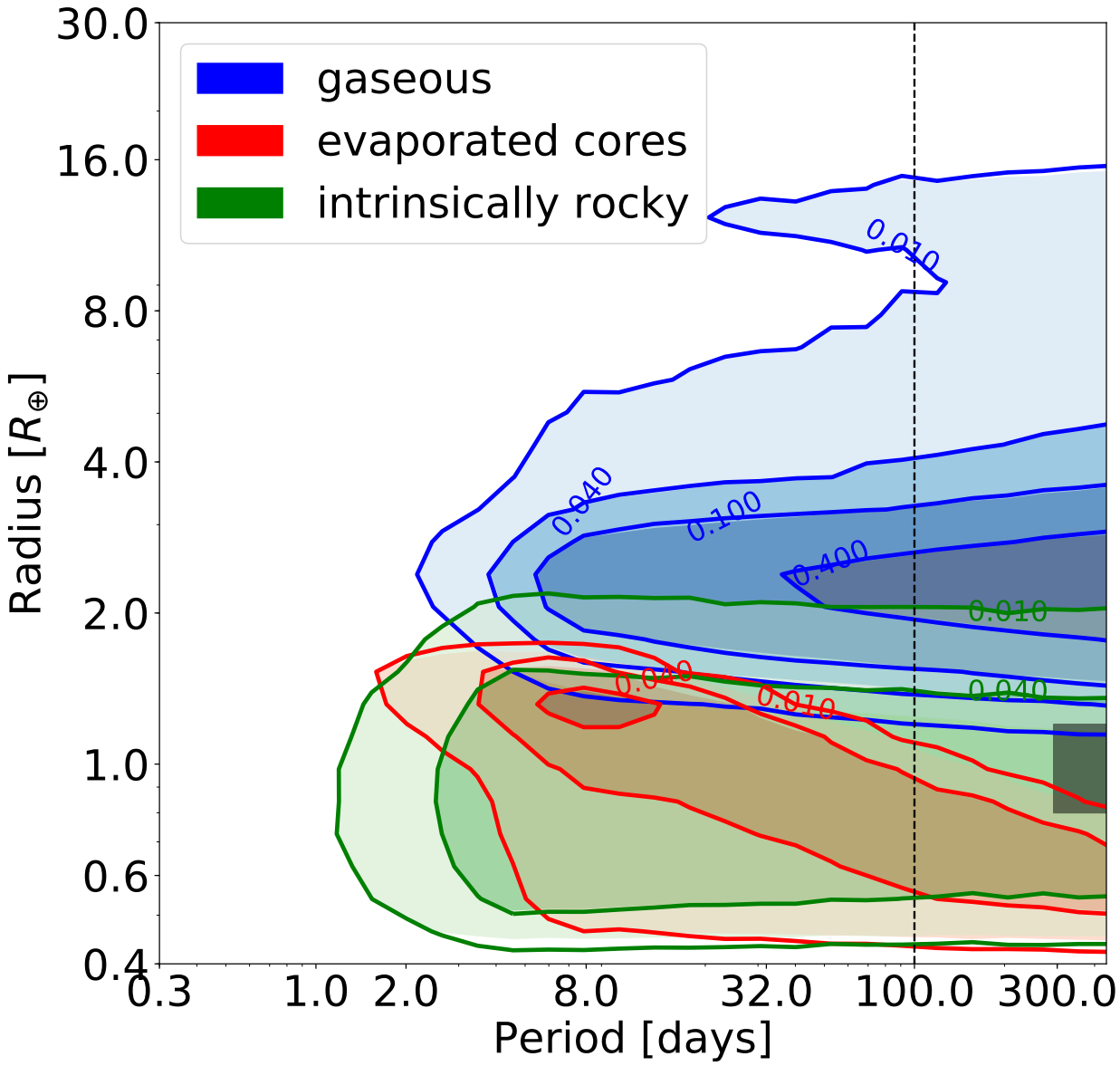

Figure 3.6: Contours of occurrence rates in radius-period space for the gaseous, evaporated core, and intrinsically rocky populations from model 4. The vertical line at 100 days shows the boundary to which we originally fitted our models, and the grey box shows the region in which we calculate $\eta_\oplus$.

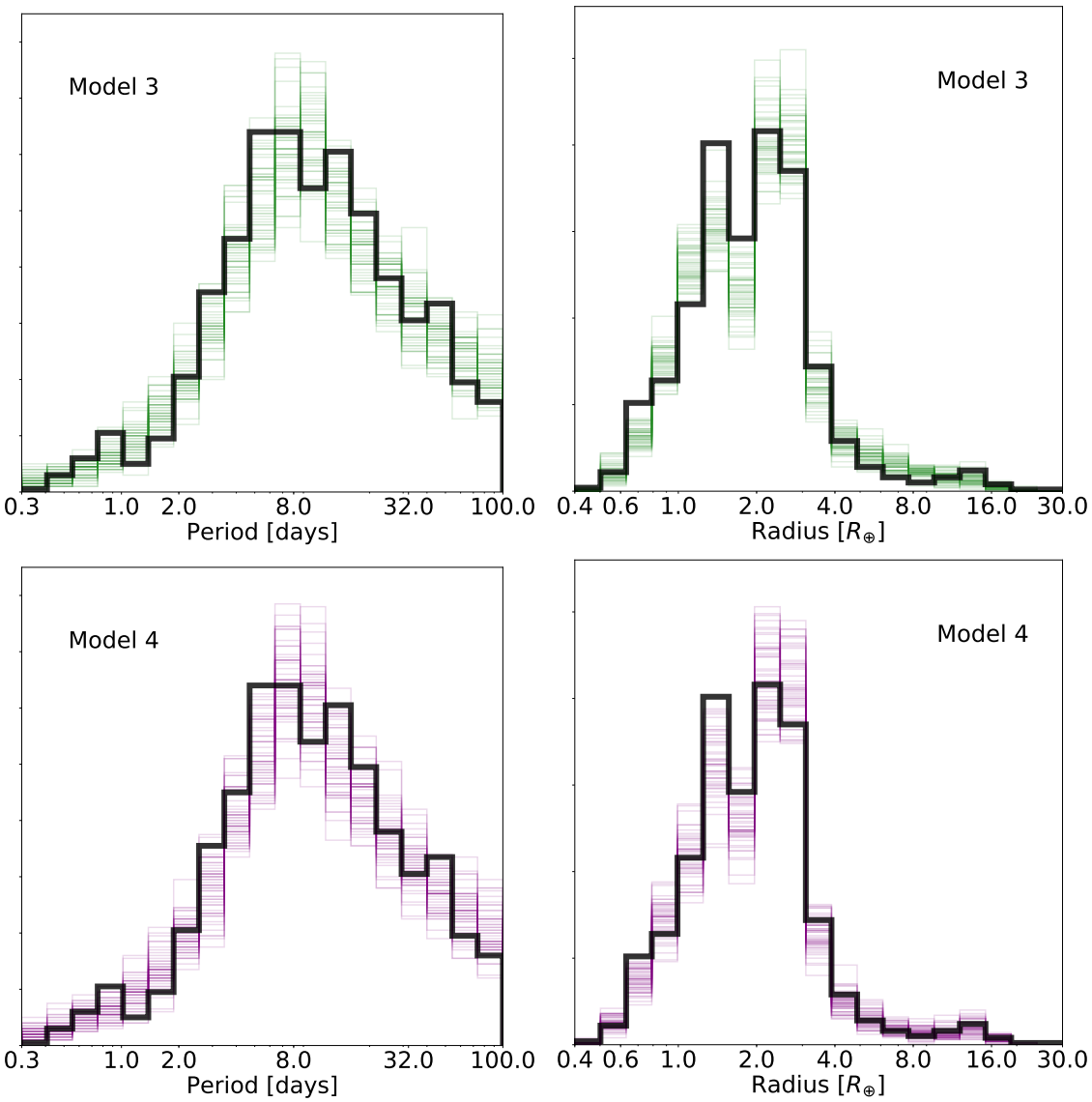

Figure 3.7: The observed period (left) and radius (right) distributions of the *Kepler* sample, shown in black, compared to the simulated observed distributions using our model 3 with three planet populations (top) shown in green, and model 4 with an extra gaseous log-normal mass component (bottom), shown in purple. We draw planets from our joint MRP distribution and then run them through the *Kepler* detection efficiency to create simulated observed catalogs of planets. We repeat this 50 times using draws from our model posteriors to show the spread in the simulated distributions.

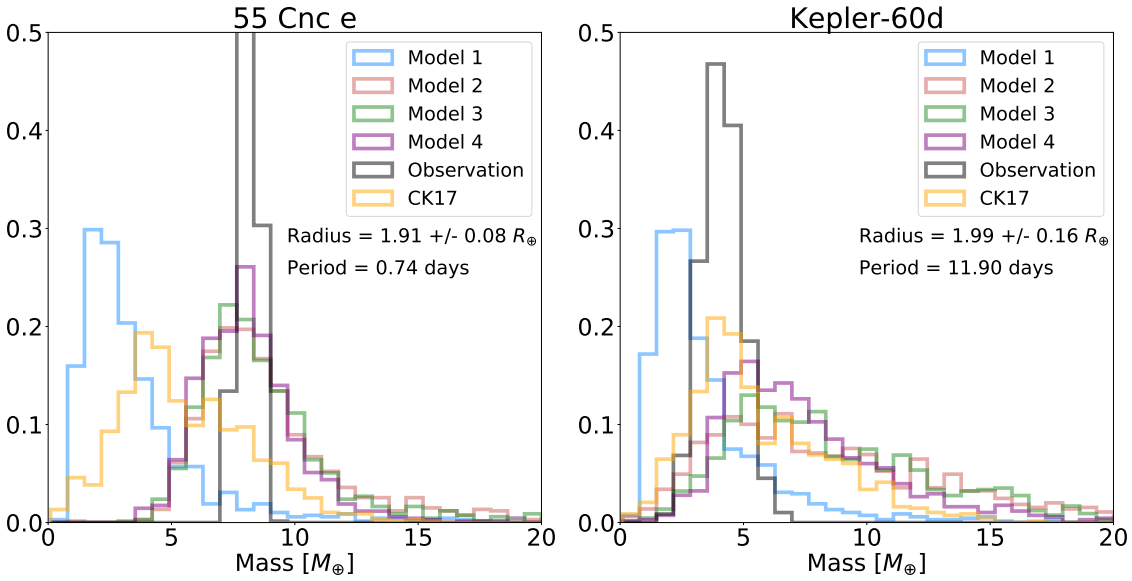

Figure 3.8: Mass prediction using our posteriors from each of our four models compared to the mass prediction from Chen & Kipping (2017) as well as the actual mass measurement, for two different planets with similar radii but dissimilar compositions. Both 55 Cnc-e and Kepler-60d have radii close to $2R_\oplus$, but 55 Cnc-e has a mass of $8.1 \pm 0.3 M_\oplus$ and a period of 0.74 days, and Kepler 60-d has a mass of $4.2 \pm 0.8 M_\oplus$ and a period of 11.9 days.

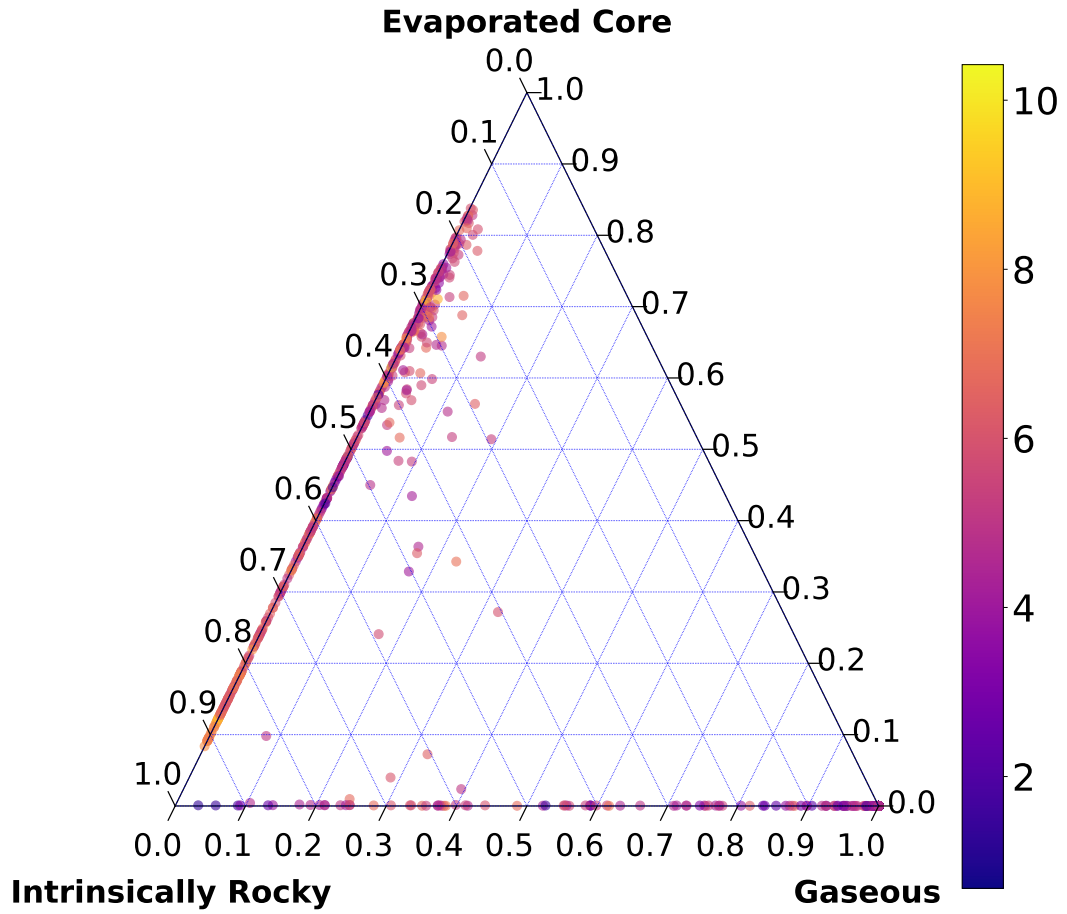

Figure 3.9: Ternary plot showing the distribution of the retrieved probabilities of the three mixture components for the 1130 planets in our sample from model 4. Each point represents a planet and are colored by incident flux (scaled to the insolation flux from the Sun at Earth's orbit). The closer a planet is to a labeled corner, the higher the probability of that planet belonging to that mixture. The axis for each mixture component is located in the clockwise direction from the mixture component's respective corner.

CHAPTER 4

EVALUATING THE EVIDENCE FOR WATER WORLD POPULATIONS USING MIXTURE MODELS

The default explanation for the Fulton radius gap has been a bifurcation in the exoplanet population between planets that have retained their envelopes and planets that have lost their envelopes. The planets that retain their envelopes are larger and less dense, sitting above the radius gap, whereas planets that lose their envelopes are smaller and more dense, sitting below the radius gap. The source of this envelope loss can be either photoevaporation or a core-powered mechanism, and both have been shown to reproduce the radius gap (e.g., Van Eylen et al., 2018; Owen & Murray-Clay, 2018; Ginzburg et al., 2018; Gupta & Schlichting, 2020). In Chapter 3 we explored photoevaporation as the source of this radius gap by incorporating multiple subpopulations of planets within a joint mass-radius-period distribution mixture model. We found that models that included a subpopulation of gaseous planets that retained their envelopes and a subpopulation of evaporated rocky cores, alongside a separate subpopulation of intrinsically rocky planets, were preferred over models that did not include photoevaporation.

Not all explanations for the Fulton radius gap rely on envelope mass loss. Zeng et al. (2019) reproduced the radius gap by modeling a subpopulation of icy composition planets, or "water worlds", alongside a subpopulation of rocky composition planets. Planets with icy cores are less dense, and therefore have a larger radius at the same mass compared to planets with rocky cores. Thus, the radius gap in this case is not due to envelope mass loss but solely through compositional diversity of the planet core.

In this chapter, we test the claims of Zeng et al. (2019) by extending the mass-radius-period mixture models developed in Chapter 3. Our models in Chapter 3 only included planet subpopulations with rocky cores: gaseous planets with rocky cores, evaporated rocky cores, and intrinsically rocky planets. In this chapter, we include three analogous subpop-

ulations with icy cores, expanding our total number of planet subpopulations to six. We then construct a total of ten separate models that include these six subpopulations in different combinations, ranging in complexity from a model with a single subpopulation to a model with all six subpopulations. We compare these models qualitatively and quantitatively through their resulting radius-period, mass-radius, and radius distributions. Additionally, we perform model selection on all six models and analyze the results.

Our cross-validation results are unable to distinguish between models of higher complexity, suggesting that both models that include photoevaporation and models that include compositional diversity in the planet core but no photoevaporation are supported by the data. However, simulations we perform to validate our model selection results show that cross-validation when applied to mixture models is still prone to overfitting and tends to prefer more complex models. This, combined with population-level degeneracies between water worlds and planets with gaseous envelopes, makes it particularly difficult to distinguish between these models. The fraction of planets with icy compositions is found to be highly model dependent, ranging from 0% to 50%.

This chapter was completed in collaboration with Jessica Liston and Leslie Rogers and has been submitted to the Astrophysical Journal, where it is currently under review at the time of writing. Jessica Liston assisted with the analysis and writing of this chapter, and produced Figures 4.1 and 4.4. This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program. This work was also completed in part with resources provided by the University of Chicago's Research Computing Center. JL thanks the University of Chicago Physics/MRSEC REU Program, and the National Science Foundation. LAR gratefully acknowledges support from NASA Habitable Worlds Research Program grant 80NSSC19K0314, from NSF FY2016 AAG Solicitation 12-589 award number 1615089, and the Research Corporation for Science Ad-

vancement through a Cottrell Scholar Award.

4.1 Abstract

Water worlds have been hypothesized as an alternative to photo-evaporation in order to explain the gap in the radius distribution of *Kepler* exoplanets. We explore water worlds within the framework of a joint mass-radius-period distribution of planets fit to a sample of transiting *Kepler* exoplanets, a subset of which have radial velocity mass measurements. We employ hierarchical Bayesian modeling to create a range of ten mixture models that include multiple compositional subpopulations of exoplanets. We model these subpopulations - including planets with gaseous envelopes, evaporated rocky cores, evaporated icy cores, intrinsically rocky planets, and intrinsically icy planets - in different combinations in order to assess which combinations are most favored by the data. Using cross-validation, we evaluate the support for models that include planets with icy compositions compared to the support for models that do not, finding broad support for both. We find significant population-level degeneracies between subpopulations of water worlds and planets with primordial envelopes. Among models that include one or more icy-core subpopulations, we find a wide range for the fraction of planets with icy compositions, with a rough upper limit of 50%. Improved datasets or alternative modeling approaches may better be able to distinguish between these subpopulations of planets.

4.2 Introduction

The discovery in the *Kepler* survey of a gap in the radius distribution of planets between $1.5 - 2.0 R_\oplus$ (Fulton et al., 2017) has sparked a flurry of investigations. This bimodal distribution, a discovery enabled by recent improvements to the characterization of the *Kepler* detection efficiency (Christiansen et al., 2015, 2020), as well as an improved stellar sample

with higher precision masses and radii (Petigura et al., 2017), suggests the existence of a subpopulation of planets below $1.5R_\oplus$ alongside a separate subpopulation of planets between $2-4R_\oplus$. Currently, the most popular explanation for this radius gap is photoevaporation, with the subpopulation below $1.5R_\oplus$ consisting of planets whose gaseous envelopes have been evaporated by irradiation from their host star, and the subpopulation above $2.0R_\oplus$ consisting of planets that have retained their gaseous envelopes (Van Eylen et al., 2018; Fulton & Petigura, 2018; Owen & Murray-Clay, 2018). Alternatively, this envelope mass loss may be explained through the mechanism of core-powered mass loss, which has been shown to equally fit the *Kepler* dataset (Ginzburg et al., 2018; Martinez et al., 2019; Gupta & Schlichting, 2020; Rogers et al., 2021).

Aside from explanations that rely on mass-loss mechanisms, there are also those that invoke compositional variation as the sole force responsible for the radius gap. Zeng et al. (2019) use a growth model and Monte Carlo simulations to assert that the radius gap can be recreated by the inclusion of a subpopulation of water worlds, modeled by the authors as planets with compositions ranging from $\frac{1}{2}$ to $\frac{2}{3}$ water by mass, and radii ranging from $2-4R_\oplus$, alongside a subpopulation of rocky-core planets at lower radius. Thus, they argue, mass-loss from photoevaporation or core-powered mechanisms are not necessary to explain the *Kepler* radius gap.

These theories, however, are not mutually exclusive, as it has also been suggested that water worlds may form through photoevaporation of mini-Neptunes that form beyond the snow line and migrate inward (Luger et al., 2015). Migration plays a large role in the water content of close-in small planets (Kuchner, 2003; Léger et al., 2004; Raymond et al., 2008; Raymond et al., 2018). Water-rich planetesimals that form at or beyond the snow line and migrate inward during the early gas phase of the disk have been shown to have lower final water content than planetesimals that migrate later during formation (Bitsch, Bertram et al., 2019). The presence of gas giants in the system can also impact water delivery, with gas giants

outside the snow line blocking water delivery to planets interior to the snow line (Bitsch, Bertram et al., 2021). Long-term retention of water is also an issue, although studies have shown that habitable-zone exoplanets can retain their surface water on Gyr timescales (Kite & Ford, 2018), and short-period rocky exoplanets can retain water-dominated atmospheres (Kite & Schaefer, 2021). An inherent difficulty in distinguishing between theories of small planet formation is the degeneracy in their composition, where a wide range of compositions can account for the same density (e.g., Adams et al., 2008; Rogers & Seager, 2010a). For example, water worlds that are sufficiently irradiated can have supercritical hydrospheres, inflating their radii to the point that they are indistinguishable by mass and radius alone from mini-Neptunes with H/He gaseous envelopes (e.g., Rogers & Seager, 2010b; Mousis et al., 2020).

Neil & Rogers (2020) (hereafter NR20) provides a framework for accounting for this degeneracy in composition and potential mass-loss of small planets when fitting the joint 3D mass-radius-period distribution by employing mixture models. They fit several mixture models, containing a range of 1 to 3 subpopulations, to the California-*Kepler* survey cross-matched with *Gaia* data, using radial velocity (RV) mass measurements of planets where available. Their models considered three separate planet subpopulations: planets with substantial gaseous envelopes, evaporated rocky cores (resulting from photoevaporation), and intrinsically rocky planets that formed without any gas. Through model selection, they found that models with all three subpopulations were preferred over simpler models with only a subset of these subpopulations. However, their models only considered rocky-core compositions, neglecting the potential for cores with substantial fractions of water-ice.

In order to investigate the mystery behind this radius gap, we incorporate various subpopulations of water worlds into the existing framework for characterizing the mass-radius-period distribution of planets established in NR20. We build a collection of 10 models that include planet subpopulations of various compositions, as well as formation history, in dif-

ferent combinations to assess the evidence for photoevaporation compared to water worlds in the current *Kepler* dataset. We then quantify the evidence for these competing models using model selection.

This paper is organized as follows. In section 4.3 we outline seven mass-radius-period distribution mixture models which incorporate water worlds into pre-existing models in different ways. In section 4.4, we present the model fits, outline major differences between models, present degeneracies between subpopulations, and put constraints on the fraction of icy composition planets. We compare our models using k-fold cross-validation in Section 4.5 and perform simulations to test the accuracy of our model selection results. We discuss these results, caveats to our methodology, and future extensions of this work in Section 4.6, and conclude in Section 4.7.

4.3 Methods

We build upon the hierarchical Bayesian modeling approach in NR20 in order to include planet subpopulations with icy-core compositions into their joint mass-radius-period distribution mixture models. We first review the breakdown of the occurrence rate density as presented in NR20, along with the parametrizations for each distribution and the envelope mass loss prescription in Section 4.3.1. We then introduce planet subpopulations with icy-core compositions in Section 4.3.2, and use these icy-core subpopulations in combination with the rocky-core composition subpopulations to formulate the ten models presented in Section 4.3.3. We highlight Model Z in Section 4.3.4 as a model that includes icy composition planets without including photoevaporation, to compare to Zeng et al. (2019). Finally, we review the planet catalog used in this work in Section 4.3.5, and review the MCMC fitting process and highlight differences in fitting methodology from NR20 in Section 4.3.6.

4.3.1 Equations and Parametrizations

The fundamental quantity that we constrain is the planet occurrence rate density $\Gamma(P, M, R)$ as a function of period, mass, and radius:

$$\Gamma(P, M, R) = \frac{dN}{dP\ dM\ dR} \tag{4.1}$$

where this occurrence rate density is defined as the number of planets per star per interval in orbital period (P), planet mass (M), and radius (R). Unless otherwise stated, the orbital period P is expressed in units of days, and the planet radius R and planet mass M are expressed in Earth units ($R_\oplus$ and $M_\oplus$, respectively). We take our boundaries in mass-radius-period space to be $0.3 - 100$ days for the period, $0.1 - 10000 M_\oplus$ in mass, and $0.4 - 30 R_\oplus$ in radius.

We incorporate multiple compositional subpopulations of planets using a mixture model and break the multidimensional distribution into component parts.

$$\begin{aligned}\Gamma(P, M, R) = &\Gamma_0 \sum_{q=0}^{N-1} \sum_{v=0}^{1} \mathrm{p}(R|M, v, q)\, \mathrm{p}(v|M, P, q) \\ &\cdot \mathrm{p}(P|q)\, \mathrm{p}(M|q)\, p(q)\end{aligned} \tag{4.2}$$

Above, Γ_0 is an overall normalization term that gives the number of planets per star within our bounds in mass-radius-period space defined above. The N mixture components are indexed by q ($0 \leq q < N$), and each represent a fraction $\mathrm{p}(q)$ of the total planet population within the mass-radius-period boundaries. The index v ($0 \leq v \leq 1$) indicates whether the subpopulation retains its gaseous envelope ($v = 0$) or loses its envelope ($v = 1$), with $\mathrm{p}(v)$ indicating the probability of either scenario, which depends on mass and period. For the case of mixtures that formed without gaseous envelopes, the summation over v reduces to $\mathrm{p}(R|M, q) = \sum_{v=0}^{1} \mathrm{p}(R|M, v, q)\, \mathrm{p}(v|M, P, q)$. For each mixture, the period distribution

$\mathrm{p}(P)$ and mass distribution $\mathrm{p}(M)$ of planets in the mixture are modeled as independent. The distribution of planet radii conditioned on mass $\mathrm{p}(R|M, v, q)$ is characterized by a mass-radius relation appropriate to the particular compositional subpopulation. All period, mass and radius distributions are normalized to 1 over their boundaries.

Across all models, and all mixtures within those models, we use a universal parametrization for the period distributions and a separate universal parametrization for the mass distributions. The orbital period distribution $\mathrm{p}(P)$ is characterized by a broken power-law with three parameters: the period break, P_{break}, and the slopes before and after the break, β_1 and β_2:

$$\begin{aligned} \mathrm{p}(P) &= AP^{\beta_1}, \ P < P_{\mathrm{break}} \\ \mathrm{p}(P) &= AP_{\mathrm{break}}^{\beta_1-\beta_2}P^{\beta_2}, \ P > P_{\mathrm{break}} \end{aligned} \tag{4.3}$$

where A is a normalization factor to ensure that the distribution is normalized to 1 over the range $0.3-100$ days. The mass distribution $\mathrm{p}(M)$ is parametrized by a truncated log-normal distribution with two parameters: the mean, μ_M, and the standard deviation, σ_M:

$$\mathrm{p}(M) = \mathrm{lnN}(M|\mu_M, \sigma_M) \ [0.1, 10000 \ M_\oplus] \tag{4.4}$$

where the limits in the brackets indicate the truncation bounds.

The radius (conditioned on mass) distribution $p(R|M)$ is characterized by a truncated normal distribution with a mean $\mu(M)$ (where μ depends on the mass M, not to be confused with the mass distribution parameter μ_M) given by a mass-radius relation that is dependent on the planet's composition, and a fractional intrinsic scatter σ, where the scatter is a fixed fraction of the mean mass-radius relation at any given mass:

$$\mathrm{p}(R|M) = \mathrm{N}(R|\mu(M), \sigma \cdot \mu(M)) \tag{4.5}$$

For planets with gaseous compositions, we use a double broken power-law mass-radius relation characterized by 9 parameters: $C, \gamma_0, \gamma_1, \gamma_2, \sigma_0, \sigma_1, \sigma_2, M_{\text{break},1}$, and $M_{\text{break},2}$. The two M_{break} parameters (with $M_{\text{break},1} \leq M_{\text{break},2}$) define the break points in mass splitting the mass-radius relationship into three regimes: low mass, intermediate mass, and high mass. C defines the radius scale, or the radius of the mean mass-radius relation at a mass of $1 M_{\oplus}$. The three γ parameters define the different slopes in the three regimes of the mass-radius relationship, and the three σ parameters define the fractional scatter in those regimes. This double broken power-law mass-radius relation is defined below:

$$\begin{aligned} \mu_0 &= C M^{\gamma_0} \\ \mu_1 &= C M_{\text{break},1}^{\gamma_0 - \gamma_1} M^{\gamma_1} \\ \mu_2 &= C M_{\text{break},1}^{\gamma_0 - \gamma_1} M_{\text{break},2}^{\gamma_1 - \gamma_2} M^{\gamma_2} \end{aligned} \tag{4.6}$$

with the subscripts 0, 1, 2 indicating the low mass range, intermediate mass range, and high mass range, respectively. The fractional scatters σ_0, σ_1, and σ_2 then apply to their respective μ with matching index. In practice, instead of abrupt transitions at the break points $M_{\text{break},1}$ and $M_{\text{break},2}$, we use a logistic function to smooth between the different power-laws (with a width of 0.2 in log-mass) and ensure continuity in the first derivatives of the radius-mass relations (which enables the use of Hamiltonian Monte Carlo to fit the models to the data, further described in Section 4.3.6).

Rather than fit a mass-radius relation to planets with rocky compositions, these planets instead are modeled as following a fixed mass-radius relation of a pure-silicate composition as calculated in Seager et al. (2007):

$$
\begin{aligned}
R &= R_1 \cdot 10^{k_1 + \frac{1}{3}\log_{10}(M/M_1) - k_2\ (M/M_1)^{k_3}} \\
R_1 &= 3.90; M_1 = 10.55 \\
k_1 &= -0.209594; k_2 = 0.0799; k_3 = 0.413
\end{aligned}
\tag{4.7}
$$

with a fixed 5% fractional scatter to account for the low amount of scatter exhibited by the current sample of low-mass planets with mass and radius measurements (Dressing et al., 2015; Zeng et al., 2015; Dai et al., 2019).

Icy composition planets, which are introduced in this work but not present in NR20, are discussed in Section 4.3.2.

For a subset of models, we include a mechanism for photoevaporation whereby a subpopulation of planets that formed with a gaseous envelope can follow either the gaseous or rocky mass-radius relations described above depending on whether or not the planet has lost its envelope. We model the probability that a planet retains its envelope, $\mathrm{p_{ret}}$, using a Bernoulli process given by the following equation:

$$
\begin{aligned}
\mathrm{p_{ret}} = \mathrm{p}(v = 0) &= \min\left(\alpha \frac{t_{\mathrm{loss}}}{\tau}, 1\right), \\
\mathrm{p_{evap}} = \mathrm{p}(v = 1) &= \ (1 - \mathrm{p_{ret}})
\end{aligned}
\tag{4.8}
$$

Here, the probability that a planet retains its envelope is $\mathrm{p_{ret}}$, the probability that a planet loses its envelope is $\mathrm{p_{evap}}$, τ is the age of the star, and α is an additional free parameter in the model. The mass-loss timescale, t_{loss}, is given by:

$$
t_{\mathrm{loss}} = \frac{G M_{\mathrm{env}}^2}{\pi \epsilon R_{\mathrm{prim}}^3 F_{\mathrm{XUV,E100}}} \frac{F_\oplus}{F_p} \tag{4.9}
$$

where M_{env} is the envelope mass of the planet, ϵ is the mass-loss efficiency, $F_{\mathrm{XUV,E100}}$ is

the XUV flux at the Earth when it was 100 My old, and F_p is the bolometric incident flux on the planet (Lopez et al., 2012). The nominal values of ϵ, $F_{\rm XUV,E100}$ and τ we take to be 0.1, 504 erg s^{-1} cm^{-2} and 5 Gyr for each star. In order to account for systematic effects in the overall normalization caused by assuming these values, the α parameter in Equation (4.8) allows this retention probability to scale up or down.

NR20 developed the above equations to encompass three compositional subpopulations of planets: planets with significant gaseous envelopes by volume (gaseous planets), rocky planets that formed with and subsequently lost their envelopes due to photoevaporation (evaporated rocky cores), and rocky planets that formed without any gaseous envelope (intrinsically rocky planets). These three subpopulations all assume a rocky composition for the core, defined by the mass-radius relation in Equation 4.7. With these three subpopulations, NR20 created four models that are further discussed in Section 4.3.3.

4.3.2 Icy-Core Subpopulations

Mirroring the three planet subpopulations with rocky cores introduced in the previous section, we develop three new subpopulations of planets: intrinsically icy planets, evaporated icy cores, and gaseous planets with icy cores. For each of these subpopulations, the period distribution is characterized by a broken power-law and the mass distribution is characterized by a log-normal, as before.

For the mass-radius relation of the intrinsically icy and evaporated icy-core subpopulations, we follow the analytic mass-radius relation of a pure water ice planet from Seager et al. (2007):

$$\begin{aligned} R &= R_1 \cdot 10^{k_1 + \frac{1}{3}\log_{10}(M/M_1) - k_2\ (M/M_1)^{k_3}} \\ R_1 &= 4.43; M_1 = 5.52 \\ k_1 &= -0.209396; k_2 = 0.0807; k_3 = 0.375 \end{aligned} \tag{4.10}$$

where the form of the equation is the same as in 4.7, but with different numerical values for the listed parameters. We also use a fixed 5% fractional intrinsic scatter for this icy mass-radius relation, the same fractional scatter used for the rocky mass-radius relation.

For gaseous planets with icy cores, we treat the mass-radius relation to be exactly the same as the mass-radius relation for gaseous planets with rocky cores in NR20, given by Equation 4.6. We neglect any potential dependence of the gaseous planet population-level mass-radius relation on the heavy element core composition. With the mass-radius relation being identical, this gaseous subpopulation then differs from the gaseous planets with rocky cores in that its remnant evaporated cores have icy compositions with radii determined by Equation 4.10, as opposed to rocky compositions with radii determined by Equation 4.7.

With these additional three subpopulations of planets, we are left with a total of six distinct subpopulations: gaseous planets with rocky cores, gaseous planets with icy cores, evaporated rocky cores, evaporated icy cores, intrinsically rocky planets, and intrinsically icy planets. The subpopulations are differentiated along two axes: composition and formation. In terms of composition, we have gaseous, rocky, and icy planets. In terms of formation, planets can either form with and retain a gaseous envelope (gaseous), form with and lose a gaseous envelope (evaporated), or form without a gaseous envelope (what we call intrinsic). Gaseous composition planets can only belong to the gaseous formation subpopulation, but icy composition and rocky composition planets can either belong to the evaporated or intrinsic formation subpopulations.

Since planets of a given core composition that form with a gaseous envelope are assumed to share a common formation history, regardless of what happens to that envelope, we have four distinct subpopulations in terms of mass and period. These four initial compositional subpopulations then comprise the six evolved compositional subpopulations that we have enumerated above. To distinguish between these initial compositional subpopulations and the evolved compositional subpopulations, we will use the term "mixture" to refer to the

initial compositional subpopulations, and "subpopulation" when referring to the evolved compositional subpopulations. Generally speaking, both "mixture" and "subpopulation" equally apply to these groupings, and we are only differentiating these terms as shorthand.

We label these four mixtures the following: "Gaseous Rocky (GR)", which includes gaseous planets with rocky cores and evaporated rocky cores; "Gaseous Icy (GI)", which includes gaseous planets with icy cores and evaporated icy cores; "Non-gaseous Rocky" (NR) which includes the intrinsically rocky planets; and "Non-gaseous Icy" (NI) which includes the intrinsically icy planets. Note that since the intrinsically icy and rocky subpopulations are assumed to not evolve over time, the initial compositional mixtures and the evolved compositional subpopulations are the same. In addition, as noted above, when we refer to the "gaseous rocky mixture" we are including evaporated rocky cores alongside the gaseous compositional subpopulation, but when we refer to the "gaseous rocky subpopulation" we are strictly referring to the evolved gaseous compositional subpopulation.

Each of these four mixtures entails six parameters: two mass distribution parameters (μ_M, σ_M), three period distribution parameters ($P_{\rm break}$, β_1, β_2), and one mixture fraction parameter (Q). To distinguish between these four mixtures, we label these parameters with the two-letter subscripts listed above. For example, the gaseous rocky mixture will have the parameters $\mu_{M,\rm GR}$, $\sigma_{M,\rm GR}$, $P_{\rm break,GR}$, $\beta_{1,\rm GR}$, $\beta_{2,\rm GR}$, $Q_{\rm GR}$. The fraction of planets belonging to an evolved subpopulation depends on both the corresponding Q parameter as well as the α parameter and mass and period distribution parameters if the mixture is gaseous.

4.3.3 Planet Population Models

With the equations and distributions listed in Section 4.3.1, NR20 created four separate models, labelled by number which increases with complexity: Models 1, 2, 3 and 4. Their first model, called Model 1, has only one population, which spans a range of compositions

from rocky to gaseous with rocky cores, and has 13 free parameters. Their second model, Model 2, accounts for photoevaporation by introducing a second subpopulation of evaporated rocky cores in addition to a subpopulation of gaseous planets with rocky cores, with both subpopulations belonging to the same initial mixture. Model 2 has two subpopulations and 16 free parameters Their third model, Model 3, added a third subpopulation (and second mixture) of intrinsically rocky planets that formed without gaseous envelopes, for a total of three subpopulations and 22 free parameters. Finally, Model 4 added an additional log-normal component to the mass distribution of the formed-gaseous mixture (gaseous planets with evaporated rocky cores). In the analysis herein, we revisit Models 1, 2 and 3, but do not include Model 4. The modification introduced in Model 4 is relatively minor and does not align with how we delineate models in this paper, where each model includes a different combination of planet subpopulations.

Building upon the models in NR20, we construct seven new models that incorporate the three additional icy-core compositional subpopulations in various ways. A graphical summary of each of the ten models in the paper (seven new models plus the three from NR20), and the subpopulations present in each of them, is shown in Figure 4.1. Six of these models build upon Model 3 to include icy composition planets, and are named with the prefix "Icy". The seventh model, called Model Z, was created in order to respond directly to Zeng et al. (2019), and includes gaseous planets with icy cores alongside intrinsically rocky planets, without incorporating photoevaporation. Because it is a special case, we will describe it more fully in Subsection 4.3.4. The other six models are described as follows.

Model Icy3a closely emulates Model 3 in NR20, with the only difference being that the intrinsically rocky subpopulation is replaced with an intrinsically icy subpopulation. Thus, the three subpopulations included in Model Icy3a are intrinsically icy, rocky evaporated cores, and gaseous with rocky cores; a mix of all three compositions and formation pathways. Since Model Icy3a has the same complexity as Model 3, it also has 22 free parameters and two

mixtures.

Model Icy3b also closely emulates Model 3 in NR20, but substitutes an icy evaporated core subpopulation for the rocky evaporated core subpopulation. Thus, the three subpopulations included in Model Icy3b are intrinsically rocky, icy evaporated cores, and gaseous with icy cores. One thing to note is that the subpopulation of gaseous planets with icy cores is implemented in exactly the same way as a subpopulation of gaseous planets with rocky cores (as in, for example Models 3 and Icy3a), because we do not make a distinction between the mass-radius relations of gaseous planets with rocky versus icy cores. However, because Model Icy3b contains a single subpopulation of evaporated core planets which are icy, we can assume the gaseous planets from which they came also had icy cores. Like Model 3 and Icy3a, Model Icy3b has 22 free parameters and two mixtures.

Model Icy4a is related to Model Icy3a in the sense of adding an intrinsically icy subpopulation to Model 3, but in this case alongside the existing intrinsically rocky subpopulation of Model 3 rather than replacing it. Thus, the four subpopulations included in Model Icy4a are intrinsically rocky, intrinsically icy, rocky evaporated cores, and gaseous planets with rocky cores. Model Icy4a has 28 free parameters and three mixtures.

In the same vein, Model Icy4b adds a subpopulation of icy evaporated core planets alongside the evaporated rocky-core planets. Thus, the four subpopulations included in Model Icy4b are intrinsically rocky, rocky evaporated cores, icy evaporated cores, and gaseous planets. We include only one subpopulation of gaseous planets, which are presumed to have a mix of both icy and rocky evaporated cores. These gaseous planets, icy evaporated cores, and rocky evaporated cores all come from the same initial mixture, and thus follow the same mass and period distributions. However, there is an additional parameter compared to Model Icy3b that determines the fraction of rocky evaporated cores compared to icy evaporated cores. Model Icy4b thus has 23 free parameters and two mixtures.

Model Icy5 combines the additions of both Model Icy4a and Icy4b and has five subpop-

ulations of planets (belonging to three initial mixtures): one gaseous, both icy and rocky evaporated cores, and both intrinsically rocky and icy subpopulations. Thus, the five subpopulations included in Model Icy5 are intrinsically rocky, intrinsically icy, rocky evaporated cores, icy evaporated cores, and gaseous planets. As in Model Icy4b, we include only one subpopulation of gaseous planets, which are presumed to have a mix of rocky and icy cores. Model 5 has 29 free parameters and three mixtures.

Finally, Model Icy6 has each one of the six subpopulations of planets that we've modeled: gaseous planets with both rocky and icy cores, evaporated cores with rocky and icy compositions, and intrinsically rocky and icy planets. In this model, there are two distinct subpopulations of gaseous planets: one with rocky cores, and one with icy cores. These two subpopulations arise from different initial mixtures, and thus have independent mass and period distributions unlike Models Icy4b and Icy5, in which the single gaseous subpopulation that encompassed both icy and rocky cores had only one mass distribution and one period distribution. Model 6 has 34 free parameters and includes all four initial mixtures.

4.3.4 Model Z

Our final model, Model Z, aims to assess the claim of Zeng et al. (2019) that planet mass loss through atmospheric escape is not necessary to explain the *Kepler* radius gap, and that the radius distribution of *Kepler* can be recreated by a mix of icy, rocky, and gaseous composition planets. Each Icy model listed in the preceding section incorporates photoevaporation alongside the addition of icy compositional subpopulations. For our Model Z, we construct a model that includes icy composition planets but does not invoke photoevaporation.

Model Z thus includes two subpopulations of exoplanets: an intrinsically rocky subpopulation, and a gaseous with icy-core subpopulation. The intrinsically rocky subpopulation is implemented the same as in Model 3. The gaseous with icy-core subpopulation, however, has a modified mass-radius relation that follows the prescription from Zeng et al. (2019) and

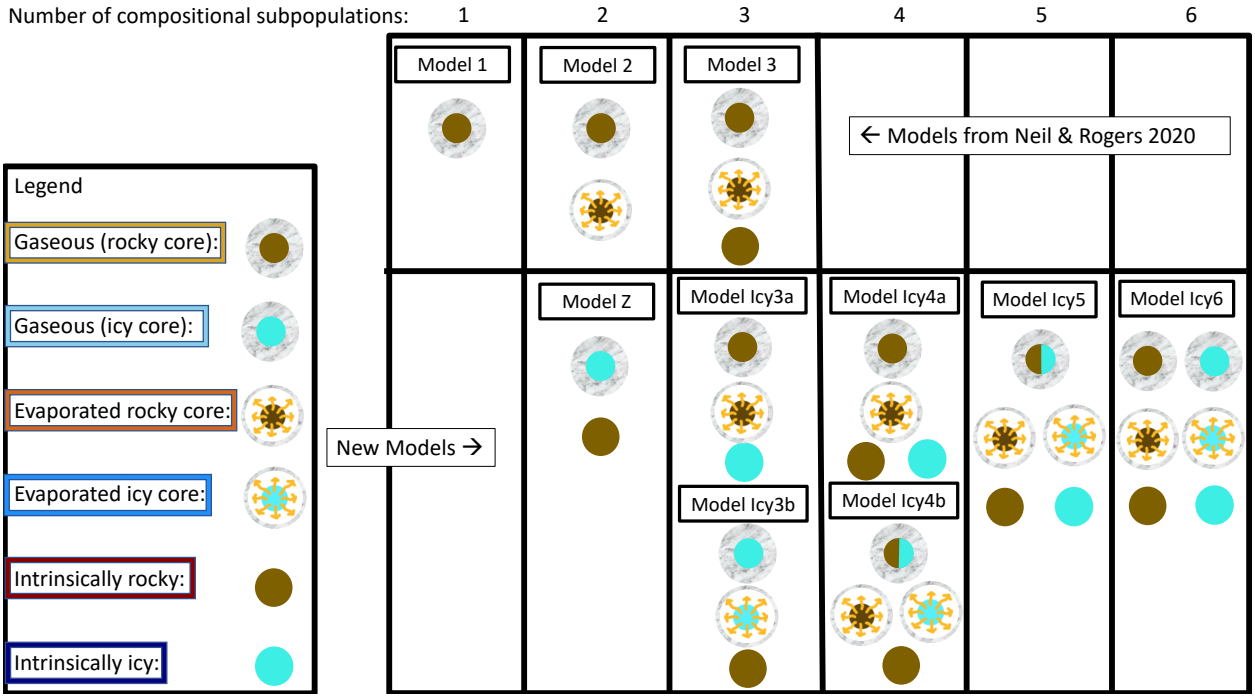

Figure 4.1: A diagram of the models developed in this paper. The legend on the left shows the six planet subpopulations included in these models, where these subpopulations are distinguished by their core composition (rocky or icy), and formation/evolution history (gaseous, evaporated, or intrinsically rocky/icy). The colored border around each subpopulation in the legend is consistent with the color scheme used in the following figures in this paper. Models in the table increase in the number of subpopulations (labeled at the top), and thus the complexity, towards the right. The top row of models, taken from NR20, only include planet subpopulations with rocky-core compositions. The models introduced in this paper, on the bottom row, incorporate a mix of rocky and icy-core compositions in different combinations. Our most complex model, Model Icy6, includes all six subpopulations of planets listed in the legend.

is similar to how Model 1 was constructed. The mass-radius relation still follows a double broken power-law with different fractional intrinsic scatter for each mass regime, but the power-law at the low-mass end is fixed to an icy composition mass-radius relation.

Zeng et al. (2019) used the following mass-radius relation for icy planets:

$$\begin{aligned} R &= fM^{1/3.7} \\ f &= 1 + 0.55x - 0.14x^2 \end{aligned} \tag{4.11}$$

where $\frac{1}{3.7}$ is the power-law slope, and f is a coefficient that depends on x, the ice mass fraction of the core (with the remainder of the core composed of silicate rock). They considered x to vary between $\frac{1}{2}$ and $\frac{2}{3}$. We take the value of x to be the mean value of these two, $\frac{7}{12}$, which leads to $f = 1.27$. We keep the fractional intrinsic scatter for this segment of the mass-radius relation fixed to 5% as in other models. This scatter is somewhat higher than the bounds of $x = \frac{1}{2}$ to $x = \frac{2}{3}$ suggest, but a lower scatter leads to issues with the Hamiltonian Monte Carlo sampler. In total, Model Z has two subpopulations and 18 free parameters.

4.3.5 Data

As in NR20, we use for our dataset the California-*Kepler* Survey (CKS), a subset of transiting planets from *Kepler* with high-resolution spectroscopic follow-up of their host stars (Petigura et al., 2017; Johnson et al., 2017), cross-matched with *Gaia* data (Gaia Collaboration et al., 2016, 2018), which reduces the uncertainty on the planet radius measurements. We follow the cuts listed in Fulton & Petigura (2018) to ensure high quality data and use the planet radii reported therein, which were calculated using both CKS spectroscopy and *Gaia* parallaxes. Our sample is limited to orbital periods between 0.3 and 100 days, and radii between 0.4 and $30R_\oplus$. Our final planet sample has 1130 planets with a median radius uncertainty of 4.8%.

We use radial velocity mass measurements where available for planets in our sample. As in NR20, we limit our mass sample to RV-measured masses only, leaving the inclusion of TTV-measured masses to future work. Our mass sample was compiled from the NASA Exoplanet Archive on July 13, 2019 and each measurement reported was manually verified by checking the original source. In a departure from NR20, we use additional mass measurements not previously included. These mass measurements are not well-constrained and mostly come from Marcy et al. (2014). Our final sample contains 68 planets with mass measurements, with a median mass uncertainty of 27%.

Finally, the detection efficiency of the *Kepler* survey as a function of radius and period that we use is identical to NR20 and we refer the reader to that work for full details on how that was calculated. Briefly, the procedure for calculating the detection efficiency follows the steps listed in Burke & Catanzarite (2017), Thompson et al. (2018), and Christiansen et al. (2020). We first apply identical cuts to the *Kepler* Q1-Q17 DR25 stellar target sample as were applied to the CKS planet candidate sample. We then use pixel-level injected light curves to fit a gamma CDF to the probability that a planet is detected, and properly dispositioned by the Robovetter, as a function of the expected multiple event statistic (MES). This fitted gamma CDF is then used in combination with the KeplerPORTS[1] Python package to calculate the detection efficiency as a function of radius and period for each target star, multiplied by the transit probability. The final detection efficiency is then the average over all target stars in our sample.

4.3.6 Fitting

We broadly follow the methodology of NR20 in terms of creating the models and fitting them with Stan[2] (Carpenter et al., 2017). Stan uses the No-U-Turn Sampler (NUTS) MCMC algo-

1. https://github.com/nasa/KeplerPORTs

2. http://mc-stan.org

rithm (Hoffman & Gelman, 2014), an extension of Hamiltonian Monte Carlo, to numerically evaluate hierarchical Bayesian models. The planet catalog is modeled as draws from an inhomogeneous Poisson process, a technique previously used to constrain the planet occurrence rate density in radius-period space (Foreman-Mackey et al., 2014). In addition to the population-level parameters, the inhomogeneous Poisson process likelihood includes as parameters in the model the true mass and radius of each planet, $M_{\rm true}$ and $R_{\rm true}$. These true parameters are sampled in the model and are conditioned on the observed mass and radius, $M_{\rm obs}$ and $R_{\rm obs}$, as well as their uncertainties, $\sigma_{M,\rm obs}$ and $\sigma_{R,\rm obs}$. Evaluating the inhomogeneous Poisson process likelihood (Eq 27 in NR20) involves calculating an integral over mass-radius-period space of the expected number of detected planets, a challenging computational task. The NUTS algorithm has the ability to efficiently handle large dimensional spaces, necessary for modeling the true radius and mass of each planet. In order to improve the performance of the MCMC sampler, we made several improvements over the methodology of NR20, which we detail below.

In order to allow for more efficient sampling of parameter space for NUTS, we reparametrize the mass-radius relation of planets with gaseous envelopes in our models. This mass-radius relation was originally parametrized in terms of the power-law slopes: γ_0, γ_1 and γ_2; along with a normalization, C, and two mass breaks, $M_{\rm break,1}$ and $M_{\rm break,2}$. We now directly sample the mean radius of the mass-radius relation at the break points and the lower and upper mass limits, to replace the three power-law slopes and the normalization. These parameters are given by: ρ_0, the radius of the mass-radius relation at the lower mass limit of $0.1M_\oplus$; ρ_1, the radius at $M_{\rm break,1}$; ρ_2, the radius at $M_{\rm break,2}$; ρ_3, the radius at the upper mass limit of $10,000M_\oplus$. The parameters ρ_0 and ρ_1 are given identical log-normal priors centered at $2.7R_\oplus$, with a spread of 0.1 dex. The parameters ρ_2 and ρ_3 are similarly given identical log-normal priors centered at $12R_\oplus$, with a spread of 0.1 dex.

This reparametrization changes the parameters that are directly sampled by NUTS and

results in different priors for the mass-radius relation parameters. However, the reparametrization is not intended to change the retrieved mass-radius relation itself. The original parameters can be obtained by simple transformations of the new parameters. These changes were implemented to reduce correlations between the parameters that are sampled, to eliminate problematic sampling behavior, and to improve the efficiency and speed of the sampling, as well as the convergence of the MCMC chains.

To further increase the efficiency, speed, and convergence of the MCMC sampler, we made improvements to the Stan code, particularly to the implementation of the integral in the inhomogeneous Poisson process likelihood. Whereas the integral was previously calculated using a fixed grid in period, mass, and radius, the grids in these three dimensions now adapt to more efficiently sample the regions in mass-radius-period space that have high posterior density, based on the values of the hyperparameters at each step in the MCMC chain.

For each of our ten models, we ran 8 MCMC chains with 2,000 iterations for each chain. The first 1,000 of these iterations are used for warm up, where the MCMC algorithm fine-tunes its internal parameters and reaches an equilibrium distribution. We are then left with 8,000 posterior samples of each parameter. To assess the convergence and independence of each chain, we look at the Gelman-Rubin convergence diagnostic, $\hat{R}$. For each parameter, we ensure that $\hat{R} < 1.01$, a standard benchmark for chains mixing well. While the effective sample size (ESS) varies between parameters, we ensure all parameters have an ESS of at least 500, with most parameters having an ESS of above 1000.

4.4 Results

4.4.1 Model Fits

Table 4.1: Posterior Summary Statistics for Each Model

Model			1	2	3	Z	Icy3a	Icy3b	Icy4a	Icy4b	Icy5	Icy6
Subpopulations			1	2	3	2	3	3	4	4	5	6
Parameters			13	16	22	18	22	22	28	23	29	34
Parameter	Units	Prior										
Γ_0	$N_{\rm pl}/N_{\rm s}$	lnN(0, 1)	$1.28^{+0.06}_{-0.05}$	$1.17^{+0.05}_{-0.05}$	$1.08^{+0.05}_{-0.05}$	$1.04^{+0.04}_{-0.04}$	$1.11^{+0.05}_{-0.05}$	$1.04^{+0.04}_{-0.04}$	$1.13^{+0.06}_{-0.05}$	$1.07^{+0.05}_{-0.04}$	$1.12^{+0.06}_{-0.05}$	$1.05^{+0.05}_{-0.04}$
C	$R_\oplus$	-[2]	-	$2.44^{+0.1}_{-0.1}$	$2.49^{+0.12}_{-0.12}$	-	$2.55^{+0.12}_{-0.11}$	$2.65^{+0.15}_{-0.14}$	$2.55^{+0.11}_{-0.11}$	$2.47^{+0.12}_{-0.11}$	$2.54^{+0.12}_{-0.12}$	$2.54^{+0.12}_{-0.11}$
γ_0	-	-[2]	-	$-0.01^{+0.03}_{-0.02}$	$0.0^{+0.03}_{-0.03}$	-	$-0.02^{+0.02}_{-0.02}$	$0.01^{+0.02}_{-0.02}$	$-0.01^{+0.03}_{-0.03}$	$0.01^{+0.03}_{-0.03}$	$-0.01^{+0.03}_{-0.03}$	$-0.02^{+0.03}_{-0.03}$
γ_1	-	-[2]	$0.42^{+0.01}_{-0.01}$	$0.74^{+0.06}_{-0.05}$	$0.74^{+0.06}_{-0.05}$	$0.82^{+0.07}_{-0.06}$	$0.71^{+0.05}_{-0.05}$	$0.75^{+0.06}_{-0.05}$	$0.65^{+0.05}_{-0.05}$	$0.72^{+0.06}_{-0.05}$	$0.65^{+0.05}_{-0.05}$	$0.62^{+0.05}_{-0.04}$
γ_2	-	-[2]	$0.01^{+0.03}_{-0.04}$	$-0.01^{+0.03}_{-0.03}$	$-0.01^{+0.03}_{-0.03}$	$-0.01^{+0.03}_{-0.03}$	$-0.01^{+0.03}_{-0.03}$	$-0.01^{+0.03}_{-0.03}$	$-0.02^{+0.03}_{-0.03}$	$-0.01^{+0.03}_{-0.03}$	$-0.02^{+0.03}_{-0.03}$	$-0.02^{+0.03}_{-0.03}$
σ_0	-	ln N(−2.8, 0.25)	$0.07^{+0.02}_{-0.01}$	$0.17^{+0.01}_{-0.01}$	$0.13^{+0.02}_{-0.02}$	-	$0.07^{+0.02}_{-0.01}$	$0.08^{+0.02}_{-0.01}$	$0.06^{+0.01}_{-0.01}$	$0.13^{+0.02}_{-0.01}$	$0.06^{+0.01}_{-0.01}$	$0.09^{+0.02}_{-0.02}$
σ_1	-	ln N(−1.3, 0.25)	$0.30^{+0.04}_{-0.03}$	$0.32^{+0.06}_{-0.05}$	$0.31^{+0.06}_{-0.05}$	$0.35^{+0.06}_{-0.05}$	$0.28^{+0.06}_{-0.05}$	$0.31^{+0.06}_{-0.05}$	$0.26^{+0.05}_{-0.05}$	$0.30^{+0.06}_{-0.05}$	$0.25^{+0.06}_{-0.05}$	$0.24^{+0.05}_{-0.04}$
σ_2	-	ln N(−2.3, 0.25)	$0.11^{+0.03}_{-0.02}$	$0.11^{+0.02}_{-0.02}$	$0.10^{+0.02}_{-0.02}$	$0.10^{+0.02}_{-0.02}$	$0.1^{+0.02}_{-0.02}$	$0.11^{+0.02}_{-0.02}$	$0.11^{+0.02}_{-0.02}$	$0.10^{+0.02}_{-0.02}$	$0.11^{+0.02}_{-0.02}$	$0.11^{+0.03}_{-0.02}$
$M_{\rm break,1}$	$M_\oplus$	ln N(2, 1)[3]	$0.46^{+0.08}_{-0.07}$	$16.2^{+2.2}_{-1.9}$	$16.5^{+2.2}_{-2.0}$	$31.1^{+2.5}_{-2.3}$	$13.3^{+1.6}_{-1.5}$	$20.8^{+2.2}_{-2.0}$	$11.2^{+1.9}_{-1.6}$	$16.0^{+2.3}_{-2.2}$	$11.4^{+2.0}_{-1.7}$	$9.6^{+1.6}_{-1.3}$
$M_{\rm break,2}$	$M_\oplus$	ln N(5, 0.25)	308^{+57}_{-46}	169^{+24}_{-20}	164^{+23}_{-19}	175^{+25}_{-20}	151^{+21}_{-17}	172^{+24}_{-20}	158^{+23}_{-19}	166^{+24}_{-19}	158^{+22}_{-18}	158^{+23}_{-19}
α	-	ln N(0, 1)	-	$7.9^{+1.3}_{-1.2}$	$8.5^{+2.2}_{-1.7}$	-	$3.0^{+1.0}_{-0.8}$	$0.3^{+0.2}_{-0.1}$	$7.4^{+3.5}_{-2.6}$	$8.5^{+4.7}_{-2.2}$	$6.9^{+3.5}_{-2.4}$	$7.2^{+2.2}_{-1.9}$
$\mu_{M,\rm GR}$	$\ln(\frac{M}{M_\oplus})$	N(1, 2)	$0.31^{+0.13}_{-0.14}$	$0.87^{+0.09}_{-0.09}$	$1.82^{+0.15}_{-0.15}$	-	$0.31^{+0.16}_{-0.18}$	-	$-0.61^{+1.01}_{-0.78}$	$1.65^{+0.19}_{-0.23}$	$-0.56^{+1.10}_{-0.84}$	$1.52^{+0.21}_{-0.21}$
$\sigma_{M,\rm GR}$	$\ln(\frac{M}{M_\oplus})$	ln N(1, 0.25)	$1.73^{+0.09}_{-0.08}$	$1.69^{+0.07}_{-0.06}$	$1.37^{+0.08}_{-0.08}$	-	$1.96^{+0.10}_{-0.09}$	-	$2.52^{+0.32}_{-0.33}$	$1.46^{+0.13}_{-0.09}$	$2.49^{+0.34}_{-0.37}$	$0.86^{+0.13}_{-0.10}$
$\beta_{1,\rm GR}$	-	N(0.5, 0.5)	$1.10^{+0.09}_{-0.09}$	$1.09^{+0.09}_{-0.09}$	$1.12^{+0.24}_{-0.17}$	-	$1.11^{+0.10}_{-0.09}$	-	$1.20^{+0.29}_{-0.20}$	$1.16^{+0.24}_{-0.17}$	$1.25^{+0.33}_{-0.22}$	$1.20^{+0.24}_{-0.19}$
$\beta_{2,\rm GR}$	-	N(−0.5, 0.5)	$-0.72^{+0.05}_{-0.05}$	$-0.79^{+0.05}_{-0.06}$	$-0.87^{+0.10}_{-0.10}$	-	$-1.06^{+0.09}_{-0.09}$	-	$-0.90^{+0.17}_{-0.13}$	$-0.87^{+0.10}_{-0.10}$	$-0.92^{+0.16}_{-0.13}$	$-1.17^{+0.20}_{-0.25}$
$P_{\rm break,GR}$	days	ln N(2, 1)	$6.0^{+0.5}_{-0.4}$	$6.1^{+0.6}_{-0.4}$	$10.6^{+1.6}_{-2.2}$	-	$5.5^{+0.4}_{-0.4}$	-	$5.6^{+0.7}_{-0.8}$	$10.7^{+1.7}_{-2.4}$	$5.7^{+0.7}_{-0.7}$	$7.9^{+3.5}_{-1.7}$
$Q_{\rm GR}$[1]	-	D(1)	-	-	$0.67^{+0.06}_{-0.06}$	-	$0.64^{+0.04}_{-0.03}$	-	$0.5^{+0.06}_{-0.09}$	$0.56^{+0.09}_{-0.14}$	$0.48^{+0.06}_{-0.09}$	$0.41^{+0.09}_{-0.07}$
$\mu_{M,\rm GI}$	$\ln(\frac{M}{M_\oplus})$	N(1, 2)	-	-	-	$2.60^{+0.10}_{-0.10}$	-	$2.18^{+0.14}_{-0.15}$	-	-	-	$2.92^{+0.87}_{-1.42}$
$\sigma_{M,\rm GI}$	$\ln(\frac{M}{M_\oplus})$	ln N(1, 0.25)	-	-	-	$1.2^{+0.08}_{-0.07}$	-	$1.37^{+0.09}_{-0.09}$	-	-	-	$2.43^{+0.74}_{-0.05}$
$\beta_{1,\rm GI}$	-	N(0.5, 0.5)	-	-	-	$1.16^{+0.17}_{-0.14}$	-	$1.22^{+0.15}_{-0.14}$	-	-	-	$1.23^{+0.40}_{-0.38}$
$\beta_{2,\rm GI}$	-	N(−0.5, 0.5)	-	-	-	$-0.74^{+0.10}_{-0.10}$	-	$-0.79^{+0.1}_{-0.1}$	-	-	-	$-0.44^{+0.18}_{-0.22}$
$P_{\rm break,GI}$	days	ln N(2, 1)	-	-	-	$11.8^{+1.2}_{-1.6}$	-	$11.9^{+1.1}_{-1.4}$	-	-	-	$3.7^{+1.1}_{-0.6}$
$Q_{\rm GI}$[1]	-	D(1)	-	-	-	$0.57^{+0.04}_{-0.04}$	-	$0.59^{+0.04}_{-0.04}$	-	$0.15^{+0.15}_{-0.09}$	$0.02^{+0.03}_{-0.01}$	$0.08^{+0.04}_{-0.03}$
$\mu_{M,\rm NR}$	$\ln(\frac{M}{M_\oplus})$	N(0, 2)	-	-	$-0.15^{+0.24}_{-0.31}$	$0.09^{+0.13}_{-0.15}$	-	$0.08^{+0.13}_{-0.16}$	$1.02^{+0.21}_{-0.71}$	$-0.15^{+0.24}_{-0.31}$	$0.95^{+0.25}_{-0.22}$	$-0.80^{+0.51}_{-0.78}$
$\sigma_{M,\rm NR}$	$\ln(\frac{M}{M_\oplus})$	ln N(0.5, 0.25)	-	-	$1.81^{+0.20}_{-0.17}$	$1.57^{+0.16}_{-0.13}$	-	$1.59^{+0.16}_{-0.13}$	$0.96^{+0.52}_{-0.24}$	$1.79^{+0.19}_{-0.17}$	$1.03^{+0.48}_{-0.31}$	$1.72^{+0.24}_{-0.23}$
$\beta_{1,\rm NR}$	-	N(0.5, 0.5)	-	-	$1.08^{+0.14}_{-0.13}$	$0.93^{+0.10}_{-0.10}$	-	$1.03^{+0.10}_{-0.10}$	$1.01^{+0.27}_{-0.22}$	$1.07^{+0.14}_{-0.13}$	$0.98^{+0.25}_{-0.22}$	$1.01^{+0.17}_{-0.16}$
$\beta_{2,\rm NR}$	-	N(−0.5, 0.5)	-	-	$-1.36^{+0.18}_{-0.22}$	$-1.43^{+0.15}_{-0.16}$	-	$-1.52^{+0.15}_{-0.18}$	$-1.42^{+0.25}_{-0.3}$	$-1.47^{+0.22}_{-0.30}$	$-1.44^{+0.25}_{-0.29}$	$-1.44^{+0.24}_{-0.27}$
$P_{\rm break,NR}$	days	ln N(2, 1)	-	-	$5.1^{+0.5}_{-0.6}$	$6.2^{+0.6}_{-0.4}$	-	$5.9^{+0.4}_{-0.4}$	$5.1^{+1.2}_{-1.0}$	$5.20^{+0.48}_{-0.65}$	$5.0^{+1.3}_{-1.1}$	$5.4^{+0.7}_{-0.7}$
$Q_{\rm NR}$[1]	-	D(1)	-	-	$0.33^{+0.06}_{-0.06}$	$0.43^{+0.04}_{-0.04}$	-	$0.41^{+0.04}_{-0.04}$	$0.15^{+0.08}_{-0.04}$	$0.30^{+0.07}_{-0.07}$	$0.14^{+0.08}_{-0.04}$	$0.27^{+0.06}_{-0.06}$

Table 4.1, continued: Posterior Summary Statistics for Each Model

Model			1	2	3	Z	Icy3a	Icy3b	Icy4a	Icy4b	Icy5	Icy6
Subpopulations			1	2	3	2	3	3	4	4	5	6
Parameters			13	16	22	18	22	22	28	23	29	34
Parameter	Units	Prior										
$\mu_{M,\mathrm{NI}}$	$\ln(\frac{M}{M_\oplus})$	$\mathrm{N}(0, 2)$	-	-	-	-	$1.87^{+0.12}_{-0.12}$	-	$1.92^{+0.14}_{-0.17}$	-	$1.93^{+0.12}_{-0.15}$	$1.94^{+0.23}_{-0.39}$
$\sigma_{M,\mathrm{NI}}$	$\ln(\frac{M}{M_\oplus})$	$\ln \mathrm{N}(0.5, 0.25)$	-	-	-	-	$1.09^{+0.10}_{-0.10}$	-	$1.08^{+0.14}_{-0.12}$	-	$1.07^{+0.12}_{-0.1}$	$1.22^{+0.27}_{-0.18}$
$\beta_{1,\mathrm{NI}}$	-	$\mathrm{N}(0.5, 0.5)$	-	-	-	-	$1.66^{+0.27}_{-0.25}$	-	$1.60^{+0.26}_{-0.24}$	-	$1.62^{+0.27}_{-0.23}$	$1.39^{+0.38}_{-0.43}$
$\beta_{2,\mathrm{NI}}$	-	$\mathrm{N}(-0.5, 0.5)$	-	-	-	-	$-0.86^{+0.13}_{-0.13}$	-	$-0.90^{+0.14}_{-0.14}$	-	$-0.89^{+0.14}_{-0.14}$	$-0.78^{+0.30}_{-0.28}$
$P_{\mathrm{break,NI}}$	days	$\ln \mathrm{N}(2, 1)$	-	-	-	-	$12.8^{+1.0}_{-1.0}$	-	$12.8^{+1.0}_{-1.2}$	-	$12.8^{+1.0}_{-1.1}$	$15.2^{+13.7}_{-2.9}$
Q_{NI}[1]	-	$\mathrm{D}(1)$	-	-	-	-	$0.36^{+0.03}_{-0.04}$	-	$0.35^{+0.04}_{-0.04}$	-	$0.35^{+0.04}_{-0.04}$	$0.24^{+0.08}_{-0.11}$

The retrieved median and 1-σ intervals (15.9% and 84.1% percentiles) for each parameter of interest for each of the ten models presented in the paper, along with their units and assumed prior. Mass and period distribution parameters are broken into four initial compositional mixtures: Gaseous Rocky (GR), containing both gaseous planets with rocky cores and evaporated rocky planets, Gaseous Icy (GI), with both gaseous planets with icy cores and evaporated icy cores, Non-gaseous (intrinsically) Rocky (NR), and Non-gaseous (intrinsically) Icy (NI). Parameters with units listed as '-' are dimensionless. For priors, N represents a normal distribution with given mean and scatter, lnN represents a lognormal distribution with given mean and scatter, and D represents a Dirichlet distribution with parameter length equal to the number of mixture components. For the case of two mixture components, the Dirichlet prior chosen is equivalent to a uniform distribution between 0 and 1.

The model fits are summarized in Table 4.1, with the posterior median and 1-σ intervals (calculated from the 15.9 and 84.1 percentiles) displayed for each parameter in each of our ten models. We briefly summarize these results for each model below, in order of increasing model complexity. We then further compare these models in Section 4.4.2 and assess the role of degeneracies between subpopulations in Section 4.4.3. As a reminder, when referring to the six evolved compositional subpopulations we will use the term "subpopulation", whereas when referring to the four initial compositional subpopulations we will use the term "mixture".

Our fits for Models 1, 2, and 3 are broadly consistent with the fits presented in NR20. All parameters are consistent within 2σ, with the majority of parameters consistent within 1σ. The most notable shift is the fraction of intrinsically rocky planets in Model 3, $Q_{\rm NR}$. This fraction increases from $0.20^{+0.06}_{-0.05}$ in NR20 to $0.33^{+0.06}_{-0.06}$ here. However, this fraction is not tightly constrained and is only discrepant by 1.5σ. Given the similarities in methodology between this work and NR20, any discrepancies must be a result of either the modification of the dataset explained in section 4.3.5, or the change to the model parametrization and Stan code explained in section 4.3.6.

Model Z, which includes a subpopulation of intrinsically rocky planets and a subpopulation of icy cores with gaseous envelopes, is best compared with Model 3. We find the fraction of intrinsically rocky planets, $Q_{\rm NR}$, to be $0.43^{+0.04}_{-0.04}$, higher by 1.4σ than the fraction retrieved by Model 3, $0.33^{+0.06}_{-0.06}$. This higher fraction compensates for the lack of evaporated rocky cores in this model. The mass and period distribution parameters for this intrinsically rocky mixture are similar to that of Model 3, each within 2σ. For the gaseous with icy-core subpopulation, the best comparison is the gaseous with rocky-core subpopulation in Model 3. In Model Z, the mass-radius relation below the first mass break, $M_{\rm break,1}$, is fixed to an icy composition rather than fitted to a power law as in Model 3. As a consequence, this mass break is significantly higher than that of the gaseous subpopulation of Model 3 at

$31.1^{+2.5}_{-2.3} M_\oplus$, compared to $16.5^{+2.2}_{-2.0} M_\oplus$. Finally, due to the lack of evaporated cores linked to this subpopulation in Model Z, the mean of the lognormal mass distribution, $\mu_{M,\mathrm{GI}}$, is shifted towards higher masses, up to $2.60^{+0.10}_{-0.10}$ ($13.5 M_\oplus$) compared to $1.82^{+0.15}_{-0.15}$ ($6.2 M_\oplus$) in Model 3.

Model Icy3a replaces the intrinsically rocky subpopulation in Model 3 with an intrinsically icy subpopulation. These intrinsically icy planets have larger overlap with the low-mass gaseous planets; as a result, the intrinsic mass-radius scatter of the low-mass gaseous planets, σ_0, is lower by 2σ at $0.07^{+0.02}_{-0.01}$ compared to $0.13^{+0.02}_{-0.02}$ for Model 3. The other mass-radius parameters for this gaseous subpopulation are largely unchanged.

Various shifts in the mass distribution and period distribution parameters of the gaseous rocky mixture in Model Icy3a serve to increase the number of evaporated planets relative to gaseous planets, to compensate for the lack of intrinsically rocky planets compared to Model 3. The scaling on the photoevaporation timescale, α, is lower by 4σ at $3.0^{+1.0}_{-0.8}$ compared to $8.5^{+2.2}_{-1.7}$, leading to more evaporated planets. The mean of the mass distribution, $\mu_{M,\mathrm{GR}}$, is shifted towards lower masses at $0.31^{+0.16}_{-0.18}$ ($1.36 M_\oplus$) with a higher scatter $\sigma_{M,\mathrm{GR}}$ of $1.96^{+0.10}_{-0.09}$, compared to $1.82^{+0.15}_{-0.15}$ ($6.2 M_\oplus$) and $1.37^{+0.08}_{-0.08}$ for Model 3. The period distribution break $P_{\mathrm{break,GR}}$ happens at roughly half the orbital period at $5.5^{+0.4}_{-0.4}$, and the slope past the break $\beta_{2,\mathrm{GR}}$ is steeper by 1.4σ at $-1.06^{+0.09}_{-0.09}$, compared to $-0.87^{+0.10}_{-0.10}$.

While the fraction of intrinsically icy planets in Model Icy3a, Q_{NI}, is $0.36^{+0.03}_{-0.04}$ and within 1σ of the fraction of intrinsically rocky planets in Model 3, the mass and period distribution parameters are significantly different. These shifts concentrate the intrinsically icy planets toward higher masses and longer orbital periods. The mass distribution has a higher mean $\mu_{M,\mathrm{NI}}$ of $1.87^{+0.12}_{-0.12}$ and lower scatter $\sigma_{M,\mathrm{NI}}$ of $1.09^{+0.10}_{-0.10}$, compared to $-0.15^{+0.24}_{-0.31}$ and $1.81^{+0.20}_{-0.17}$ for the intrinsically rocky mixture in Model 3. Furthermore, the period break $P_{\mathrm{break,NI}}$ is shifted higher, at $12.8^{+1.0}_{-1.0}$ compared to $5.1^{+0.5}_{-0.6}$, with a steeper slope before the break $\beta_{1,\mathrm{NI}}$ of $1.66^{+0.27}_{-0.25}$ compared to $1.08^{+0.14}_{-0.13}$, and a shallower slope after the break $\beta_{2,\mathrm{NI}}$

of $-0.86^{+0.13}_{-0.13}$ compared to $-1.36^{+0.18}_{-0.22}$.

Whereas Model Icy3a replaces the intrinsically rocky subpopulation with an equivalent intrinsically icy subpopulation, Model Icy3b retains the intrinsically rocky subpopulation but modifies the evaporated core subpopulation, giving these evaporated planets an icy composition. Like Icy3a, this model has an increased number of evaporated planets relative to gaseous planets when compared to Model 3. Model Icy3b retrieves a lower σ_0 of $0.08^{+0.02}_{-0.01}$ compared to $0.13^{+0.02}_{-0.02}$ for Model 3, and an even lower α of $0.3^{+0.2}_{-0.1}$ compared to $3.0^{+1.0}_{-0.8}$.

This mixture of planets that formed with gaseous envelopes has similar mass and period distribution parameters to the analogous mixture in Model 3, except with the mean of the mass distribution $\mu_{M,\mathrm{GI}}$ higher by 1.7σ at $2.18^{+0.14}_{-0.15}$ ($8.8M_\oplus$) compared to $1.82^{+0.15}_{-0.15}$ ($6.2M_\oplus$). The fraction of intrinsically rocky planets Q_{NR} is $0.41^{+0.04}_{-0.04}$, closer to the value retrieved in Model Z than that of Model 3. Similarly, the mass and period distribution parameters for this mixture are similar to both Models 3 and Z, although closer that of Model Z.

Similar to Model Icy3a, Model Icy4a introduces an intrinsically icy subpopulation. However, rather than replace the intrinsically rocky subpopulation of Model 3, the intrinsically icy subpopulation is added alongside it for a total of four subpopulations. Compared to Icy3a, the characteristics of the intrinsically icy subpopulation are largely the same, but the gaseous rocky mixture is shifted to accommodate the intrinsically rocky planets. The fractions of these intrinsically icy planets Q_{NI} is $0.35^{+0.04}_{-0.04}$, a value consistent within 0.2σ of Model Icy3a. The mass and period distribution parameters of this mixture are similarly all within 0.3σ of their retrieved values in Model Icy3a. Thus, the fraction of intrinsically rocky planets Q_{NR}, retrieved to be $0.15^{+0.08}_{-0.04}$, serves to reduce the fraction of gaseous and evaporated planets by an equivalent amount relative to Model Icy3a, with Q_{GR} changing from $0.64^{+0.04}_{-0.03}$ in Model Icy3a to $0.5^{+0.06}_{-0.09}$ in Model Icy4a. This intrinsically rocky mixture has similar period distribution parameters to Model 3, but the mass distribution is shifted towards a higher mean $\mu_{M,\mathrm{NR}}$ at $1.02^{+0.21}_{-0.71}(2.8M_\oplus)$ and a lower scatter $\sigma_{M,\mathrm{NR}}$ of $0.96^{+0.52}_{-0.24}$,

compared to $-0.15^{+0.24}_{-0.31}$ and $1.81^{+0.20}_{-0.17}$ in Model 3. Conversely, relative to both Models 3 and Icy3a, the gaseous rocky mixture shifts toward a lower mean for the mass distribution $\mu_{M,\mathrm{GR}}$ to $-0.61^{+1.01}_{-0.78}$ ($0.5 M_\oplus$), with 2.4σ and 0.9σ shifts relative to Models 3 and Icy3a, and a higher $\sigma_{M,\mathrm{GR}}$ of $2.52^{+0.32}_{-0.33}$, which are 3.4σ and 1.6σ shifts. The retrieved period distribution parameters are closely similar to those of Model Icy3a, with all parameters within 1σ.

Model Icy4b introduces evaporated icy cores alongside evaporated rocky cores to Model 3, except the two subpopulations share the same underlying mass and period distribution. The resulting fit for this model has a high degree of similarity to the fit to Model 3. The retrieved values for the mass-radius relation parameters, as well as the mass and period distribution parameters for both the intrinsically rocky and formed-gaseous mixtures, are all within 1σ of the corresponding values in Model 3. The fraction of planets that formed gaseous with icy cores, Q_{GI}, is $0.15^{+0.15}_{-0.09}$, which is low but is loosely constrained with large uncertainty. In the case of $Q_{\mathrm{GI}} = 0$, Model Icy4b reduces to to Model 3. The formed-gaseous with icy-core planets mostly replace formed-gaseous with rocky-core planets, as Q_{GR} reduces to $0.56^{+0.09}_{-0.14}$ from $0.67^{+0.06}_{-0.06}$ in Model 3. The fraction of intrinsically rocky planets Q_{NR} is slightly reduced, but within 0.4σ of the value in Model 3.

Model Icy5 introduces intrinsically icy planets to Model Icy4b, or equivalently, introduces evaporated icy-core planets to Model Icy4a. The resulting model fit ends up looking similar to another model fit, that of Model Icy4a. Explicitly, when the fraction of formed-gaseous with icy-core planets, Q_{GI}, is zero, Model Icy5 reduces to Model Icy4a. Indeed, this mixture is found to have an extremely low fraction Q_{GI} of $0.02^{+0.03}_{-0.01}$, consistent with zero within 2σ. The mass-radius relation parameters, together with the mass and period distribution parameters for each mixture, are all retrieved at highly similar values to those of Model Icy4a, well within 1σ. We find little support for including this subpopulation of evaporated icy-core planets, assuming identical mass and period distribution to the evaporated rocky-core

planets.

Finally, Model Icy6 builds upon Model Icy5, where the formed-gaseous with icy-core mixture is given separate mass and period distributions from the formed-gaseous with rocky-core mixture. In the fit to the *Kepler* data, this newly independent mixture is concentrated towards high masses and long orbital periods, and constitutes the majority of gas giants in the sample. Additionally, these mass and period distributions result in very low numbers of evaporated icy cores. The fraction of planets belonging to this formed-gaseous with icy-core mixture $Q_{\rm GI}$ increases relative to Models Icy4b and Icy5 to $0.08^{+0.04}_{-0.03}$, though is still consistent with zero within 3σ. Its mass distribution has a high mean $\mu_{M,\rm GI}$ of $2.92^{+0.87}_{-1.42}$ ($18.5 M_\oplus$) and a high scatter $\sigma_{M,\rm GI}$ of $2.43^{+0.74}_{-0.05}$. Its period distribution has a break $P_{\rm break,GI}$ at a short orbital period of $3.7^{+1.1}_{-0.6}$, and a shallow slope past the break $\beta_{2,\rm GI}$ of $-0.44^{+0.18}_{-0.22}$.

The retrieved properties of the remaining mixtures in Model 6 take on values close to those of the analogous mixtures in other simpler models, with some minor perturbations. The gaseous rocky mixture remains the mixture with the highest fraction $Q_{\rm GR}$, coming at $0.41^{+0.09}_{-0.07}$. Due to the addition of the separate mass distribution of the gaseous icy mixture, the gaseous rocky mixture in Model Icy6 has a mass distribution with a lower scatter of $0.86^{+0.13}_{-0.10}$, lower by 3σ relative to Models 3 and Icy4b. Planets that formed gaseous with rocky cores are also slightly shifted toward shorter orbital periods, with a steeper slope past the break $\beta_{2,\rm GR}$ of $-1.17^{+0.20}_{-0.25}$. The remainder of the mass and period distribution parameters for this mixture are retrieved at values consistent within 1σ of their values in Models 3 and Icy4b. For the intrinsically icy mixture, the mass and period distribution properties are within 1σ of Models Icy3a, Icy4a, and Icy5, although with a reduced mixture fraction $Q_{\rm NI}$ of $0.24^{+0.08}_{-0.11}$. Aside from the mixture fraction, this intrinsically icy mixture has remarkably consistent retrieved parameters across all models that include it. Finally, the intrinsically rocky mixture has properties most similar to those of Models 3 and Icy4b, which did not include an intrinsically icy mixture. It has a mixture fraction $Q_{\rm NR}$ of $0.27^{+0.06}_{-0.06}$. The

mass and period distribution parameters are within 1σ of their values in Models 3 and Icy4b, except the mean of the mass distribution which is shifted toward a lower mean $\mu_{M,\mathrm{NR}}$ of $-0.80^{+0.51}_{-0.78}$ ($0.45M_\oplus$) compared to $-0.15^{+0.24}_{-0.31}$ ($0.86M_\oplus$) in Model 3.

4.4.2 Comparison of Inferred Underlying Planet Mass-Radius-Period Distributions

In order to illustrate the differences between the underlying planet populations inferred by fitting each of the models to the *Kepler* dataset, we present the 2D projections in radius-period space, shown in Figure 4.2, the 2D projection in mass-radius space, shown in Figure 4.3, as well as the 1D projection in radius space, shown in Figure 4.4. These figures show the corresponding 2D and 1D distributions inferred by each model, separated by their component subpopulations. For each of the plots in this section, we sample from the posterior predictive distribution. To do so, we marginalize over the posteriors by simulating a 1000-planet population using a posterior sample s of the population-level parameters θ, with the individual sample denoted by θ^s. We repeat this sampling S times, with $S = 500$ in this case, and average over these S posterior samples.

Compared to Model 3, Model Z lacks photoevaporation and evaporated cores, and fixes the mass-radius relation below the first mass break to an icy composition. In radius-period space, the intrinsically rocky subpopulation in Model Z looks similar to that of Model 3. The gaseous subpopulation similarly follows the distribution in Model 3, except it extends down to lower radii with the mass-radius relation transitioning to an icy composition. This leads to substantial overlap between the rocky and icy planets between $1.5R_\oplus$ and $2.0R_\oplus$, also seen in the mass-radius distribution and 1D radius distribution, which has the effect of washing out the radius gap seen more clearly in other models.

The intrinsically icy subpopulation, included in Models Icy3a, Icy4a, Icy5 and Icy6, largely overlaps with the gaseous subpopulations present in these models. Noteworthy dif-

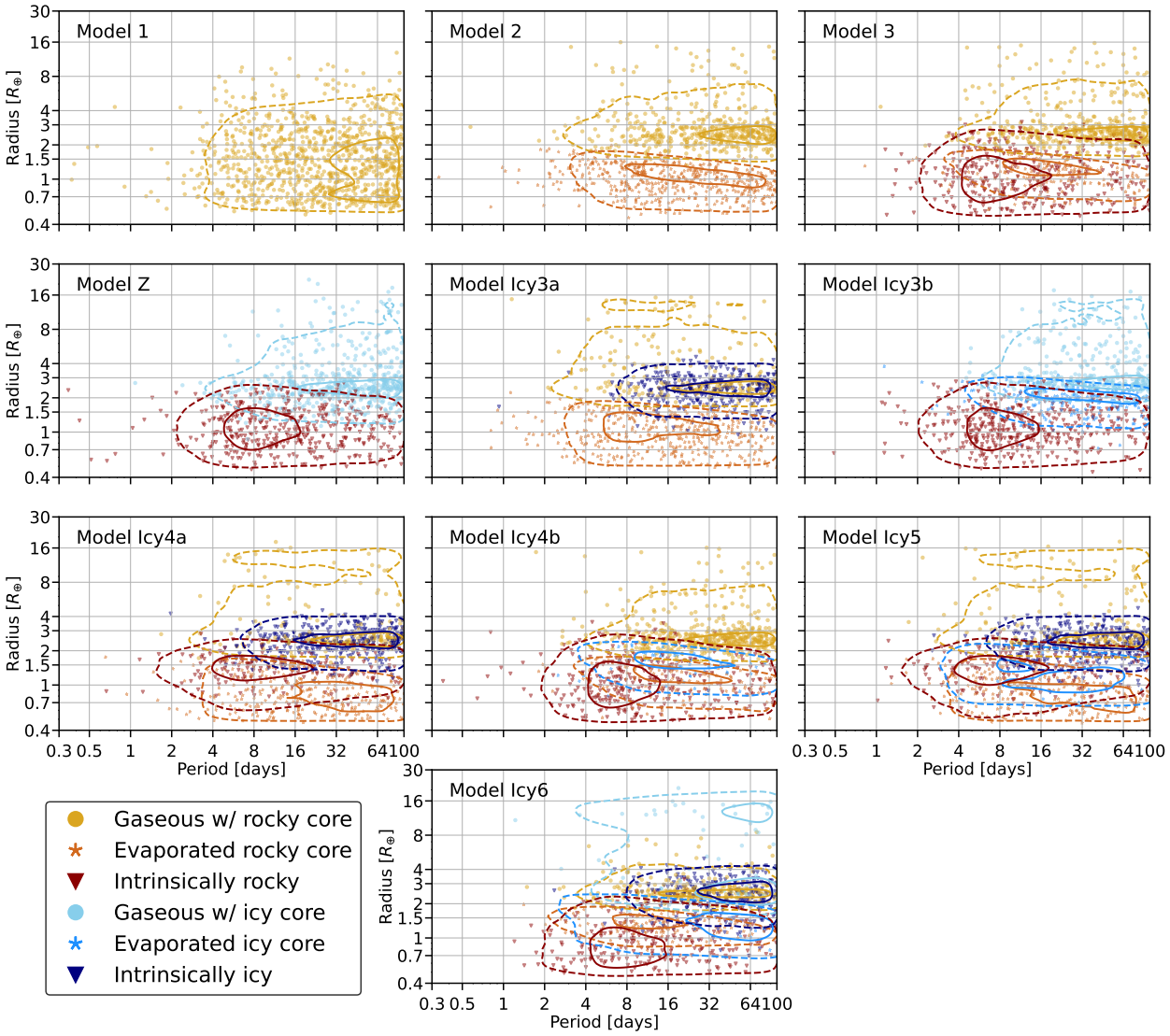

Figure 4.2: Projected 2D radius-period distributions for each of the ten models presented in this work. The contours show the underlying true distribution in radius-period space for each subpopulation within a model. Contours were generated by simulating planet populations from the posterior predictive distribution, then using a Gaussian kernel density estimation to estimate the 2D probability density function (PDF). The inner contour (solid line) contains 25% of planets belonging to a subpopulation, while the outer contour (dashed line) contains 90%. The scatter points represent one realization of a 1000-planet sample, marginalized over the posterior samples. Contours and points are colored according to their subpopulation, denoted by the shared legend in the bottom-left corner.

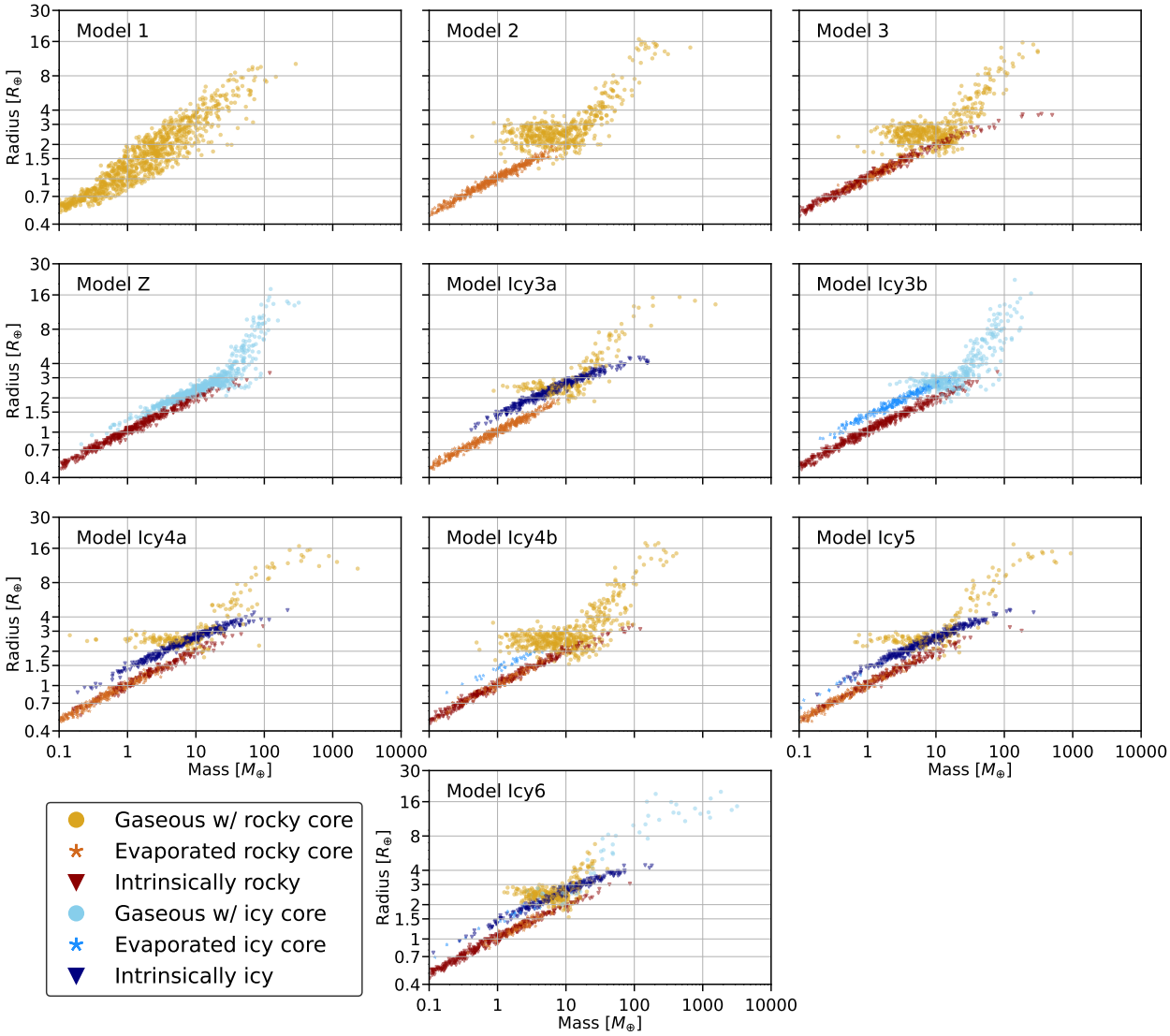

Figure 4.3: Projected 2D mass-radius distributions for each of the ten models presented in this work. The scatter points represent one realization of a 1000-planet sample, drawn from the posterior predictive distribution. The points are colored by which subpopulation they belong to, according to the shared legend in the bottom-left corner.

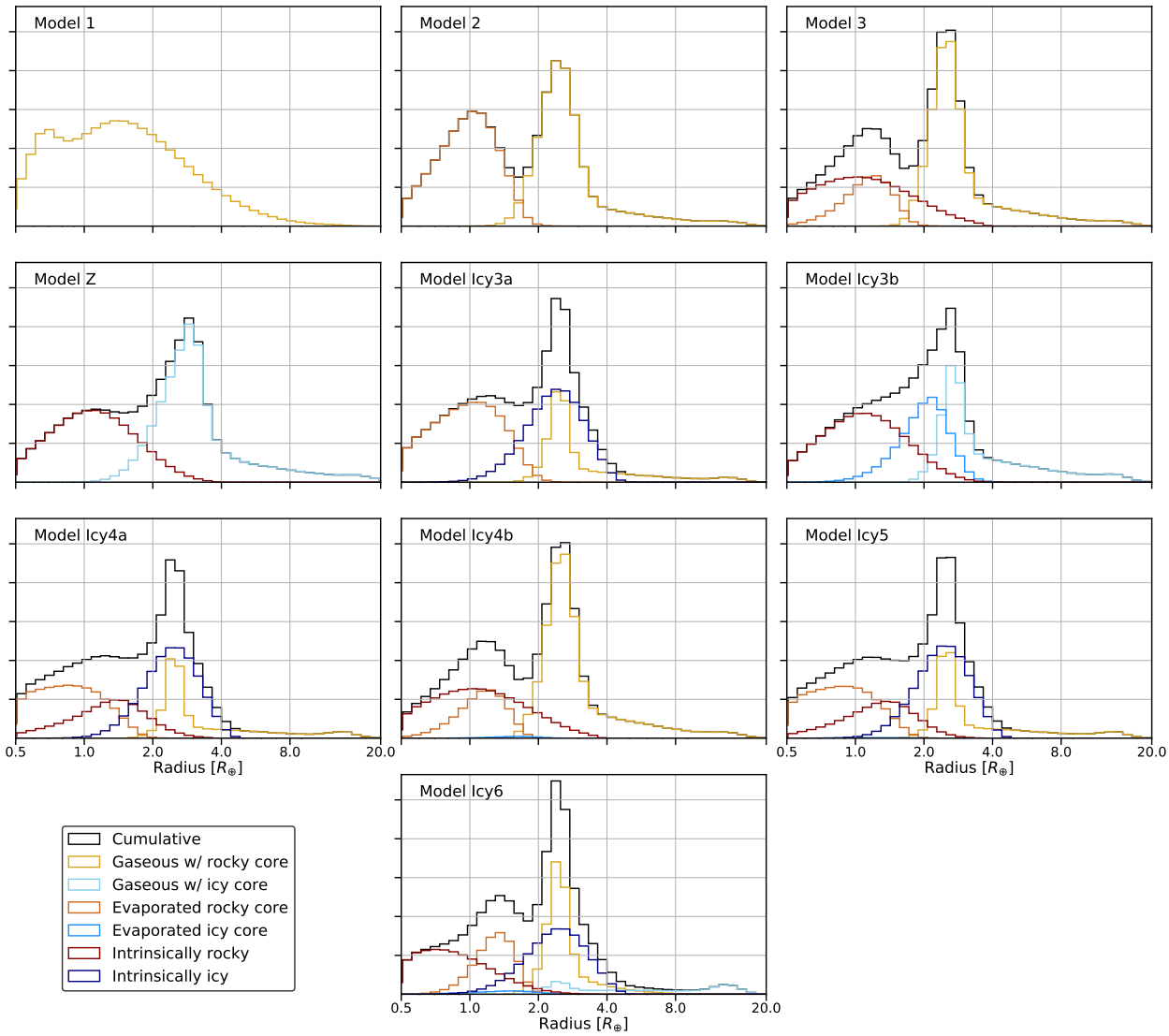

Figure 4.4: Projected 1D radius distributions for each of the ten models presented in this work. The solid black line shows the overall radius distribution for a given model, while the colored lines show the radius distributions of the component subpopulations, listed in the shared legend in the bottom-left corner. These distributions were generated by simulating planet radii from the posterior predictive distribution.

ferences exist, however, between these two compositional subpopulations. The intrinsically icy subpopulation extends to lower radii, with the 2σ contours in Figure 4.2 extending below $1.5R_\oplus$, whereas the gaseous subpopulation is mostly contained above $2.0R_\oplus$. The intrinsically icy subpopulation also doesn't extend to as short orbital periods as the gaseous subpopulation for Models Icy6 and Icy3a. This overlap between the intrinsically icy and gaseous subpopulations is apparent by the reduced numbers of gaseous planets retrieved when fitting models that include intrinsically icy planets compared to fits to models that omit them.

The properties of the evaporated icy-core subpopulation vary significantly depending on the model. The posterior fit to Model Icy3b has the highest fraction of these planets compared to other models, as adding in additional subpopulations in later models tends to reduce the amount of evaporated icy cores. Compared to the evaporated rocky-core subpopulation in Model 3, the evaporated icy-core subpopulation in Model Icy3b is shifted towards larger radii and longer orbital periods. These evaporated icy cores span the radius gap between $1.5R_\oplus$ and $2.0R_\oplus$, as well as above and below it. In Model Icy3b they overlap significantly in radius-period space with the gaseous subpopulation, and as a result the gaseous subpopulation has fewer planets at low masses compared to Model 3. This overlap also leads to a complete washing out of the radius gap in Model Icy3b, apparent in Figure 4.4, where the overall radius distribution monotonically decreases from its peak at around $3R_\oplus$ as you go towards smaller radii. In comparison to the fit to Model Icy3b, in the fits to Models Icy4b and Icy5, the evaporated icy planets more closely follow the evaporated rocky planets in radius-period space, except shifted towards higher radii. This is a result of the two subpopulations sharing the same mass and period distributions, as well as the same photoevaporation prescription. Including both evaporated subpopulations in this way, as in Model Icy4b and Icy5, leads to very few evaporated icy planets overall.

Separating the mass and period distributions of the planets that formed with a rocky core

and gaseous envelope from the planets that formed with an icy core and gaseous envelope, as in Model Icy6, significantly changes the distribution of both evaporated core subpopulations. As shown by the radius-period plot in Figure 4.2, the evaporated icy planets concentrate towards longer orbital periods, with 1σ contours between 30 and 100 days, whereas the evaporated rocky planets are shifted towards shorter orbital periods, with the bulk falling between 8 and 20 day periods. Additionally, the two subpopulations have a more similar distribution in radius space compared to when they share a mass and period distribution as in Models Icy4b and Icy5. The overall number of evaporated icy planets in Model Icy6 is still very low despite these changes.

Separating these two subpopulations that formed with a gaseous envelope in Model Icy6 also creates a distinct gaseous subpopulation that have icy cores rather than rocky cores. The gaseous subpopulation with rocky cores are compressed in radius and mass space, contained mostly within $2.0R_{\oplus}$ and $4.0R_{\oplus}$, below the first break in the mass-radius relation. By comparison the mass and radius distribution of the gaseous subpopulation with icy cores is broader, and as a result the gas giants above $8.0R_{\oplus}$ are entirely composed of gaseous planets with icy cores. The period distributions are also different, with the icy-core gaseous planets concentrated towards longer orbital periods compared to the rocky-core gaseous planets. The overall number of icy-core gaseous planets is much lower compared to the number of rocky-core gaseous planets. However, given the very low number (roughly between 10-40 planets in a 1000-planet simulated *Kepler*-like sample) of evaporated icy cores associated with this subpopulation, these gaseous planets need not physically have icy cores. Instead, the model could be using this population to better fit the most massive gaseous planets that have undergone runaway gas accretion.

Finally, the distributions in mass-radius-period space of the evaporated rocky planets compared to the intrinsically rocky planets seem to shift depending on the model. In Models 3, Icy4b, and Icy6, the evaporated rocky cores are found at higher radii and slightly

longer orbital periods compared to the intrinsically rocky planets, with most planets below $1.0R_\oplus$ belonging to the intrinsically rocky subpopulation. These evaporated planets share mass and period distributions with the gaseous planets, so if the gaseous planets are concentrated towards longer orbital periods and higher masses, this would reflect on the evaporated subpopulation. However, in Models Icy4a and Icy5, which include intrinsically icy planets alongside both subpopulations of rocky planets, but without an evaporated icy-core subpopulation that is independent in mass-period space from the evaporated rocky-core subpopulation, the distributions are significantly altered. In these models, the evaporated cores comprise the majority of the low radius planets below $1.0R_\oplus$, with the intrinsically rocky planets shifting towards larger radii. Additionally, the evaporated cores are found at even longer orbital periods than in the other models.

The shifts in the evaporated rocky-core and intrinsically rocky subpopulations are a consequence of the addition of the intrinsically icy subpopulation. The intrinsically icy subpopulation in Models Icy4a and Icy5 occupy a similar radius range, between 1.5-4.0$R_\oplus$, to the bulk of the planets belonging to the gaseous with rocky-core and evaporated rocky-core subpopulations in other models. In order to reduce the occurrence of the gaseous rocky mixture in this radius range (in compensation for the presence of the intrinsically icy subpopulation), the mass distribution of the gaseous rocky mixture in Models Icy4a and Icy5 is shifted towards lower masses, but with a higher scatter in order to still fit the gas giant regime. The intrinsically rocky population then shifts toward higher masses to compensate for the higher occurrence at lower masses of the evaporated rocky cores. When the formed-gaseous with icy-core mixture is given its own mass and period distribution in Model Icy6, this additional flexibility allows this mixture to fit the gas giants without necessitating shifts in the formed-gaseous with rocky-core and intrinsically rocky mixtures. These shifts reveal the degeneracies between these subpopulations, and the issue of labeling planets as belonging to one subpopulation or the other. This is further discussed in the next section and in Section

4.4.4.

4.4.3 Mixture Fractions and Degeneracies

The relative fractions of planets belonging to each subpopulation are loosely constrained and vary significantly between models. These fractions are shown for Model Icy6 in Figure 4.5. As reflected in the radius-period and mass-radius distributions shown in Section 4.4.2, the gaseous with rocky core, intrinsically rocky, and intrinsically icy subpopulations are the most numerous. Although there are wide uncertainties, these fractions each fall between 20 and 30% of the total underlying planet population, within the bounds we set in mass, radius, and period (see Section 4.3.1). Of the less numerous subpopulations, the evaporated rocky-core subpopulation contains slightly over 10%, the gaseous with icy-core subpopulation includes about 5%, and the evaporated icy-core subpopulation is the least numerous at a few percent. Compared to Model 3, which only has rocky-core compositional subpopulations, the fraction of intrinsically rocky and evaporated rocky planets are only slightly reduced. In contrast, the gaseous with rocky-core subpopulation is greatly reduced, as it has a large degree of overlap with the intrinsically icy subpopulation. These mixture fractions, however, are not tightly constrained and have substantial uncertainties. For example, the intrinsically icy subpopulation ranges from 13% to 32% within the 1-σ confidence interval obtained by sampling over the posteriors of the population-level parameters. The large error bars on the mixture fractions demonstrate the degeneracy between these subpopulations and reflect the overlap seen in the mass-radius and radius-period distributions.

To further show the degeneracy between these compositional subpopulations, we present the sub-population membership probabilities of each planet in the *Kepler* dataset fit to Model Icy6 (Figure 4.6). The retrieved subpopulation membership probabilities are derived from Equation 4.2, using samples from the retrieved population-level parameters and the retrieved true masses and radii. The left-hand ternary plots show the retrieved membership

probabilities divided by formation pathway: planets that formed with and retained a gaseous envelope, formed with but lost a gaseous envelope through photoevaporation, and formed without a gaseous envelope (intrinsically rocky/icy). The right-hand ternary plots show retrieved membership probabilities divided by present composition - gaseous, rocky, and icy. The ternary plots on the top have points colored by incident flux, whereas the ternary plots on the bottom have points colored by the inferred true planet radius (retrieved by the fit to the *Kepler* dataset).

In the formation pathway ternary plot, planets are broadly concentrated into two clusters: one corresponding to "not-evaporated" planets, and a second corresponding to "not-currently-gaseous" planets. The first cluster of planets, located along the bottom axis, has near-zero probability of belonging to evaporated subpopulations, with between 40% and 80% probability of belonging to currently gaseous subpopulations, and between 20% and 60% probability of belonging to the formed non-gaseous subpopulations. The other cluster is located along the left-hand axis, with near-zero probability of belonging to the currently gaseous subpopulations, between 25% and 100% probability of belonging to the formed non-gaseous subpopulations, and between 0% and 75% probability of belonging to the evaporated subpopulations. As expected, given the period-dependent and mass-dependent photoevaporation prescription incorporated in the model, the separation of these two clusters is dependent on the radii and incident flux of the planets. Less-irradiated and larger planets (on the larger end of the radius gap) tend to belong to the not-evaporated cluster, whereas highly-irradiated and smaller planets belong to the not-currently-gaseous cluster.

While the majority of planets in the formation pathway ternary plot are found in the two clusters described above, there is a substantial fraction (comprising about 20% of the *Kepler* sample) that have a non-negligible ($> 5\%$) probability of belonging to each of the three formation pathway categories. These planets tend to be around $2R_\oplus$ in size, with relatively high incident flux of at least $10^2 S_\oplus$. As seen in Figures 4.2 and 4.3, these planets are large

enough to be intrinsically icy or small gaseous planets, but with high enough incident flux to possibly be evaporated. These planets with substantial probability of belonging to each category do not have significantly higher radius or mass uncertainties than planets belonging to the two clusters. Rather, they belong to a region of mass-radius-period space where degeneracies between subpopulations is especially high, given our model parametrization.

Some features in the formation pathway ternary plot may reflect our choices of model parameterization, especially with regard to the formed non-gaseous subpopulations. Planets trend towards smaller radii as you go down the left-hand axis toward the corner with 100% probability of belonging to the formed non-gaseous subpopulations. This is due to the intrinsically rocky planets concentrating towards shorter orbital periods and smaller radii compared to the evaporated rocky planets in the fit to the *Kepler* data. As discussed in Section 4.4.2, this feature is dependent on the combination of subpopulations included in the model, and does not appear in every model. Additionally, there is a dearth of planets with less than 20% probability of belonging to the formed non-gaseous subpopulations; planets only have a low formed non-gaseous membership probability if their probability of belonging to the evaporated subpopulations is also near 0%. This is a consequence of the broad lognormal mass distributions of the intrinsically rocky and icy subpopulations; only the largest planets definitively do not belong to these formed non-gaseous subpopulations.

Turning now to the categorization of planets based on their inferred current composition (right-hand side of Figure 4.6), we also find planets are clustered with a clear separation in both planet radius and incident flux. The first cluster of planets, along the bottom axis, comprises planets with $< 5\%$ probability of belonging to the rocky subpopulations, and substantial probabilities of belonging to either the gaseous or icy subpopulations. The second cluster, towards the top corner, comprises planets with $< 5\%$ of belonging to the gaseous subpopulations, $> 80\%$ probability of belonging to the rocky subpopulations, and up to 20% probability of belonging to the icy subpopulations. This separation is highly radius

dependent, with planets in the first "not-rocky" cluster having radii $> 2R_\oplus$, and planets in the second "not-gaseous" cluster having radii $< 2R_\oplus$. The separation in incident flux is less sharp, but generally planets in the first cluster have lower incident fluxes compared to planets in the second cluster. As with the breakdown by formation, planets with non-negligible probability of belonging to each composition category (rocky, icy, and gaseous) tend to have radii close to $2R_\oplus$, with moderate incident fluxes (roughly between 10^2 and $10^3 S_\oplus$). While there are many planets with $> 90\%$ probability of belonging to the rocky subpopulations, and several with $> 90\%$ probability of belonging to the gaseous subpopulations, there are zero planets with similar high probability of belonging to the icy subpopulations. This is a consequence of the icy subpopulations occupying an intermediate position in mass-radius-period space compared to the other two compositions, with significant overlap, as seen in Figures 4.2 and 4.3.

Planets with mass measurements have additional information that may help constrain their formation and composition membership compared to planets that only have radius and period information and no mass measurement. However, this does not appear to be the case for formation membership, as there does not appear to be a trend in the location of planets in the formation ternary plot with whether or not the planet has a radial velocity mass measurement. Having mass measurements does appear to significantly constrain composition, as nearly all planets with mass measurements fall along or close to the three axes in this ternary plot. The lack of planets with mass measurements in the center of the ternary diagram, where all three compositions have non-negligible probabilities, indicates that the presence of a mass measurement combined with radius and period information is enough to rule out at least one composition for a given planet. This demonstrates that mass is a primary driver behind composition, whereas the breakdown of planets into currently gaseous, evaporated, and formed non-gaseous categories is more dependent on radius and period.

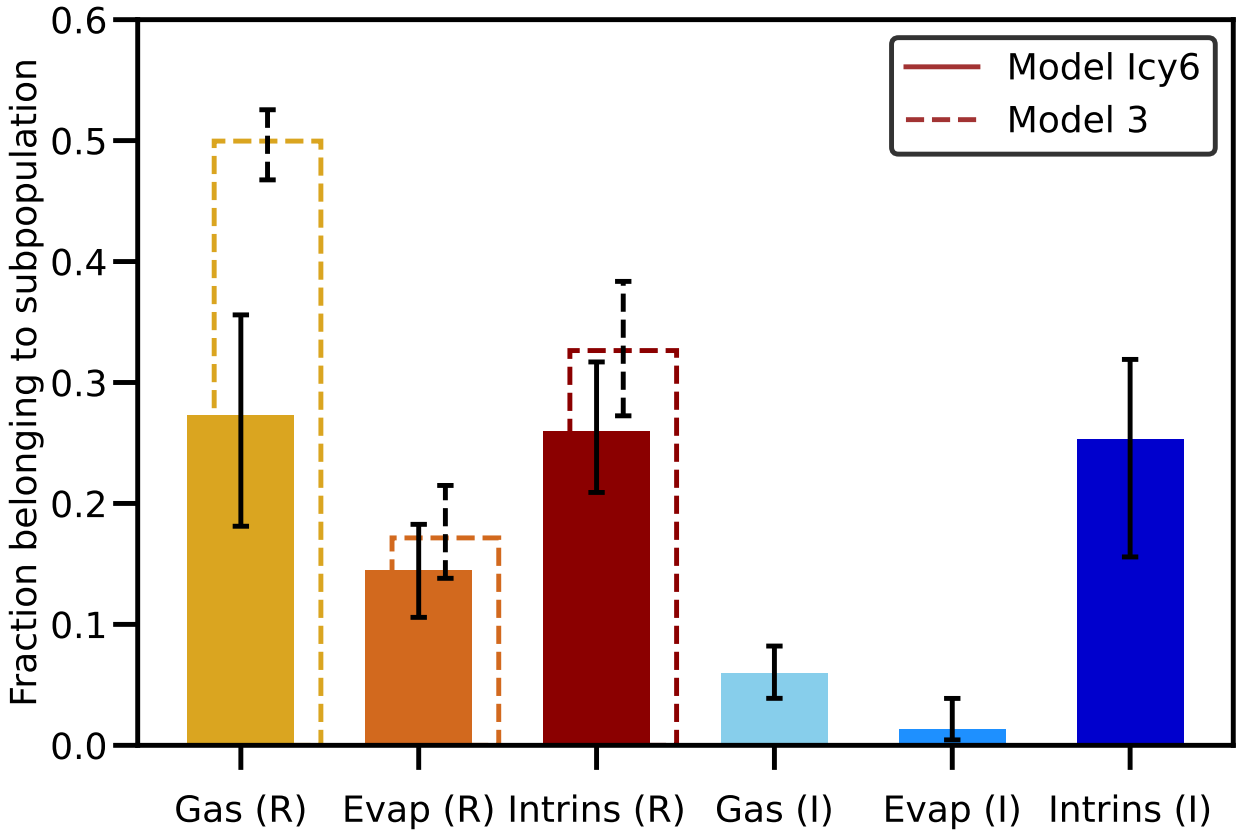

Figure 4.5: The fraction of planets belonging to each subpopulation in Model Icy6 (shown by the solid lines), as compared to Model 3 (shown by the dashed lines). The subpopulations from left to right are: gaseous with rocky core, evaporated rocky core, intrinsically rocky, gaseous with icy core, evaporated icy core, and intrinsically icy. These fractions were generated by forward modeling a sample of planets, and sampling over the posteriors. The height of the bar represents the median of the distribution obtained by sampling over the posteriors, while the error bars represent the 1-σ (15.9% and 84.1%) percentiles. The errors are not independent and are correlated with the error bars for the other subpopulations.

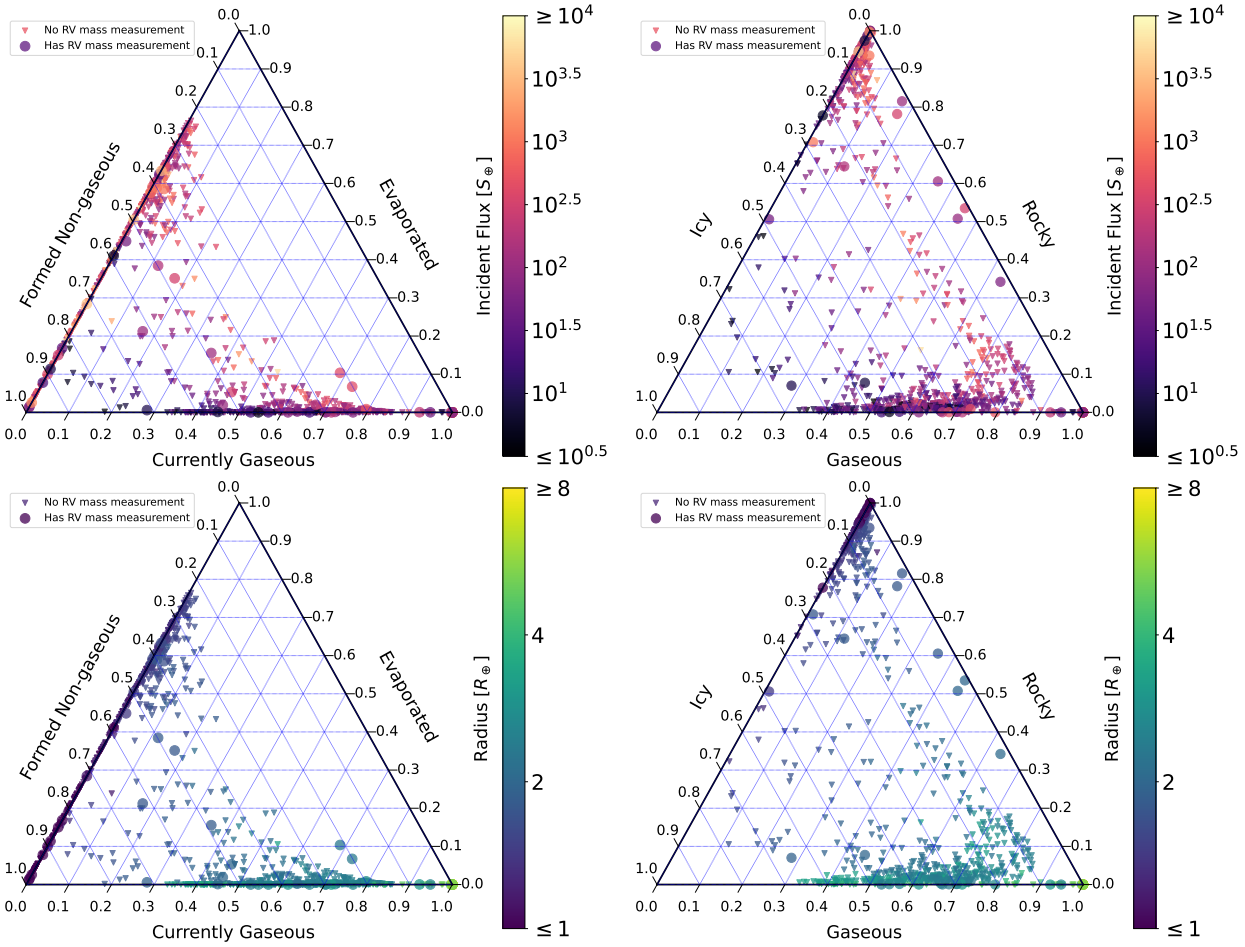

Figure 4.6: Ternary plot showing the distribution of retrieved subpopulation membership probabilities for planets in the *Kepler* sample using Model Icy6. The closer a point is to a corner, the higher the probability of that planet belonging to the corresponding category (comprised of two subpopulations). Triangles represent planets without RV mass measurements, whereas circles represent planets with RV mass measurements. The points are colored by their incident flux in the top two panels, and colored by radius in the bottom two panels. Left: The three membership probabilities represent the three formation scenarios: formed with and retained a gaseous envelope (currently gaseous), formed with but lost a gaseous envelope due to photoevaporation (evaporated), and formed without a gaseous envelope (formed not-gaseous). Each of these categories combines subpopulations of planets with icy cores and planets with rocky cores. Right: Membership probabilities represent the current composition of the planet: gaseous, icy, or rocky. Each of these three categories combines two different subpopulations in the model - gaseous planets with icy or rocky cores, and icy/rocky planets that either formed that way (intrinsic) or formed with and lost a gaseous envelope (evaporated).

4.4.4 The Fraction of Water Worlds

We find the fraction of icy-composition planets for models that include them to be dependent on the combination of icy and rocky subpopulations included in the model. Figure 4.7 shows the distribution of the inferred fraction of planets belonging to icy compositional subpopulations (for models that include them), marginalized over the model posteriors. On the low end, Model Icy4b has an icy composition fraction consistent with zero, with an upper limit around 10%. Model Z has the highest icy composition fraction, ranging from about 30% to 60%, peaking at around 45%. Models Icy3a, Icy4a, and Icy5, which all include intrinsically icy planets, have similar distributions, peaking between 35% and 40%, and spanning from 25% and 50%. Model Icy3b, which includes evaporated icy planets but omits intrinsically icy planets, has a lower fraction which peaks at 25% and ranges from 20% to 40%. Finally, Model Icy6 peaks at 25% but shows a wider range than other models, from 0% to 45%.

Overall, for models that include planets with icy compositions alongside planets with rocky compositions, we find an upper limit of 50% of planets belonging to these icy compositional subpopulations. This fraction is not tightly constrained for any model. When including all subpopulations considered in this paper in Model Icy6, the icy fraction becomes even more uncertain compared to most other less complex models, with considerable posterior mass approaching 0%. Thus, for models that include icy compositional subpopulations, our lower limit on the fraction of icy planets is 0%.

We now turn to the question of model selection. With the large range in the fraction of icy compositional planets retrieved by these models, which models are preferred and which of these estimates should we trust?

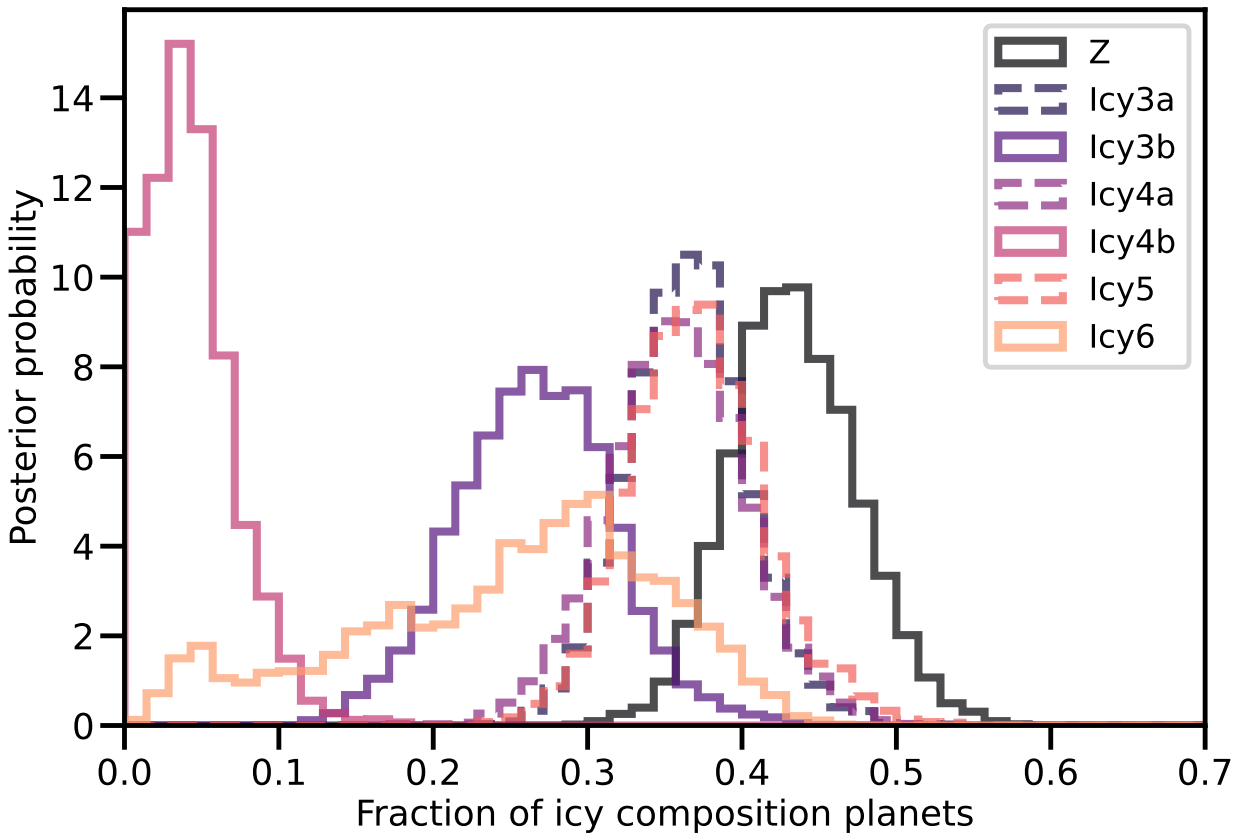

Figure 4.7: The distribution of the fraction of icy composition planets (as opposed to rocky or gaseous composition) for each model that includes icy composition planets. For Model Z, this only includes planets belonging to the gaseous with icy-cores subpopulation whose masses fall below the first mass break, below which the mass-radius relation is fixed to an icy (with no gas) composition. For all other models, this includes either the evaporated icy or intrinsically icy planets, or both when applicable (Models Icy5 and Icy6). These distributions were generated by repeatedly simulating 1000-planet samples, while simultaneously sampling from the posteriors. Sampling uncertainty and posterior uncertainty are thus both accounted for in these distributions.

4.5 Model Selection

4.5.1 K-fold cross-validation

With the introduction of six new models, each including one or several subpopulations of planets with icy compositions, we require an objective measure of model performance to assess whether or not the current data support the inclusion of such planets. As in NR20, we use 10-fold cross-validation to evaluate each model. Cross-validation estimates the predictive accuracy of a model when exposed to new data not used in fitting the model, by withholding data from the sample to use as a validation set. Cross-validation penalizes over-fitting, as models with high degrees of freedom will find spurious correlations in the data that are not present in the general population, and thus perform worse when exposed to new data. Given that we have 10 models ranging from one subpopulation to six subpopulations of planets, our higher complexity models are plausibly susceptible to over-fitting. Refer to NR20 and Vehtari et al. (2017) for more details on cross-validation; we briefly summarize the process below for one model.

We first divide the sample into ten subsets, labeled with the subscript j. For each of these subsets, we fit our model to the dataset with that subset excluded, resulting in a set of posteriors denoted θ^j (where θ represents the set of population-level parameters) which has S total posterior samples, with $\theta^{j,s}$ denoting an individual sample. Our aim is to calculate the expected log predictive density $\widehat{\text{elpd}}$, a measure of the predictive accuracy of the posterior predictive distribution when exposed to new data (the hat indicates that it is a computed estimate of the quantity). Specifically, we want to calculate the $\widehat{\text{elpd}}_k$ for each planet k in the dataset using the model fit j where that planet was excluded. To start, we draw Z samples of the planet's true mass and radius from the underlying mass and radius distribution for each mixture q and evaporated/non-evaporated subpopulation v using individual posterior samples $\theta^{j,s}$:

$$
\begin{aligned}
M_{\text{true},k}^{z,q,\theta^{j,s}} &\sim \text{p}(M|q,\theta^{j,s}) \\
R_{\text{true},k}^{z,q,v,\theta^{j,s}} &\sim \text{p}(R|M_{\text{true},k}^{z,q,\theta^{j,s}},q,v,\theta^{j,s})
\end{aligned}
\tag{4.12}
$$

where z indicates an individual mass or radius sample, and q indicates the mixture that this sample belongs to. We note that the true mass and radius samples above are distinct from the true mass and radius implemented as parameters in the model. Since we are calculating the $\widehat{\text{elpd}}_k$ of a planet using the model fit where that model was excluded, there are no true mass and radius samples of the corresponding planet from the model fit. We can then use these individual drawn samples of the true mass and radius of planet k to calculate its $\widehat{\text{elpd}}_k$:

$$
\begin{aligned}
\widehat{\text{elpd}}_k = \log \Bigg[& \frac{1}{SZ} \sum_{s=1}^{S} \sum_{q=0}^{N_{\text{mix}}-1} \sum_{z=1}^{Z} \text{p}(M_{\text{obs},k}|M_{\text{true},k}^{z,q,\theta^{j,s}}) \\
& \cdot \sum_{v=0}^{1} p(R_{\text{obs},k}|\text{det},R_{\text{true},k}^{z,q,v,\theta^{j,s}})\,\text{p}(\text{det}|R_{\text{true},k}^{z,q,v,\theta^{j,s}},P_k) \\
& \cdot \text{p}(v|M_{\text{true},k}^{z,q,\theta^{j,s}},P_k,\theta^{j,s},q)\,\text{p}(P_k|\theta^{j,s},q)\,\text{p}(q|\theta^{j,s}) \Bigg]
\end{aligned}
\tag{4.13}
$$

where we average over S posterior samples from the model fit and Z samples of the true mass and radius. The term $p(M_{\text{obs},k}|M_{\text{true},k}^{z,q,\theta^{j,s}})$ and the equivalent term with R_{obs} are calculated using a normal distribution with σ equal to the measurement error in the mass or radius. For planets without mass measurements, the corresponding term $p(M_{\text{obs},k}|M_{\text{true},k}^{z,q,\theta^{j,s}})$ is removed. The total expected log predictive density of the model, $\widehat{\text{elpd}}$, is then the sum of the individual $\widehat{\text{elpd}}_k$ from each planet:

$$
\widehat{\text{elpd}} = -2 \sum_{k=1}^{K} \widehat{\text{elpd}}_k \tag{4.14}
$$

where the sum is multiplied by -2 to put the output on the "deviance" scale, a convention

when calculating information criteria and other model selection metrics. On this scale, lower numbers (closer to zero) are preferred over higher numbers. Finally, the error in $\widehat{\text{elpd}}$ is given by the standard error:

$$\text{se}\left(\widehat{\text{elpd}}\right) = 2\sqrt{K\, V_{k=1}^{K} \widehat{\text{elpd}}_k} \tag{4.15}$$

where V indicates the variance, and we multiply by 2 again to conform to the deviance scale. Equation 4.13 differs from Equation 29 in NR20 in that we are more explicit about how $\widehat{\text{elpd}}_k$ is calculated in practice, and have corrected for the detection probability of each planet. This correction is necessary to properly calculate the probability of planets in the *Kepler* sample from the inferred distribution of detected planets, rather than the inferred distribution of the underlying population. This has the effect of weighting each planet's contribution to the total $\widehat{\text{elpd}}$ by its detection probability, making it more beneficial to more accurately predict planets with a high probability of being detected, i.e. larger planets on shorter orbital periods. The overall effect on the $\widehat{\text{elpd}}$ with this correction term is small, and does not significantly affect the results presented in NR20.

4.5.2 Cross-validation Results

We present our cross-validation results of the computed expected log predictive density $\widehat{\text{elpd}}$ of each model in Figure 4.8. As in NR20, we find that Models 1 and 2 perform significantly worse than Model 3, and this extends to every model introduced in this work - they all have a score higher than Models 1 and 2. The best performing models are Models Z, Icy3b and Icy6. The error bars, representing the standard error for the $\widehat{\text{elpd}}$ of a model given in Equation 4.15, indicate large discrepancies of $\widehat{\text{elpd}}_k$ between planets. As a result, these higher-performing models (Z, Icy3b, Icy6) are all consistent with Model 3 within error. Model Icy4b performs on par with Model 3, whereas Models Icy3a, Icy4a, and Icy5 all perform worse, but again within error of Model 3.

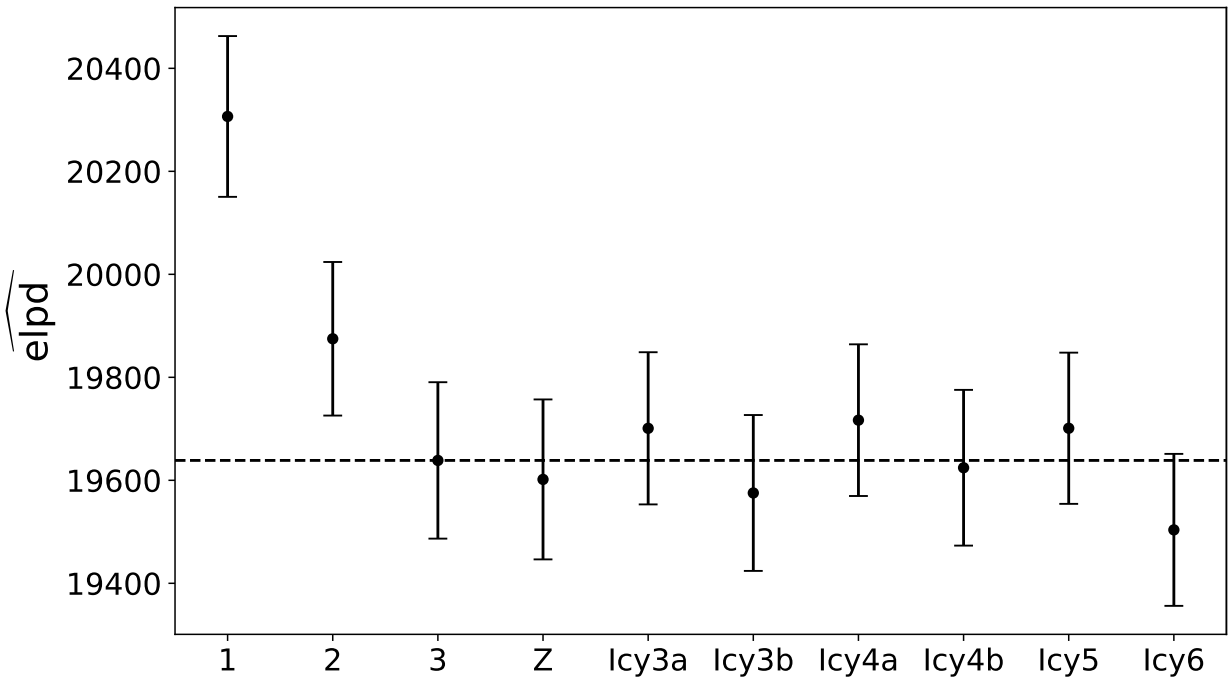

Figure 4.8: Cross-validation results for the *Kepler* dataset used in this paper. The points show the computed expected log predictive density ($\widehat{\text{elpd}}$), the metric calculated by the k-fold cross-validation from Equations (4.12) to (4.14), for each model. Higher scores indicate worse performance, and lower scores indicate better performance. The error bars show the standard error, given in Equation 4.15. The dashed horizontal line shows the $\widehat{\text{elpd}}$ of Model 3, for ease of comparison.

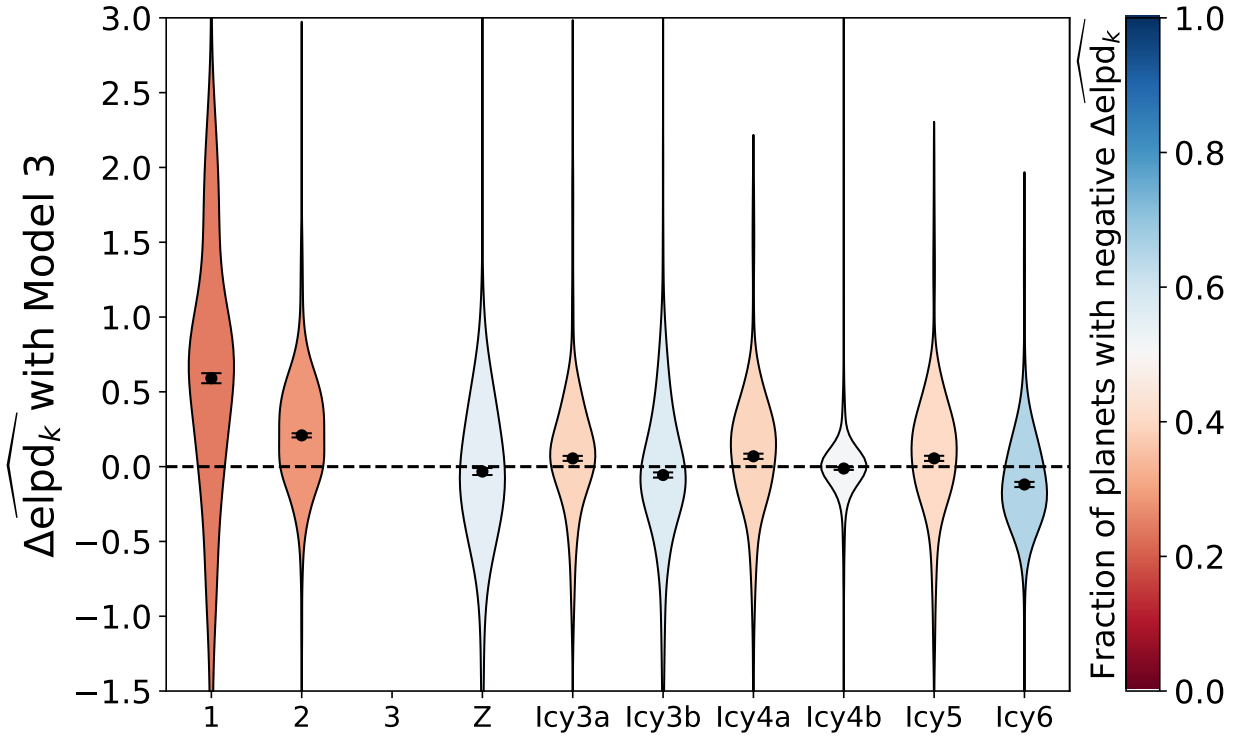

Figure 4.9: Further cross-validation results for the *Kepler* dataset used in this paper. Each model is represented by a violin and is compared to Model 3, the 'standard' model from NR20. The violin shows the distribution in the difference in computed expected log predictive density ($\widehat{\Delta\text{elpd}_k}$) between the model in question and Model 3 across all planets in the *Kepler* dataset. Negative numbers indicate the given model is performing better than Model 3. The points within each violin represent the difference in mean $\widehat{\text{elpd}_k}$, with the error bars representing the standard error in this difference. The violins are colored according to the fraction of planets which have a negative $\widehat{\Delta\text{elpd}_k}$, indicating the fraction of planets that the model predicts better than Model 3. These fractions are displayed along the top axis for each model.

We present a more detailed planet-by-planet picture of these cross-validation results in Figure 4.9, where we compare each model's performance to Model 3, the default model from NR20 with gaseous planets, evaporated rocky cores, and intrinsically rocky planets. Additionally, the violin plots in Figure 4.9 show the distribution of the difference in $\widehat{\text{elpd}_k}$, $\Delta\widehat{\text{elpd}}_k$, defined below:

$$\Delta\widehat{\text{elpd}}_k^{AB} = -2\left(\widehat{\text{elpd}}_k^{A} - \widehat{\text{elpd}}_k^{B}\right) \tag{4.16}$$

where in this case B represents Model 3, A represents the model we are comparing to Model 3, and we multiply the difference by -2 to follow convention. A negative $\Delta\widehat{\text{elpd}}_k$ indicates that the model in question predicts planet k better than Model 3. We show the fraction of planets that each model predicts better than Model 3 along the top axis of Figure 4.9.

Consistent with the results for the $\widehat{\text{elpd}}$ in Figure 4.8, we find that Model Icy6 performs the best, predicting 65% of planets better than Model 3. Models Z and Icy3b predict 55% and 57% of planets better than Model 3. On the other end of the spectrum, Models 1 and 2 only predict 24% and 28% of planets better than Model 3, although Model 1 has a wide distribution in Δelpd with long tails in both directions compared to the other models. Note that the error bars in this figure are much smaller than in Figure 4.8, as here we are taking the standard error in the difference in mean $\widehat{\text{elpd}_k}$ between two models, rather than the standard error in the $\widehat{\text{elpd}}$ (Equation 4.15). Correlations between the $\widehat{\text{elpd}_k}$ of a planet in one model and the $\widehat{\text{elpd}_k}$ in the second model lead to a tighter variance in this difference than in the raw $\widehat{\text{elpd}}$.

Looking at these results, we can see some interesting patterns. First, due to model degeneracies several more complex models reduce to lower complexity models and obtain similar cross-validation scores. For example, the posteriors for Model Icy5 show a low fraction of gaseous/evaporated icy planets, consistent with zero, as shown in Section 4.4.1. This effec-

tively reduces the model to Model Icy4a, with no gaseous/evaporated icy subpopulation, and the two models have very similar cross-validation scores and $\Delta\widehat{\text{elpd}}_k$ distributions. Similarly, Model Icy4b also has a low fraction of gaseous/evaporated icy planets, effectively reducing to Model 3, and the two have nearly identical cross-validation scores and consequently a tight $\Delta\widehat{\text{elpd}}_k$ distribution.

The best performing models (Z, Icy3b, and Icy6) all share the following features: they have an intrinsically rocky subpopulation, and include planets with icy compositions of some kind. The large difference between Model 2 and all of the more complex models (including Model Z) suggest this cross-validation score favors the inclusion of intrinsically rocky planets over evaporated cores. Indeed, Model Z, a model with no photoevaporation, performs on par or better than many of the more complex models with photoevaporation. This suggests that a model without photoevaporation is as consistent with the current dataset as models with photoevaporation, and higher-quality, higher-quantity data is necessary to resolve this difference. We also note that Icy6 performs significantly better than its predecessors, suggesting that adding a subpopulation of icy cores that formed with gaseous envelopes is more supported if those icy cores have a distinct mass and period distribution from the subpopulation of rocky cores that formed with gaseous envelopes. Alternatively, this could just be evidence that additional mass and period distribution flexibility is required to sufficiently fit the subpopulation of gaseous planets that have undergone runaway gas accretion, as discussed in Section 4.4.2. Models Icy4b and Icy5, which include both subpopulations of gaseous planets but with common mass and period distributions, perform on par or worse than Model 3, and significantly worse than Model Icy6.

Based on these cross-validation results, our main takeaway is not that Models Z, Icy3b and Icy6 are preferred over the others. Rather, given the large error bars on the $\widehat{\text{elpd}}$, all of these models (with the possible exceptions of Models 1 and 2) are broadly consistent with the data, and there is not a single definitive model that performs best. Models that include

photoevaporation but not icy compositional subpopulations, or models that include icy compositional subpopulations but not photoevaporation, or models that include both, perform comparably. With the mass-radius-period mixture models that we have formulated, the current *Kepler* dataset does not have sufficient information to strongly distinguish between models that include different combinations of these compositional and formational subpopulations. We explore the broader nature of the cross-validation method and the possibility of distinguishing between these models in the following subsections.

4.5.3 Simulated Catalogs for Model Selection

In order to better interpret and assess these cross-validation results, we perform cross-validation on four sets of simulated planet catalogs. Each set differs in the properties of the planet catalogs — the total number of planets, number of planets with mass measurements, and mass and radius uncertainties — as well as the true parameters of the models used to generate the catalogs. For each of these four sets, we generate simulated planet catalogs from the following four models: Model 1 (as the simplest model), Model 3 (as the preferred model from NR20), Model Icy3b (as a model with icy planets to compare to Model 3), and Model Icy6 (as the most complex model). Then, with each of these four simulated catalogs, we perform cross-validation using each of the ten models presented in the paper. With these cross-validation simulations, we remove model misspecification as a confounding factor, since one of the models is a perfect parametrization of the simulated planet catalog that we use to fit the models.

The first set of simulated planet catalogs aims to answer the question: if one of our models was a perfect description of the true planet distribution, would we be able to distinguish this true "generative" model from the other models, using a planet catalog that is similar to *Kepler* (both in terms of the quantity and quality of the dataset, and the properties of those planets)? To generate this set, we use the median model posteriors from the runs presented

in this paper, reported in Table 4.1, as the true model parameters for the generative model. We generate the same total number of planets (1130), and keep the same number of planets with mass measurements (68). For mass and radius uncertainties, we randomly sample from the relative uncertainties of *Kepler* planets, weighted towards planets similar to a simulated planet in mass-radius-period space. For a given simulated planet, we calculate the Euclidean distance in log-radius−log-period space between that planet and each planet in the *Kepler* sample, where both axes are normalized to their corresponding range of values. We then select a random planet in the *Kepler* sample, where the probability for a given planet is the inverse of that distance, normalized so that the probabilities sum to 1. The simulated planet is then given the same relative radius uncertainty as the chosen *Kepler* planet. A similar process is used to choose the mass uncertainty, except in log-mass−log-period space, and only using planets in the *Kepler* sample with mass measurements.

The second set aims to answer the question: If the data is similar to *Kepler* in quantity and quality, but the population-level parameters of the generative model are chosen to accentuate the differences between subpopulations within a model, and between models themselves, would we be able to distinguish the generative model from the others? For this set, we use the same total number of planets, number of planets with mass measurements, and method for generating uncertainties as in the first set. However, for the true parameters for the generative models, we start with the *Kepler* model posteriors as a base and modify them to make each subpopulation within a model more distinct from each other. For example, we set the mixture fractions to be more evenly distributed, and change the mass and period distribution parameters so that the differences between subpopulations is greater. This enhanced separation between subpopulations within a model then translates to greater differences between models with different numbers of mixtures.

The third set aims to answer the question: If we had a higher-quality dataset from a future survey such as *PLATO*, would that improve our ability to distinguish between

these models? For this set, we use the same true model parameters as the first set of simulations. However, we increase the total number of planets to 4000, the minimum science goal set by the *PLATO* team (PLATO Study Team, 2017). We also increase the number of mass measurements to 400, in line with the target number of radial velocity followup measurements. We simulate higher precision mass and radius measurements to meet the target of a median radius uncertainty of 3% and a median mass uncertainty of 10% by scaling down the *Kepler* uncertainties sampled in the first two simulations by a factor of 1.6 for radius, and 2.6 for mass. We still use the *Kepler* detection efficiency function, as simulating a *PLATO* detection efficiency is beyond the scope of this paper.

Finally, the fourth set of simulations aims for a near-perfect dataset, to see what the best case scenario of the model selection performance would be. While we still limit the number of planets to 4000 for computational reasons, we include mass measurements for each of these 4000 planets, and give each planet a radius and mass uncertainty of 1%. Additionally, we use the modified input parameters as in the second set of simulations, to create greater separation between subpopulations.

4.5.4 Simulated Catalog Results

Overall, each set of simulations that we performed failed to consistently identify the generative model as the preferred model. K-fold cross-validation was generally successful in disfavoring models with lower complexity than the generative model. However, models with higher complexity than the generative model performed on par with the generative model.

The cross-validation method used here remains prone to overfitting, as models with higher degrees of freedom were able to emulate lower complexity models without any penalty to their $\Delta\widehat{\text{elpd}}_k$ score. This tendency to overfit did not diminish by upgrading to an improved dataset from *PLATO*. Modifying the true parameters as in the second set of simulations did improve the results for Model 1 simulated catalogs, but not for others. Even an idealized

PLATO dataset with complete mass information and extremely low errors did not resolve these issues.

The cross-validation simulation results for the first and third set of simulations (for *Kepler* and *PLATO* planet catalogs, respectively) are shown in Figure 4.10, where we show the $\Delta\widehat{\text{elpd}}_k$ distribution as in Figure 4.9. For the first set of simulations, which emulates the *Kepler* dataset used in this paper and uses the retrieved median values of the population-level parameters recorded in Table 4.1 as the input values for the generative model, the cross-validation fails to consistently identify the generative model. This is shown in the left-hand side of the violin plots in Figure 4.10. With the simulated Model 1 catalog, higher complexity models with three or more subpopulations perform slightly worse than Model 1, but Model 2 performs significantly better and Models 3 and Icy3a perform about the same as Model 1. With the simulated Model Icy3b and Model 3 catalogs, the lower complexity Models 1 and 2 perform significantly worse, but the higher complexity models with four or more subpopulations perform on par with the generative model. Finally, for the simulated Model Icy6 catalog, there is no clear pattern and most models perform on par or slightly better than Model Icy6.

Simulations emulating *PLATO*, shown by the right-hand side of the violin plots in Figure 4.10, do not fare much better when considering the $\Delta\widehat{\text{elpd}}_k$ on a planet-by-planet basis. The results for the Model 1 simulated catalog are about the same as the *Kepler* simulations, except higher complexity models perform slightly better in comparison, on par with the generative model. Similarly, for the Model 3 simulated catalog, the results are similar but with the higher complexity models performing better than the generative model, rather than performing similarly. The results for the Model Icy3b simulated catalog are about the same as the results for the corresponding *Kepler* simulated catalog. Some lower complexity models have $\Delta\widehat{\text{elpd}}_k$ distributions shifted lower compared to the *Kepler* simulations, indicating better comparative performance for the generative model. Finally, Model Icy6 is the only

set of simulations with marked improvement, with the generative model preferred over all other models, albeit slightly in some cases.

While we have been primarily focused on the planet-wise $\Delta\widehat{\text{elpd}}_k$ when presenting these simulations, we do note improvements in the summed $\widehat{\text{elpd}}$ with this *PLATO* dataset. Even if the distributions in the $\Delta\widehat{\text{elpd}}_k$ are the same, since the sample size is roughly four times as large, the $\widehat{\text{elpd}}$ will have larger differences between models. Additionally, the relative error on this total would be lower by a factor of $\sqrt{N}$, or in this case by a factor of 2. This difference is enough to select the generative model in some cases where it would not be enough for a *Kepler*-like catalog. However, a similar $\Delta\widehat{\text{elpd}}_k$ distribution does indicate that a larger dataset does not on average result in superior predictions of planet properties.

Modifying the input parameters to create a larger separation between subpopulations compared to the retrieved parameters from the model fits to the real *Kepler* data, as in the second set of simulations, leads to significant improvement in correctly selecting low complexity generative models. For higher complexity generative models, the results are comparable to the first set of simulations, with no improvement. The fourth set of simulations, which uses idealized *PLATO* planet catalogs, showed improvement over the realistic *PLATO* dataset shown in Figure 4.10 in some cases, and regressed performance in others. The main improvements are that the lower complexity models performed even worse when the generative model was of high complexity. Conversely, the higher complexity models had distributions shifted toward higher $\Delta\widehat{\text{elpd}}_k$ across all simulated catalogs, leading to worse ability to rule out higher complexity models.

Our simulations provide valuable context to inform the interpretation of the model selection results presented in Section 4.5.2. The simulations confirm the inherent difficult in performing model selection with mixture models. We discuss this further in Section 4.6.1.

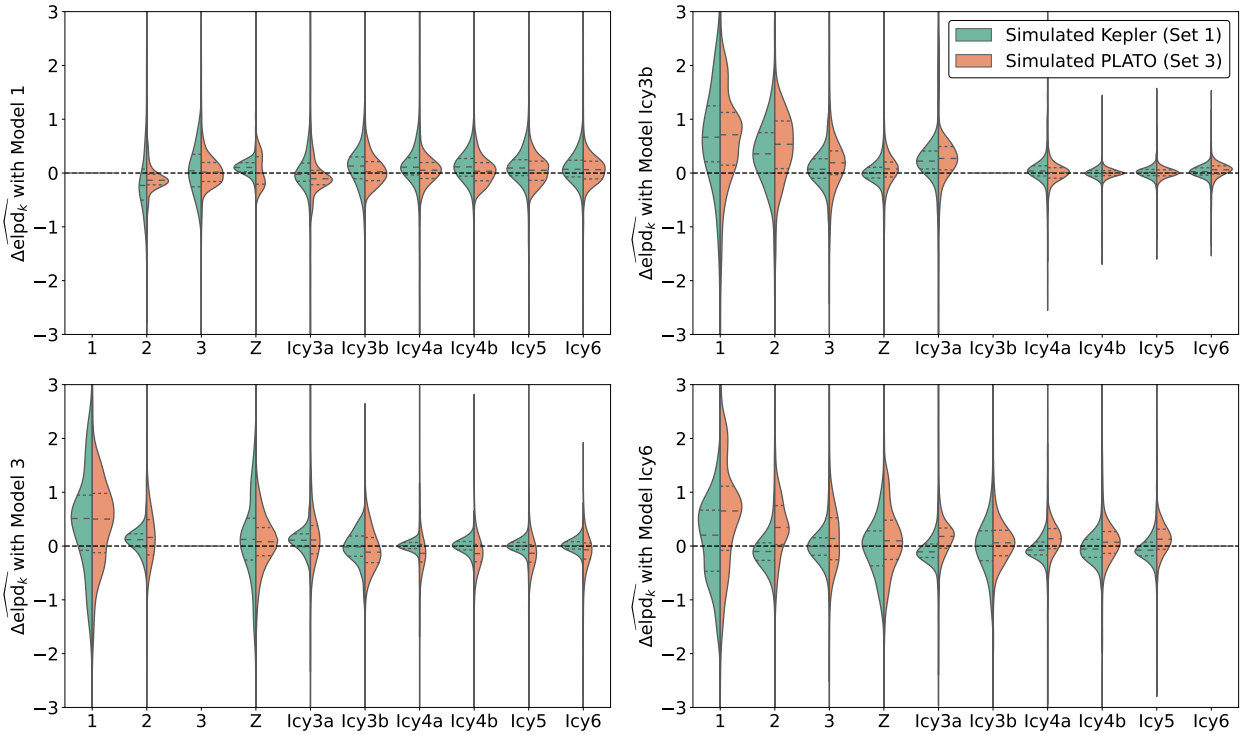

Figure 4.10: Cross-validation results for *Kepler* and *PLATO* simulated datasets. These are the first and third sets of simulations out of the four described in Section 4.5.3. Each panel shows a different simulated catalog, generated by a different model: Model 1 (top left), Model 3 (bottom left), Model Icy3b (top right), Model Icy6 (bottom right). Within each panel we compare the cross-validation results for each model compared to the model that was used to generate the data. The results for the *Kepler* datasets are represented by the left half of the curve in the violin plots (in teal), and the *PLATO* datasets are represented by the right half of the violin (in orange). Negative numbers indicate better performance relative to the generative model, and positive numbers indicate worse performance, measured by the difference in the computed expected log predictive density, $\widehat{\Delta\text{elpd}}_k$. The dashed lines within each violin represent the median, and 15.9% and 84.1% quantiles of the distributions.

4.6 Discussion

4.6.1 Difficulties with Model Selection

We presented model selection results using k-fold cross validation in section 4.5.2. However, simulations presented in section 4.5.3 showed that k-fold cross validation was unable to consistently select the generative model as the preferred model. As a result, our model selection results are inconclusive.

A comprehensive and systematic study of the performance of k-fold cross validation as applied to Growth Mixture Modeling was performed by He & Fan (2019). While there are several notable differences between the specific models they studied and the models presented in this paper, the results are applicable to mixture models in general. They examined five factors and their effects on the ability of k-fold cross validation to successfully select the number of classes (i.e. mixtures) used to generate the data: mixing ratio (how evenly the data is split between the different classes), the size of the dataset, the number of folds, model parametrization, and class separation (how similar or dissimilar the classes are). For each permutation of the listed five factors, they generated 100 datasets using a 3-class model, and subsequently fit these datasets to models ranging from 1 to 5 classes. They also investigated four different methods of selecting the preferred model. The fraction of replications which correctly identified the generative 3-class model as the preferred model demonstrates how well the k-fold cross validation had performed.

For nearly all combinations of the five factors, the results were consistently poor. The 3-class model was only selected between roughly 20 and 30% of the time. Depending on the method used to select the preferred model, models with lower or higher complexity were predominantly chosen. Consistent with our own simulations, increasing the sample size (in their case from 500 to 2000) had a small but relatively insignificant effect on the performance. The only conditions where the model selection performed relatively well was the case of high

class separation, reaching up to 78% of replications preferring the generative model. In the context of our models, class separation is difficult to define. Between gaseous and evaporated rocky planets, for instance, the two share a mass and period distribution (no class separation), but are separated in mass-radius space as a function of the photoevaporation timescale, which strongly depends on mass and period (high separation). As shown by the ternary diagrams in Figure 4.6, it is difficult to distinguish between evaporated or intrinsically rocky planets, which indicates low class separation. Rocky and icy planets may appear similar in mass-period space, but are cleanly separated in mass-radius space.

There are numerous aspects of our models that distinguish them from the work of He & Fan (2019), preventing a direct comparison. We are working with three-dimensional data with observational uncertainties, which is incomplete due to not detecting or lacking mass measurements for a subset of planets. Standard mixture models like those presented in He & Fan (2019) have mixtures that are identical in form and parametrization. While our mixtures all use a log-normal parametrization for the mass distribution and a broken power-law for the period distribution, the mass-radius relation takes a different form depending on the composition and the photoevaporation timescale. However, these differences are additional complexities compared to their simulations, and if the k-fold cross validation fails in even the simpler case, it is unlikely to be effective here.

While He & Fan (2019) focused on k-fold cross validation, there is no clear alternative that would be effective as a method of model selection for mixture models. Likelihood ratio test (LRT) methods cannot be applied consistently to our suite of models, as they rely on comparing two models where one model is a subset of the other. While this is true for certain pairs of models in our work, other pairs cannot be compared this way. As an example, Model Z, which has no photoevaporation, is not contained within another model nor is any other model a subset of Model Z.

Another commonly used class of model selection techniques are various information crite-

ria, such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC) among others. Usami (2014) performed a similar simulation study to assess the performance of various information criteria in estimating the number of classes in growth mixture models. No one criterion showed consistent performance across the various factors they tested. As shown in the k-fold cross validation study by He & Fan (2019), they found a strong dependence of the performance on class separation. Additionally, model misspecification as well as model complexity, both a concern for our models, were shown to significantly negatively impact performance.

Altogether, model selection applied to finite mixture models is a problem with no clear solution. This has been shown in various simulation studies in other fields, as well as in the simulations presented in this paper. The specifics of identifying compositional subpopulations in exoplanet transit survey data — three-dimensional data, a hierarchical model with observational uncertainties, non-uniform mixtures, model misspecification — increase the severity of the challenges. With these challenges in mind, care must be taken when first formulating mixture models, implementing physical principles and constraints wherever possible. Model selection techniques should be considered one aspect of evaluating models, rather than the end-all and be-all, and validated with simulations where possible. As these simulations have brought to light, data-driven models can easily lead to overfitting and over-interpretation of results, even when physical principles are implemented.

4.6.2 Do Water Worlds exist?

Our paper aims to assess the claims of the existence of water worlds in the sample of *Kepler* exoplanets, as put forth by Zeng et al. (2019) and others. In order to do so, we compile a range of models that either incorporated icy composition planets alongside rocky composition planets, without invoking photoevaporation (Model Z), or included planets with icy compositions within models that include photoevaporation (Models Icy3a - Icy6).

While we cannot definitively prove the existence of water worlds, we find that both models that include water worlds and those that do not are valid fits to the *Kepler* sample. Our model selection results, while subject to the caveats discussed in Sections 4.5.3 and 4.6.1, suggest that Model 1 and 2 are disfavored. The remainder of the models, however, are all within 1σ of each other in terms of their expected log predictive density, with no clear best model. This suggest that models without water worlds (Model 3), models with water worlds and no photoevaporation (Model Z), and models with water worlds and photoevaporation (Models Icy3a through Icy6) are all valid fits to the data. In fact, if you select the best model through Method 2 of He & Fan (2019), discussed in Section 4.6.1, where you choose the model with the fewest parameters within 1σ of the model with the highest score, Model Z would be the selected model, as it has the fewest parameters among comparable models.

With the validity of both models that include and do not include water worlds, we assessed the constraint on the fraction of icy composition planets in Section 4.4.4. We found a wide range across all models, with a general upper limit of 50%, and a lower limit of 0%. Including all compositional and formational subpopulations of planets in Model Icy6 resulted in a broad posterior distribution of the fraction of icy planets, within the limits of 0% to 50% and peaking at roughly 25%.

The wide uncertainty in the fraction of icy composition planets is reflected in the large degree of degeneracy between planet subpopulations, as demonstrated by the ternary diagrams in Figure 4.6. While there is a high density of planets that are definitively rocky, there is significant overlap between planets that could be rocky or gaseous, and an even larger overlap between planets that could be icy or gaseous. This overlap is demonstrated in the radius-period and mass-radius plots of Figures 4.2 and 4.3 as well. Intrinsically icy planets and gaseous planets below $4R_\oplus$ occupy similar territory in mass-radius-period space. Distinguishing between these two subpopulations for planets between 2 and $4R_\oplus$ is difficult.

This difficulty is primarily due to three different facets: the dataset, the model, and

inherent degeneracies in composition. The majority of planets in the dataset lack mass measurements, and only contain radius and period information. With the high degree of overlap in radius-period space between intrinsically icy planets and gaseous planets, these planets are extra difficult to identify as belonging to one subpopulation over the other. For planets with mass measurements, the uncertainties on the mass and radius measurements are high enough that this ambiguity in the identification of an individual planet's composition still persists. Even for planets with perfect mass and radius information, our model parametrization and inherent degeneracies will make this identification difficult.

In our model parametrization, we have abstracted the complexity inherent to planet compositions by choosing three broad categories of composition: rocky, icy, and gaseous, choosing a mean mass-radius relation for each, and adding intrinsic scatter to simulate compositional diversity within these subpopulations. The mass-radius relation for the gaseous subpopulation is retrieved in the model rather than fixed to a theoretical curve as it is for the rocky and icy subpopulations. As a result, there is significant overlap in mass-radius space between the icy and gaseous subpopulations. While overlap between these subpopulations is expected, the retrieved scatter in the gaseous mass-radius relation is likely higher than physically expected. The envelope mass fraction and core composition is not directly modeled in the gaseous subpopulation, and as a result it is possible to have gaseous planets in the model with higher densities than pure ice planets. Furthermore, each subpopulation uses the same functional form for the mass and period distributions. These distributions also tend to be broad, with the consequence that it is difficult to distinguish between these subpopulations by mass and period alone.

Regardless of our parametrization, there is also the fact that overlap between gaseous and icy planets is inherent to some regions of parameter space, leading to degeneracies in composition when assessing any individual mass-radius-period point. As a result, even with a perfect model and a perfect dataset, some planets will necessarily have substantial proba-

bilities of belonging to either subpopulation when considering mass-radius-period alone. The question of compositional categorization is a probabilistic one, given the broad distributions of these subpopulations in mass-radius-period space.

One possible solution to break the degeneracy (as much as we can) is to increase the amount of physical constraints and information into the model, by shifting the framework to model composition directly rather than purely dealing with the abstractions of mass and radius. With core composition, core mass, and envelope mass fraction as fundamental parameters, we avoid the situation of gaseous planets with higher density than rocky or icy planets, although some overlap with icy planets will persist. The three compositional subpopulations will have greater separation in mass-radius-period space. With this greater separation, we may be able to better constrain the underlying mass and period distributions of these compositions. Constraining the core mass and envelope mass fraction distributions will yield some important insights on their own as well. Incorporating a more sophisticated mass loss prescription may help break these compositional degeneracies, as low-density planets with short mass loss timescales present a possible region of parameter space to identify water worlds.

4.6.3 Model Caveats

As mentioned in the previous section, model misspecification is a significant limiting factor in our ability to perform model selection on this collection of models. Metrics such as k-fold cross-validation have a tendency to over-select high complexity models, and this problem is exacerbated by model misspecification. Given the nature of our models as three-dimensional mixture models with up to six subpopulations of planets, the potential for model misspecification is high.

For each subpopulation, we assume the mass distribution to be characterized by a log-normal distribution, and the period distribution to be characterized by a broken power-law.

While these were chosen based on their ability to fit the data better than several other alternatives, such as a normal distribution for the mass distribution, and a power-law for the period distribution, there are many possibilities that we did not explore. With the log-normal only having two free parameters, and the broken power-law having only three free parameters, there is the potential for additional complexity in these distributions that these parameterizations cannot account for. For example, due to the log-normal distribution having long, symmetric tails in log-space, the intrinsically rocky subpopulations extend to higher masses than is physically likely. An exponential cutoff might be a feature of the mass distribution that could limit the masses of intrinsically rocky planets, as above a certain mass these planets will have accreted significant hydrogen-helium gaseous envelopes. Our assumption of a period distribution independent from mass and radius (aside from photoevaporation) may break down if we extended our period range beyond 100 days.

Additionally, in our models there is the added assumption that each mixture has a mass and period distribution that follows the same form. It could be the case that a lognormal mass distribution is a decent approximation for one subpopulation, but inadequately describes another subpopulation. The possibility of different mixture subpopulations being described by distinct distributions compounds the potential for model misspecification.

Our models assume three distinct compositions for exoplanets: rocky, icy, or gaseous. Of course, compositions are better described by a smooth spectrum rather than discrete options. In order to account for this spectrum, we incorporate a scatter in the mass-radius relation, so that a given mass planet for a given composition can have a range of radii. This is a convenient abstraction, but a better way to incorporate the complexities of composition would be to directly model the composition of exoplanets, rather than the mass and radius.

The prescription we use to incorporate photoevaporation into our models, given by Equations (4.8) and (4.9), is another potential source of model misspecification. This photoevaporation timescale depends on several factors, including the typical age of a star, the XUV

flux, and the mass-loss efficiency, for which we choose nominal values but recognize them to be highly uncertain. To compensate for this uncertainty, we include a free parameter α to scale the mass-loss timescale. However, this parameter is strongly influenced by the prior, and takes disparate values depending on the model, ranging from 0.3 to 8.5. This indicates that this parameter acting more to scale the number of evaporated planets rather than modeling a specific physical process. Furthermore, there are additional sources of mass loss besides hydrodynamic, XUV-driven mass loss such as core-powered mass loss that we are not accounting for here.

Finally, as discussed in NR20, the mass-radius-period distribution of exoplanets is just one 3D projection of the many-dimensional distribution of exoplanets. Exoplanet occurrence rates can also depend on the multiplicity of planetary systems, host star spectral type and metallicity, and the galactic location, among other star-planet system parameters. These additional dimensions will tend to smear out these mass-radius-period distributions even more, adding more sources of model misspecification.

While there remain many avenues for improvement, this work has made several important steps forward compared to the models in NR20. We have explored a much wider range of models, considering models with up to six subpopulations of exoplanets. We have included subpopulations with icy-core compositions alongside subpopulations with rocky-core compositions, opening the door for a broader range of compositional diversity. Careful tests of model selection using simulated planet catalogs have shown that overfitting remains a problem when considering high complexity models. As a result, several of these models remain viable fits to the *Kepler* dataset, and higher quality datasets or improved modeling techniques may be necessary to distinguish between the competing theories of planet formation processes. Though our results were inconclusive in ruling a subpopulation of water worlds in or out, our study presents an in-depth case study on the challenges of model selection applied to mixture models and the population-level degeneracies that persist when interpreting

a statistical sample of planet mass, radius, and orbital period measurements.

4.6.4 Future Prospects

Our ability to fit multidimensional distributions including multiple subpopulations of planets is limited by the quality of the data that we use. As seen in Section 4.5.3, the quality and quantity of data is not a complete solution to the difficulties of model selection with mixture models. However, improving the dataset (as in our *PLATO* simulations, compared to *Kepler* simulations), did help distinguish between models when the model complexity was high. Improving the dataset will thus allow us to uncover additional complexities in the mass-radius-period distribution, just as improving the *Kepler* sample with higher precision stellar radius measurements led to the discovery of the *Kepler* radius gap (Fulton et al., 2017). These improvements would include more transiting planets in the sample, more planets with mass measurements, lower mass and radius uncertainties, and improved sensitivity to smaller planets and longer period planets.

TESS is an ongoing all-sky transit survey looking at the brightest nearby stars that has identified 2241 exoplanet candidates to date, with over 100 confirmed exoplanets (Guerrero et al., 2021). *TESS* also aims to have follow-up radial velocity measurements of 50 small planets below $4R_\oplus$. With its two-year Prime Mission, its sensitivity is mostly limited to planets with periods shorter than 100 days. Additionally, it is limited in its ability to detect smaller planets below $1.5R_\oplus$, currently having detected less than 100 candidates out of 2241. Nonetheless, *TESS* offers a dataset comparable in size to *Kepler*, and will add considerably to the sample of small planets with radial velocity mass measurements. *TESS* will also enable more detailed studies of host star dependence for planet occurrence rates, given the increased fraction of planets around low-mass stars.

PLATO is a space telescope scheduled for launch in 2026, that is designed to detect planetary transits around bright stars. The *PLATO* mission has a goal of detecting 7000

transiting systems, with radial velocity mass measurements of 400 of those planets to a rough precision of 10%, in addition to TTV mass measurements of an undetermined subset of planets. *PLATO* also pushes the bounds in radius and period, with the goal of detecting $1R_{\oplus}$ planets orbiting in the habitable zone of G-type stars to 3% precision (PLATO Study Team, 2017). *PLATO*'s nominal science mission lifetime of 4 years, with a possible extension to 8 years, will enable it to better study long-period planets compared to *Kepler* and *TESS*. *PLATO* will represent a significant upgrade to *Kepler* by offering a larger dataset with higher precision radii, a massively expanded RV sample with higher precision masses, and improved sensitivity to smaller planets on longer orbits.

In the future, we will have several high-quality datasets available to us, from *Kepler*, *TESS*, *PLATO*, and others. Rather than fitting a mass-radius-period distribution to each dataset independently, a better way to leverage these datasets would be to combine them into one single fit. The inhomogeneous Poisson process we use to model this distribution offers a possible way to incorporate multiple surveys. Our ability to include multiple compositional subpopulations in our mass-radius-period models will thus be greatly improved with the advent of additional *Kepler*-like surveys.

4.7 Conclusion

We include various subpopulations of planets with icy compositions into our joint mass-radius-period distribution modelling. We create a suite of ten mixture models of varying complexity that combine six fundamental subpopulations of planets - gaseous planets with rocky or icy cores, evaporated rocky or icy cores, and intrinsically rocky or icy planets - in different permutations in order to assess the evidence for water worlds in the *Kepler* dataset. We find that these icy compositional subpopulations have significant overlap with existing gaseous and rocky subpopulations, leading to large degeneracies in identifying any particular planet as belonging to a specific subpopulation. We apply k-fold cross-validation as a form

of model selection, and find that while the lowest complexity models are disfavored, we cannot strongly distinguish between the majority of our ten models. Models that include icy composition planets either with photoevaporation, or without photoevaporation, or models that include photoevaporation but not icy composition planets, all have support in the data. We set a rough upper limit of 50% on the fraction of icy composition planets, although this fraction varies widely between models. We propose that improved modeling approaches, such as fundamentally modeling composition, and improved datasets, such as those from *PLATO* or *TESS*, could further probe the existence of water worlds.

CHAPTER 5

A COMPREHENSIVE GRID OF PLANET EVOLUTIONARY MODELS

The models we have presented thus far in Chapters 2, 3 and 4 have explored exoplanet distributions across the dimensions of mass, radius and period. The mass-radius relations we use in these chapters have taken different forms. In Chapter 2, the mass-radius relation we fit to the sample of planets with mass and radius measurements below 8 $R_\oplus$ takes the form of a power-law, with optional dependence on host star mass and incident flux. In Chapters 3 and 4, we use a similar formulation for the mass-radius relation of the subpopulations of planets with gaseous envelopes, with the addition of two breaks to account for changes in the mass-radius relation at higher planet radii. By contrast, the subpopulations without gaseous envelopes follow the fixed mass-radius relations of rocky or icy cores derived from planet interior structure models. Ideally, we'd be able to extend our use of planet structure models to planets with gaseous envelopes, without having to rely on empirically fitted mass-radius relations.

A mass-envelope mass fraction-flux distribution of exoplanets would move beyond a mass-radius-period distribution to instead directly model the properties of exoplanets more fundamental to their structure. Such a distribution would require a mapping from mass, envelope mass fraction and flux to the observables of mass, radius and period. In this chapter, we develop a comprehensive and dense grid of planet evolution models across planet mass, envelope mass fraction, incident flux, and age that can provide this mapping. We evolve our grid using Modules for Experiments in Stellar Astrophysics (MESA), a stellar evolution code that has been adapted for the evolution of planets with hydrogen-helium envelopes. Our grid spans four orders of magnitude in planet mass, five orders of magnitude in incident flux, envelope mass fractions from 0.001 to 0.999, and ages from 100 Myr to 10 Gyr, with a total size of nearly $100,000$ grid points. We provide tabulated radii across these four dimensions,

as well as initial and final MESA evolution files for further study. We also explore behaviors in the flux-radius relation that have dependence on planet mass, envelope mass fraction and age, whereby a breakpoint in the flux-radius relation at $6\,F_\oplus$ can arise.

This chapter was completed in collaboration with Isaac Malsky and Leslie Rogers. The code to evolve planet models in MESA relies on prior development by Phil Arras, Howard Chen, and Isaac Malsky. This work was also completed in part with resources provided by the University of Chicago's Research Computing Center.

5.1 Abstract

Interpretation of exoplanet mass and radius measurements relies on comparing to theoretical mass-radius curves generated by planet evolution models. Existing grids of planet evolution models either only cover specific regions of parameter space, or offer sparse coverage. We evolve a dense grid of planet models without mass loss using Modules for Experiments in Stellar Astrophysics (MESA) across four orders of magnitude in planet mass and five orders of magnitude in incident flux, with envelope mass fractions from 0.001 to 0.999 and ages from 0.1 to 10 Gyr. We offer radii as a function of these four dimensions as a benchmark for comparison, as well as MESA initial and final evolution files for the evolution of these planets for further studies. We also note several behaviors in the flux-radius relation that depend on planet mass, envelope mass fraction and age, notably a breakpoint around $6\,F_\oplus$. Our dense and comprehensive grid can serve as a base for further studies that explore the effects of envelope composition, core composition, mass loss and inflation, among other physical processes. Additionally, our grid can enable population studies that require translation from observables of mass, radius and period to envelope mass fraction across a wide range of parameter space.

5.2 Introduction

Observational surveys of exoplanets have revealed that hydrogen-helium envelopes are not unique to the cold gas giants found in our solar system, but rather exist around exoplanets with a diverse range of masses and orbital periods. Combining radii of transiting exoplanets from the *Kepler* survey (Borucki et al., 2011a; Batalha et al., 2013) with mass measurements from radial velocity follow-up (Marcy et al., 2014) has shown that most planets above a threshold radius of $1.6\,R_\oplus$ have a bulk density that requires a significant volatile-rich envelope (Rogers, 2015). Statistical studies of *Kepler* have also revealed a bimodal radius distribution (Fulton et al., 2017), thought to be a result of envelope mass-loss due to either photoevaporation (Fulton & Petigura, 2018; Van Eylen et al., 2018; Owen & Murray-Clay, 2018) or core-powered heating (Ginzburg et al., 2018; Martinez et al., 2019; Gupta & Schlichting, 2020; Rogers et al., 2021), although there are also explanations that do not require mass-loss (Zeng et al., 2019). The survivability of the envelope primarily depends on its mass and composition, as well the flux received from the planet's host star (e.g., Lopez et al., 2012). Despite this, hydrogen-helium envelopes have been shown to persist for planets down to at least $2\,M_\oplus$ and high incident fluxes of around $100\,F_\oplus$, as in the *Kepler*-11 system (Lissauer et al., 2013).

Understanding the thermal evolution of exoplanets with hydrogen-helium envelopes is crucial to correctly interpreting the mass, radius, and period demographics derived from observational surveys. The current sample of small planets with both mass and radius measurements exhibit a large degree of scatter, both due to observational uncertainties and intrinsic scatter in the mass-radius relation (Wolfgang et al., 2016; Ning et al., 2018). Additionally, studies of the mass-radius relation have found that flux can have a non-negligible impact, particularly for Jupiter-sized gas giants (Sestovic et al., 2018; Ma & Ghosh, 2021). The broader mass-radius-flux distribution has been shown to be consistent with multiple subpopulations of planets, including evaporated cores that have lost their envelopes to pho-

toevaporation (Neil & Rogers, 2020). This diversity in mass-radius-flux space is reflective of the compositional diversity of exoplanets and their envelopes, as an envelope that is only a small percentage of the planet by mass can have a significant effect on the planet's radius (e.g., Rogers et al., 2011; Lopez & Fortney, 2014).

The thermal evolution of planetary interiors and their hydrogen-helium atmospheres has been achieved by adapting the open source, 1D stellar evolution code Modules for Experiments in Stellar Astrophysics (MESA) (Paxton et al., 2011, 2013, 2015, 2018, 2019). Initial studies of the evolution of low-mass, gas-dominated planets using MESA focused on the structure and impact of evaporation on the modestly irradiated sub-Neptunes discovered by *Kepler* (Batygin & Stevenson, 2013; Owen & Wu, 2013). MESA has been implemented to study the process of Roche-lobe overflow on short-period gaseous planets, whereby gas giants are stripped of their envelopes and become remnant planetary cores (Valsecchi et al., 2014, 2015; Jackson et al., 2016, 2017). Chen & Rogers (2016) extended the application of MESA to planet masses down to $1\,M_{\oplus}$, showing how the envelope evaporation timescale varies non-monotonically with envelope mass fraction, peaking at 1 to 3% and resulting in convergent mass loss evolution. They additionally showed that planet radii show little hysteresis, further supporting the idea that planet radii can be considered a proxy for the planet's current composition (Lopez & Fortney, 2014). Malsky & Rogers (2020) coupled the thermal and mass-loss evolution of the planetary envelope with its compositional evolution, demonstrating how preferential loss of hydrogen can lead to helium-enhanced atmospheres. Additional studies with MESA have focused on giant planet evolution (Buhler et al., 2016; Berardo & Cumming, 2017; Dederick & Jackiewicz, 2017; Valletta & Helled, 2020), the accretion of planetesimals by fully-formed planets (Chatterjee & Chen, 2018), and the coupling of the thermal evolution of the planet atmosphere with more realistic mass-loss prescriptions (Kubyshkina et al., 2020).

In this work, we build upon the algorithms developed by Chen & Rogers (2016) and

Malsky & Rogers (2020) (themselves built upon MESA's make_planets test suite (Paxton et al., 2013)) to employ MESA to evolve a grid of models across a broad range of planet masses, envelope mass fractions, incident fluxes, and ages. Compared to previous studies that computed evolutionary models across these dimensions (e.g. Fortney et al., 2007b; Lopez & Fortney, 2014; Howe & Burrows, 2015), our grid covers a wider range of parameter space and is more densely sampled. Unlike many of the aforementioned studies, our models do not include any prescription for envelope mass loss. Given the lack of hysteresis shown by planetary radii (Chen & Rogers, 2016), our models can be applied to study the present-day composition of exoplanets and their envelopes.

This paper is organized as follows. In Section 5.3 we give an overview of the parameters of the grid of planet models as well as their execution with MESA. We present a summary table and associated mass-radius diagrams of these models, as well as compare to past work, in Section 5.4. We discuss potential applications of this grid of models and conclude in Section 5.5.

5.3 Methods

5.3.1 MESA Routines

We follow the methodology of Malsky & Rogers (2020) (hereafter MR20) in adapting Modules for Experiments in Stellar Astrophysics (MESA), an open-source Fortran library for evolving stars, to evolving planets. We use MESA release version 12778, released in March 2020. We first summarize the MESA routines from MR20 used in this work, and then detail differences in methodology for our specific purposes.

Broadly, the process starts with an initial model that is relaxed through various stages in order to bring the model to the desired set of initial conditions, including the following: total mass M, core mass $M_{\rm core}$, envelope composition $\{X, Y, Z\}$, initial entropy S, and incident

flux F. For planets below $100 M_{\oplus}$, we follow the 13 steps in Section 2.1 of MR20 in order to set these initial conditions. For planets above $100 M_{\oplus}$, we skip several steps as they are unnecessary for planets with higher masses. Namely, we use an initial planet file with the specified final mass of the planet, rather than starting with a larger planet and reducing it to the final mass over several stages. This skips steps 2, 4, and 8 in MR20. We still follow steps 3, 5 and 9, whereby a planetary core is initially inserted with a low mass and gradually brought to its final core mass in two stages.

After the total planet mass, core mass, and envelope composition have been set, we reinflate the planet to simulate a "hot start" formation scenario. We insert an artificial core luminosity in order to heat the planet to a specified hot-start initial interior entropy. The chosen hot-start initial entropies are a function of total planet mass and interpolated from a composite of two sources, for two separate mass regimes. For giant planets ($\geq 1\, M_J$), we use hot-start initial entropies from the thermal evolution models from Marley et al. (2007) that span from $1\, M_J$ to $10\, M_J$. We include a constant positive offset of $0.47\, k_{\rm B}$ baryon^{-1} in order to match the cooling curves from the evolution models of Marleau & Cumming (2014) as well as cooling curves generated with MESA. This offset partially arises from the omission of a quantum-mechanical term (related to the statistical weight of the spin states of the proton in the hydrogen partition function) in the equation of state (EOS) tables used in a number of works (Marleau & Cumming, 2014). For low-mass planets, we use the initial entropies from the evolutionary models of Linder et al. (2019) that range from $5\, M_{\oplus}$ to $318\, M_{\oplus}$ ($1\ M_J$). Since the initial entropy at $1\, M_J$ is $10.04\, k_{\rm B}$/baryon using the models from Marley et al. (2007) but $9.2\, k_{\rm B}$/baryon using the models from Linder et al. (2019), we increase the entropies from Linder et al. (2019) by a constant offset of $10.04 - 9.2 = 0.84$ in order to match the initial entropies at $1 M_J$. For masses between the grid points provided by these two sets of models, we linearly interpolate between the grid points. At the bounds of our mass range, our initial entropies are $7.47\, k_{\rm B}$/baryon at $1\, M_{\oplus}$ and $13.32\, k_{\rm B}$/baryon at

$10,000\,M_{\oplus}$.

In order to inflate the planets to their specified hot-start initial entropies, we insert an artificial luminosity to the core of the planet. The artificial core luminosity necessary to reach the desired entropy is dependent on the total mass and envelope mass fraction of the planet. In order to find the luminosity that heats the planet to the target entropy without over-heating it and causing it to lose its envelope, we first try an artificial luminosity of 300 times the resulting luminosity at the end of the previous evolution step. The planet evolves with this artificial luminosity for 10^8 years, or until the central entropy reaches the target entropy. If the target entropy is reached, we iteratively decrease the artificial luminosity by 10% until the entropy is lower than the target entropy, and use the prior model that just met the target entropy. Conversely, if the target entropy is not initially reached, we iteratively increase the artificial luminosity by 10% until the entropy surpasses the target entropy, and use that model. We ensure that the resulting models from this step do not over-inflate the planet and cause the planet envelope to be blown off.

As in MR20, we use MESA's built-in function for heating the planet envelope to a specified column depth through the incident flux received by the planet (Paxton et al., 2013). The column depth to irradiation Σ_* for a given model is found by solving for the zone i where Σ_*, determined by the atmospheric opacity κ_ν, matches the mass column density in that zone, $\Sigma_{*,i}$. The atmospheric opacity in a given zone, $\kappa_{\nu,i}$, is found by interpolating within the opacity tables from Freedman et al. (2014) (using the Planck mean opacities for a Solar metallicity and a stellar effective temperature $T_{\rm eff} = 6000\,K$) at the equilibrium temperature of the envelope $T_{\rm eq}$ and the pressure within the zone P_i.

Unlike MR20, our models do not include a prescription for envelope mass loss. With the range of incident fluxes and envelope mass fractions that we evolve, a substantial fraction of our planet models are susceptible to envelope mass loss through photoevaporation. However, our models remain applicable to determining the present-day composition of exoplanets, as

Parameter	Units	Range	Spacing	Number of Points
Mass	$M_\oplus$	[1 - 10,000]	Logarithmic	41
Envelope mass fraction	-	[0.001 - 0.398] / [0.5] / [0.602 - 0.999]	Log / - / Log in 1-envf	14 / 1 / 14 (29 total)
Incident flux	$F_\oplus$	[0.01 - 1) / [1 - 1000]	Log / Log	4 / 16 (20 total)
Age	Gyr	[0.1, 10]	Log	9
Envelope helium fraction (Y)	-	0.24	-	1
Envelope metallicity (Z)	-	0.02	-	1
Core composition	-	70% silicate, 30% iron	-	1
Total grid size	-	-	-	$128,340$ ($97,944$ after cuts)

Table 5.1: Initial Parameters for our Grid of MESA Planet Evolution Models
In the range of the parameters, brackets are inclusive and parentheses are exclusive. The total grid size, shown in the bottom right corner, takes into account limitations on the core mass discussed in Section 5.3.2. The final number of planet models after applying cuts and removing failed models as discussed in Section 5.4.2 is indicated in parentheses.

radius is a proxy for composition and shows little hysteresis (Chen & Rogers, 2016). A systematic error of $\leq 1\%$ in planet radius is expected in applying a mapping calculating without mass loss from planet mass, envelope mass fraction, incident flux and age to planet radius for evaporating planets (Chen & Rogers, 2016). We leave the creation of a dense grid of planet models evolved with envelope mass loss to future work.

5.3.2 Grid Parameters

The ranges of our initial parameters for our grid of planet models is summarized in Table 5.1. We explain our choice of initial parameters below.

For the purpose of a comprehensive grid of planet evolution models, we aim to cover a wide range of planet masses, from small sub-Neptunes to gas giants. Between the fundamental properties that determine the mass of the planet's core and envelope, we choose to structure the grid in terms of total planet mass and envelope mass fraction. The core mass, envelope mass, and core mass fraction can then be calculated from these two quantities. We choose total planet mass rather than core mass as total planet mass is the most natural unit to use when comparing to observations of detected planets. We model planet masses M from $1M_\oplus$ to $10,000M_\oplus$, with 41 grid points logarithmically spaced between these two extremes. Our lower limit of $1M_\oplus$ was chosen due to MESA having difficulty evolving planet masses below $1M_\oplus$. The upper limit of $10,000M_\oplus$ is around the mass of the most massive exoplanets

discovered to date (Sahlmann et al., 2013; Konopacky et al., 2016), and while it extends into brown dwarf territory, this upper limit ensures we fully encompass the planet regime.

The envelope mass fraction $f_{\rm env}$ similarly spans a wide range in order to capture the diversity inherent to the planet population. We model envelope mass fractions from a lower limit of 0.001 to an upper limit of 0.999. In order to capture the variation that can happen both at low envelope mass fractions (sub-percent level) and high envelope mass fractions (above 0.99), we structure the grid points to be symmetric in envelope mass fraction and core mass fraction ($1 - f_{\rm env}$), with 14 points logarithmically spaced from 0.001 to 0.398 for each. Additionally, we add a grid point at 0.5, for a total of 29 grid points in envelope mass fraction. In terms of envelope mass fraction alone, we first have 14 points from 0.001 to 0.398, log-spaced; a point at 0.5; 14 points from 0.602 to 0.999, log-spaced in $1 - f_{\rm env}$).

With our grid defined in terms of total planet mass and envelope mass fraction, we also place limits on the mass of the planetary core such that arbitrary combinations of total planet mass and envelope mass fraction do not lead to unphysical core masses. We use a minimum core mass of $0.1 M_\oplus$ and a maximum core mass of $100 M_\oplus$. For giant planets, the difference in radius between a core mass of $0.1 M_\oplus$ and no core is negligible, whereas for small planets a core mass of $0.1 M_\oplus$ implies a high envelope mass fraction that will be likely be unbound or unstable. Core masses above our upper limit of $100 M_\oplus$ are physically unlikely, with the maximum core mass likely closer to $25 M_\oplus$ (Johansen & Lambrechts, 2017; Bitsch, Bertram et al., 2019; Otegi et al., 2020). As a result, from the $41x29$ grid points in total planet mass and envelope mass fraction, we throw out 476 grid points, bringing our total grid points in this 2D space from 1189 to 713.

The third dimension of our grid is incident flux F, and we use a wide range of fluxes to capture the variation inherent to the planet population, from observations of exoplanets to planets in the solar system. We model incident fluxes from $0.01\,F_\oplus$ to $1000\,F_\oplus$. Since the variation as a function of incident flux is higher for high fluxes than for low fluxes, we

more densely sample the range from $1 - 1000\,F_\oplus$ compared to $0.01 - 1\,F_\oplus$. We use 4 points between 0.01 and $1\,F_\oplus$, log-spaced, not including the point $1\,F_\oplus$, and 16 points between 1 and $1000\,F_\oplus$, including $1\,F_\oplus$, for a total of 20 points. This takes our 3D grid in planet mass, envelope mass fraction, and incident flux to 14260 total points.

Each of these 14060 grid points results in an individual planet evolved over time for a total of 10 Gyr. We provide the radius at a number of different ages, although higher resolution in time is available through the corresponding output MESA evolve files. The ages we present in Table 5.2 span from 100 Myr to 10 Gyr, with 9 points logarithmically spaced between, for a total of 128,340 grid points in the four-dimensional space of planet mass, envelope mass fraction, incident flux, and age. We also provide the output evolve files where the full information of the evolution of these planets is available.

For each of these planet models, we fix several planetary and stellar properties as the dependence of planetary evolution on these properties is not investigated in this work. We use Sun-like properties for the host star ($M_* = 1M_\odot$, $R_* = 1R_\odot$, $T_{\rm eff,*} = 6000$K) and Solar composition for the planetary envelopes ($X = 0.74$, $Y = 0.24$, $Z = 0.02$). The planetary heavy-element cores have an Earth-like composition (70% silicate and 30% iron). We use a Bond albedo of $A = 0.2$ in order to compute the equilibrium temperature of the planet's envelope from the incident flux on the planet:

$$F = T_{\rm eff,*} \sqrt{\frac{R_*}{2a}} (1 - A)^{1/4} \tag{5.1}$$

where a is the semimajor axis of the planet's orbit. The planet radius for our purposes is defined at a pressure of 1 mbar (100 Pa), which roughly corresponds to the transit radius of a planet without high-altitude clouds at low pressures (Miller et al., 2009).

Age [Gyr]	Flux [$F_\oplus$]	Mass [$M_\oplus$]	0.001	0.0016	0.0025	0.004	0.0063	0.01	...	0.99	0.9937	0.996	0.9975	0.9984	0.999
0.1	0.01	1.0	1.15	1.19	1.24	1.3	1.38	1.48	...	COB	COB	COB	COB	COB	COB
0.1	0.01	1.26	1.22	1.25	1.3	1.36	1.43	1.53	...	COB	COB	COB	COB	COB	COB
0.1	0.01	1.58	1.28	1.32	1.36	1.42	1.49	1.59	...	COB	COB	COB	COB	COB	COB
0.1	0.01	2.0	1.36	1.39	1.43	1.49	1.56	1.65	...	COB	COB	COB	COB	COB	COB
0.1	0.01	2.51	1.43	1.46	1.51	1.56	1.63	1.72	...	COB	COB	COB	COB	COB	COB
0.1	0.01	3.16	1.51	1.54	1.58	1.64	1.7	1.79	...	COB	COB	COB	COB	COB	COB
0.1	0.01	3.98	1.6	1.63	1.67	1.72	1.79	1.87	...	COB	COB	COB	COB	COB	COB
0.1	0.01	5.01	1.68	1.71	1.75	1.81	1.87	1.96	...	COB	COB	COB	COB	COB	COB
0.1	0.01	6.31	1.78	1.81	1.85	1.9	1.96	2.05	...	COB	COB	COB	COB	COB	COB
0.1	0.01	7.94	1.87	1.9	1.94	1.99	2.06	2.15	...	COB	COB	COB	COB	COB	COB
0.1	0.01	10.0	1.97	2.0	2.04	2.09	2.16	2.25	...	12.2	COB	COB	COB	COB	COB
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
0.1	0.01	1000	COB	COB	COB	COB	COB	COB	...	13.8	13.9	13.9	13.9	13.9	13.9
0.1	0.01	1259	COB	COB	COB	COB	COB	COB	...	13.9	13.9	13.9	13.9	13.9	13.9
0.1	0.01	1585	COB	COB	COB	COB	COB	COB	...	13.9	13.9	13.9	13.9	13.9	13.8
0.1	0.01	1995	COB	COB	COB	COB	COB	COB	...	13.8	13.8	13.8	13.8	13.8	13.8
0.1	0.01	2512	COB	COB	COB	COB	COB	COB	...	13.7	13.7	13.7	13.7	13.7	13.6
0.1	0.01	3162	COB	COB	COB	COB	COB	COB	...	13.6	13.6	13.5	13.5	13.5	13.5
0.1	0.01	3981	COB	COB	COB	COB	COB	COB	...	13.4	13.4	13.4	13.4	13.3	13.3
0.1	0.01	5012	COB	COB	COB	COB	COB	COB	...	13.3	13.3	13.2	13.2	13.2	13.1
0.1	0.01	6310	COB	COB	COB	COB	COB	COB	...	13.3	13.2	13.2	13.1	13.1	13.1
0.1	0.01	7943	COB	COB	COB	COB	COB	COB	...	13.4	13.4	13.3	13.3	13.3	13.2
0.1	0.01	10000	COB	COB	COB	COB	COB	COB	...	13.7	13.7	13.6	13.6	13.5	13.5
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
0.1	1000	100	FTC	3.13	3.18	3.25	3.34	3.46	...	15.3	15.4	15.4	15.4	15.5	15.5
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
10	0.01	100	FTC	ECS	ECS	ECS	ECS	ECS	...	10.3	10.3	10.4	10.4	10.4	10.4
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
10	1000	10000	COB	COB	COB	COB	COB	COB	...	10.9	10.8	10.7	10.6	10.5	10.4

Table 5.2: Planet Radii Table
Columns in the top row represent values for the envelope mass fraction. Planet radii in these columns are shown in units of $R_\oplus$. Acronyms for missing values are explained in Section 5.4.2.

5.4 Results

5.4.1 Tables and Comparisons

We present planet radii as a function of our input parameters of planet mass, envelope mass fraction, incident flux, and age, in Table 5.2. This table contains 97,944 total entries with planet radius data. Gaps in the table where data is missing are explained in Section 5.4.2.

We compare our models to the 1.0 to 20.0 $M_\oplus$ models of Chen & Rogers (2016) (Table 2) at fluxes of 0.1, 10 and 1000 $F_\oplus$, envelope mass fractions of 0.001, 0.01 and 0.1, and an age of 10 Gyr. Both our models and those of Chen & Rogers (2016) were evolved with a silicate-iron core and without mass loss and MESA. We find general agreement within 5%, with our planet radii typically slightly larger. Chen & Rogers (2016) compared their models to Lopez & Fortney (2014) and found radii to be consistent within 5%, deviating at the most at low planet radii and high envelope mass fraction. We find similar consistency with Lopez & Fortney (2014).

For planets above a mass of 100 $M_\oplus$, we compare to the models of Müller & Helled (2021), who used MESA to model the evolution of giant planets. We use their publicly available code planetsynth which interpolates between their grid of models to generate predicted planet radii at our specific grid points. We test a mass range of 100 to 3981 $M_\oplus$, envelope mass fractions from 0.5 to 0.98, fluxes from 0.01 to 630 $F_\oplus$, and ages from 0.1 to 100 Gyr. We find that our radii generally agree within 10%, with our radii consistently higher than theirs by about 5%. We note that they assumed a proto-Solar hydrogen-helium ratio for the planetary envelope ($X = 0.705$, $Y = 0.275$), and used a modified EOS for masses below $5M_J$, which may account for these differences in radii. Defining the radius at 1 mbar rather than at a radial optical depth of $\frac{2}{3}$ may also cause our radii to be systematically higher than both Chen & Rogers (2016) and Müller & Helled (2021).

We also compare our giant planet models to those of Fortney et al. (2007b), in areas where our grids overlap. As defined by their grid (Tables 2 and 3), we compare in the following parameter ranges: from orbital separations of 0.045 to 1 AU, core masses from 0 to 100 $M_\oplus$, ages from 300 Myr to 1 Gyr, and masses from 125 to 774 $M_\oplus$. We find our radii to be within 10%, although our radii are generally lower, rather than higher as they were compared to Chen & Rogers (2016) and Müller & Helled (2021). We note differences in the core composition (they model a 50% rock and 50% ice mixture) as well as differences in how the heating from the flux incident upon the envelope is applied as possible reasons for these radii discrepancies.

Mass-radius relations for planets of varying envelope mass fraction at fixed fluxes and an age of 10 Gyr are shown in Figure 5.1, with each panel showing a different flux. Similar mass-radius relations for a fixed flux of 10 $F_\oplus$ and for four different ages are shown in Figure 5.2. We note that these mass-radius relations do not capture the entirety of the grid, as many flux and age combinations are not shown.

As shown in Figure 5.1 through the four panels, as a function of flux, the largest changes

in radius occur at low masses. Giant planets have inflated radii at high fluxes, but show little variation from 0.1 to 100 $F_\oplus$. By contrast, the lowest masses between 1 and 3 $M_\oplus$ have radii up to a factor of 2 larger from 0.1 to 100 $F_\oplus$. Additionally, while most mass-radius relations at fixed envelope mass fraction and flux are monotonically increasing, at high fluxes the mass-radius relation decreases from roughly 1 to 3 $M_\oplus$ at fixed envelope mass fraction. This is most significant for an envelope mass fraction of about 0.01. We also note that due to envelopes becoming unbound, at higher fluxes we are limited to lower and lower envelope mass fractions at a given mass, visualized as the mass-radius points shifting to the right as flux increases in Figure 5.1.

Whereas increasing flux changes the radius most significantly in the regime of low planet mass, increasing age changes the radius most significantly in the regime of high envelope mass fraction. This is apparent in the 10 to 100 $M_\oplus$ range for envelope mass fractions greater than 0.1 from ages 0.1 to 1 Gyr in Figure 5.2. The radius at a fixed envelope mass fraction can decrease by a factor of up to 1.5 from the lowest mass to an intermediate mass at 0.1 Gyr. At 3.2 Gyr, by contrast, these mass-radius relations at fixed envelope fraction are nearly flat.

All other parameters fixed, we find that radius as a function of age monotonically decreases, whereas radius as a function of flux monotonically increases. We note that there are a handful of cases where radius decreases as flux increases at fixed mass, envelope mass fraction and age. These cases occur at giant planet masses and high envelope mass fractions, where radius as a function of flux is nearly flat. These spurious non-monotonicities result in sub-percent level decreases in planet radii and may be due to interpolation within the EoS tables. Radius is not monotonically increasing or decreasing as a function of mass or envelope mass fraction.

5.4.2 Gaps in Grid

Not every planet model evolves for the full 10 Gyr duration, some planet models fail to evolve at all, and others have their envelopes blown apart. In Table 5.2, we have labelled missing values with four separate acronyms that give the reason for the missing data.

Core Outside Bounds (COB) indicates planet mass and envelope mass fraction combinations that result in a core mass beyond the limits that we set in Section 5.3.2: below $0.1\,M_\oplus$ or above $100\,M_\oplus$. Unlike the other acronyms, COB does not represent failed models; they are simply outside the grid. These account for 85680 missing values in the table.

Failed To Converge (FTC) indicates planets that failed at some point in the 13-step process of evolving planets with MESA listed in MR20. The affected planet models all have planet mass $\leq 100\,M_\oplus$, and fail either during the inflation to the hot start entropy in steps 11 and 12, or while the irradiation is relaxed to its final value in step 13. There is a cluster of models with planet masses between 25 and $63\,M_\oplus$ with envelope mass fractions ≥ 0.9 that fail to converge during the inflation steps. There are a total of 10431 FTC grid points.

Evolution Cut Short (ECS) indicates planets that evolved successfully for earlier ages, but failed to evolve to the listed age. These failures are due to planet interior conditions approaching boundaries in the EOS tables. Affected planets fall between 3 and $100\,M_\oplus$, with envelope mass fractions below 0.1, at ages longer than 5 Gyr. There are 1625 ECS grid points.

Envelope Blown Apart (EBA) indicate planet envelopes that became unbound due to high incident flux. This mostly affects planets with low total mass, high envelope mass fraction, and high flux, with a boundary in this 3D space separating planets with bound and unbound envelopes. To remove these planet models, we filter any planets that exceed their Roche limit at any point or have a radius greater than $100\,R_\oplus$ at the specified age. There are 18340 EBA grid points.

Combining the FTC, ECS and EBA planet models, there are 30,396 failed models out

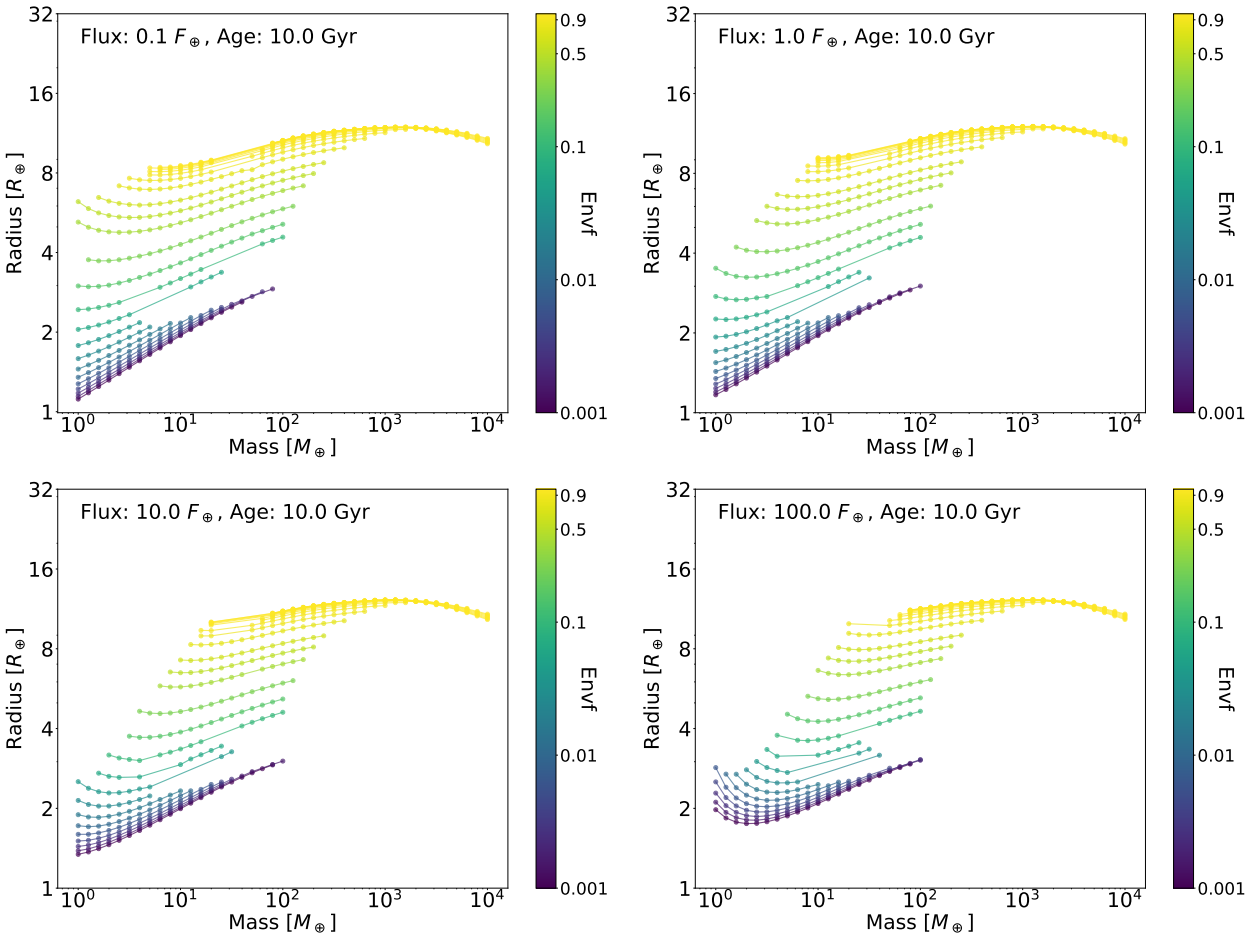

Figure 5.1: Mass-radius diagrams for planets of varying envelope mass fractions (denoted by the colorbar) at a fixed age of 10 Gyr. The four panels show four different fluxes labeled in the top left corner, ranging from 0.1 to 100$F_\oplus$. The points show planet models in the grid, while the lines linearly interpolate between these points at a fixed envelope mass fraction. Gaps in the grid are explained in Section 5.4.2.

of a total of 128,340 4D grid points, with 97,944 remaining planet models. The FTC and ECS failed models may be fixed by improvements to MESA's EOS tables, whereas the EBA models have a real physical origin. These gaps in the table can be compensated for by interpolating between successful planet models in the 4D parameter space of mass, envelope mass fraction, incident flux and age.

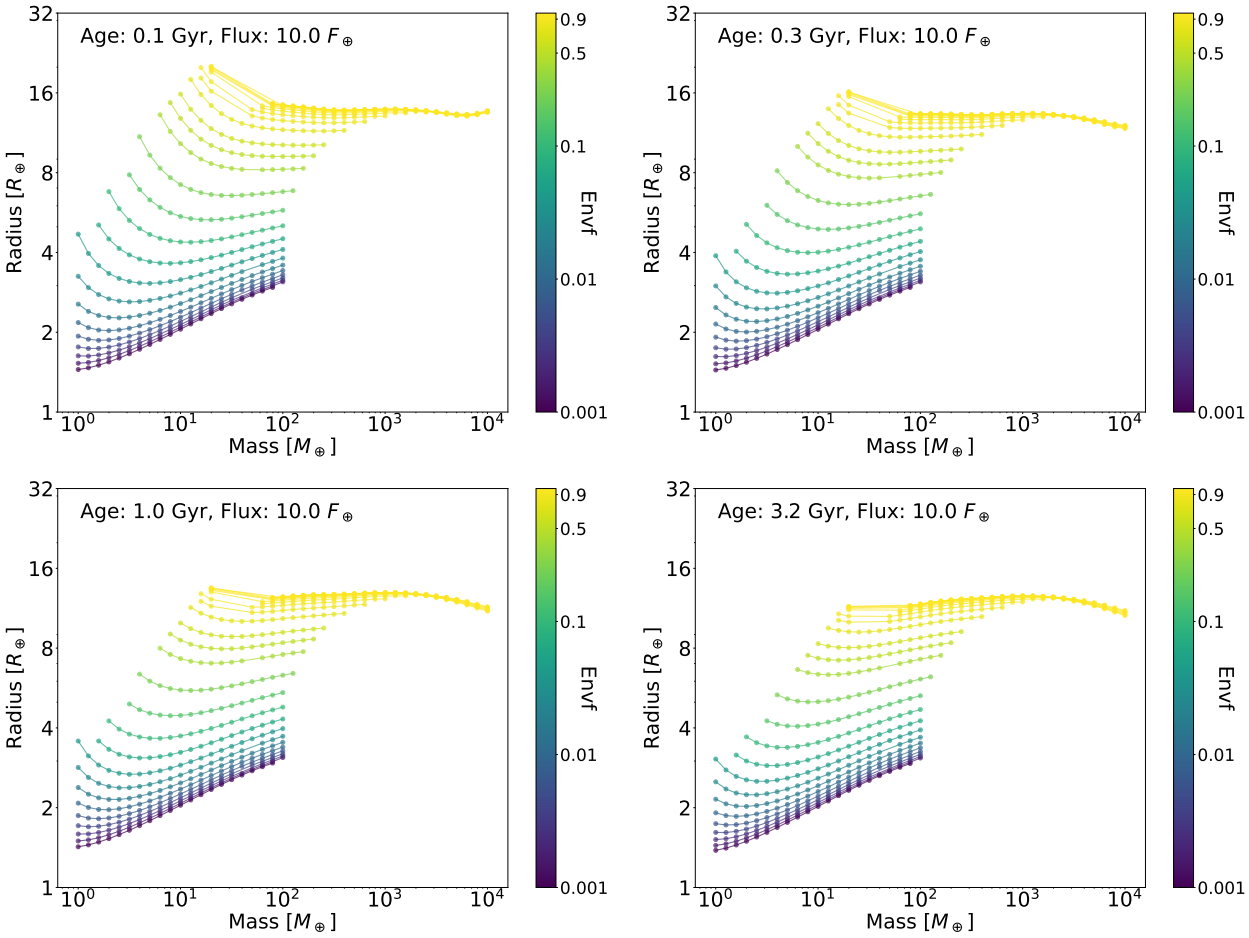

Figure 5.2: Same as Figure 5.1, except the four panels show different ages (from 0.1 to 3.2 Gyr) at a fixed flux of 10 $F_\oplus$.

5.4.3 Flux-radius relation

The planet radius as a function of incident flux exhibits different behavior depending on the planet's mass, age, and envelope mass fraction. This change in behavior is demonstrated in Figure 5.3, where the flux-radius relation is shown for planets of varying masses and ages at fixed envelope mass fraction. For some combinations of planet mass and age, the flux-radius relation has a clear break point between $1\,F_\oplus$ and $10\,F_\oplus$, where the slope ($d\ln R/d\ln F$) changes significantly. At low fluxes, before the breakpoint, the slope is shallow and gradually increases as you go to higher fluxes. At the breakpoint around $6\,F_\oplus$, the slope flattens out significantly, with an eventual steepening at high fluxes. For other combinations of mass and age, the flux-radius relation is either flat with near-zero slope until steepening at high flux, or has a nearly constant, positive slope throughout.

Whether or not the flux-radius relation has a breakpoint depends on the planet's mass and age. As shown in Figure 5.3, planets with a mass of $1000\,M_\oplus$ exhibit this breakpoint behavior at later ages, but not at earlier ages. On the contrary, planets with a mass of $100\,M_\oplus$ exhibit this behavior only at earlier ages, whereas at later ages the slope of the flux-radius relation becomes more constant. As the mass increases beyond $1000\,M_\oplus$, this breakpoint arises at later and later ages, until it disappears completely at $10000\,M_\oplus$. Similarly, as mass decreases beyond $100\,M_\oplus$, the breakpoint is more apparent at earlier and earlier ages, until the behavior ceases completely. The envelope mass fraction also has an effect, with the breakpoint generally stronger at higher envelope mass fractions. We note that we also find this behavior in the giant planet models of Müller & Helled (2021), although the flux breakpoint is consistently lower by a factor of two, at around 2 to $3\,F_\oplus$ rather than $6F_\oplus$ as in our calculations.

This behavior is also reflected in the structure of the planetary envelope. Figure 5.4 shows the pressure-temperature profiles of the planet envelope across a range of fluxes for planets of mass $1000\,M_\oplus$, envelope mass fraction of 0.99, and ages of 0.1 and 10 Gyr. The left panel of

Figure 5.4 shows planet models with an age of 0.1 Gyr that do not show a breakpoint in the flux-radius relation. Close to the surface of the planet, at low temperatures and pressures, the pressure-temperature profiles diverge as a function of flux, with some models at lower fluxes developing an intermediate convective region (shown by the dotted line, identified where the convective velocity is nonzero). Deeper into the planet, at temperatures hotter than 1400 K, the profiles converge and show vary little variation. The inner convective region also begins at the same pressure, at around 3×10^6 Pa. The largest deviation between profiles at high pressure occurs at the highest flux, 1000 $F_\oplus$, where the temperature is hotter at the same pressure compared to the profiles of lower fluxes. Mirroring the behavior of these profiles at higher temperatures, the flux-radius relation for this planet mass, envelope mass fraction and age has near-zero slope until the highest flux, when the slope increases.

The right panel of Figure 5.4 shows planet models of the same mass and envelope fraction but at a later age of 10 Gyr where the flux-radius relation exhibits a breakpoint. At temperatures hotter than 1400 K, the profiles at high fluxes $> 6\,F_\oplus$ converge similarly to their convergence at 0.1 Gyr. However, profiles for fluxes lower than the breakpoint at 6 $F_\oplus$ start to diverge, exhibiting lower temperatures at a given pressure. In addition, while the profiles at higher fluxes all become convecting at the same pressure, the profiles at lower fluxes become convecting at much lower pressures. These trends are reflected in the flux-radius relation, which is mostly flat from $6\,F_\oplus$ to $1000\,F_\oplus$, with a steeper slope below $6\,F_\oplus$.

These shifts in the pressure-temperature profiles and envelope structure as a function of flux are related to shifts in other physical properties of the envelope. We compared our pressure-temperature profiles to the EOS tables used by MESA for hydrogen and helium and found no significant difference between profiles before or after the flux break. Similarly, adiabatic indices of profiles below and above the break were not significantly different. However, we did identify a larger shift in opacity between profiles before or after the break. High-flux models whose pressure-temperature profiles converge at high pressure also converge in

opacity, and the low-flux models diverge in opacity at those same pressures. At pressures between $10^7 - 10^8$ Pa, prior to the convective boundary of high-flux models, opacities of low-flux models below the break are up to three times higher than those of high-flux models beyond the break. However, at pressures beyond 10^8 Pa, after the convective boundary of high-flux models, opacities of low-flux models are consistently up to a factor of two lower than those of high-flux models.

The overall effect of these different flux-radius behaviors is relatively small and likely beyond the sensitivity of planet radius measurements. Using the examples in Figure 5.3, the radius changes by at most 20% from $0.01\,F_\oplus$ to $1000\,F_\oplus$ for planets with a mass of $100\,M_\oplus$ at 0.1 Gyr, whereas the radius changes by only 4% across the same range for planets with a mass of $1000\,M_\oplus$ at 10 Gyr. Fitting a power-law to the flux-radius relation with a mass of $1000\,M_\oplus$ at 10 Gyr yields a maximum difference with the breakpoint relation of 0.6% in radius, below the typical sensitivity of transit surveys. Additionally, this is an ideal case for planets across a range of fluxes at fixed mass, envelope mass fraction and age, whereas the observational sample of planets will vary across each of these dimensions. Any break that we see here will likely be washed out by these variations with other parameters. This feature may also depend on the envelope metallicity or other factors that we do not vary in our models, further complicating its detectability.

5.5 Discussion

5.5.1 Using the Grid

Our grid provides a dense coverage across a wide range of parameter space in planet mass, envelope mass fraction, incident flux, and age. The ranges of the parameters that we include are intended to be exhaustive and cover extremes on both ends. Our grid can thus be used as a self-consistent suite of planet evolution models, useful when studying a wide range of

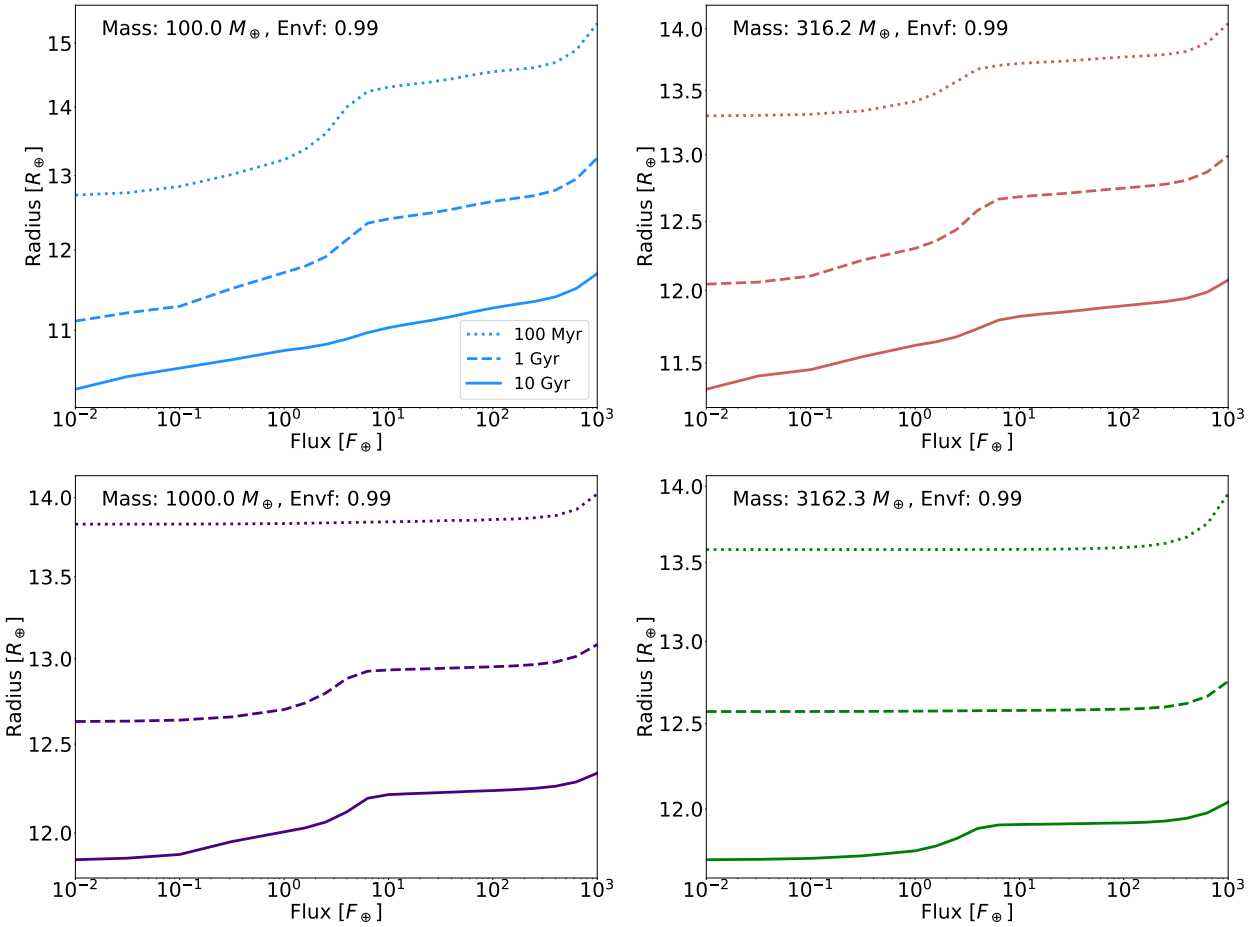

Figure 5.3: Planet radius as a function of incident flux for planets of varying mass and age, at a fixed envelope mass fraction of 0.99. The four panels show four different planet masses as labeled in the top-left corner, ranging from 100 to 3162 $M_\oplus$. The dotted line (top) shows the flux-radius relationship for an age of 100 Myr, the middle (dashed) line for an age of 1 Gyr, and the solid (bottom) line for an age of 10 Gyr. The lines interpolate between the 20 flux grid points.

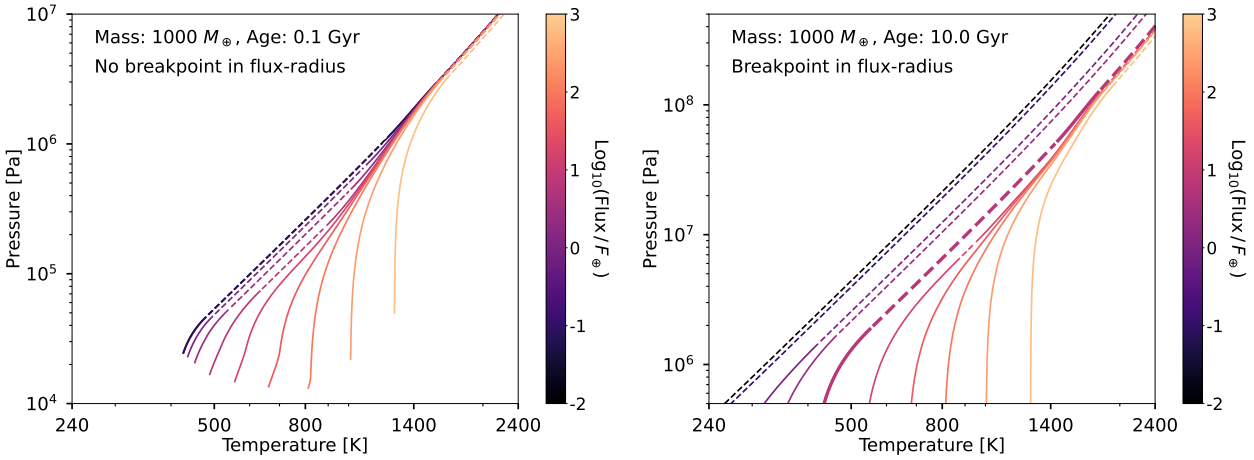

Figure 5.4: Pressure-temperature profiles for the planetary interior for varying ages and fluxes at a fixed planet mass of 1000 $M_\oplus$ and envelope mass fraction of 0.99. The color of the lines denotes the flux of the planet, given by the colorbar on the right. The two panels show profiles at ages of 0.1 Gyr (left) and 10 Gyr (right). The left panel of this figure corresponds to the dotted line of the lower-left panel in Figure 5.3, and the right panel corresponds to the solid line of the lower-left panel. Dotted lines in this figure indicate convective regions, whereas solid lines indicate radiative regions. Each pressure-temperature profile is convective throughout at higher temperatures and pressures beyond the axis bounds. In the right-hand panel, the two profiles with the lowest flux become radiative beyond the axis bounds, at pressures of 10^6 Pa ($10^{-2}\,F_\oplus$) and 3×10^6 Pa ($10^{-1}\,F_\oplus$). The right panel shows planet models that exhibit a breakpoint in the flux-radius relation as shown in Figure 5.3, whereas the left panel shows models that do not. In the right panel, the line corresponding to the flux at the breakpoint is thicker. While the y-axis bounds are different between the two panels, both axes span three orders of magnitude.

planetary properties. A single grid is preferable to using a patchwork of grids that cover different regions of parameter space.

Our grid can be interpolated to provide planet radii as a function of any combination of our input parameters. The grid can also be used to identify the boundary in this 4D space where the envelopes of highly irradiated, small planets become unbound or unstable. We note that our assumptions of a Solar composition envelope with an Earth-like core composition and no mass-loss must be taken into consideration when applying our results. We also note that we do not include any extra inflation in our models, so our results should be not applied to Hot Jupiters with $T_{\rm eq} \gtrsim 1500$ (Thorngren & Fortney, 2018). While our suite of planet evolution models are "vanilla" in the sense of not exploring higher-order effects, they provide a benchmark against which future model grids that include these additional physical ingredients can be compared.

5.5.2 Extensions to the Grid

The 4D grid we present in this paper explores variation in planet mass, envelope mass fraction, incident flux, and age. In order to limit our exploration to these four quantities, we fixed various other parameters that impact the evolution of planetary envelopes. We briefly review these below, and leave this expansion of the grid to future work.

We assumed a Solar composition for the envelope ($X = 0.74$, $Y = 0.24$, $Z = 0.02$). However, envelope metallicities for low-mass, low-density planets can potentially reach up to $Z = 0.9$ (Fortney et al., 2013), and envelopes can become helium-enhanced through preferential loss of hydrogen, reaching metallicities of $Z = 0.23$ and helium fractions of $Y = 0.4$ (Malsky & Rogers, 2020). These drastic changes in envelope composition can have similarly drastic effects on the radius of the planet. We also assumed an Earth-like composition of the core of the planet (70% silicate and 30% iron). Several studies have explored cores composed of mixtures of ice and rock, with large effects on the radii compared

to Earth-like core compositions (e.g., Chen & Rogers, 2016).

Our models do not include any prescription for envelope mass loss. As such, our models can be used to study the current-day composition of exoplanets. Including envelope mass loss into our models (as in Chen & Rogers (e.g., 2016); Malsky & Rogers (e.g., 2020) would allow us to study the properties of planets at their formation as well as their present stage. Envelope mass loss would also allow the composition of our envelope to change over time, rather than fixed to its initial composition. We also do not include any extra sources of inflation for highly irradiated gas giants, e.g. ohmic dissipation, which are necessary to match Hot Jupiter radii to observations (Batygin & Stevenson, 2010). Including these inflation mechanism would extend the applicability of our models to this region of parameter space. Finally, we assume Sun-like properties for the host star and its radiation incident upon the planet. This affects the column depth to irradiation through the gaseous mean opacities interpolated within the tables from Freedman et al. (2014), which may have an effect on the output planet radii, particularly for planets orbiting M dwarfs.

5.6 Conclusion

We have evolved a dense grid of planet models across an exhaustive range of planet masses, envelope mass fractions, incident fluxes, and ages. We provide planet radii for our nearly $100,000$ grid points as well as their evolution files generated by MESA. We also identified several trends in the flux-radius relationship, notably the existence of a breakpoint between 1 and $10\,F_{\oplus}$ that depends on planet mass, envelope mass fraction, and age. This grid can be used to study the current composition of exoplanets across a wide range of parameter space, whereas several separate grids might have been previously been necessary. The grid we present may be expanded in the future to accommodate an even wider range of parameter space, including variations in envelope composition, core composition, and mass loss.

CHAPTER 6

CONCLUSION

In this chapter, we first review the limitations of the datasets used in this thesis as well as the capabilities of future surveys in Section 6.1. In Section 6.2, we briefly demonstrate how to extend our joint mass-radius-period distributions developed in Chapters 3 and 4 into a composition distribution using the evolution models presented in Chapter 5. Finally, we summarize the research presented in this thesis in Section 6.3.

6.1 Future Surveys

The current state of exoplanet demographics is data-limited: despite our sample size of thousands of exoplanets, the current data often lacks the ability to distinguish between competing theories or models. In Chapter 2, we were unable to find a dependence on host star mass in the mass-radius relation of exoplanets. While such a dependence may yet exist, the dataset we used was likely too limited to detect a difference. In Chapter 4, we could not distinguish between multiple models that either included or excluded photoevaporation, and either included or excluded water worlds. While this was partly due to issues with model selection methods, we also showed through simulations that improving the dataset could improve our ability to discriminate between models.

The dataset can be improved in several ways. One issue is the lack of coverage in certain regions of parameter space due to low efficiency of detection, for example small planets below $1\,R_\oplus$ or long-period planets with ≥ 1 year periods. In Chapter 3, we extrapolated $\eta_\oplus$ from our models and found highly model-dependent estimates. A direct constraint of $\eta_\oplus$ without the need for extrapolation would reduce the variation in our estimates, but is not currently viable due to the low detection efficiency of *Kepler* at Earth's properties. Our mass-radius dataset in Chapter 2 only had six planets around low-mass stars, due to the low sensitivity

to *Kepler* and other early transit surveys to planets around faint stars. This inhibited our ability to detect a dependence of the mass-radius relation on host star mass, because most of our dataset consisted of planets orbiting FGK dwarfs.

Another issue is the simple lack of data, more relevant for the mass plane than for the radius plane. For example, in Chapters 3 and 4 our transiting planet catalog consisted of 1130 planets with radius measurements, but only 68 with mass measurements, an order of magnitude lower. In Chapter 2, the dataset of planets with both radius and mass measurements (RV combined with TTV) had only 143 planets. If every transiting planet in our sample had a radial velocity mass measurement, our ability to constrain the mass-radius plane would be greatly improved.

Finally, uncertainties on mass and radius measurements are still too high to definitely distinguish between possible compositions. In Chapters 3 and 4, the median mass uncertainty was 27% and the median radius uncertainty was 4.8%, which contributed to the degeneracies between our compositional subpopulations. The power of lowering measurement uncertainties has already been demonstrated through the discovery of the Fulton gap, which was enabled by using the CKS sample, consisting of planets with lower radius uncertainties than typical of the full *Kepler* sample.

Ongoing and future transit surveys will serve to increase the size of our sample of transiting planets, increase our sensitivity to planets orbiting low-mass stars, as well as provide higher precision radii. The *K2* mission made use of the *Kepler* spacecraft after its nominal mission to observe 18 different fields, identifying 747 planet candidates, 366 of them previously undiscovered (Zink et al., 2021). Since it only observed each field for 80 days, as well as having to deal with additional detrending compared to *Kepler*, *K2* has poorer sensitivity to longer-period and smaller planets than *Kepler*. This sample nonetheless increases the total number of transiting planets, and has a greater proportion of planets orbiting low-mass stars as well as planets in different regions of the galactic sub-structure.

TESS is an ongoing all-sky transit survey looking at the brightest nearby stars that will have increased sensitivity to planets orbiting low-mass stars. At the time of writing, *TESS* has identified over 4000 exoplanet candidates, with over 150 confirmed exoplanets (Guerrero et al., 2021). *TESS* has also driven the campaign to characterize more transiting planets through radial velocity, aiming to have characterize 50 small planets below $4R_\oplus$. Launched in 2018 and extended to run through 2022, *TESS* will have generally worse ability to detect long-period planets than *Kepler*, and has a similar drop-off in detection efficiency below $1\,R_\oplus$. Despite these limitations, *TESS* offers a dataset comparable in size to *Kepler*, will dramatically increase the sample of planets orbiting low-mass stars, as well as the sample of small planets with radial velocity mass measurements.

Looking to the near future, *PLATO* is a space telescope scheduled for launch in 2026, designed to detect planetary transits around bright stars. The *PLATO* mission has a goal of detecting 7000 transiting systems, with radial velocity mass measurements of 400 of those planets to a rough precision of 10%, in addition to TTV mass measurements of an undetermined subset of planets. *PLATO* also pushes the bounds in radius and period compared to *Kepler*, with the goal of detecting $1\,R_\oplus$ planets orbiting in the habitable zone of G-type stars to 3% precision (PLATO Study Team, 2017). This increased sensitivity to long-period planets compared to *Kepler* and *TESS* is due to its nominal mission lifetime of 4 years, with a possible extension to 8 years. *PLATO* will represent a next-generation upgrade to *Kepler* by offering a larger dataset with higher precision radii, a massively expanded RV sample with higher precision masses, and improved sensitivity to smaller planets on longer orbits.

Unlike telescopes designed for large-scale transit surveys like *Kepler* and *TESS*, radial velocity spectrographs do not require space telescopes and can be installed on ground-based telescopes. As such, there are a greater number of current and planned spectrographs that intend to push the envelope of radial velocity measurements to higher precision. The radial velocity signal of a G dwarf star hosting an Earth-sized planet in the habitable zone is roughly

$10\,\mathrm{cm\,s^{-1}}$, a benchmark yet to be reached by the capabilities of current instruments. Modern spectrographs, such as MAROON-X, can push down below $30\,\mathrm{cm\,s^{-1}}$ single-measurement precision in ideal cases (Seifahrt et al., 2020), but may be limited by the level of stellar activity in the radial velocity signal (Gilbertson et al., 2020). A host of extremely precise radial velocity spectrographs are currently online or under development including ESPRESSO (Pepe et al., 2013), EXPRES (Jurgenson et al., 2016), NEID (Schwab et al., 2016), KPF (Gibson et al., 2016), and G-CLEF (Szentgyorgyi et al., 2016), among others. Having a larger sample size of planet mass measurements, with greater precision measurements and increased sensitivity to low-mass planets, will significantly improve our ability to distinguish between degenerate compositions in the mass-radius-period plane.

Alongside transit surveys and radial velocity campaigns, microlensing surveys offer a separate sample of planets that has unique sensitivity to regions of parameter space not explored by the other techniques. The Nancy Grace Roman Space Telescope (Roman) will include a microlensing survey that is predicted to discover roughly 1400 planets (Penny et al., 2019). Microlensing is sensitive to planets of much lower mass than transits or radial velocity, with Roman expected to have sensitivity down to $0.02\,M_\oplus$, roughly the mass of Ganymede. Additionally, whereas transits and radial velocities have greater sensitivity towards shorter orbital periods, microlensing is sensitive to cold exoplanets beyond orbits of 1 AU. Since microlensing is not sensitive to the light from the host star, it also has greater ability to detect planets around the lowest-mass stars. Finally, microlensing has the ability to discover free-floating planets, with Roman estimated to detect about 250 free-floating planets with masses as low as Mars (Johnson et al., 2020). Roman will characterize the exoplanet population in a completely new region of parameter space, centered around the snow-line at several AU and down to lower planet masses, complementing the capabilities of transit surveys and radial velocity campaigns.

This thesis primarily used the catalog from a single exoplanet survey, *Kepler*. At the

time of writing, *Kepler* provided the largest sample of planets with a well-characterized detection pipeline, with no comparable alternative. With the advent of additional catalogs from surveys like *K2*, *TESS*, *PLATO*, and Roman, we will have multiple large catalogs of planets for the purposes of statistical analysis. While each catalog can be analyzed on its own, combining these catalogs in a self-consistent manner will yield greater insights. Doing so is non-trivial. Each survey will have its own detection efficiency function that requires separate characterization and treatment. Constraining exoplanet populations from a combination of catalogs may be more amenable to a forward-modeling approach than the inhomogeneous Poisson process likelihood technique used in Chapters 3 and 4 (Clanton & Gaudi, 2016). The sheer size of the combined catalog may also strain the analysis techniques used in this thesis. Regardless, combined constraints from complementary surveys will push the field even further.

6.2 Composition Distribution

The joint mass-radius-period distribution that we constrained in Chapters 3 and 4 is a distribution in the observable plane of exoplanet properties: mass obtained from radial velocity and transit timing variations, and radius and period obtained from transits. Mass is a fundamental quantity of a planet, similar to how mass is fundamental to a star. However, an exoplanet's radius is a consequence of processes during its formation and evolution, and is dependent on its composition as well as the light it receives from its host star. A more fundamental joint distribution of exoplanets would directly constrain the axes of composition and incident flux, rather than radius and period.

In order to constrain this composition distribution, we require some way to map between the fundamental plane (mass, composition, incident flux) to the observable plane (mass, radius, period). Theoretical mass-radius curves serve just this purpose. Planet interior structure models have long been used to produce mass-radius relations for planet cores of

varying composition (e.g., Seager et al., 2007), where radius is largely independent of flux and age. Planets with significant hydrogen-helium envelopes, on the other hand, have radii that largely depend on the evolution of their envelopes over time, and are significantly affected by the amount of light the envelope receives. In Chapter 5 we evolved a dense, comprehensive grid of models across mass, envelope mass fraction, incident flux, and age. These models will be able to serve as the mapping from the fundamental to the observable plane for planets with gaseous envelopes.

In addition to this grid of evolution models, we require an interpolating function that takes as input any combination of input parameters and outputs a radius. While not presented in this thesis, we have developed a Gaussian process emulator that serves this purpose. The Gaussian process emulator returns credibility bounds for a given input, allowing us to quantify levels of uncertainty in the interpolation. We have also incorporated monotonicity constraints in the emulator, given that radius is shown to be monotonically decreasing with age and monotonically increasing with incident flux in our models. The hyperparameters of the emulator are fit using a separate training set that is randomly sampled from our input parameter space.

With this emulator, we will be able to construct a composition distribution that builds upon the mass-radius-period models we developed in Chapters 3 and 4. The mass dimension, shared between the two distributions, can be modeled with a log-normal distribution as before. Instead of a mass-radius relation, we can constrain a mass-envelope mass fraction relation. More massive planets (those that have undergone runaway gas accretion) typically have greater envelope mass fractions, whereas low-mass planets can only sustain an envelope of a certain mass before it is blown away. Incident flux is largely analogous to period and can be constrained in a similar way. The composition distribution can also be modeled using mixture models, with subpopulations that do not possess gaseous envelopes able to be modeled largely as before.

In order to incorporate a statistical emulator into our modeling, we may need to move beyond the inhomogeneous Poisson process likelihood used in Chapters 3 and 4. Approximate Bayesian computation (ABC) is an alternative, having been applied to constrain *Kepler* occurrence rates in previous studies (Kunimoto & Bryson, 2020; Hsu et al., 2020). However, these studies have not applied any model selection methods, and given our simulations in Chapter 4, care must be taken to ensure any model selection techniques we use are valid when used with ABC.

In addition to exploring the composition distribution, our work can be extended to constrain exoplanet distributions as a function of host star properties by including data from future surveys with greater sensitivity to planets around low-mass stars. Expanding from three dimensions as in our models in Chapters 3 and 4 to four dimensions is a nontrivial step. The integral in the inhomogeneous Poisson process likelihood used in Chapters 3 and 4 is a significant computational bottleneck and may not be able to withstand an additional dimension. This provides additional motivation to use alternative techniques to constrain occurrence rates that can more readily handle high dimensionality. ABC, introduced as a possibility for constraining the composition distribution, may also be relevant for this purpose.

6.3 Summary

This thesis has explored the imprint that exoplanet composition leaves on the mass-radius-period plane through the use of hierarchical Bayesian modeling applied to observational catalogs of transiting exoplanets. In Chapter 2, we investigate whether the host star mass affects average planet composition through the mass-radius relation, using the dataset of transiting planets from *Kepler* and other small-scale transit surveys that have radial velocity or transit timing variation measurements on their mass. We find that the dataset is consistent with no host star mass dependence, and use this relation that allows for a dependence to

confirm that the amount of planetary heavy-element mass in short orbits is higher for M dwarf stars than FGK stars.

In Chapters 3 and 4, we move on from the mass-radius relation to constrain the joint mass-radius-period distribution, using the CKS sample of *Kepler* planets with high-precision radius constraints, a subset of which have radial velocity mass measurements. We employ mixture models that incorporate multiple compositional subpopulations as well as photoevaporation to fit this 3D distribution. In Chapter 3, we model three subpopulations: planets with gaseous envelopes, evaporated rocky cores, and intrinsically rocky planets. We use k-fold cross-validation to perform model selection and find that models that include all three are preferred over models that only include one or two of these subpopulations. Additionally, we apply our models to calculate exoplanet occurrence rates and find that our calculations are highly model dependent, with $\eta_\oplus$ specifically decreasing by an order of magnitude when photoevaporation is included in the model.

In Chapter 4, we expand upon the three subpopulations from Chapter 3 to further include planets with icy core compositions, bringing the total number of subpopulations to six. We create a suite of ten models that include these ten subpopulations in different combinations, ranging in complexity from a model with a single population to a model with all six subpopulations. We find that k-fold cross-validation is unable to distinguish between models with or without photoevaporation, and with or without icy composition subpopulations. This is in part due to the degeneracies inherent to these subpopulations in the mass-radius-period plane, but also in part due to issues applying model selection techniques to mixture models. We find the fraction of icy composition planets to be highly model-dependent, ranging from 0% to 50% depending on which subset of subpopulations are included in the model.

Moving beyond the mass-radius-period distribution to more directly model composition necessitates the use of planetary evolution models to translate between the properties inherent to composition (mass, envelope mass fraction, incident flux, age) and the observable

plane (mass, radius, period). In Chapter 5, we use MESA to evolve a dense, comprehensive grid of planet models that can produce such a mapping. Our 4D grid spans four orders of magnitude in planet mass, five orders of magnitude in incident flux, envelope mass fractions from 0.001 to 0.999, and ages from 100 Myr to 10 Gyr. We provide tabulated radii across this 4D grid for nearly $100,000$ grid points. Additionally, we identify a breakpoint in the flux-radius relation between 1 and $10\,F_{\oplus}$ that depends on planet mass, envelope mass fraction and age. Our grid can be used to model the current composition of exoplanets across a wide range of parameter space and can serve as a point of comparison for evolution models that include higher-order effects.

The ability to constrain the mass-radius relation, the mass-radius-period distribution, or the composition distribution, is reliant upon the statistical sample available to us. Future instruments and surveys, including transit surveys like *TESS* and *PLATO*, the Nancy Grace Roman Space Telescope microlensing survey, and extremely precise radial velocity spectrographs like *EXPRES* and *KPF* will enhance our ability to constrain these distributions. Combining catalogs from multiple surveys will allow us to cover a wider range of parameter space across mass, radius, period, and host star mass. In addition to the incorporation of improved datasets, our work can be extended by modeling the planet composition distribution through the use of a statistical emulator to interpolate between planet evolution models.

REFERENCES

Adams, E. R., Seager, S., & Elkins-Tanton, L. 2008, ApJ, 673, 1160

Alibert, Y., & Benz, W. 2017, A&A, 598, L5

Andrews, S. M., Rosenfeld, K. A., Kraus, A. L., & Wilner, D. J. 2013, ApJ, 771, 129

Ansdell, M., Ioannou, Y., Osborn, H. P., et al. 2018, ApJ, 869, L7

Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, ApJS, 204, 24

Batygin, K., & Stevenson, D. J. 2010, ApJ, 714, L238

Batygin, K., & Stevenson, D. J. 2013, The Astrophysical Journal, 769, L9

Berardo, D., & Cumming, A. 2017, The Astrophysical Journal, 846, L17

Berta, Z. K., Irwin, J., Charbonneau, D., Burke, C. J., & Falco, E. E. 2012, AJ, 144, 145

Bitsch, Bertram, Raymond, Sean N., Buchhave, Lars A., et al. 2021, A&A, 649, L5

Bitsch, Bertram, Raymond, Sean N., & Izidoro, Andre. 2019, A&A, 624, A109

Bodenheimer, P., & Lissauer, J. J. 2014, ApJ, 791, 103

Bonfils, X., Delfosse, X., Udry, S., et al. 2013, A&A, 549, A109

Bonfils, X., Almenara, J. M., Jocou, L., et al. 2015, in Proc. SPIE, Vol. 9605, Techniques and Instrumentation for Detection of Exoplanets VII, 96051L

Borucki, W. J., Koch, D. G., Basri, G., et al. 2011a, ApJ, 728, 117

—. 2011b, ApJ, 736, 19

Buhler, P. B., Knutson, H. A., Batygin, K., et al. 2016, ApJ, 821, 26

Burdanov, A., Delrez, L., Gillon, M., & Jehin, E. 2017, ArXiv e-prints

Burke, C. J., & Catanzarite, J. 2017, Planet Detection Metrics: Per-Target Detection Contours for Data Release 25, Tech. rep.

Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, ApJ, 809, 8

Burt, J., Holden, B., Wolfgang, A., & Bouma, L. G. 2018, AJ, 156, 255

Carpenter, B., Gelman, A., Hoffman, M. D., et al. 2017, Journal of Statistical Software, Articles, 76, 1

Chabrier, G., & Baraffe, I. 2007, ApJ, 661, L81

Chatterjee, S., & Chen, H. 2018, The Astrophysical Journal, 852, 58

Chen, H., & Rogers, L. A. 2016, ApJ, 831, 180

Chen, J., & Kipping, D. 2017, ApJ, 834, 17

Chiang, E., & Laughlin, G. 2013, MNRAS, 431, 3444

Chiang, E., & Youdin, A. N. 2010, Annual Review of Earth and Planetary Sciences, 38, 493

Christiansen, J. L. 2017, Planet Detection Metrics: Pixel-Level Transit Injection Tests of Pipeline Detection Efficiency for Data Release 25, Tech. rep.

Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., et al. 2012, PASP, 124, 1279

Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2015, ApJ, 810, 95

—. 2020, AJ, 160, 159

Christille, J.-M., Bernagozzi, A., Bertolini, E., et al. 2013, in European Physical Journal Web of Conferences, Vol. 47, European Physical Journal Web of Conferences, 17001

Ciardi, D. R., von Braun, K., Bryden, G., et al. 2011, AJ, 141, 108

Ciesla, F. J., Mulders, G. D., Pascucci, I., & Apai, D. 2015, ApJ, 804, 9

Clanton, C., & Gaudi, B. S. 2016, ApJ, 819, 125

Dai, F., Masuda, K., Winn, J. N., & Zeng, L. 2019, ApJ, 883, 79

Dederick, E., & Jackiewicz, J. 2017, The Astrophysical Journal, 837, 148

Dong, S., Xie, J.-W., Zhou, J.-L., Zheng, Z., & Luo, A. 2018, Proceedings of the National Academy of Science, 115, 266

Dressing, C. D., & Charbonneau, D. 2015, ApJ, 807, 45

Dressing, C. D., Charbonneau, D., Dumusque, X., et al. 2015, ApJ, 800, 135

Foreman-Mackey, D. 2016, The Journal of Open Source Software, 24

Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, ApJ, 795, 64

Fortney, J. J., Marley, M. S., & Barnes, J. W. 2007a, ApJ, 668, 1267

—. 2007b, ApJ, 659, 1661

Fortney, J. J., Mordasini, C., Nettelmann, N., et al. 2013, ApJ, 775, 80

Freedman, R. S., Lustig-Yaeger, J., Fortney, J. J., et al. 2014, ApJS, 214, 25

Fressin, F., Torres, G., Charbonneau, D., et al. 2013, ApJ, 766, 81

Fulton, B. J., & Petigura, E. A. 2018, AJ, 156, 264

Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, AJ, 154, 109

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1

Gaidos, E., Mann, A. W., Kraus, A. L., & Ireland, M. 2016, MNRAS, 457, 2877

Gelman, A., Hwang, J., & Vehtari, A. 2014, Statistics and Computing, 24, 997

Gibson, S. R., Howard, A. W., Marcy, G. W., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, ed. C. J. Evans, L. Simard, & H. Takami, 990870

Gilbertson, C., Ford, E. B., Jones, D. E., & Stenning, D. C. 2020, ApJ, 905, 155

Ginzburg, S., Schlichting, H. E., & Sari, R. 2018, MNRAS, 476, 759

González Hernández, J. I., Pepe, F., Molaro, P., & Santos, N. 2017, ArXiv e-prints

Guerrero, N. M., Seager, S., Huang, C. X., et al. 2021, arXiv e-prints, arXiv:2103.12538

Gupta, A., & Schlichting, H. E. 2019, MNRAS, 487, 24

—. 2020, MNRAS, 493, 792

Hastie, T., Tibshirani, R., & Friedman, J. 2001, The Elements of Statistical Learning, Springer Series in Statistics (New York, NY, USA: Springer New York Inc.)

He, J., & Fan, X. 2019, Structural Equation Modeling: A Multidisciplinary Journal, 26, 66

Hoffman, M. D., & Gelman, A. 2014, Journal of Machine Learning Research, 15, 1593

Holman, M. J., Fabrycky, D. C., Ragozzine, D., et al. 2010, Science, 330, 51

Howard, A. W., Marcy, G. W., Johnson, J. A., et al. 2010, Science, 330, 653

Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, ApJS, 201, 15

Howe, A. R., & Burrows, A. 2015, ApJ, 808, 150

Hsu, D. C., Ford, E. B., Ragozzine, D., & Ashby, K. 2019, arXiv e-prints, arXiv:1902.01417

Hsu, D. C., Ford, E. B., Ragozzine, D., & Morehead, R. C. 2018, AJ, 155, 205

Hsu, D. C., Ford, E. B., & Terrien, R. 2020, MNRAS, 498, 2249

Ida, S., & Lin, D. N. C. 2005, ApJ, 626, 1045

Ikoma, M., & Hori, Y. 2012, ApJ, 753, 66

Jackson, B., Arras, P., Penev, K., Peacock, S., & Marchant, P. 2017, ApJ, 835, 145

Jackson, B., Jensen, E., Peacock, S., Arras, P., & Penev, K. 2016, Celestial Mechanics and Dynamical Astronomy, 126, 227

Johansen, A., & Lambrechts, M. 2017, Annual Review of Earth and Planetary Sciences, 45, 359

Johnson, J. A., Petigura, E. A., Fulton, B. J., et al. 2017, AJ, 154, 108

Johnson, S. A., Penny, M., Gaudi, B. S., et al. 2020, AJ, 160, 123

Jontof-Hutter, D., Lissauer, J. J., Rowe, J. F., & Fabrycky, D. C. 2014, ApJ, 785, 15

Jurgenson, C., Fischer, D., McCracken, T., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, ed. C. J. Evans, L. Simard, & H. Takami, 99086T

Kains, N., Bramich, D. M., Sahu, K. C., & Calamida, A. 2016, MNRAS, 460, 2025

Kennedy, G. M., & Kenyon, S. J. 2008, ApJ, 673, 502

Kipping, D. M., & Lam, C. 2017, MNRAS, 465, 3495

Kite, E. S., & Ford, E. B. 2018, ApJ, 864, 75

Kite, E. S., & Schaefer, L. 2021, ApJ, 909, L22

Konopacky, Q. M., Rameau, J., Duchêne, G., et al. 2016, ApJ, 829, L4

Kraus, A. L., Ireland, M. J., Huber, D., Mann, A. W., & Dupuy, T. J. 2016, AJ, 152, 8

Kubyshkina, D., Vidotto, A. A., Fossati, L., & Farrell, E. 2020, MNRAS, 499, 77

Kuchner, M. J. 2003, ApJ, 596, L105

—. 2004, ApJ, 612, 1147

Kunimoto, M., & Bryson, S. 2020, Research Notes of the American Astronomical Society, 4, 83

Léger, A., Selsis, F., Sotin, C., et al. 2004, Icarus, 169, 499

Linder, E. F., Mordasini, C., Mollière, P., et al. 2019, A&A, 623, A85

Lissauer, J. J. 2007, ApJ, 660, L149

Lissauer, J. J., Fabrycky, D. C., Ford, E. B., et al. 2011, Nature, 470, 53

Lissauer, J. J., Jontof-Hutter, D., Rowe, J. F., et al. 2013, ApJ, 770, 131

Lopez, E. D., & Fortney, J. J. 2013, ApJ, 776, 2

—. 2014, ApJ, 792, 1

Lopez, E. D., Fortney, J. J., & Miller, N. 2012, ApJ, 761, 59

Loredo, T. J. 2004, in American Institute of Physics Conference Series, Vol. 735, American Institute of Physics Conference Series, ed. R. Fischer, R. Preuss, & U. V. Toussaint, 195–206

Luger, R., Barnes, R., Lopez, E., et al. 2015, Astrobiology, 15, 57

Ma, Q., & Ghosh, S. K. 2021, MNRAS, 505, 3853

Malsky, I., & Rogers, L. A. 2020, ApJ, 896, 48

Mandel, I., Farr, W. M., & Gair, J. R. 2019, MNRAS, 486, 1086

Mandel, K., & Agol, E. 2002, The Astrophysical Journal Letters, 580, L171

Marcy, G., Isaacson, H., Howard, A. W., et al. 2014, ApJS, 210, 20

Marleau, G. D., & Cumming, A. 2014, MNRAS, 437, 1378

Marley, M. S., Fortney, J. J., Hubickyj, O., Bodenheimer, P., & Lissauer, J. J. 2007, ApJ, 655, 541

Martinez, C. F., Cunha, K., Ghezzi, L., & Smith, V. V. 2019, ApJ, 875, 29

Mayor, M., Marmier, M., Lovis, C., et al. 2011, arXiv e-prints, arXiv:1109.2497

Miller, N., & Fortney, J. J. 2011, ApJ, 736, L29

Miller, N., Fortney, J. J., & Jackson, B. 2009, ApJ, 702, 1413

Millholland, S., Wang, S., & Laughlin, G. 2017, ApJ, 849, L33

Mills, S. M., & Mazeh, T. 2017, ApJ, 839, L8

Montet, B. T., Yee, J. C., & Penny, M. T. 2017, Publications of the Astronomical Society of the Pacific, 129, 044401

Mousis, O., Deleuil, M., Aguichine, A., et al. 2020, ApJ, 896, L22

Moutou, C., Boisse, I., Hébrard, G., et al. 2015, in SF2A-2015: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, ed. F. Martins, S. Boissier, V. Buat, L. Cambrésy, & P. Petit, 205–212

Muñoz, D. J., & Perets, H. B. 2018, AJ, 156, 253

Mulders, G. D., Pascucci, I., & Apai, D. 2015a, ApJ, 814, 130

—. 2015b, ApJ, 798, 112

Mulders, G. D., Pascucci, I., Apai, D., & Ciesla, F. J. 2018, AJ, 156, 24

Mulders, G. D., Pascucci, I., Apai, D., Frasca, A., & Molenda-Żakowicz, J. 2016, AJ, 152, 187

Mullally, F., Coughlin, J. L., Thompson, S. E., et al. 2015, ApJS, 217, 31

Müller, S., & Helled, R. 2021, MNRAS

Neil, A. R., & Rogers, L. A. 2018, ApJ, 858, 58

—. 2020, ApJ, 891, 12

Ning, B., Wolfgang, A., & Ghosh, S. 2018, ApJ, 869, 5

Otegi, J. F., Bouchy, F., & Helled, R. 2020, A&A, 634, A43

Owen, J. E., & Murray-Clay, R. 2018, MNRAS, 480, 2206

Owen, J. E., & Wu, Y. 2013, ApJ, 775, 105

Pascucci, I., Mulders, G. D., Gould, A., & Fernandes, R. 2018, ApJ, 856, L28

Paxton, B., Bildsten, L., Dotter, A., et al. 2011, ApJS, 192, 3

Paxton, B., Cantiello, M., Arras, P., et al. 2013, ApJS, 208, 4

Paxton, B., Marchant, P., Schwab, J., et al. 2015, ApJS, 220, 15

Paxton, B., Schwab, J., Bauer, E. B., et al. 2018, ApJS, 234, 34

Paxton, B., Smolec, R., Schwab, J., et al. 2019, ApJS, 243, 10

Penny, M. T., Gaudi, B. S., Kerins, E., et al. 2019, ApJS, 241, 3

Pepe, F., Cristiani, S., Rebolo, R., et al. 2013, The Messenger, 153, 6

Petigura, E. A., Howard, A. W., Marcy, G. W., et al. 2017, AJ, 154, 107

Petigura, E. A., Marcy, G. W., Winn, J. N., et al. 2018, AJ, 155, 89

PLATO Study Team. 2017, PLATO Defintion Study Report, Tech. rep., ESA

Quirrenbach, A., Amado, P. J., Caballero, J. A., et al. 2014, in Proc. SPIE, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, 91471F

Raymond, S. N., Barnes, R., & Mandell, A. M. 2008, Monthly Notices of the Royal Astronomical Society, 384, 663

Raymond, S. N., Boulet, T., Izidoro, A., Esteves, L., & Bitsch, B. 2018, MNRAS, 479, L81

Raymond, S. N., & Cossou, C. 2014, MNRAS, 440, L11

Raymond, S. N., Scalo, J., & Meadows, V. S. 2007, The Astrophysical Journal, 669, 606

Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

Rogers, J. G., Gupta, A., Owen, J. E., & Schlichting, H. E. 2021, arXiv e-prints, arXiv:2105.03443

Rogers, L. A. 2015, ApJ, 801, 41

Rogers, L. A., Bodenheimer, P., Lissauer, J. J., & Seager, S. 2011, ApJ, 738, 59

Rogers, L. A., & Seager, S. 2010a, ApJ, 712, 974

—. 2010b, ApJ, 716, 1208

Sahlmann, J., Lazorenko, P. F., Ségransan, D., et al. 2013, A&A, 556, A133

Sanchis-Ojeda, R., Rappaport, S., Winn, J. N., et al. 2014, ApJ, 787, 47

Scalo, J., Kaltenegger, L., Segura, A. G., et al. 2007, Astrobiology, 7, 85

Schwab, C., Rakich, A., Gong, Q., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, ed. C. J. Evans, L. Simard, & H. Takami, 99087H

Seager, S., Kuchner, M., Hier-Majumder, C. A., & Militzer, B. 2007, ApJ, 669, 1279

Seifahrt, A., Bean, J. L., Stürmer, J., et al. 2016, in Proc. SPIE, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, 990818

Seifahrt, A., Stürmer, J., Bean, J. L., & Schwab, C. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10702, Ground-based and Airborne Instrumentation for Astronomy VII, ed. C. J. Evans, L. Simard, & H. Takami, 107026D

Seifahrt, A., Bean, J. L., Stürmer, J., et al. 2020, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 11447, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 114471F

Sestovic, M., Demory, B.-O., & Queloz, D. 2018, A&A, 616, A76

Steffen, J. H. 2016, MNRAS, 457, 4384

Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., et al. 2015, ApJ, 809, 77

Suzuki, D., Bennett, D. P., Sumi, T., et al. 2016, ApJ, 833, 145

Szentgyorgyi, A., Baldwin, D., Barnes, S., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, ed. C. J. Evans, L. Simard, & H. Takami, 990822

Thompson, S. E., Coughlin, J. L., Hoffman, K., et al. 2018, ApJS, 235, 38

Thorngren, D. P., & Fortney, J. J. 2018, AJ, 155, 214

Thorngren, D. P., Fortney, J. J., Murray-Clay, R. A., & Lopez, E. D. 2016, ApJ, 831, 64

Usami, S. 2014, Journal of the Japanese Society of Computational Statistics, 27, 17

Valencia, D., O'Connell, R. J., & Sasselov, D. 2006, Icarus, 181, 545

Valletta, C., & Helled, R. 2020, ApJ, 900, 133

Valsecchi, F., Rappaport, S., Rasio, F. A., Marchant, P., & Rogers, L. A. 2015, ApJ, 813, 101

Valsecchi, F., Rasio, F. A., & Steffen, J. H. 2014, ApJ, 793, L3

Van Eylen, V., Agentoft, C., Lundkvist, M. S., et al. 2018, MNRAS, 479, 4786

Vehtari, A., Gelman, A., & Gabry, J. 2017, Statistics and Computing, 27, 1413

Wang, J., & Fischer, D. A. 2015, AJ, 149, 14

Wang, J., Xie, J.-W., Barclay, T., & Fischer, D. A. 2014, ApJ, 783, 4

Watanabe, S. 2010, J. Mach. Learn. Res., 11, 3571

Watanabe, S. 2012, ArXiv e-prints

Weiss, L. M., & Marcy, G. W. 2014, ApJ, 783, L6

Weiss, L. M., Marcy, G. W., Rowe, J. F., et al. 2013, ApJ, 768, 14

Weiss, L. M., Marcy, G. W., Petigura, E. A., et al. 2018, AJ, 155, 48

Wolfgang, A., Rogers, L. A., & Ford, E. B. 2016, ApJ, 825, 19

Wu, Y., & Lithwick, Y. 2013, ApJ, 772, 74

Zeng, L., Sasselov, D., & Jacobsen, S. 2015, ArXiv e-prints

Zeng, L., Jacobsen, S. B., Sasselov, D. D., et al. 2019, Proceedings of the National Academy of Science, 116, 9723

Zink, J. K., Hardegree-Ullman, K. K., Christiansen, J. L., et al. 2021, arXiv e-prints, arXiv:2109.02675