PROBABILISTIC MASS-RADIUS RELATIONSHIP FOR SUB-NEPTUNE-SIZED PLANETS

Angie Wolfgang[1,2,6,7], Leslie A. Rogers[3,4,5], and Eric B. Ford[6,7,8]
Submitted to the Astrophysical Journal on April 27, 2015

ABSTRACT

The Kepler Mission has discovered thousands of planets with radii < 4 R$_\oplus$, paving the way for the first statistical studies of the dynamics, formation, and evolution of these sub-Neptunes and super-Earths. Planetary masses are an important physical property for these studies, and yet the vast majority of *Kepler* planet candidates do not have theirs measured. A key concern is therefore how to map the measured radii to mass estimates in this Earth-to-Neptune size range where there are no Solar System analogs. Previous works have derived deterministic, one-to-one relationships between radius and mass. However, if these planets span a range of compositions as expected, then an intrinsic scatter about this relationship must exist in the population. Here we present the first probabilistic mass-radius relationship (M-R relation) evaluated within a Bayesian framework, which both quantifies this intrinsic dispersion and the uncertainties on the M-R relation parameters. We analyze how the results depend on the radius range of the sample, and on how the masses were measured. Assuming that the M-R relation can be described as a power law with a dispersion that is constant and normally distributed, we find that $M/M_\oplus = 2.7(R/R_\oplus)^{1.3}$, a scatter in mass of $1.9 M_\oplus$, and a mass constraint to physically plausible densities, is the "best-fit" probabilistic M-R relation for the sample of RV-measured transiting sub-Neptunes ($R_{pl} < 4$ R$_\oplus$). More broadly, this work provides a framework for further analyses of the M-R relation and its probable dependencies on period and stellar properties.

Keywords: planets and satellites: composition — methods: statistical

1. INTRODUCTION

The *Kepler Mission* has found thousands of planetary candidates with sizes between that of Earth and Neptune (Mullally et al. 2015; Rowe et al. 2015; Burke et al. 2014; Batalha et al. 2013; Borucki et al. 2011). The emergence of this population poses fundamental questions about the typical compositional constituents of planets within a few times Earth's size. As bulk densities offer some insight into this problem, mass and radius measurements of individual planets have provided observational constraints for theoretical composition studies performed on a per-planet basis (e.g. Valencia et al. 2010; Rogers & Seager 2010; Lopez et al. 2012). Recently these studies have shifted to considering the available planets as a statistical ensemble (e.g. Rogers 2015; Wolfgang & Lopez 2015 sans mass constraints), which motivates detailed analyses of the observed mass-radius distribution.

The joint planetary mass-radius distribution, which is

akw5014@psu.edu
[1] Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064, USA
[2] NSF Graduate Research Fellow; NSF Astronomy & Astrophysics Postdoctoral Fellow
[3] Department of Astronomy and Division of Geological and Planetary Sciences, California Institute of Technology, MC249-17, 1200 East California Boulevard, Pasadena, CA 91125, USA
[4] Department of Astronomy, University of California Berkeley, 501 Campbell Hall #3411, Berkeley, CA 94720, USA
[5] Hubble Fellow; Sagan Fellow
[6] Center for Exoplanets and Habitable Worlds, 525 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802, USA
[7] Department of Astronomy and Astrophysics, The Pennsylvania State University, 525 Davey Laboratory, University Park, PA 16802, USA
[8] Center for Astrostatistics, 417 Thomas Building, The Pennsylvania State University, University Park, PA 16802, USA

often couched in terms of the mass-radius "relationship" (M-R relation), is also highly relevant for dynamical and formation studies of the *Kepler* planet candidates (PCs). Mass measurements for individual PCs are often unavailable, as the majority orbit stars too faint for Doppler follow-up (Batalha et al. 2010) and only $\sim 6\%$ exhibit transit timing variations (TTVs) at high signal-to-noise ratios (Mazeh et al. 2013). Therefore, a statistical "conversion" is necessary to map observed radii to the masses these studies need.

To date, several M-R relations have been posed in the exoplanet literature. To solve the practical issue described above, Lissauer et al. (2011) fit a power law to Earth and Saturn and found $M = R^{2.06}$, where M and R are in Earth units. Wu & Lithwick (2013) derived masses using the amplitudes of sinusoidal TTVs for 22 planet pairs, and found $M = 3R$. More recently, Weiss & Marcy (2014), hitherto WM14, fit a power law to masses and radii available in the literature, which was dominated by the 42 planets chosen by the *Kepler* team to be followed up with radial velocity measurements (Marcy et al. 2014); they found $M = 2.69R^{0.93}$ for planets with $1.5 < R < 4$ R$_\oplus$.

All of these results were produced via basic least squares regression, which is commonly used in astronomy to fit lines through points. However, this classic technique does not properly account for several issues that are relevant to the small-planet M-R relation: measurement uncertainty in the independent variable (i.e. planet radii), non-detections and upper limits, and intrinsic, astrophysical scatter in the dependent variable (i.e. planet masses). Thankfully, there are solutions to these problems in both the Bayesian and frequentist statistics literature (see §1 of Kelly (2007) for a concise overview). We present an example of one of these techniques which

can be executed using existing numerical algorithms and code (§4), which is effectively a simplified implementation of the Kelly (2007) linear regression scheme.

Of particular interest is the intrinsic scatter that has not been previously characterized. Theoretical work on planet compositions suggest this scatter should exist: thermally evolved rock-hydrogen sub-Neptune internal structure models yield radii mostly independent of mass (Lopez & Fortney 2014), which produces significant mass-radius scatter when a distribution of gaseous mass fractions is present in the population (Wolfgang & Lopez 2015). Furthermore, the diversity of choices for exoplanets' internal structures produces a range of radii at a given mass due only to differences in the layers' compositions (e.g. Seager et al. 2007; Fortney et al. 2007; Rogers et al. 2011).

These theoretical findings motivate us to move beyond deterministic, one-to-one mappings, which are in a sense "mean" relationships. This average behavior is insufficient and inappropriate if one's aim is to argue for a particular physical process based on full distributions of parameters (versus qualitative comparison to observations), or if the purpose is to rule out parts of parameter space, which requires knowledge of the full mass-radius distribution.

Realizing the need to move beyond deterministic mass-radius relations for their own theoretical work, Chatterjee & Tan (2015) derived a piecewise probabilistic M-R relation by fitting the density distribution of planets in four mass bins, and then fit a continuous, yet still deterministic, relation to those results. However, they stop short of computing a relation which is both continuous and probabilistic (which they admit would be ideal), and do not incorporate measurement error, which is significant for small planets. With the hierarchical Bayesian modeling that we employ here, we do both. In the process, we also more fully characterize the uncertainty in the M-R relation based on the current data. The effort to understand this uncertainty is important, as quantifying how well constrained the M-R relation parameters are will be a key metric by which we measure the improvement in our understanding of the M-R distribution, especially as TESS and its follow-up observations produce more individual mass and radius measurements.

In this paper we show how a probabilistic M-R relation can be constructed (§2) and constrained (§4) using any subset of planetary masses and radii (§3). We also highlight the observational evidence for this expected intrinsic scatter and quantify it in a statistically robust way that includes uncertainties on the M-R relation parameters (§5). We discuss the correct usage and some major implications of these findings in §6.

2. MODELING THE M-R RELATION

Power laws are often used to parameterize the M-R relation because they are conceptually and computationally simple and can be easily fit to data using the familiar tool of linear regression. We continue with this choice to facilitate more direct comparisons with previous work and to illustrate how a hierarchical framework enables straightforward extensions to entire families of M-R relations. In addition, we cast this in terms of $M(R)$ instead of $R(M)$ to address the practical problem of estimating masses from *Kepler* radii.
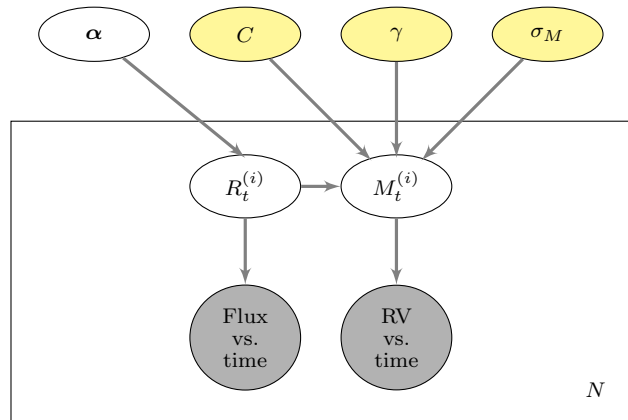

Figure 1. Graphical model used to find the best-fit parameters for the probabilistic mass-radius relationship in Eqn 2. These parameters of interest are yellow while the observed data are gray (see §3) and unobserved parameters are white; definitions are below. In practice, we summarize each planet's "RV vs. time" dataset as the mass measurement $M_{ob}^{(i)}$ and the uncertainty in that measurement, $\sigma_{Mob}^{(i)}$; similarly, the "Flux vs. time" dataset is summarized as the radius measurement $R_t^{(i)}$ and its uncertainty $\sigma_{Rob}^{(i)}$. §6.3 contains further discussion of this choice. Our full hierarchical model, which includes the details of the probability distributions from which each parameter is drawn, is displayed in Equation 4.

$\boldsymbol{\alpha}$ = population-wide radius distribution parameters
C = constant in mean M-R relation
γ = power law index of mean M-R relation
σ_M = intrinsic dispersion in planet masses at a given radius
$R_t^{(i)}$ = true radius of the ith planet
$R_{ob}^{(i)}$ = observed radius of the ith planet
$\sigma_{Rob}^{(i)}$ = measurement uncertainty in $R_{ob}^{(i)}$
$M_t^{(i)}$ = true mass of the ith planet
$M_{ob}^{(i)}$ = observed mass of the ith planet
$\sigma_{Mob}^{(i)}$ = measurement uncertainty in $M_{ob}^{(i)}$

In particular, we consider three power law-based M-R relations (Eqns 1-3). The first is the form used by most prior studies (see §1):

$$\frac{M}{M_\oplus} = C\left(\frac{R}{R_\oplus}\right)^\gamma \tag{1}$$

where M is the mass of the planet, R is the planetary radius, and C and γ are the parameters to be fit to the data. This relation is deterministic in the sense that only one mass is allowed for a given radius.

If instead we want to allow for a range — that is, if we want to incorporate the expected intrinsic scatter — then we need to create an M-R relation which specifies how those masses should be distributed at a given input radius. Again, taking the most simple, familiar, and analytically tractable approach, we choose a Gaussian distribution, where the mean population mass μ is given by the above power-law relation and where the standard deviation σ_M (units of $M_\oplus$) parameterizes the intrinsic scatter in planet masses:

$$\frac{M}{M_\oplus} \sim \text{Normal}\left(\mu = C\left(\frac{R}{R_\oplus}\right)^\gamma, \sigma = \sigma_M\right) \tag{2}$$

Note that $\sim$ means "drawn from the distribution", thereby marking the difference between a deterministic

and a probabilistic M-R relation. Figure 1 is the graphical model corresponding to Eqn 2, and includes Gaussian error bars on the measured masses and radii (see §4 for all details of the model).

Generalizing further, the width of the intrinsic scatter may change as planets increase in size, so we consider a probabilistic M-R relation that allows the standard deviation itself to vary as a function of radius via the slope β (units of $M_\oplus^2 R_\oplus^{-1}$):

$$\frac{M}{M_\oplus} \sim \text{Normal}\Big(\mu = C\Big(\frac{R}{R_\oplus}\Big)^\gamma, \sigma = \sqrt{\sigma_{M1}^2 + \beta\tilde{R}}\Big) \quad (3)$$

where $\tilde{R} = R/R_\oplus - 1$ and σ_{M1} is now the standard deviation in planet masses at 1 $R_\oplus$ ($\tilde{R} = 0$).

3. DATA

With the statistical M-R relations defined, we turn to the problem of identifying which observational dataset to use. Optimally we would use a subset of mass and radius measurements that is uniform and complete, as any systematic biases present in the sample will manifest as biased M-R parameter values. Unfortunately, the available masses and radii are far from this ideal, with mass measurements made with two fundamentally different methods by many different pipelines and chosen for follow-up by a complex, poorly documented selection function. There is significant work to be done to understand how these systematics affect the M-R relation, but it is outside the scope of this paper, as our main purpose is to show how a probabilistic M-R relation can be derived from whichever dataset one wishes to use. Therefore, we choose a baseline dataset consisting of radial velocity-measured masses, which somewhat reduces the heterogeneity of the sample while preserving a fairly large number of data points.

Table 2 shows all of the masses and radii that we consider, with our baseline dataset denoted with a label of 0; the list was constructed by starting with the WM14 dataset and identifying new planets and updates in the NASA Exoplanet Archive (last accessed 1/30/2015). We manually double-checked each planet to verify that the reported measurements were correct and most up-to-date, paying particular attention to which methods and stellar parameters were used (data denoted by a label of 1 were present in and haven't changed since WM14). Given the above concerns with dataset heterogeneity, when both TTV and RV masses are independently available for a single planet, we choose the RV-measured masses. In practice, only Kepler-18b (Cochran et al. 2011) provide strong enough mass constraints from both methods to require a choice to be made, and even then the two mass measurements are consistent. The TTV dataset (label of 2) contains only the sub-Neptune-sized planets that have had their transit timing variations fit with N-body integrations, as these masses are the best constrained and therefore provide the most information for the sub-Neptune M-R relation; neither circumbinary planets nor unconfirmed planets were included, again to try to keep a somewhat more homogeneous dataset. Finally, to enable easier comparison with previous work, we continued the error treatment of WM14: if asymmetric upper and lower uncertainties were reported, we used

their average as a symmetric 1σ error bar[9]. 2σ upper limits were included if they were < 80 $M_\oplus$ for $R < 4$ $R_\oplus$ and < 300 $M_\oplus$ for $4 < R < 8$ $R_\oplus$.

4. FITTING THE M-R RELATIONS

We use hierarchical Bayesian modeling (HBM) to fit the M-R relations in §2 to the data described in §3. This statistical method is described in detail in Wolfgang & Lopez (2015) in the context of exoplanet compositions; further pedagogical discussion and examples of HBM in the astronomical literature is provided by Loredo (2013). A very similar approach to this HBM-enabled linear regression was detailed in Kelly (2007); we refer the reader to that paper for an in-depth discussion of the general advantages and improvements of this approach over the commonly used χ^2 analysis for linear regression.

For the problem at hand, HBM (or the analogous frequentist methods for multi-level modeling) is necessary for a number of reasons:

- It allows us to directly model and fit the astrophysical dispersion in the population as an explicit parameter.

- It allows us to self-consistently incorporate uncertainties on the independent variable (radii in this case), without the need for elaborate bootstrapping schemes.

- Most sub-Neptune mass uncertainties are large, and some are realistically only upper limits. HBM is able to simultaneously use all likelihood distributions no matter their width or shape, which increases the information content of the resulting M-R relation and decreases the biases that binning or weighting schemes introduce when these likelihoods are asymmetric.

- Relatedly, HBM allows us to introduce the true masses and radii as latent (unobserved) parameters; this enables us to restrict the masses to physically allowed parameter space (such as $M > 0$ or $\rho < \rho_{iron}(M)$) while preserving all of the information in the observations, including when the "best-fit" masses happen to be negative.

- As with all Bayesian methods, HBM produces posterior distributions, allowing us to easily see the uncertainties in the M-R relation parameters. Most of the M-R relations currently reported and used in the literature have no published uncertainties.

The hierarchical model for our baseline M-R relation (Eqn 2) is displayed in Figure 1 to clarify the structural relationships between parameters and observables. This structure is also present in the written version below, along with details of the distributions we used ("N" represents a normal distribution with the listed parameters in order of μ and σ; "U" represents a uniform distribution with the listed numbers bounding the interval; and

[9] Future work using HBM can improve on this error treatment by using the full posteriors of the mass and radius measurements, if these posteriors are made available in the literature (see §6.3).
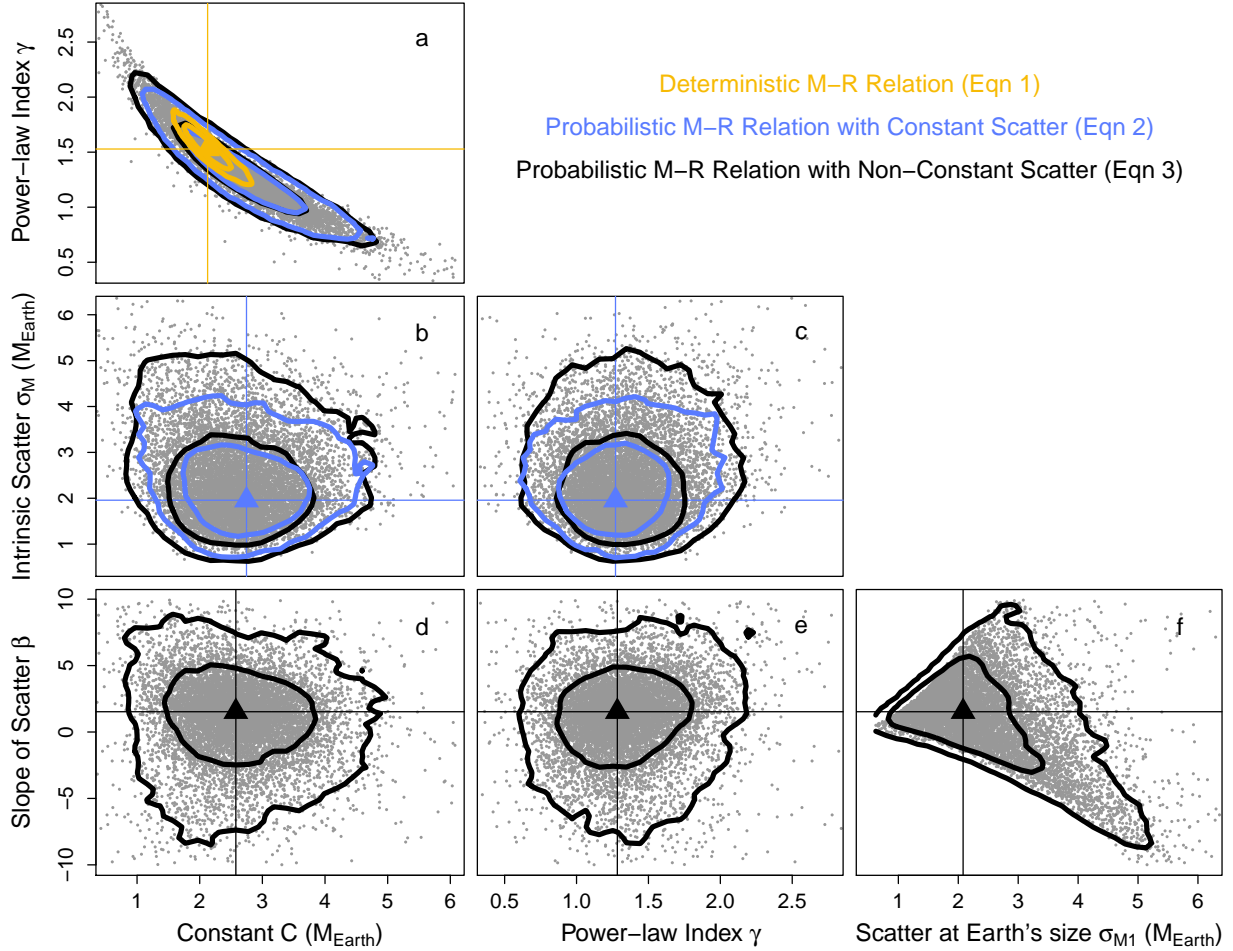
Figure 2. Posteriors for the parameters in our family of M-R relations (row 1: Equations 1-3; row 2: Equations 2-3; row 3 and gray posterior samples in all panels, which contain all 10,000 saved samples from our thinned MCMC chains: Equation 3) when fit to our baseline dataset. 68% and 95% contours are shown for each, and demarcate the uncertainties on these M-R relation parameters; the triangles denote best-fit values. Panels b and c show that $\sigma_M = 0$ is strongly excluded for $R < 4$ R$_\oplus$, and so astrophysical scatter exists in the sub-Neptune M-R relation. Therefore, theoretical studies which require an M-R relation should use a probabilistic one like that of Eqn 2 with one of the sets of parameter values in Table 1.

"|" means "given", i.e. the parameter to the left depends on the parameters to the right):

$$\gamma \sim \mathrm{N}(1,1)$$
$$\ln(C) \sim \mathrm{U}(-3,3)$$
$$\log(\sigma_M^2) \sim \mathrm{U}(-4,2)$$
$$R_t^{(i)} \sim \mathrm{U}(\alpha_1 = 0.1, \alpha_2 = 10)$$
$$\mu_M^{(i)} | R_t^{(i)}, C, \gamma = \gamma \ln(R_t^{(i)}) + \ln(C)$$
$$M_t^{(i)} | R_t^{(i)}, C, \gamma, \sigma_M \sim \mathrm{N}\left(e^{\mu_M^{(i)}}, \sigma_M\right)$$
$$R_{ob}^{(i)} | R_t^{(i)}, \sigma_{Rob}^{(i)} \sim \mathrm{N}(R_t^{(i)}, \sigma_{Rob}^{(i)})$$
$$M_{ob}^{(i)} | M_t^{(i)}, \sigma_{Mob}^{(i)}, R_t^{(i)}, C, \gamma, \sigma_M \sim \mathrm{N}(M_t^{(i)}, \sigma_{Mob}^{(i)}) \quad (4)$$

For the deterministic M-R relation of Eqn 1, Eqn 4 remains the same except there is no σ_M parameter, and

$$M_t^{(i)} | R_t^{(i)}, C, \gamma = e^{\mu_M^{(i)}}$$

while for the M-R relation of Eqn 3, there was an addi-

tional parameter β such that:

$$\beta \sim \mathrm{U}(-10, 10)$$
$$M_t^{(i)} | R_t^{(i)}, C, \gamma, \sigma_{M1}, \beta \sim \mathrm{N}\left(e^{\mu_M^{(i)}}, \sqrt{\sigma_{M1}^2 + \beta \tilde{R}_t^{(i)}}\right)$$

Note that the normal distributions in the last two lines of the model (collectively Eqn 4) are the same likelihoods that are assumed when using χ^2 to perform linear regression.

For all M-R relations we consider, we truncated the $M_t^{(i)}$ distribution such that $0 < M_t^{(i)} < M_{t,pureFe}^{(i)}$ where $M_{t,pureFe}^{(i)}$ was computed using the 0% rock mass fraction analytic fits to the Fortney et al. (2007) rock-iron internal structure models:

$$log(M_{t,pureFe}^{(i)}) = \frac{-b + \sqrt{b^2 - 4a(c - R_t^{(i)})}}{2a} \quad (5)$$

where a = 0.0975, b= 0.4938, and c=0.7932 (Fortney et al. 2007b). This truncation was imposed to

restrict the planet masses to physically plausible densities given the planet's true radius.

We also performed a prior sensitivity analysis on our population-wide parameters to assess the degree to which our results depend on our priors. As given in Eqn 4, we used a wide normal distribution for the power law index; we chose this prior to be wide with a variance of 1 to minimize the amount of information in the posterior that comes from the prior, and we chose a mean of 1 because we expected *a priori* that near-linear relationships were physically plausible. This choice was admittedly arbitrary, so we tested the sensitivity of our results to these assumptions. To do this, we ran the MCMC again, but with two other priors: a uniform distribution for γ and a uniform distribution for s where $\gamma = \tan(s)$, which corresponds to a uniform distribution in the slope of the power law[10]. Under each prior, we find that the posterior modes (the "best fits") for C, γ, and σ_M differed by no more than 0.05 from the best-fits listed in Table 1, which is below our reported precision. Additionally, we tested several end member cases for the R_t distribution, and the choice for this prior had a similarly negligible effect on the result, primarily because R_{ob} is fairly well constrained throughout the sample. Therefore, we conclude that our results are robust to our choice of priors.

To produce the results shown in §5, we evaluate each model with JAGS (Just Another Gibbs Sampler; Plummer 2003), an R code for numerically evaluating hierarchical Bayesian models with MCMC. For each set of posteriors in Figures 2 and 3, we ran 10 chains consisting of 500,000 iterations each. The first half of each chain is discarded as "burn-in", and the resulting half is thinned by a factor of 250, such that we retain 10,000 posteriors samples of each parameter.

To assess the independence of these samples, we compute the effective sample size (ESS), which accounts for the autocorrelation still present within these thinned Markov chains (ESS = 10000 indicates perfect independence). The ESS is > 4000 for each parameter listed in Table 1, with two exceptions. The ESS of the deterministic relation parameters are around 230, an order of magnitude lower than all the probabilistic relations we tested. The difficulty this model had with convergence reflects the challenges of applying a physically inappropriate model to data and is another indication that a deterministic relation does not fit the observed masses and radii well. Less concerning yet not quite as well converged compared to the others were the parameters for the probabilistic M-R relation fit only to the smallest radii (ESS = $1500 - 4000$). This occurs because these small planets have the largest mass uncertainties; this causes the maximum mass restriction in Eqn 5 to severely truncate most of the likelihoods, which results in high autocorrelation in the MCMC chains. Given the small ESS for these two sets of parameters, we caution against over-interpretation of their results: their "best-fit" values in Table 1 are ~ 6 and ~ 2 times more uncertain than the others (corresponding to the precision of the posterior mean with the square root of the sample size [the ESS]), and the boundaries of their 95% credible regions in Figure 3 are poorly estimated. We do not spend time

[10] Uniform γ places high probability at steep power laws, which are highly unlikely on physical grounds.

Table 1
Best-Fit Parameters of the M-R Relation

Equation	Dataset	C	γ	σ_M	β
1	baseline: RV only, < 4 R$_\oplus$	2.1	1.5	—	—
2	baseline: RV only, < 4 R$_\oplus$	2.7	1.3	1.9	—
2	N-body TTVs only, < 4 R$_\oplus$	0.6	1.7	1.7	—
2	Weiss (< 4 R$_\oplus$)	2.8	0.9	2.5	—
2	RV only, < 1.6 R$_\oplus$	1.4	2.3	0.0	—
2	RV only, < 8 R$_\oplus$	1.6	1.8	2.9	—
3	baseline: RV only, < 4 R$_\oplus$	2.6	1.3	2.1	1.5

Note. — These "best fit" values correspond to the mode of the joint posterior distributions; see code and posterior samples in the github repository dawolfgang/MRrelation to account for the full uncertainty in the parameters that is contained the posteriors (see §5.3 for more details on this). Also, when using these M-R relations to generate masses from planet radii, one should apply the density constraint given by Eqn 5.

running these simulations longer, as they were performed for the sake of comparison, and we display them for this purpose only.

For the main result — the baseline dataset evaluated with the probabilistic M-R relation — the ESS of C and γ are 10,000, and the ESS of σ_M is 5,300. Furthermore, the between-chain convergence diagnostic $\hat{R}$ of Gelman & Rubin (1992) is ≤ 1.002 for all parameters in our probabilistic models (except again for the smallest radii planets, for which $\hat{R} = 1.008$ at its worst). Together, these two tests provide no evidence that the posteriors have not converged, and we proceed with the usual amount of confidence (given that no one can ever prove convergence).

5. RESULTS

Table 1 shows the results of our modeling: it displays the best-fit parameters for each of the various M-R relations and datasets that we consider. The first entry corresponds to the deterministic M-R relation; entries $2 - 6$ correspond to our probabilistic M-R relation for various datasets (see §3, §5.2); and the last entry corresponds to the probabilistic M-R relation with non-constant scatter. In particular, the second entry lists the best-fit values for Eqn 2 using our baseline dataset. All were computed with the density restriction given by Eqn 5; this constraint should also be applied to the masses generated from these M-R relations when these relations are used in forward modeling.

In all cases the reported "best fit" values correspond to the mode of the joint posterior distribution, and are denoted by the triangles in Figures 2-3. The uncertainties in the parameters are represented by the displayed 68% and 95% posterior contours, with the contours corresponding to our baseline dataset colored blue. The gray points are the 10,000 saved posterior samples from our thinned MCMC chains using the baseline dataset.

5.1. *Deterministic vs. Probabilistic M-R Relations*

The primary motivation for this paper was to assess the observational evidence for intrinsic scatter in the sub-Neptune M-R relation, and to characterize this scatter if warranted. To do so, we compare the posteriors for our three M-R relations in Figure 2 (note that not all relations have all parameters: for example, the deterministic
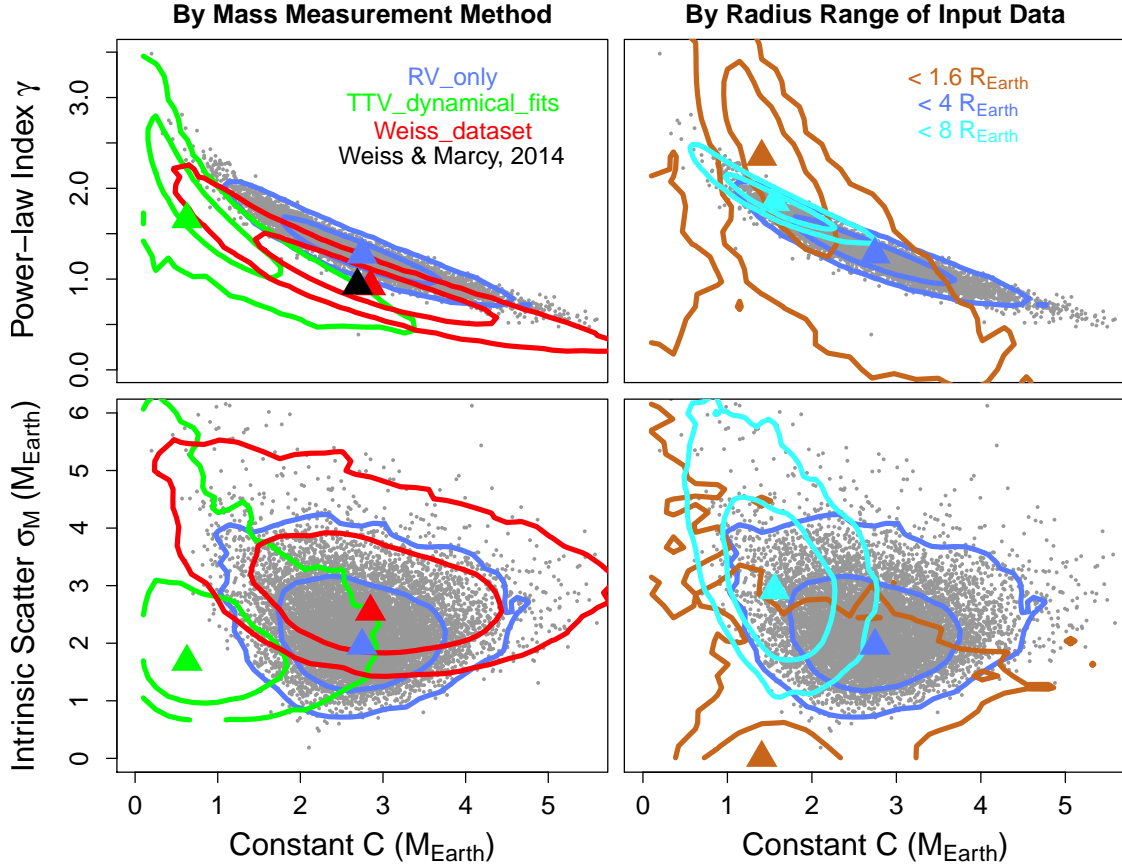
Figure 3. Posteriors for Eqn 2's M-R relation parameters when we change the input dataset (68% and 95% contours shown; triangles are best-fit values). The blue contours represent the baseline dataset and are the same as those in panels a and b of Figure 2; the gray points are the 10,000 saved posterior samples from our MCMC chains using this baseline dataset. The green TTV M-R relation is systematically shifted downward (lower C) compared to the baseline M-R relation, while the red WM14 dataset, a hybrid of the two, produces a posterior which falls between them (the black point is the WM14 result itself). When we consider different radius ranges, we see that $R_{obs} < 8$ $R_\oplus$ (cyan) produces a slightly down-shifted, steeper, and more dispersed M-R relation than the default $R_{obs} < 4$ $R_\oplus$ (lower C and higher γ, σ_M, although the posteriors do overlap), while the M-R relation for $R_{obs} < 1.6$ $R_\oplus$ (orange) is not well constrained (although $\sigma_M \approx 0$ for reasonable values of C).

M-R relation of Eqn 1 is described only by C and γ, so it only appears in panel a). Panels b and c show that this intrinsic scatter exists: because the posteriors lie away from zero, $\sigma_M = 0$ is strongly excluded by the data, even with the currently large individual mass error bars. This is not a result of our choice of priors: the parameterization in Eqn 4 is equivalent to $\sigma_M^2 \sim 1/\sigma_M^2$, which is strongly weighted toward zero, in contrast to the posterior we compute.

Comparing the different M-R relations, we see that the C, γ posterior for the model given by Eqn 1 is much tighter than that for Eqns 2-3. This is expected: when we keep the dataset fixed but add more parameters, especially one like σ_M that by construction allows wiggle room around a deterministic relation, the observational information content per parameter decreases, and the posteriors widen. Given this expectation, what is arguably more notable are the small differences between Eqn 2 and 3's model posteriors for the parameters they have in common: most of the extra width of Eqn 3's joint posterior is contained in the new parameter β (Figure 2, panels d-f), which spans zero. There is therefore not enough evidence in the current dataset to justify an intrinsic scatter that changes as a function of radius, at

least not under our model assumptions[11]. For the best-fit values of these parameters, which correspond to the triangles in Figure 2, see Table 1.

5.2. *Changing the Dataset*

The results in §5.1 are for our baseline dataset, an RV-only sample with $R_{obs} < 4$ $R_\oplus$. However, all Bayesian results depend on the data that are used, so it is important to carefully consider what the dataset contains. To demonstrate this, we present some illustrative examples of the M-R relation posteriors under different mass and radius selection functions (Figure 3).

The left side of Figure 3 displays results for samples of planets that have had their masses measured in different ways. A number of prior studies (e.g. Jontof-Hutter et al. 2014, WM14) have noted that planets with masses measured from their high SNR TTVs tend to be systematically less dense than planets with

---

[11] While outside the scope of this paper, future analyses of the M-R relation can address this and other questions of model selection more quantitatively by computing posterior Bayes factors. Regardless, the results for the statistical models represented by Eqns 1 and 3 can serve as a sensitivity test for that of Eqn 2, as we describe.
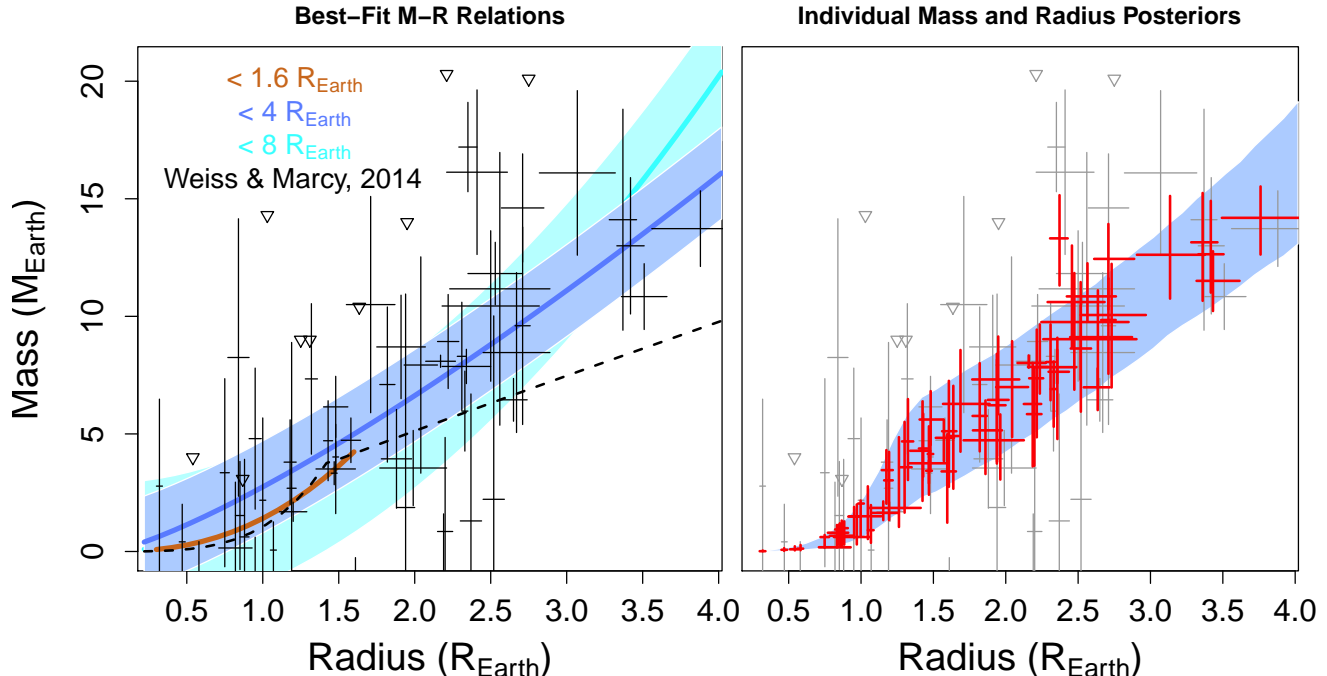
Figure 4. *Left:* the best-fit M-R relations from the right column of Figure 3. For each, the solid line denotes the mean relation μ_M while the faded region denotes the standard deviation of the intrinsic scatter (vertical height of region to either side of line = σ_M; note $\sigma_M = 0$ for the smallest planets). The M-R relation of WM14 is the dashed black line while the baseline dataset is overplotted as the thin black lines with triangles for the 2-σ upper limits (note that WM14 was calculated with a dataset that includes TTV planets). *Right:* the baseline M-R relation (the second entry in Table 1) marginalized over the corresponding posterior distribution and subjected to our physical mass range restriction. The blue region now corresponds to the central 68% of planet masses that were drawn at a given radius. The 68% coverage interval of the posterior true masses and radii of individual planets are plotted red; the same R_{ob} and M_{ob} as on the left are plotted in gray for comparison.

RV-measured masses[12]. The origin of this difference is still unclear, and is not something our current modeling is able to address (see §6.3): it could be due to either an intrinsic difference in the densities of these two populations, or to observational bias, as TTVs for larger (and thus less dense) planets are easier to detect while RVs for more massive (and thus more dense) planets are easier to measure. In any case, our results confirm the existence and nature of this discrepancy, if not the reason: the green TTV-only posterior is shifted towards lower C with similar γ and σ_M, which produces on average lower masses for a given radius. Furthermore, the hybrid WM14 dataset yields the red posterior, which falls between the TTV-only and RV-only posteriors yet peaks at lower γ, illustrating that posterior modes (Bayesian "best fits") for joint datasets are not necessarily averages of the modes for separate subsets. This behavior can be understood when one considers that these TTV-measured planets are preferentially larger than the RV-measured planets: this pulls the joint M-R relation down

---

[12] This appears to be a population-level effect: there are few planets with independently analyzed, strongly constrained TTV and RV masses, and they all yield measurements that are consistent between the two methods [Kepler-18b & c (Cochran et al. 2011), Kepler-88c (Nesvorný et al. 2013), and Kepler-117 (Bruno et al. 2015); note that only Kepler-18b appears in our table]. There are other systems where the two methods have been used in concert to infer planet masses (e.g., Kepler-9, Kepler-10 Kepler-11, Kepler-89; see table for references), but the constraints from one or both methods are relatively weak, making them insensitive tests of a TTV-RV difference on a per-planet basis.

at higher radii because the TTV-measured planets there have lower masses (which lowers γ) but affects the relation at lower radii very little because there are few small planets in our TTV-measured sample (which keeps C roughly the same).

The right side of Figure 3 displays results for samples of planets spanning different radius ranges, illustrating the effect that a somewhat arbitrary radius cut can have on one's results. Compared to our default sub-Neptune range, a $R_{obs} < 8$ $R_\oplus$ cut produces an M-R relation that is overall shifted down, is steeper, and has more intrinsic scatter (the cyan posterior has lower C and higher γ, σ_M). This is consistent with the Lissauer et al. (2011) fit to Earth and Saturn over a similar radius range, although neither of these Solar System planets were included in our dataset. Meanwhile, the M-R relation is poorly constrained for the $R_{obs} < 1.6$ $R_\oplus$ sample, the radius range outside of which rocky planets likely do not occur (Rogers 2015). This is because our $0 < M_t^{(i)} < M_{t,pureFe}^{(i)}$) restriction is most severe for these small planets, allowing only a small range of physically plausible masses. This range is completely spanned by most of the mass measurements (see right side of Figure 4), so there is little empirical extrasolar information for $R_{obs} < 1.2$ $R_\oplus$, and the orange posteriors are dominated by the few larger planets with well measured masses. With this sample, there is not currently enough observational evidence in this radius range to rule out a deterministic relation. This does not mean, however, that these small planets all have the same composition, as the

posterior 68% contour spans power-law indexes between 1 and 3 and a constant population-wide Earth-like rocky composition would have a power-law index of around 3.7 (Valencia et al. 2006). More data will be needed to yield a better estimate of the power-law index, and therefore of the compositional diversity of small planets.

5.3. *Visualizing the M-R Relation*

While the posterior contours in Figures 2-3 show the best-fit M-R relation parameters and their uncertainties, visualizing the M-R relation itself requires that they be mapped from parameter space to mass, radius space. There are at least two ways to do this with Bayesian analysis, and they are displayed in Figure 4.

First, one can simply take the best-fit values and plot the resulting relation, as was done in the left panel. Here the 1σ width of the probabilistic relation, as parameterized by σ_M, is denoted by the faded colored region while the mean relation, as parameterized by C and γ, is the thick line of the same color. Note that the mean M-R relations extend into unphysical regimes for $R < 1$ R$_\oplus$; this is because the mass observations span the physically allowed region, as discussed in §5.2, leaving the M-R relation to be constrained primarily by the locations of the larger, higher mass planets and our model assumptions. The presence of intrinsic scatter in our M-R relation nevertheless allows physically realistic masses to be assigned to the smallest planets; to force this requirement, we recommend adding a density constraint to Eqn 2 such that the probability of a planet being drawn outside this range is 0 (the constraint we used is given in Eqn 5), or to use a different M-R relation for sub-Earth-sized planets. The different colors in the left panel correspond to the M-R relations in the right column of Figure 3; these mostly overlap in the sub-Neptune regime. Note that the RV-only dataset produces a steeper relation than one which also contains high SNR TTV planets (i.e. the black dashed WM14 relation), as discussed in §5.2.

While these best-fit M-R relations are easy to use, they do not take into account the fact that the posteriors have non-zero width and therefore a range of M-R relation parameters are allowed by any one dataset. A more thorough implementation of these results would account for these uncertainties by integrating over all of the posterior samples. This marginalization, which also incorporates the physical restrictions on M_t as given by Eqn 5, is displayed on the right: now the blue region corresponds to the central 68% of planet masses that were drawn for a given radius. Note that this region is wider than that on the left and that the masses no longer extend into unphysical regimes. The 68% coverage interval of the posterior true masses and radii of individual planets in the baseline sample are plotted red, while the same R_{ob} and M_{ob} as on the left are plotted in gray. As expected (see the end of §6.2), the posteriors have "shrunk" toward the mean relation within the uncertainties provided by the data. Furthermore, one can readily see that the data are qualitatively consistent with the modeled M-R relation: the red lines fall within and immediately around the blue region (see §6.2 for a quantitative treatment of model checking).

6. DISCUSSION

6.1. *Using the M-R Relation to Predict Masses*

The most straightforward and computationally simple way to map a sub-Neptune's radius to a mass while accounting for intrinsic scatter is to adopt Eqn 2 with one of the sets of parameters in Table 1 and impose a density constraint like Eqn 5 for the smallest planets. This best-fit M-R relation is analytic and represents a substantial improvement over the previous deterministic relationships in capturing the full mass-radius distribution. However, it does not incorporate uncertainties in the M-R relation parameters or uncertainties in the measured planet radius itself. Depending on how detailed one's analysis needs to be, a more accurate predictive mass distribution may be needed.

To account for these issues, one must compute the posterior predictive M-R relation, which marginalizes over both the posteriors displayed here and the radius posterior produced by one's light curve modeling. This mass distribution will be wider than that produced by simply applying Eqn 2 (see right side of Figure 4) because it incorporates the above sources of uncertainty and thus more accurately reflects our state of knowledge about these planets' masses. Kepler-452 b (Jenkins et al. 2015) provides an example of an individual planet's posterior predictive mass distribution that has been calculated with this probabilistic M-R relation; because its computation requires the numerical posterior samples that we have produced, the resulting mass distribution is also numerical in nature. To enable more calculations like this one, we have posted our posterior samples in the github repository dawolfgang/MRrelation along with R code that uses them to compute and plot the posterior predictive mass distribution for individual planets.

6.2. *Model Checking*

The purpose of the right panel in Figure 4 is to provide a qualitative comparison between the data and the baseline probabilistic M-R relation; this visual check immediately verifies that, in the broadest sense, our model is a reasonable description of the data (see §5.3). However, no model perfectly describes nature. A more in-depth look is warranted, to both understand the limitations of the current model and to identify areas for improvement in future work.

Quantitatively, we can check the data-model consistency by computing a "hierarchical p-value", which yields the fraction of all possible datasets which are more discrepant from the model than the observed dataset. This calculation necessarily involves sampling from our forward model (Eqn 4) and using those samples to calculate a statistic which quantifies "discrepant". The identification of robust and useful hierarchical statistics is still an active area of statistical research (see, for example, Bayarri & Castellanos 2007), as the multi-level nature of hierarchical modeling offers a number of choices that test different aspects of the model. As a result, model checking in practice can be an involved process that requires investigations into multiple parts of the problem. We provide two such investigations below to illustrate some of the subtlety of this endeavor.

Regardless of the details, most forms of Bayesian model checking use the posterior predictive distribution. Conceptually, this distribution defines the probability of ob-

serving a certain data value given the currently observed dataset and the model. Mathematically, the posterior predictive distribution for a new observation x_{new} is defined as:

$$p(x_{new}|\hat{\boldsymbol{x}}, M_1) = \int p(x_{new}|\boldsymbol{\theta}, M_1)p(\boldsymbol{\theta}|\hat{\boldsymbol{x}}, M_1)d\boldsymbol{\theta} \quad (6)$$

where $p(x_{new}|\boldsymbol{\theta}, M_1)$ is the likelihood of a new, currently unobserved data point given the parameters $\boldsymbol{\theta}$ of model M_1, and $p(\boldsymbol{\theta}|\hat{\boldsymbol{x}}, M_1)$ is the posterior of model M_1, i.e. the joint probability of all of M_1's parameters given the observed dataset $\hat{\boldsymbol{x}}$. The integral denotes the process of marginalization over the parameters, so that the probability of a new observation incorporates the model uncertainty allowed by the current observations.

Next, one draws hypothetical data from this posterior predictive distribution until a dataset of the same size N as the observed dataset is achieved:

$$\hat{\boldsymbol{x}}_{\boldsymbol{new}} \underset{N}{\sim} p(x_{new}|\boldsymbol{x}, M_1) \quad (7)$$

where $\hat{\boldsymbol{x}}_{\boldsymbol{new}} = \{\hat{x}_{new,1}, \hat{x}_{new,2}, ..., \hat{x}_{new,N}\}$ and "$\underset{N}{\sim}$" means "draw from that same distribution N times". This dataset is then used to compute the statistic of choice (see discussion below). Repeating this process thousands of times (i.e. bootstrapping the statistic) generates a distribution of the statistic which can then be compared to the value that was calculated from the observed dataset. If the observed statistic falls within this distribution, the model is consistent with the data (see Figure 5).

In practice, MCMC simulations (§4) provide samples from the posterior rather than the analytic form of the posterior as required in Eqn 6; therefore, the posterior predictive distribution is not directly calculated. Instead, Eqns 6 and 7 are combined by using the posterior samples:

$$\hat{\boldsymbol{\theta}} \sim p(\boldsymbol{\theta}|\hat{\boldsymbol{x}}, M_1)$$

to define the likelihood that one then draws from:

$$\hat{\boldsymbol{x}}_{\boldsymbol{new}} \underset{N}{\sim} p(x_{new}|\hat{\boldsymbol{\theta}}, M_1) \quad (8)$$

Performing these two steps repeatedly produces an ensemble of datasets drawn from Eqn 6. Applying Eqn 8 to our M-R relation requires evaluating the lower levels of the forward model described in Equation 4.

Part of the subtlety of checking data-model consistency arises because our model is hierarchical. In particular, $\hat{\boldsymbol{\theta}}$ of Eqn 8 includes both the population-wide parameters C, γ, σ_M and the individual parameters $M_t^{(i)}, R_t^{(i)}$; we can use the posterior samples of either of these groups of parameters to calculate $\hat{\boldsymbol{x}}_{\boldsymbol{new}}$. We show the result of both choices in Figure 5, using the posterior samples from the baseline M-R relation (second line of Table 1). Conceptually, using samples from the individual true mass and radius posteriors ($\hat{\boldsymbol{\theta}} = \left\{M_t^{(i)}, R_t^{(i)}\right\}$) evaluates the fit of the model to the currently observed set of planets (the green histograms on the left), while using posterior samples for the M-R relation parameters ($\hat{\boldsymbol{\theta}} = \{C, \gamma, \sigma_M\}$) evaluates the fit of the model to altogether new sets of planets (the blue histograms on the right). For the former case, the same $\sigma_{Mob}^{(i)}$ and $\sigma_{Rob}^{(i)}$ as

the observed dataset are used; for the latter case, $\sigma_{Mob}^{(i)}$ is drawn from the distribution of σ_{Mob} for the observed planets that have a similar mass, and $\sigma_{Rob}^{(i)}$ is drawn from the observed dataset's full distribution of σ_{Rob} without controlling for radius (the size of the radius error bars are fairly constant across the dataset).

The second aspect of hierarchical model-checking which requires some effort is the identification of a robust statistic to quantify the discrepancy between the observed data and the model-generated data. Optimally this statistic would test the fit of every part of the modeled probability distribution, including the "average" behavior of the model, the "extreme" behavior out on the tails of the distribution, and for hierarchical models, the accuracy of the grouping and relational structure that is illustrated by graphs like Figure 1. Due to high dimensional parameter space, this proves to be very difficult, so one must identify several statistics which test such aspects separately. To illustrate the problem, we choose two: $f_{1\sigma}$, the fraction of a given dataset's simulated mass, radius observations which fall within the 68% coverage interval of our probabilistic, baseline M-R relation (blue region in Figure 4), and f_μ, the fraction of a given dataset's mass measurements whose 1σ error bars cross the mean relation μ (see Eqn 2) of that same model. Therefore, $f_{1\sigma}$ tests how well the width of our probabilistic M-R relation fits the data, and f_μ tests how tightly grouped the data are around the mean compared to the normal distribution of our model.

The results of these tests are displayed in Figure 5. The observed data's $f_{1\sigma}$ statistic (top) falls at the 63rd and 83rd percentiles (i.e. within "one sigma") of the distributions for datasets generated from the individual posteriors and the population-wide posteriors, respectively. Therefore, the model is fully consistent with the data for the aspect of the model that this statistic tests: the inferred intrinsic scatter of the M-R relation. On the other hand, the observed data's f_μ statistic falls at the 2nd and 0.2th percentile of the distributions. This test is sensitive to the shape of the distribution around the mean relation, and thereby probes the appropriateness of our assumption of a normal distribution in Eqn 2. The fact that the data are marginally inconsistent with this aspect of the model reveals that this is an area of improvement for future work (see §6.3 for a further discussion about this).

The differences in the model-data fit implied by these two statistics illustrates how careful one must be in performing model checking and interpreting the results. Instead of asking "is the model consistent with the data?", a more well-posed question would be "in what ways are the model consistent with the data?", especially as these statistical models become more complex to accommodate more sources of uncertainty and more realistic physics. For the model at hand, we address this question by returning to our purpose. We wanted to explore the need for intrinsic scatter in the M-R relation; therefore, we care most about the quality of the fit with respect to the spread of the M-R relation, i.e. the data-model consistency as quantified by $f_{1\sigma}$. Because this fit is good, we are satisfied that our main result holds up to this further scrutiny. The f_μ data-model discrepancy has implications for our choice to use a normal distribution for the
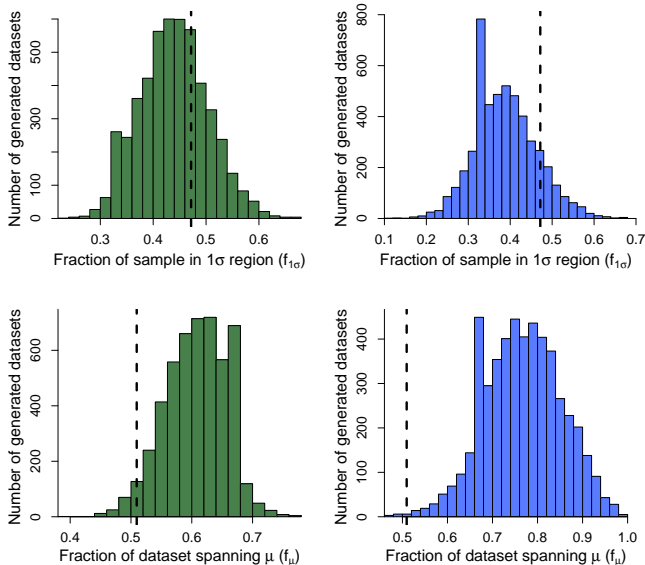
Figure 5. Checking the model against the observed data, using two different statistics to quantify the data-model consistency (see §6.2). All panels contain 5,000 hypothetical datasets generated from the baseline posterior predictive distribution (Eqns 6-8), with the statistic for the observed dataset (Table 2) as the vertical black dashed line. The green histograms on the left show the consistency of the model with the current dataset, while the blue histograms on the right show the consistency with new hypothetical datasets. The top histograms show the distribution of the statistic $f_{1\sigma}$. The observed $f_{1\sigma}$ falls at the 63rd and 83rd percentiles of the distribution (left and right, respectively), indicating that the width of the model's M-R relation is consistent with the data. The bottom histograms show the distribution of the statistic f_μ. The observed f_μ falls at the 2nd and 0.2th percentiles of the distribution, indicating that there is room for future improvement on the choice of what distribution to use for the probabilistic M-R relation.

probabilistic M-R relation, which we discuss in §6.3.

To wrap up our discussion on hierarchical Bayesian model checking, we note that the reader may find it surprising that the distribution of $f_{1\sigma}$ statistics peaks around ~ 0.4, given that the statistic was defined by the "one-sigma" central region of the M-R relation. The reason why the statistic does not instead cluster around 0.68 is due to a well-known feature of hierarchical modeling called shrinkage. First worked out by Stein (1955) and developed within a Bayesian framework by Good (1965), shrinkage refers to the tendency for individual parameter values at intermediate levels of hierarchical models (such as $M_t^{(i)}, R_t^{(i)}$) to cluster more closely together, or "shrink" toward their mean, than if the parameters had been analyzed completely independently of each other (we see this shrinking visually by comparing the gray and red lines in the right panel of Figure 4). This occurs because, by design, the hierarchical structure of the model causes these individual parameters to be related to each other, which then enables information about one parameter to influence our inference for another parameter. This additional information, provided solely by the hierarchical structure, causes the variance among the population to decrease relative to the case where the structure, and thus the information, was not available. This decrease in variance is perceived as a "shrinking" towards the mean.

We can understand how shrinkage manifests for the M-

R relation by returning to the forward model defined in Eqn 4. First we note that it is the true masses and radii, not the observed masses and radii, which are drawn from the M-R relation. Therefore, it is the $M_t^{(i)}$ and $R_t^{(i)}$ values which would produce $f_{1\sigma} = 0.68$. However, we don't actually observe $M_t^{(i)}$ and $R_t^{(i)}$ outright; we observe them convolved with some error $\sigma_{ob}^{(i)}$. Because the convolution of a normal distribution with another normal produces a wider normal distribution, we would indeed expect more of the observed mass and radius points to fall outside of the "one-sigma" bounds defined by the top level of the model. With this additional insight into the nature of hierarchical models, we understand that this behavior is not only consistent with what we see, but expected.

6.3. Caveats and Future Work

As discussed in §2-3, we made a number of assumptions and modeling choices to facilitate a straightforward investigation into the need for an intrinsic scatter term in the sub-Neptune M-R relation. Some of these choices, such as parameterizing the relation with a power law or using a normal distribution to characterize the scatter, were driven by convenience and familiarity rather than by physics. In particular, we chose to use a power law in order facilitate direct comparisons between our results and those in the literature, and we chose the normal distribution because it directly parameterizes the intrinsic scatter in the population rather than relying on a transformation from a more obscure distribution to derive the population variance. In general, our philosophy was to limit the number of free parameters in our model as much as possible in order to maximize the information content per parameter.

That said, these parameterizations are by no means the only ones that could be reasonably used. These approximations can and should be revisited in future work, especially as more data becomes available and we begin to describe more subtle features in the M-R relation. An important aspect of these studies will be model-checking to inform choices for parameterizing the M-R relation (see §6.2) and performing quantitative model comparisons. For now, we emphasize that the inclusion of any kind of probability distribution to account for intrinsic scatter represents a significant improvement over prior work, and we leave testing different distributions for future studies.

A related parameterization issue is our choice not to include any other planet properties into this M-R relation, such as a dependence on orbital period. It is entirely plausible on both theoretical and observational grounds that the M-R relation at short periods may be different from that at long periods. Theoretically, photoevaporation is likely to have eroded planets on short orbits (see, for example, Murray-Clay et al. 2009; Lopez et al. 2012; Howe & Burrows 2015), thereby causing the population of highly irradiated planets to be denser on average; alternatively, migration could have produced a mass-dependent stopping location given a certain structure for the inner regions of the disk (e.g. Benítez-Llambay et al. 2011) or tidal circularization could have produced a density-period correlation for the shortest orbits (Barnes 2014). Observationally, we see a suggestive dearth of > 3 $R_\oplus$ *Kepler* planet candidates and lower-mass RV planets

at short orbital periods (Beaugé & Nesvorný 2013) and a period dependence in the marginalized, bias-corrected *Kepler* radius distribution (Youdin 2011; Howard et al. 2012; Morton & Swift 2014). These observations hint at the potential for interesting features in the joint density-period distribution.

While these results provide strong motivation for further investigations into period-dependent M-R relations, there will be numerous details to address in fitting the more complex period-dependent statistical model to the data. We choose to leave these analyses for future work given the number of tests that we have already performed here (choice of parameterization for the intrinsic scatter, RV vs. TTV masses, and different radius ranges). In particular, this future work will need to robustly account for the systematic biases between the different mass measurement methods to make sure that our inferences are not driven by arbitrary choices for the data we use (or if these choices prove to be unavoidable, to quantify their effect). In general, future studies will need to perform model comparison to identify the most useful parameterizations given the data. These two efforts pose cutting-edge statistical challenges that are worthy of separate investigations in and of themselves, and so we reserve them for follow-up studies that build on the results presented here.

Returning to our current statistical model for the M-R relation, our third major modeling choice after the power law and Gaussian astrophysical scatter was using Gaussian errors for the observed mass and radius measurements, as denoted in the last line of our statistical model (Eqn 4). The most accurate hierarchical analyses include the actual likelihood used to infer these planetary parameters from the lower-level photometric and spectroscopic data, rather than assume such a heavily simplified functional form as we did. Incorporating observers' true likelihoods is important to capture important correlations that exist in the measurement analysis and to use all of the information contained in the lower-level data. Ideally, future planet discovery papers will make their full, joint likelihood distributions available in addition to the reported "best-fit" value; depending on the type of analysis that observers use to make measurements from their data, this means providing either log-likelihood (e.g. χ^2) contours over all of the parameters that were considered or providing posterior samples along with detailed information about the choice of priors. Unfortunately, this information is not publicly available, and so we cannot use it. Therefore, we follow the convention established by WM14 in using a normal distribution to represent the marginalized mass or radius likelihood.

Another notable approximation that we have made arises from the difference between the graphical model describing our M-R relation (Figure 1) and the way in which we have implemented it (Eqn 4). The difference between the two is subtle: by conditioning on $\sigma_{Mob}^{(i)}$ in the lowest level of the model, we are assuming that the observed mass measurement uncertainty is independent from the measurement itself. In reality, the calculation of both the uncertainty and the measured value are produced by the same modeling process and thus are correlated in a potentially nontrivial way[13]. In practice, we had no other choice than to assume independence because the nature of this correlation is rarely, if ever, published.

An important caveat about our results is that we ignore selection effects, primarily due to the inability to model the human decisions which affected the ground-based follow-up observing campaigns. As discussed in §3, we would ideally have a uniform sample of masses and radii which were analyzed in the same way, and have well-characterized the selection effects and detection efficiencies. Unfortunately, this is simply not possible at the present moment. The sample is highly heterogeneous, a compilation of many observing teams' programmatic, yearly, and nightly priorities that are not communicated in the literature and therefore cannot be accurately modeled. There is significant, difficult work still to be done to understand the extent to which the follow-up process shapes the observed mass-radius space and therefore influence our inferences for the underlying M-R relation.

This is particularly important for interpreting the apparent discrepancy between the population of TTV-measured masses and the population of RV-measured masses. In §5.2 we corroborate the systematic density difference between TTV and RV planets that had been noticed by Jontof-Hutter et al. (2014) and Weiss & Marcy (2014), as we find a ∼ two "sigma" offset in the two datasets' M-R relation parameters. Since we do not attempt to model the selection effects that are involved, we cannot distinguish how much of this offset is created by observational bias and how much is due to an intrinsic difference in the density distribution. If the latter turns out to be the driving factor behind this apparent discrepancy, disentangling the underlying reason for this astrophysical density difference from the current list of features which distinguish the two samples will have numerous implications for planet formation and evolution.

7. CONCLUSIONS

In this paper we have defined and constrained a probabilistic mass-radius relationship for sub-Neptune planets (Eqn 2 with parameter values in Table 1 and the density constraint provided in Eqn 5). In particular, we demonstrate that there is intrinsic, astrophysical scatter in this relation, and that, except for the smallest planets, this scatter is nonzero for all considered datasets. For the first time in the exoplanet literature, we display the uncertainties in the M-R relation parameters through posterior distributions and explain how to properly incorporate these uncertainties into a predictive distribution of masses for individual planets. This M-R relation will be useful for anyone who wishes to perform large-scale dynamical or planet formation studies with the *Kepler* planet candidates.

More broadly, this work provides a framework for fur-

---

[13] At first glance this may seem like an inconsequential difference considering the other modeling assumptions we have made, but the assumption of independence becomes problematic when the processes of detection and measurement use the same data and when the population analyses use regions of parameter space where the detection efficiency starts to drop (see Loredo & Wasserman (1995) for the technical details). We do not correct for detection bias in this paper, so are unaffected by this problem.

ther analyses of the M-R relation and its probable dependencies on period and stellar properties. Here we have demonstrated how to develop and apply a statistical model that incorporates measurement uncertainties into population-wide inference of the sub-Neptune M-R relation and that directly produces estimates of the uncertainty of the inferred parameters. This method is advantageous because it is quantitative and easily generalizable to include additional variables which may be important in the underlying M-R relation that we are trying to model, such as incident flux (§6.3) or different stellar masses or metallicities. We do not investigate these possibilities here, as we wish to begin this broader effort with the simplest reasonable statistical model. Nevertheless, searching for additional dependencies in the M-R relation is an important endeavor in order to understand which physical processes shape the super-Earth population. With this work we establish both a point of comparison and a framework for these further studies.

REFERENCES

Bakos, G. Á., Torres, G., Pál, A., et al. 2010, ApJ, 710, 1724
Ballard, S., Fabrycky, D., Fressin, F., et al. 2011, ApJ, 743, 200
Barnes, R. 2014, International Journal of Astrobiology, FirstView, 1
Barros, S. C. C., Almenara, J. M., Deleuil, M., et al. 2014, A&A, 569, A74
Batalha, N. M., et al. 2010, ApJ, 713, L109
Batalha, N. M., Borucki, W. J., Bryson, S. T., et al. 2011, ApJ, 729, 27
Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, ApJS, 204, 24
Bayarri, M. J., & Castellanos, M. E. 2007, Statistical Science, 22, 322
Beaugé, C., & Nesvorný, D. 2013, ApJ, 763, 12
Benítez-Llambay, P., Masset, F., & Beaugé, C. 2011, A&A, 528, A2
Biddle, L. I., Pearson, K. A., Crossfield, I. J. M., et al. 2014, MNRAS, 443, 1810
Bonfils, X., Gillon, M., Udry, S., et al. 2012, A&A, 546, A27
Bonomo, A. S., Sozzetti, A., Lovis, C., et al. 2014, A&A, 572, A2
Bordé, P., Bouchy, F., Deleuil, M., et al. 2010, A&A, 520, A66
Borucki, W. J., Koch, D. G., Brown, T. M., et al. 2010, ApJ, 713, L126
Borucki, W. J., Koch, D. G., Basri, G., et al. 2011, ApJ, 736, 19
Borucki, W. J., Agol, E., Fressin, F., et al. 2013, Science, 340, 587
Bruno, G., Almenara, J.-M., Barros, S. C. C., et al. 2015, European Physical Journal Web of Conferences, 101, 06014
Burke, C. J., Bryson, S. T., Mullally, F., et al. 2014, ApJS, 210, 19
Butler, R. P., Vogt, S. S., Marcy, G. W., et al. 2004, ApJ, 617, 580
Carter, J. A., Winn, J. N., Holman, M. J., et al. 2011, ApJ, 730, 82
Carter, J. A., Agol, E., Chaplin, W. J., et al. 2012, Science, 337, 556
Charbonneau, D., Berta, Z. K., Irwin, J., et al. 2009, Nature, 462, 891
Chatterjee, S., & Tan, J. C. 2015, ApJ, 798, LL32
Cochran, W. D., Fabrycky, D. C., Torres, G., et al. 2011, ApJS, 197, 7
Dragomir, D., Matthews, J. M., Eastman, J. D., et al. 2013, ApJ, 772, L2
Dressing, C. D., Charbonneau, D., Dumusque, X., et al. 2015, ApJ, 800, 135
Dumusque, X., Bonomo, A. S., Haywood, R. D., et al. 2014, ApJ, 789, 154
Fortney, J. J., Marley, M. S., & Barnes, J. W. 2007, ApJ, 659, 1661
Fortney, J. J., Marley, M. S., & Barnes, J. W. 2007, ApJ, 668, 1267
Fressin, F., Torres, G., Rowe, J. F., et al. 2012, Nature, 482, 195
Gautier, T. N., III, Charbonneau, D., Rowe, J. F., et al. 2012, ApJ, 749, 15
Gelman, A., & Rubin, D. 1992, Statistical Science, 7, 457
Gillon, M., Demory, B.-O., Benneke, B., et al. 2012, A&A, 539, A28
Good, I. J. 1965, The Estimation of Probabilities: An Essay on Modern Bayesian Methods, M.I.T. Press
Hartman, J. D., Bakos, G. Á., Kipping, D. M., et al. 2011, ApJ, 728, 138
Howard, A. W., Johnson, J. A., Marcy, G. W., et al. 2011, ApJ, 730, 10
Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, ApJS, 201, 15
Howard, A. W., Sanchis-Ojeda, R., Marcy, G. W., et al. 2013, Nature, 503, 381
Howe, A. R., & Burrows, A. 2015, ApJ, 808, 150
Howell, S. B., Rowe, J. F., Bryson, S. T., et al. 2012, ApJ, 746, 123
Jenkins, J. M., Twicken, J. D., Batalha, N. M., et al. 2015, in review.
Jontof-Hutter, D., Lissauer, J. J., Rowe, J. F., & Fabrycky, D. C. 2014, ApJ, 785, 15
Kelly, B. C. 2007, ApJ, 665, 1489
Kipping, D. M., Nesvorný, D., Buchhave, L. A., et al. 2014, ApJ, 784, 28
Lanotte, A. A., Gillon, M., Demory, B.-O., et al. 2014, A&A, 572, A73
Léger, A., Rouan, D., Schneider, J., et al. 2009, A&A, 506, 287

Lissauer, J. J., Ragozzine, D., Fabrycky, D. C., et al. 2011, ApJS, 197, 8

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

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

Lopez, E. D., & Fortney, J. J. 2014, ApJ, 792, 1

Loredo, T. J., & Wasserman, I. M. 1995, ApJS, 96, 261

Loredo, T. J. 2013, *Astrostatistical challenges for the new astronomy*, 15

McArthur, B. E., Endl, M., Cochran, W. D., et al. 2004, ApJ, 614, L81

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

Mazeh, T., Nachmani, G., Holczer, T., et al. 2013, ApJS, 208, 16

Morton, T. D., & Swift, J. 2014, ApJ, 791, 10

Moutou, C., Almenara, J. M., Díaz, R. F., et al. 2014, MNRAS, 444, 2783

Mullally, F., Coughlin, J. L., Thompson, S. E., et al. 2015, ApJS, 217, 31

Murray-Clay, R. A., Chiang, E. I., & Murray, N. 2009, ApJ, 693, 23

Nelson, B. E., Ford, E. B., Wright, J. T., et al. 2014, MNRAS, 441, 442

Nesvorný, D., Kipping, D., Terrell, D., et al. 2013, ApJ, 777, 3

Plummer, M. 2003, Proceedings of the 3rd International Workshop on Distributed Statistical Computing, http://www.r-project.org/conferences/DSC-2003/Proceedings/, ISSN:1609-395X

Queloz, D., Bouchy, F., Moutou, C., et al. 2009, A&A, 506, 303

Rogers, L. A., & Seager, S. 2010, ApJ, 716, 1208

Rogers, L. A., Bodenheimer, P., Lissauer, J. J., & Seager, S. 2011, ApJ, 738, 59

Rogers, L. A. 2015, ApJ, 801, 41

Rowe, J. F., Coughlin, J. L., Antoci, V., et al. 2015, ApJS, 217, 16

Sanchis-Ojeda, R., Fabrycky, D. C., Winn, J. N., et al. 2012, Nature, 487, 449

Sanchis-Ojeda, R., Rappaport, S., Winn, J. N., et al. 2013, ApJ, 774, 54

Sanchis-Ojeda, R., Winn, J. N., Marcy, G. W., et al. 2013, ApJ, 775, 54

Schmitt, J. R., Agol, E., Deck, K. M., et al. 2014, ApJ, 795, 167

Seager, S., Kuchner, M., Hier-Majumder, C. A., & Militzer, B. 2007, ApJ, 669, 1279

Stein, C. 1955, Proceedings of the Third Berkeley Symposium, 1, 197

Valencia, D., O'Connell, R. J., & Sasselov, D. 2006, Icarus, 181, 545

Valencia, D., Ikoma, M., Guillot, T., & Nettelmann, N. 2010, A&A, 516, A20

Vanderburg, A., Montet, B. T., Johnson, J. A., et al. 2015, ApJ, 800, 59

Weiss, L. M., Marcy, G. W., Rowe, J. F., et al. 2013, ApJ, 768, 14

Weiss, L. M., & Marcy, G. W. 2014, ApJ, 783, L6 (WM14)

Wolfgang, A., & Lopez, E. D.. 2015, in press.

Wu, Y., & Lithwick, Y. 2013, ApJ, 772, 74

Youdin, A. N. 2011, ApJ, 742, 38

Table 2
Masses and Radii of Small Planets

Planet Name	Period (days)	M_{obs} ($M_\oplus$)	σ_{Mobs} ($M_\oplus$)	R_{obs} ($R_\oplus$)	σ_{Robs} ($R_\oplus$)	First Reference	Mass, Radius Reference	Note
55 Cnc e	0.737	8.09	0.26	2.17	0.098	McArthur(2004)	Nelson(2014), Gillon(2012)	0
CoRoT-7 b	0.854	4.73	0.95	1.58	0.064	Queloz(2009); Leger(2009)	Barros(2014)	0
GJ 1214 b	1.580	6.45	0.91	2.65	0.09	Charbonneau(2009)	Carter(2011)	0,1
GJ 3470 b	3.337	13.73	1.61	3.88	0.32	Bonfils(2012)	Biddle(2014)	0
HD 97658 b	9.491	7.87	0.73	2.34	0.16	Howard(2011)	Dragomir(2013)	0,1
HIP 116454 b	9.12	11.82	1.33	2.53	0.18	Vanderburg(2015)	Vanderburg(2015)	0
Kepler-10 b	0.837	3.33	0.49	1.47	0.02	Batalha(2011)	Dumusque(2014)	0
Kepler-10 c	45.294	17.2	1.9	2.35	0.06	Batalha(2011)	Dumusque(2014)	0
Kepler-19 b	9.287	—	20.3	2.21	0.048	Borucki(2011)	Ballard(2011)	0,4
Kepler-20 b	3.696	8.7	2.2	1.91	0.16	Borucki(2011)	Gautier(2012)	0
Kepler-20 c	10.854	16.1	3.5	3.07	0.25	Borucki(2011)	Gautier(2012)	0
Kepler-20 d	77.612	—	20.1	2.75	0.23	Borucki(2011)	Gautier(2012)	0,4
Kepler-20 e	6.098	—	3.08	0.868	0.08	Borucki(2011)	Fressin(2012)	0,4
Kepler-20 f	19.58	—	14.3	1.03	0.11	Borucki(2011)	Fressin(2012)	0,4
Kepler-21 b	2.786	—	10.4	1.635	0.04	Borucki(2011)	Howell(2012)	0,4
Kepler-25 b	6.239	9.60	4.20	2.71	0.05	Borucki(2011)	Marcy(2014)	0,1
Kepler-37 b	13.367	2.78	3.70	0.32	0.02	Borucki(2011)	Marcy(2014)	0,1
Kepler-37 c	21.302	3.35	4.00	0.75	0.03	Borucki(2011)	Marcy(2014)	0,1
Kepler-37 d	39.792	1.87	9.08	1.94	0.06	Borucki(2011)	Marcy(2014)	0,1
Kepler-48 b	4.778	3.94	2.10	1.88	0.10	Borucki(2011)	Marcy(2014)	0,1
Kepler-48 c	9.674	14.61	2.30	2.71	0.14	Borucki(2011)	Marcy(2014)	0,1
Kepler-48 d	42.896	7.93	4.60	2.04	0.11	Borucki(2011)	Marcy(2014)	0,1
Kepler-62 b	5.715	—	9	1.31	0.04	Borucki(2011)	Borucki(2013)	0,4
Kepler-62 c	12.44	—	4	0.54	0.03	Borucki(2013)	Borucki(2013)	0,4
Kepler-62 d	18.164	—	14	1.95	0.07	Borucki(2011)	Borucki(2013)	0,4
Kepler-62 e	122.39	—	36	1.61	0.05	Borucki(2011)	Borucki(2013)	0,4
Kepler-62 f	267.29	—	35	1.41	0.07	Borucki(2013)	Borucki(2013)	0,6
Kepler-68 b	5.399	5.97	1.70	2.33	0.02	Borucki(2011)	Marcy(2014)	0,3
Kepler-68 c	9.605	2.18	3.50	1.00	0.02	Batalha(2013)	Marcy(2014)	0,3
Kepler-78 b	0.354	1.69	0.41	1.20	0.09	Sanchis-Ojeda(2013a)	Howard(2013)	0,1
Kepler-89 b	3.743	10.50	4.60	1.71	0.16	Borucki(2011)	Weiss(2013)	0,1
Kepler-93 b	4.727	4.02	0.68	1.48	0.019	Borucki(2011)	Dressing(2015)	0
Kepler-94 b	2.508	10.84	1.40	3.51	0.15	Borucki(2011)	Marcy(2014)	0,1
Kepler-95 b	11.523	13.00	2.90	3.42	0.09	Borucki(2011)	Marcy(2014)	0,1
Kepler-96 b	16.238	8.46	3.40	2.67	0.22	Borucki(2011)	Marcy(2014)	0,1
Kepler-97 b	2.587	3.51	1.90	1.48	0.13	Borucki(2011)	Marcy(2014)	0,1
Kepler-98 b	1.542	3.55	1.60	1.99	0.22	Borucki(2011)	Marcy(2014)	0,1
Kepler-99 b	4.604	6.15	1.30	1.48	0.08	Borucki(2011)	Marcy(2014)	0,1
Kepler-100 b	6.887	7.34	3.20	1.32	0.04	Borucki(2011)	Marcy(2014)	0,1
Kepler-100 c	12.816	0.85	4.00	2.20	0.05	Borucki(2011)	Marcy(2014)	0,1
Kepler-100 d	35.333	-4.36	4.10	1.61	0.05	Borucki(2011)	Marcy(2014)	0,1
Kepler-101 c	6.03	—	9	1.25	0.18	Borucki(2011)	Bonomo(2014)	0,5
Kepler-102 d	10.312	3.80	1.80	1.18	0.04	Borucki(2011)	Marcy(2014)	0,1
Kepler-102 e	16.146	8.93	2.00	2.22	0.07	Borucki(2011)	Marcy(2014)	0,1
Kepler-102 f	27.454	0.62	3.30	0.88	0.03	Borucki(2011)	Marcy(2014)	0,1

Table 2 — *Continued*

Planet Name	Period (days)	M_{obs} ($M_\oplus$)	σ_{Mobs} ($M_\oplus$)	R_{obs} ($R_\oplus$)	σ_{Robs} ($R_\oplus$)	First Reference	Mass, Radius Reference	Note
Kepler-102 b	5.287	0.41	1.60	0.47	0.02	Borucki(2011)	Marcy(2014)	0,1
Kepler-102 c	7.071	-1.58	2.00	0.58	0.02	Borucki(2011)	Marcy(2014)	0,1
Kepler-103 b	15.965	14.11	4.70	3.37	0.09	Borucki(2011)	Marcy(2014)	0,1
Kepler-106 b	6.165	0.15	2.80	0.82	0.11	Borucki(2011)	Marcy(2014)	0,1
Kepler-106 c	13.571	10.44	3.20	2.50	0.32	Borucki(2011)	Marcy(2014)	0,1
Kepler-106 d	23.980	-6.39	7.00	0.95	0.13	Batalha(2013)	Marcy(2014)	0,1
Kepler-106 e	43.844	11.17	5.80	2.56	0.33	Borucki(2011)	Marcy(2014)	0,1
Kepler-109 b	6.482	1.30	5.40	2.37	0.07	Borucki(2011)	Marcy(2014)	0,1
Kepler-109 c	21.223	2.22	7.80	2.52	0.07	Borucki(2011)	Marcy(2014)	0,1
Kepler-113 b	4.754	7.10	3.30	1.82	0.05	Borucki(2011)	Marcy(2014)	0,1
Kepler-113 c	8.925	-4.60	6.20	2.19	0.06	Borucki(2011)	Marcy(2014)	0,1
Kepler-131 b	16.092	16.13	3.50	2.41	0.20	Borucki(2011)	Marcy(2014)	0,1
Kepler-131 c	25.517	8.25	5.90	0.84	0.07	Batalha(2013)	Marcy(2014)	0,1
Kepler-406 b	2.426	4.71	1.70	1.43	0.03	Borucki(2011)	Weiss(2014)	0,1
Kepler-406 c	4.623	1.53	2.30	0.85	0.03	Batalha(2013)	Weiss(2014)	0,1
Kepler-407 b	0.669	0.06	1.20	1.07	0.02	Borucki(2011)	Marcy(2014)	0,1
Kepler-409 b	68.958	2.69	6.20	1.19	0.03	Batalha(2013)	Marcy(2014)	0,1
Kepler-4 b	3.213	24.47	3.81	4.00	0.21	Borucki(2010)	Borucki(2010)	
GJ 436 b	2.64	25.4	2.1	4.10	0.16	Butler(2004)	Lanotte(2014)	
Kepler-89 c	10.42	15.6	10.6	4.32	0.41	Batalha(2013)	Weiss(2013)	
HAT-P-11 b	4.888	25.74	2.86	4.73	0.157	Bakos(2010)	Bakos(2010)	
CoRoT-22 b	9.756	—	35	4.88	0.28	Moutou(2014)	Moutou(2014)	4
Kepler-103 c	179.61	36.1	25.2	5.14	0.14	Borucki(2011)	Marcy(2014)	
Kepler-101 b	3.488	51.1	4.9	5.77	0.82	Borucki(2011)	Bonomo(2014)	
Kepler-63 b	9.43	—	95	6.1	0.2	Borucki(2011)	Sanchis-Ojeda(2013b)	6
HAT-P-26 b	4.235	18.75	2.23	6.33	0.58	Hartman(2011)	Hartman(2011)	
CoRoT-8 b	6.212	69.92	9.53	6.39	0.22	Borde(2010)	Borde(2010)	
Kepler-89 e	54.32	35	23	6.56	0.62	Batalha(2013)	Weiss(2013)	
Kepler-11 b	10.304	1.90	1.2	1.80	0.04	Lissauer(2011)	Lissauer(2013)	1,2
Kepler-11 c	13.024	2.90	2.3	2.87	0.06	Lissauer(2011)	Lissauer(2013)	1,2
Kepler-11 d	22.684	7.30	1.2	3.12	0.07	Lissauer(2011)	Lissauer(2013)	1,2
Kepler-11 f	46.689	2.00	0.9	2.49	0.06	Lissauer(2011)	Lissauer(2013)	1,2
Kepler-11 g	118.38	—	25	3.33	0.07	Lissauer(2011)	Lissauer(2013)	2,4
Kepler-18 b	3.505	6.9	3.4	2.00	0.100	Borucki(2011)	Cochran(2011)	1,2
Kepler-30 b	29.334	11.3	1.4	3.90	0.200	Borucki(2011)	Sanchis-Ojeda(2012)	1,2
Kepler-36 b	13.840	4.45	0.30	1.486	0.035	Carter(2012)	Carter(2012)	1,2
Kepler-36 c	16.239	8.08	0.53	3.679	0.054	Borucki(2011)	Carter(2012)	1,2
Kepler-79 b	13.485	10.9	6.7	3.47	0.07	Borucki(2011)	Jontof-Hutter(2014)	1,2
Kepler-79 c	27.403	5.9	2.1	3.72	0.08	Borucki(2011)	Jontof-Hutter(2014)	1,2
Kepler-79 e	81.066	4.1	1.2	3.49	0.14	Batalha(2013)	Jontof-Hutter(2014)	1,2
Kepler-88 b	10.954	8.7	2.5	3.78	0.38	Borucki(2011)	Nesvorny(2013)	2
Kepler-138 c	13.782	3.83	1.39	1.610	0.160	Borucki(2011)	Kipping(2014)	2
Kepler-138 d	23.089	1.01	0.38	1.610	0.160	Borucki(2011)	Kipping(2014)	2
Kepler-289 b	34.545	7.3	6.8	2.15	0.1	Borucki(2011)	Schmitt(2014)	2
Kepler-289 d	66.063	4.0	0.9	2.68	0.17	Borucki(2011)	Schmitt(2014)	2

Note. —

0. Included in baseline dataset, which consists of RV masses (see §3).

1. Mass, radius values and their error bars are unchanged (within rounding error) from WM14.

2. Mass measured by fitting the observed TTVs to N-body integrations of the system.

3. The Kepler-68 planets were repeated twice in the WM14 dataset, so we use the Marcy et al. (2014) values.

4. The σ_{Mobs} column contains the 2σ upper limit as reported in the second reference.

5. Only a 1σ upper limit of 3.78 was given, and no posteriors were shown; in this analysis, we set the 2σ upper limit at 9 $M_\oplus$ to include 1.8 m/s uncertainty quoted in RV semi-amplitude for the larger Kepler-101 b.

6. The 2σ upper limit is interpolated from given 1σ and 3σ upper limits.