The Radius Cliff is a Waterfall: Explaining Sub-Neptune Exoplanets with Steam Worlds

Aritra Chakrabarty,[1,2,*] Gijs D. Mulders,[3] Artyom Aguichine,[4,5] and Natalie Batalha[4]

[1] *NASA Ames Research Center, Moffett Field, CA, USA*
[2] *Oak Ridge Associated Universities*
[3] *Instituto de Astrofísica, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 7820436 Macul, Santiago, Chile*
[4] *Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA, USA*
[5] *Instituto de Astronomía, Universidad Nacional Autónoma de México, Apartado Postal 106, CP 22800 Ensenada, Baja California, México*

(Received 2025 September 30; Revised 2026 January 8; Accepted 2026 January 23)

ABSTRACT

The demographics of Kepler planets provide a key testbed for models of planet formation and evolution, particularly for explaining the *radius valley* separating super-Earths and sub-Neptunes. A primordial interpretation based on differences in bulk densities—where rocky and water-rich planets form via migration pathways—offers an alternative to atmospheric loss scenarios. Updated interior structure models of water worlds with adiabatic steam atmospheres reproduce the observed valley near $\sim 2, R_\oplus$ more accurately. Furthermore, migration models from our Genesis library suggest that these formation pathways can also account for the distinct period distributions of super-Earths and sub-Neptunes, as well as the emergence of the hot Neptune desert. Motivated by this, we develop a Bayesian hierarchical mixture model for close-in Kepler planets ($P < 100$ days), combining rocky planets and water worlds without H/He envelopes. The inferred mass distributions of rocky and water-rich planets peak at $\sim 2.6 M_\oplus$ and $\sim 7 M_\oplus$, respectively, with the water mass fraction of water worlds peaking at $\sim 41\%$. Water worlds provide a good representation of the Kepler sub-Neptune population, with the *radius cliff* emerging as a "waterfall"—a sharp decline in their occurrence. However, our mass–radius analysis shows that water worlds alone cannot explain planets with $R \gtrsim 3, R_\oplus$, implying that at least $\sim 20\%$ of sub-Neptunes in the sample are enriched in H/He gas.

1. INTRODUCTION

The Kepler mission has revealed a rich and diverse population of small, close-in exoplanets. The distinct patterns observed within this population reflect underlying differences in planetary composition, formation, or evolutionary history, and offer critical constraints for models of planet formation and atmospheric evolution. Among the most striking features are the *radius valley*, a deficit of planets $\sim 1.8\ R_\oplus$ (Fulton et al. 2017; Fulton & Petigura 2018; Van Eylen et al. 2019; Ho & Van Eylen 2023), which divides the population into super-Earths and sub-Neptunes, and the *radius cliff* near $4\ R_\oplus$ that marks a sharp drop in the occurrence of small planets (Fulton et al. 2017; Fulton & Petigura 2018; Dattilo et al. 2023; Dattilo & Batalha 2024). Another closely related feature is the lack of ultra-short-period ($P \lesssim 10$ days) sub-Neptunes and Neptune-sized planets, known as the *hot Neptune desert* (Mazeh et al. 2016).

Atmospheric loss models (Owen & Wu 2017; Ginzburg et al. 2018; Gupta & Schlichting 2019; Owen & Campos Estrada 2020; Rogers & Owen 2021) explain the radius valley as a consequence of evolutionary processes, wherein initially gaseous sub-Neptunes, also called gas dwarfs, lose their H/He envelopes and are stripped down to bare rocky cores. While successful in many respects, these models work well with an *in situ* formation scenario (Owen & Wu 2017; Rogers & Owen 2021), which is broadly inconsistent with planet formation theories (Ogihara et al. 2015). Furthermore, they require additional mechanisms to account for the radius cliff near $\sim 4\ R_\oplus$, such as spontaneous mass loss of H/He envelopes (Ikoma & Hori 2012; Ginzburg et al. 2016; Owen & Wu 2016), sequestration of H_2 in magma oceans (Kite et al. 2019), among others. Alternatively, formation theories propose that the observed dichotomy may arise from intrinsic differences in bulk composition—namely, rocky super-Earths versus water-ice-rich sub-Neptunes— shaped by accretion and migration processes in the early protoplanetary disk (Mordasini et al.

Corresponding author: Aritra Chakrabarty
aritra.astrophysics@gmail.com

* NASA Postdoctoral Program Fellow

2009; Izidoro et al. 2017; Venturini et al. 2020; Mulders et al. 2020; Izidoro et al. 2022). This removes the need for all sub-Neptunes to possess significant H/He-rich envelopes to begin with.

Initial water world models assumed water as a condensed, high-pressure ice layer (Zeng et al. 2019), resulting in planets less "puffy" than gas dwarfs. Population models based on such structures often struggle to match observations, typically placing water worlds either at the radius valley (Venturini et al. 2020; Burn et al. 2024) or in the super-Earth regime (Chakrabarty & Mulders 2024a). In contrast, models invoking hot, adiabatic, supercritical steam layers for close-in water worlds offer a more natural explanation for the radius valley, as steam atmospheres substantially inflate planetary radii (Turbet et al. 2020; Aguichine et al. 2021). Burn et al. (2024) showed that their migration-based planet formation model can reproduce the $\lesssim 4\ R_\oplus$ Kepler population when it includes water worlds with mixed steam and hydrogen atmospheres undergoing photoevaporation. However, in the absence of atmospheric escape, the model underproduces rocky super-Earths and fails to generate a sufficient number of water worlds beyond $\sim 3\ R_\oplus$ (see Figure 2c of Burn et al. (2024)).

This raises an important question for formation models: Can a primordial population of rocky planets and water-rich worlds reproduce the observed Kepler demographics without invoking post-formation atmospheric evolution? Moreover, can formation and migration processes inherently reproduce the transition from the hot Neptune desert regime to the warm sub-Neptune-dominated regime across orbital periods (Fulton et al. 2017; Fulton & Petigura 2018; Petigura et al. 2018; Pascucci et al. 2019; Bergsten et al. 2022)—a feature that atmospheric loss models naturally reproduce, providing compelling evidence in their favor (Owen & Wu 2017; Rogers & Owen 2021)? Addressing these questions can help guide the development of future formation-based population synthesis models.

In this work, we present a hierarchical population model of rocky planets and water-rich steam worlds devoid of H/He envelopes. Using a Bayesian inference framework, we analyze the period–radius distribution of close-in Kepler planets (orbital periods < 100 days) from the California-Kepler Survey (CKS, Johnson et al. 2017; Petigura et al. 2017), constraining the underlying period, mass, and water mass-fraction distributions that best reproduce the observed features. We further assess our inferred model by comparing it with the observed mass–radius distribution and with mass distributions derived from independent radial velocity (RV) surveys. Our results offer constraints for future formation models aiming to explain the observed exoplanet demographics and highlight how future observations can help narrow down the processes that likely shaped the Kepler population.

§2 outlines our methodology, including the definition of the mass, orbital period, and water mass-fraction distributions for rocky planets and water worlds, and how we infer the model parameters by maximizing the likelihood of detecting the Kepler planets under our framework. §3 presents the results of our Bayesian inference. §4 discusses how our model can inform and guide future observational efforts. Finally, we summarize the key conclusions in §5.

2. METHOD

We develop a Bayesian hierarchical model of planet populations based on the theory of *ab initio* dichotomy in the planetary bulk compositions, to represent the small, short-period planets observed by Kepler. Accordingly, we assume that the Kepler super-Earths are primarily rocky, and the sub-Neptunes are composed of water-ice-rich cores. For the water worlds, we adopt an updated internal structure model that extends the previous isothermal condensed water-layer model (e.g., Zeng et al. 2019) by incorporating supercritical and extended steam atmospheres. These "steam worlds" (Aguichine et al. 2021; Burn et al. 2024) are physically larger than their condensed counterparts and represent promising candidates for explaining the observed radii of Kepler sub-Neptunes. All planets in our model are assumed to be bare, i.e., devoid of primordial H/He envelopes.

We then use a Bayesian framework to fit our model to the observed period–radius (P-R_P) distribution of Kepler planets with radii between 0.9–6 $R_\oplus$ and period (P) < 100 days. Through this inference, we aim to constrain the following aspects of the planet population:

- The mass distributions of rocky planets (M_r) and water worlds (M_w).
- The water mass-fraction (WMF) distribution of the water worlds.
- The period distributions of the rocky planets (P_r) and water worlds (P_w).
- The occurrence rate of planets per star (f_*) within the range of $0.9\ R_\oplus < R_P < 6\ R_\oplus$ and $P < 100$ days.

2.1. *Distributions of Period, Mass, and Water Mass-Fraction in the Population Model*

As a first step in our hierarchical mixture model, we define the distributions of orbital period, mass, and

WMF from which we draw planet samples. Each of the rocky and water-rich populations is assigned its own conditional distributions of period, mass, and WMF, as explained in the following subsections.

2.1.1. *Orbital period distributions*

To define the period distribution of the ensemble population, we adopt the smooth broken power-law form from Rogers & Owen (2021), which fits the observed Kepler occurrence rates:

$$\frac{dN}{d\log P} \propto \frac{1}{(\frac{P}{P_0})^{-k_1} + (\frac{P}{P_0})^{-k_2}} \tag{1}$$

The overall period distribution is partitioned into composition-based conditional distributions for rocky planets and water worlds, motivated by the observed trend that super-Earths dominate at shorter orbital periods while sub-Neptunes become more common at longer periods (Fulton & Petigura 2018; Petigura et al. 2018; Pascucci et al. 2019; Bergsten et al. 2022). Our model assumes that this difference is a primordial feature shaped by formation and migration processes, which is further discussed and justified in §3.1. We therefore adopt empirical fits to the bias-corrected Kepler period distributions reported by Bergsten et al. (2022). These fits provide a functional form for the fractional occurrence of super-Earths as a function of period, $G(P)$, from which we define the conditional period distributions of rocky planets and water worlds as:

$$\begin{aligned} \frac{dN}{d\log P_r} &\propto G\frac{dN}{d\log P} \\ \frac{dN}{d\log P_w} &\propto (1-G)\frac{dN}{d\log P} \end{aligned} \tag{2}$$

G(P) is defined by Bergsten et al. (2022) via a hyperbolic tangent transition function:

$$t(P) = 0.5 - 0.5\tanh\left(\frac{\log P - \log P_{\rm cen}}{\log s}\right) \tag{3}$$

with

$$G(P) = t(P)\cdot\chi_1 + (1-t(P))\cdot\chi_2 \tag{4}$$

Here, P_{cen} is the transition center, s controls the smoothness of the transition, and χ_1 and χ_2 are the limiting fractions of super-Earths at short and long periods, respectively.

Since G(P) is defined as the fractional occurrence of super-Earths within the ensemble, it also allows us to calculate the global fractions of rocky planets (r_f) and water worlds (1-r_f), where r_f can be calculated as

$$r_f = \frac{\int G(P)\frac{dN}{d\log P}d\log P}{\int \frac{dN}{d\log P}d\log P} \tag{5}$$

2.1.2. *Mass Distributions*

Choosing an appropriate mass distribution for the Kepler planets is challenging. While a subset of Kepler planets have mass measurements, and some broader trends in planetary masses have emerged (Howard et al. 2010; Cassan et al. 2012; Marcy et al. 2014; Fulton et al. 2021, etc.) from independent radial velocity and microlensing surveys (often probing different planet populations), a detailed and completeness-corrected mass distribution specifically for Kepler super-Earths and sub-Neptunes remains unavailable. Several studies have instead used forward population modeling to infer the mass distribution indirectly. For example, Wu (2019); Rogers & Owen (2021) showed that the observed Kepler period-radius distribution can be reproduced using a log-normal distribution of masses.

However, we find that applying a single log-normal mass distribution across the full planet population leads to poor convergence of our model, primarily due to contamination of the rocky component with implausibly large masses. We therefore adopt separate log-normal mass distributions for rocky planets (M_r) and water worlds (M_w), allowing greater flexibility in capturing the underlying physical differences between the two populations:

$$\begin{aligned} \frac{dN}{d\log M_r} &= \mathcal{N}(\log\mu_{mr}, \sigma_{mr}), \\ \frac{dN}{d\log M_w} &= \mathcal{N}(\log\mu_{mw}, \sigma_{mw}) \end{aligned} \tag{6}$$

This choice of separate mass distributions not only improves the stability of the model fit but is also supported by observed differences in the mass–radius relationships of super-Earths and sub-Neptunes (Wolfgang et al. 2016; Chen & Kipping 2017). The resulting combined mass distribution remains of particular interest and is further discussed in §3.3.

2.1.3. *WMF Distributions*

The WMF is set to zero for rocky planets, consistent with a dry, Earth-like bulk composition. For the water worlds, however, the WMF distribution is not constrained by models or observations. Therefore, the first three models we test correspond to three different WMF distributions for the water worlds:

- Model I: A truncated normal distribution, $\mathcal{N}(\mu_{\rm WMF}, \sigma_{\rm WMF})$, restricted to [0, 0.5];
- Model II: A uniform distribution, $\mathscr{U}(a_{\rm WMF}, b_{\rm WMF})$, with $b_{\rm WMF} < 0.5$;

- Model III: A non-parametric model based on a fifth-order Bernstein polynomial expansion, normalized over the range (0, 0.5).

We truncate the WMF at 0.5 as a pragmatic, physically motivated upper limit. Condensation chemistry and the elemental budget of a solar-composition nebula suggest that water ice constitutes $\lesssim 50\%$ of the solid mass (Lodders 2003; Marboeuf et al. 2014). In addition, processes such as radial transport (pebble drift and sublimation at the snowline; Booth et al. 2017) and post-accretion evolution (e.g., differentiation and volatile loss; Genda & Abe 2003) typically reduce the water ultimately incorporated into planets to well below this ceiling.

Furthermore, to assess whether this truncation on WMF affects the MCMC convergence, we also develop Model IV, which considers a normal distribution of WMF like Model I but untruncated, i.e., $0 <$ WMF < 1. We evaluate each of the proposed WMF models within our Bayesian framework, as detailed in the following sections. The rationale behind our final choice of WMF parameterization is discussed in §2.6.

2.2. *Properties of the host stars*

We adopt Sun-like central stars, in both mass and luminosity, for Models I–IV. To investigate the effect of host star properties, we develop Model V, in which the planets from Model I are placed around Kepler-like host stars. The stellar mass (M_*) and luminosity (L_*) for these stars are directly drawn from the CKS stellar sample via bootstrapping (with replacement).

2.3. *Calculating Size of the Planets*

We assume an Earth-like rocky composition for the rocky planets with 32.5% iron and 67.5% silicates by mass. We then use the mass-radius relation from Zeng et al. (2019), $m = r^{3.7}$, where both the mass (m) and size (r) are in Earth units. For the water worlds, we assume the same core composition and consider an adiabatic layer of H_2O (steam) on top. To calculate the sizes of these water (steam) worlds, we adopt the equations of state and the static internal structure model of Aguichine et al. (2021), in which the entropy of the steam envelope does not evolve. This model provides mass–radius relationships for such planets as a function of core fraction, WMF, and irradiation temperature. We interpolate from their publicly available model grid[1] to calculate the radii of our water worlds. Figure 1 illustrates the mass–radius models for rocky planets and water worlds with WMF = 0.5, where steam atmospheres expand relative to the condensed-water model with increasing equilibrium temperature.

[1] https://archive.lam.fr/GSP/MSEI/IOPmodel/mr_all.dat

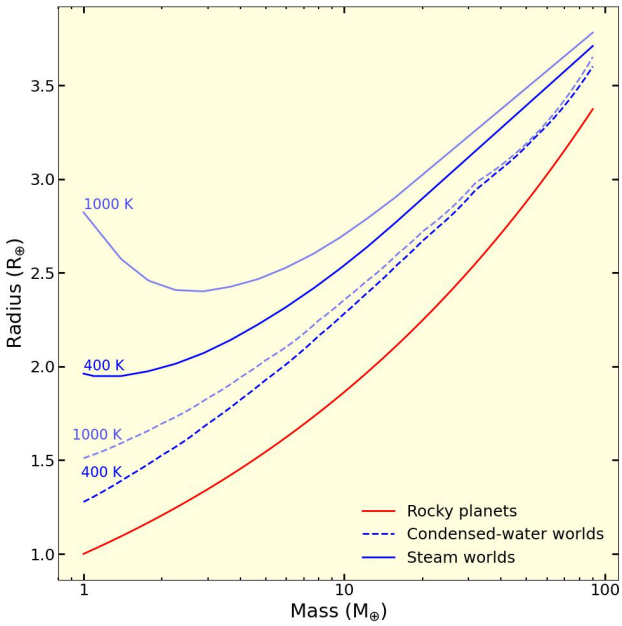

Figure 1. Mass–radius models for rocky planets (Zeng et al. 2019) and water worlds with WMF = 0.5, having condensed-water layers (Zeng et al. 2019) and steam atmospheres (Aguichine et al. 2021).

2.4. *Calculation of Likelihood for Bayesian Inference*

To calculate the occurrence rates from our synthetic planet population, we first compute the probability density function (PDF) of occurrence as a function of radius and orbital period of the planets using 2D Gaussian Kernel Density Estimation (KDE) technique (implemented via the Python package *scipy.stats.gaussian_kde*). We calculate the KDE of the population (p_{kde}) in the period-radius space on a logarithmic scale over a defined grid of periods and radii. The KDE bandwidth is set to the package's baseline value, determined using Scott's rule. We additionally explore alternative bandwidths by scaling this baseline up and down, and find that our results are largely insensitive to the choice of bandwidth within a reasonable range, as discussed in Section 2.6. We then calculate the occurrence rate per star, $\lambda(P, R_P)$, by multiplying p_{kde} with f_*, the average number of planets per star within the range $0.9\ R_\oplus < R_P < 6\ R_\oplus$ and $P < 100$ days.

We then introduce biases associated with the CKS with the help of its completeness map. We follow the approach of Fulton et al. (2017); Fulton & Petigura (2018)

and partly use their code *cksgaia*[2] for this purpose. Accordingly, we take the product of the transit probability and the detection probability to calculate the completeness factor at the same grid points over which we calculate the KDE. *cksgaia* calculates the transit probability as

$$p_{transit} = b_{cut}\frac{R_*}{a} \tag{7}$$

R_* is the stellar radius, a is the semi-major axis, and $b_{cut} = 0.7$ is the cutoff in the transit parameter (Fulton et al. 2017; Fulton & Petigura 2018). The mean probability of detection resulting from the Kepler pipeline efficiency is then calculated as

$$p_{detection} = \frac{1}{N_*}\sum_{k}^{N_*} C(m_k) \tag{8}$$

N_* denotes the number of stars in the CKS sample, and $C(m_i)$ denotes the fraction of injected planets that can be recovered using the Kepler pipeline as a function of its MES, i.e., the injected signal-to-noise ratio. $C(m_i)$ has been calculated using a Gamma cumulative distribution function (see Equations 2 and 3 of Fulton et al. (2017)). The total completeness function is given by:

$$\eta = p_{transit} \cdot p_{detection} \tag{9}$$

Thus, the average number of detectable planets in a grid cell (i, j), with a grid spacing of ΔP and ΔR_P, is given by

$$\Lambda_{ij} = N_* \lambda_{ij} \eta_{ij} \Delta P \ \Delta R_P \tag{10}$$

We assume that the exoplanets can be detected independently from each other (Bryson et al. 2020, 2021; Rogers & Owen 2021). Hence, we treat the detection of an exoplanet as a Poisson point process. The likelihood of detecting the ensemble of planets (π_{cks}) from the CKS sample given the free parameters (Θ) can be expressed as (see Bryson et al. (2020))

$$\begin{aligned} p(\pi_{cks}|\Theta) = &\prod_{i,j \in \pi_{cks}} Poisson(n_{ij} = 1, \Lambda_{ij}(\Theta)) \\ &\times \prod_{i,j \notin \pi_{cks}} Poisson(n_{ij} = 0, \Lambda_{ij}(\Theta)) \end{aligned} \tag{11}$$

n_{ij} denotes the number of planets detected in the grid cell (i, j). $Poisson(n, \Lambda)$ denotes the Poisson probability density function with mean Λ.

Thus, finding the free parameters consistent with the CKS sample translates to maximizing the posterior probability that the current set of parameters represents the population in the CKS sample, $p(\Theta|\pi_{cks})$. Following Bayes' theorem, this can be expressed, as

$$p(\Theta|\pi_{cks}) \propto p(\pi_{cks}|\Theta)\ p(\Theta) \tag{12}$$

$p(\Theta)$ is a prior on Θ, which is explained in detail in the following subsection.

2.5. *Maximizing the posterior likelihood using MCMC*

We maximize Equation 12 using Markov Chain Monte Carlo (MCMC), implemented via the Python package *emcee*, to obtain the posterior distributions of the free parameters (Θ). The parameters P_0, k_1, and k_2 in Equation 1 only dictate the period distribution of the ensemble population. Similarly, the parameter $P_{\rm cen}$ in Equation 3 controls only the period at which the transition from super-Earth–dominated to sub-Neptune–dominated populations occurs. These parameters can therefore be constrained directly from the period distribution alone, and including them in the joint fit to the period–radius distribution only slows the convergence of the MCMC across all models without affecting our conclusions. Accordingly, rather than re-inferring them, we adopt values from the literature. We take P_0, k_1, and k_2 from Rogers & Owen (2021), who demonstrated that these parameters successfully reproduce the Kepler period distribution. We fix $P_{\rm cen} = 11.33$ days, corresponding to the median value reported by Bergsten et al. (2022) for stars in the mass range of 0.56–1.63 $M_\odot$(see Table 2 of Bergsten et al. (2022)).

Thus, the free parameters (Θ) of our model are:

- Parameters associated with the mass distributions: μ_{mr}, σ_{mr}, μ_{mw}, and σ_{mw}
- Parameters associated with the WMF distributions: i) $\mu_{\rm WMF}$ and $\sigma_{\rm WMF}$ for the normal distribution, ii) $a_{\rm WMF}$ and $b_{\rm WMF}$ for the uniform distributions, and iii) the Bernstein coefficients, $\beta_{k,\rm WMF}$ $(k = 1, .., 5)$, for the Bernstein polynomial-based distribution.
- Parameters governing the period distributions: s, χ_1, and χ_2
- Parameter related to the ensemble: f_*

We assign uniform priors to all free parameters. The priors for the free parameters associated with the mass distributions are chosen empirically. The WMF distribution parameters are also chosen empirically depending on the distribution type, with the constraint $0 <$ WMF < 0.5 ($0 <$ WMF < 1 in the case of Moel IV). The prior for the parameter s is guided by Bergsten et al. (2022), while χ_1 and χ_2 are assigned uniform priors

[2] https://github.com/California-Planet-Search/cksgaia

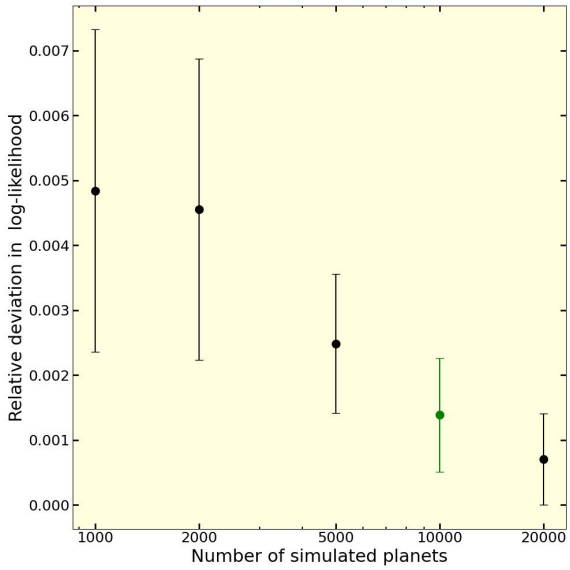

Figure 2. Relative deviation of the log-likelihood values for different numbers of simulated planets with respect to the median log-likelihood obtained from 100,000 simulated planets. The green point denotes our chosen number of simulated planets, i.e., 10,000, with which we get <0.2% deviation while maintaining a manageable computational cost.

over the full range (0, 1). The prior for f_* is informed by a rough estimate obtained by dividing the sum of completeness weights ($1/\eta$) computed for the planets in the CKS sample within our region of interest by N_*, thereby providing a physically motivated constraint.

We simulate 10,000 planets in each run to accurately compute the detection likelihood from our model. Figure 2 shows the relative deviation of the log-likelihood ($\log p(\pi_{cks}|\Theta)$) from the median value obtained from 100,000 simulated planets, plotted as a function of the number of simulated planets. Using 10,000 planets achieves adequate accuracy without excessive computational expense.

2.6. *Comparison of Predictive Performance of the Models*

We evaluate the performance of models I-V presented in §2.1 and §2.2 by computing the expected log pointwise predictive density (ELPD) using the widely applicable information criterion (WAIC) method, along with its uncertainty, with the help of the Python package *ArviZ*. ELPD provides a robust estimate of a model's predictive accuracy by summing the expected log-likelihoods for individual data points. WAIC offers a Bayesian approximation to ELPD by incorporating the full posterior distribution and penalizing model complexity, making it suitable for comparing models in a probabilistic framework (Watanabe 2010; Vehtari et al. 2017). Figure 3 shows the 1σ range of the posterior WMF distributions from the four models (top) and a comparison of their ELPD values and uncertainties (bottom).

The results indicate that the ELPD values for all five models largely overlap. Notably, the 1σ range of WMF distributions from Model II resembles a normal distribution similar to that from Model I, rather than a uniform distribution, indicating a preferred WMF value inferred from the data. Again, the untruncated normal WMF model in Model IV shows that the peak location and overall WMF distribution closely resemble those of Model I, except for being truncated at the right edge. The ELPD value is also similar, suggesting that the applied truncation on the WMF distribution—motivated by disk chemistry and accretion processes—barely affects the MCMC convergence. Again, adopting Kepler-like central stars in Model V causes the WMF distribution peak at a slightly lower value than Model I with sun-like stars, albeit with 1σ overlap with Model I. However, their predictive performances, measured by the ELPD, are statistically indistinguishable.

Furthermore, we find that the predictive performance of Model I is robust to the choice of KDE bandwidth within a reasonable range. We repeat the analysis for Model I by scaling the baseline bandwidth by factors of 0.25, 0.5, and 2, and find that both the ELPDs and the inferred parameters remain very close to their baseline values. For a bandwidth scaling factor of 4, however, the MCMC chains fail to converge, producing a NaN value for the ELPD. Accordingly, we adopt the baseline KDE bandwidth for all results presented in the following section.

3. RESULTS AND DISCUSSION

The posterior mean and uncertainties of the free parameters from MCMC for Models I-V are shown in Table 1. The inferred free parameters are highly consistent across models. Although Models I–IV differ by construction in their WMF distribution parameterizations, the inferred WMF distributions are nonetheless broadly similar across models, with the exception of Model III (see Figure 3). We therefore adopt Model I as our fiducial model and present all subsequent results based on Model I. The full MCMC posterior distribution for Model I is shown in Appendix Figure A1.

3.1. *The period distributions*

We directly adopt the observed period distribution of Kepler planets as fixed input to our model and so, we

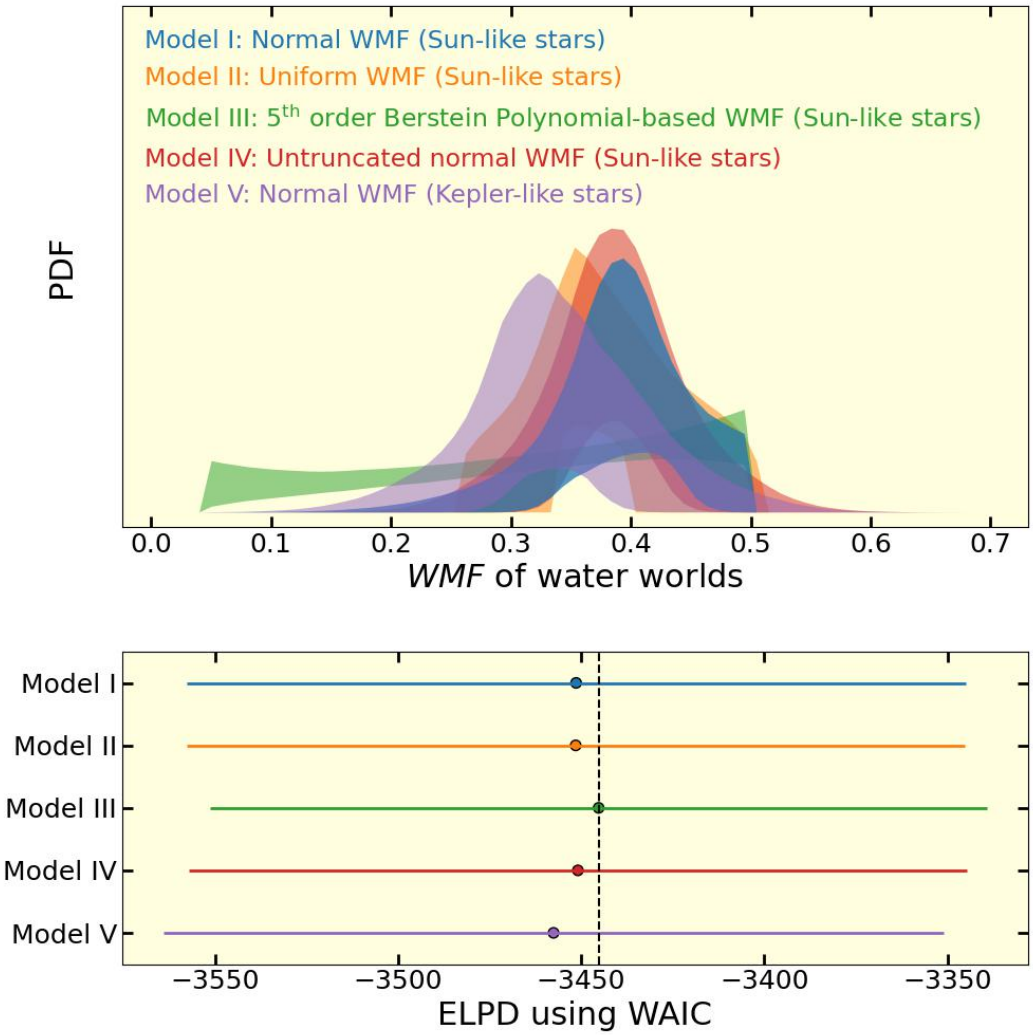

Figure 3. Top: Shaded regions indicate the 1-σ range of WMF distributions from the different models. Bottom: ELPD values and their associated uncertainties for the different models. The ELPD values show significant overlap, with Model I slightly outperforming the others.

do not compare the ensemble period distribution. Instead, we focus on comparing the period distributions of the individual rocky and water-rich populations with the Kepler super-Earths and sub-Neptunes, respectively. As shown in the top panel of Figure 4, Model I reproduces the observed distributions closely, even though the individual period distributions were not explicitly fit. The inferred parameters χ_1 and χ_2 are slightly higher and lower, respectively (deviating by more than 1σ but within 2σ), compared to the values reported by Bergsten et al. (2022), while s is marginally higher but shows 1σ overlap. These differences can be attributed to our joint modeling of the period and radius distributions.

The distinct period distributions assumed for the rocky and water-rich planets align with predictions from planet formation and migration models. Rocky planets that form early and inside the snowline tend to migrate inward rapidly, populating the inner regions of the system (Mulders et al. 2020; Venturini et al. 2020). In contrast, the water-rich planets originating beyond the snowline often migrate more slowly and may stall at slightly longer orbital periods. This behavior may be influenced by dynamical processes such as trapping in mean-motion resonances with interior planets (Mulders et al. 2020), the outward movement of the magnetospheric cavity during disk dispersal (Liu et al. 2017), among others.

To further assess the consistency, we use the migration-based population models from the *Genesis* library[3] (Mulders et al. 2020) for comparison. *Genesis*

[3] https://github.com/GijsMulders/Genesis

Table 1. Estimates of the Model Parameters Derived from the MCMC Posterior Samples for All Five Models Considered

Parameters	Model I	Model II	Model III	Model IV	Model V
μ_{mr}	2.65 ± 0.14	2.62 ± 0.14	2.52 ± 0.15	2.65 ± 0.14	2.62 ± 0.15
σ_{mr}	0.74 ± 0.06	0.74 ± 0.06	0.71 ± 0.06	0.74 ± 0.06	0.73 ± 0.07
μ_{mw}	6.99 ± 2.36	$7.76^{+2.13}_{-2.78}$	$9.41^{+1.95}_{-3.8}$	$7.37^{+2.01}_{-2.64}$	$6.2^{+2.97}_{-2.07}$
σ_{mw}	$1.82^{+0.22}_{-0.16}$	$1.8^{+0.22}_{-0.16}$	$1.71^{+0.29}_{-0.16}$	$1.81^{+0.24}_{-0.17}$	$1.83^{+0.14}_{-0.17}$
$\mu_{\rm WMF}$	0.41 ± 0.05	–	–	0.39 ± 0.03	0.34 ± 0.05
$\sigma_{\rm WMF}$	0.06 ± 0.05	–	–	0.04 ± 0.03	0.04 ± 0.04
$a_{\rm WMF}$	–	0.31 ± 0.05	–	–	–
$b_{\rm WMF}$	–	0.45 ± 0.06	–	–	–
$\beta_{0,\rm WMF}$	–	–	$0.2^{+0.35}_{-0.15}$	–	–
$\beta_{1,\rm WMF}$	–	–	$0.19^{+0.33}_{-0.15}$	–	–
$\beta_{2,\rm WMF}$	–	–	$0.34^{+0.38}_{-0.25}$	–	–
$\beta_{3,\rm WMF}$	–	–	$0.63^{+0.27}_{-0.38}$	–	–
$\beta_{4,\rm WMF}$	–	–	$0.74^{+0.19}_{-0.33}$	–	–
f_*	1.12 ± 0.05	1.12 ± 0.04	1.25 ± 0.05	1.11 ± 0.04	1.11 ± 0.05
s	$3.29^{+0.7}_{-0.62}$	$3.4^{+0.73}_{-0.6}$	$3.24^{+0.5}_{-0.58}$	$3.29^{+0.75}_{-0.63}$	$3.29^{+0.95}_{-0.61}$
χ_1	$0.97^{+0.02}_{-0.04}$	$0.98^{+0.02}_{-0.03}$	$0.97^{+0.02}_{-0.05}$	$0.97^{+0.02}_{-0.04}$	$0.96^{+0.03}_{-0.04}$
χ_2	0.22 ± 0.06	0.21 ± 0.06	0.20 ± 0.06	$0.21^{+0.07}_{-0.04}$	0.22 ± 0.06
r_f	0.47 ± 0.04	0.47 ± 0.04	0.46 ± 0.03	0.47 ± 0.04	0.47 ± 0.04

NOTE—r_f is derived from the posterior sample of the free parameters for each model using Equation 5

Population Synthesis[4] (GPS, Chakrabarty & Mulders 2024b) identifies the rocky and water-rich populations from the ensemble population from the *Genesis* library by calculating the WMF of those planets as a function of the snowline of the disk. The bottom panel of Figure 4 compares the mean period distributions of the rocky and water worlds from Model I to those from *GPS*. Although the distributions are not identical, the transition in the relative occurrence of rocky planets and water worlds with orbital period is broadly consistent between our current model (Model I) and *GPS*. This transition is further found to be sensitive to the assumed location of the snowline in the protoplanetary disk, which in turn depends on the properties of the host star, such as the stellar mass (Mulders et al. 2015) and stellar metallicity (Dominguez 2020). This may potentially explain the increased occurrence of warm sub-Neptunes with increasing stellar metallicity (Mulders et al. 2016; Petigura et al. 2018). The bottom panel of Figure 4 presents the case for a snowline at 2.7 AU for a G-type star, which produces the closest agreement with the transition period observed in the Kepler sample.

[4] https://github.com/aritrachakrabarty/GPS

3.2. *The radius histogram and period-radius distribution: primordial radius valley from dichotomy in bulk composition*

Figure 5 shows that our mixture model of rocky and water-rich populations from Model I efficiently reproduce both the radius histogram (top left panel) and the period–radius distribution (top right panel) of small Kepler planets. The top left panel displays the mean radius histogram and its 1σ uncertainty band (in green), along with the individual contributions from the rocky planets (in red) and water worlds (in blue) to the total distribution. These are compared to the completeness-corrected radius histogram of Kepler planets. The top right panel presents the period–radius distributions of the model rocky and water-rich populations, computed from the MCMC posterior sample and weighted by their Kepler detection completeness and posterior likelihood. The separation between the two populations in the period–radius plane closely matches the estimated slope of the Kepler radius valley (Van Eylen et al. 2019; Gupta & Schlichting 2019). Furthermore, to illustrate the joint distribution, the radius histograms for $P < 30$ days and $P < 60$ days are shown in the bottom panels, demonstrating good agreement with the observed planets in

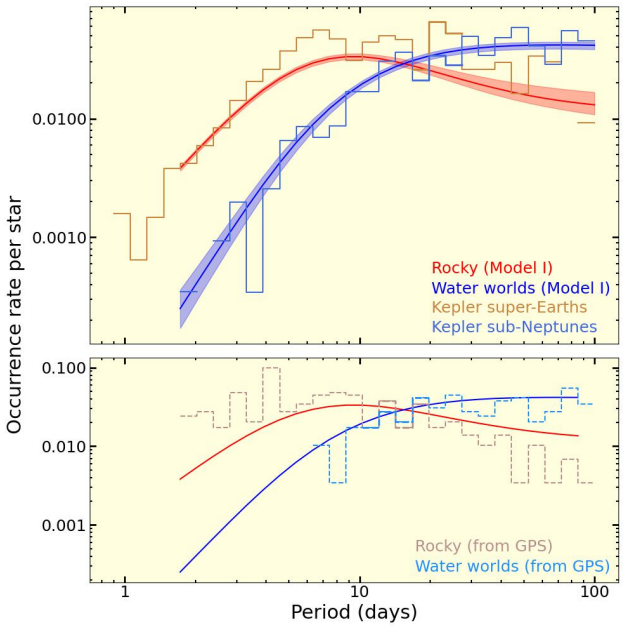

Figure 4. Top: The period distributions of the rocky planets and water worlds from Model I compared to the completeness-corrected period distributions of the Kepler super-Earths and sub-Neptunes, respectively. The model period distributions are resampled from the MCMC posterior using Equations 1-4. The solid lines and the shaded regions (red and blue) denote the mean distributions and their 1σ uncertainty ranges, respectively. Bottom: The mean period distributions from Model I are compared with those from a migration-based formation model from *GPS* (Chakrabarty & Mulders 2024a; Mulders et al. 2020), assuming a snow-line location of 2.7 AU in the protoplanetary disk. Both the present model and the *GPS* migration model suggest a consistent transition period.

the same period ranges. The occurrence rates are peak-normalized, as f_* in these period ranges is not known.

Our results reinforce the notion that an *ab initio* compositional dichotomy can account for the radius valley, as proposed by previous studies (Izidoro et al. 2022; Burn et al. 2024). In particular, we demonstrate that a model based solely on steam worlds is capable of reproducing the entire observed radius histogram and period–radius distribution of Kepler sub-Neptunes, without invoking H/He envelopes and atmospheric loss models, for a reasonable range of model parameters. We notice a drop in the occurrence of such planets at 3-4 $R_\oplus$, since such large water worlds would require extremely high mass or WMF, suggesting the observed *radius cliff* to be a result of this "waterfall". Planets larger than this are likely to possess significant H/He envelopes, indicating a distinct H/He-rich population that extends toward the regime of giant planets.

3.3. *The mass and WMF distributions*

We aim to infer the mass and WMF distributions of the small exoplanets from our model (Model I) that can reproduce the period-radius distribution of the Kepler short-period planets. The top panel of Figure 6 shows the marginalized mass distributions for the rocky planets and water worlds. These distributions peak at approximately 2.6 $M_\oplus$ and 7 $M_\oplus$, respectively, indicating distinct populations. This separation is consistent with trends observed in the mass–radius distribution of exoplanets. We also investigate whether the combined mass distribution exhibits a bimodality—analogous to the radius distribution—that would further support the presence of two distinct populations. However, the bottom left panel of Figure 6 shows that the total mass distribution is nearly unimodal, resembling that predicted by photoevaporation models of rocky cores with H/He envelopes (Wu 2019; Rogers & Owen 2021). This suggests that distinguishing between a planetary population shaped by a primordial compositional dichotomy and one sculpted by atmospheric evolution may remain challenging based on the observed mass and radius distributions.

The distribution of WMF, shown in the bottom right panel of Figure 6, peaks at ∼41%, which is broadly consistent with the canonical 50% value typically assumed for water worlds and with the water–ice fraction observed in the solar system (Lodders 2003). The spread in the WMF distribution may reflect a range of factors, including variations in disk chemistry, differences in host star metallicity, or diverse accretion histories as planetary embryos migrate through the protoplanetary disk. It is also consistent with the model uncertainties in estimating the ice fractions in disks, e.g., the 38% value for water ice among all condensates in the solar system from Lodders (2003), versus the 48% estimated by Anders & Grevesse (1989).

3.4. *Mass-Radius relationship: A more realistic size limit of the steam worlds*

We compare our model (Model I) to the observed mass–radius distribution of small planets around G-type stars, defined by effective temperatures between 5200 K and 6500 K. We select planets from the TEPCat catalog with radii $\lesssim 4$ $R_\oplus$ and mass and radius measurements accurate to better than 25% and 8%, respectively, as shown with error bars in Figure 7. Figure 7 also presents the mass–radius distributions of simulated rocky planets and water worlds, computed from the MCMC posterior and weighted by their posterior likelihood and Kepler detection completeness based on their radii and orbital periods. In the top panel, the model distribution

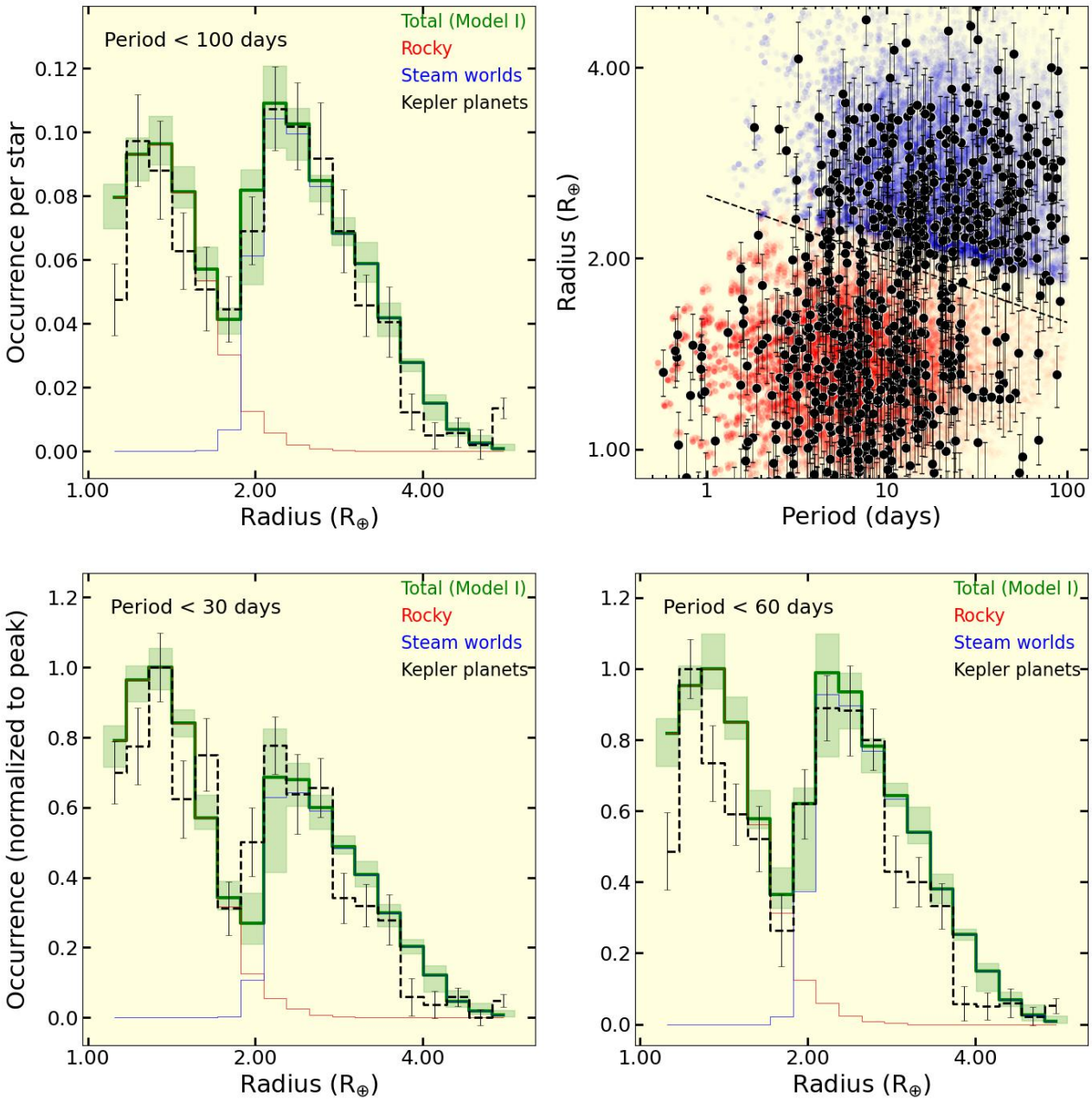

Figure 5. Top left: Occurrence rates (per star) of the simulated rocky population, water world population, and total population from Model I, compared with the completeness-corrected occurrence rates of the Kepler planets from the CKS catalog in the complete analysis range of orbital periods ($P < 100$ days). The green solid line denotes the total occurrence rate corresponding to the median parameters, while the green shaded region indicates the 1-σ uncertainty range. Top right: Period–radius distribution of the planets from Model I compared with that of Kepler planets from the CKS catalog. The model periods and radii are calculated from the MCMC posterior sample of the free parameters, with each point weighted by the Kepler detection completeness and its posterior likelihood. The dashed black line denotes the observed slope of radius valley over the orbital period (Van Eylen et al. 2019). Bottom panels: Same as the top left panel, but restricted to $P < 30$ days and $P < 60$ days, to better illustrate the period dependence of the radius distribution. The occurrence rates are normalized to their peaks, since we did not estimate f_*, i.e., the average occurrence per star in these period ranges.

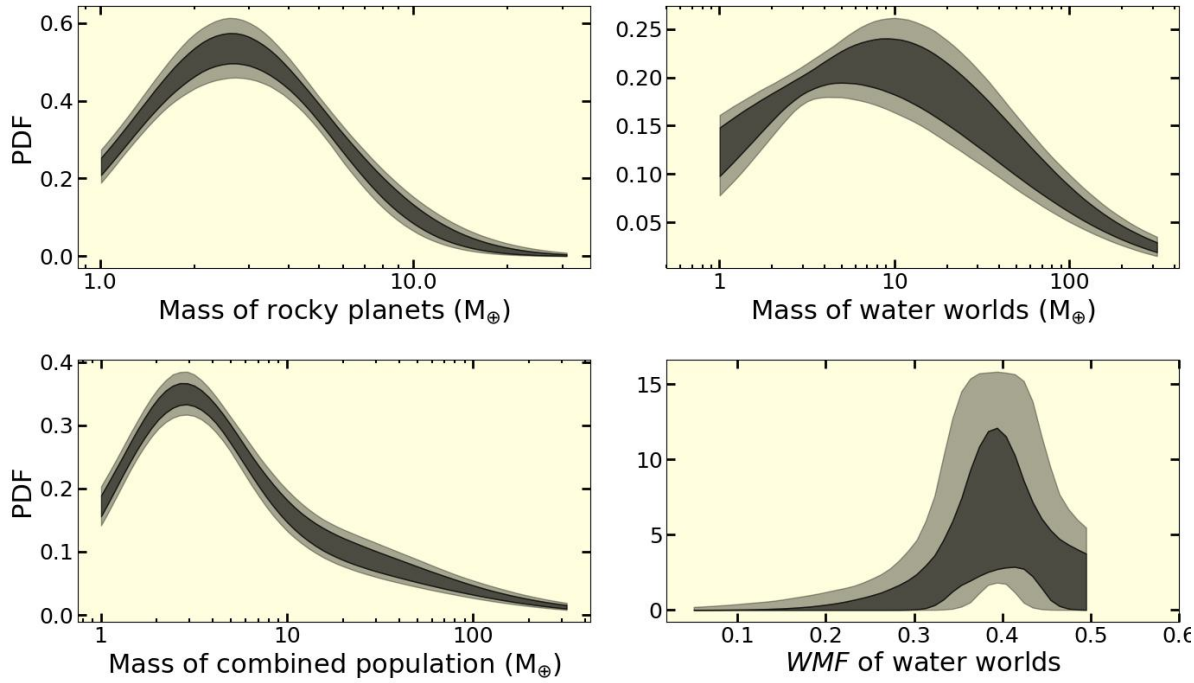

Figure 6. Mass distributions of the rocky planets (top left), the water worlds (top right), and the total ensemble (bottom left), along with the WMF distribution of the water worlds (bottom right) from Model I. The dark and light bands represent the 1σ and 2σ uncertainty ranges, as computed from the MCMC posterior sample.

is shown with a color map of WMF, but the effect of the weights is not visually apparent. Therefore, in the bottom panel we plot the same weighted mass–radius distribution without the color map, making the weight contours explicit.

Although steam worlds can account for the radius distribution up to $\sim 4\ R_\oplus$, the model underpredicts the observed low-density planets beyond $\sim 3\ R_\oplus$, indicating a growing prevalence of gas dwarfs with substantial H/He envelopes at larger radii. We check whether the 1σ error bars of the TEPCat planets fall within the water-world contour drawn using Model I (at 0.001 level, shown in grey dashes in Figure 7), thereby classifying them as bare rocky, bare water-rich, and H/He-rich planets, shown in Figure 7 as brown, cyan, and grey error-bars, respectively. We estimate their overall occurrence fractions as 26.2% rocky, 59% water-rich, and 14.8% H/He-rich, suggesting that at least $\sim 20\%$ of the sub-Neptunes with both mass and radius measurements may be H/He-rich.

Our model predicts a significant population of low-mass sub-Neptunes that are underrepresented in the observed distribution due to the bias of RV surveys against detecting such planets. Conversely, although the outer grey contour of the water-world mass-radius distribution extends beyond $\sim 50 M_\oplus$, suggesting a finite simulated population of massive sub-Neptunes that are not observed, the probability density significantly drop in this regime (see also the bottom-left panel of Figure 6). In fact, such planets could instead be H/He-rich, in which case lower masses would be sufficient to explain their observed radii. Nevertheless, within regions of high density, the weighted model distribution broadly agrees with the observed population for planets with radii $\lesssim 3R_\oplus$, noting that this upper limit is actually mass dependent.

This highlights the importance of benchmarking models against the observed joint mass–radius–period distribution, underscoring the need for future surveys that provide comprehensive mass measurements with well-characterized detection completeness for small planets in the Kepler size range.

4. GUIDANCE TO FUTURE OBSERVATIONS

A detailed demographic study across the full mass–radius–period parameter space can provide valuable constraints on planet formation and evolution models. Since a bias-corrected mass distribution for the Kepler population is not yet available, we use the computed total mass distribution of the simulated ensemble from Model I to help guide future observations aiming to measure the masses of small planets in the Kepler size range. We compare our model's mass distribution with the completeness-corrected minimum mass distribution ($m \sin i$, where i is the orbital inclination) from Howard et al. (2010), who estimated planet occurrence

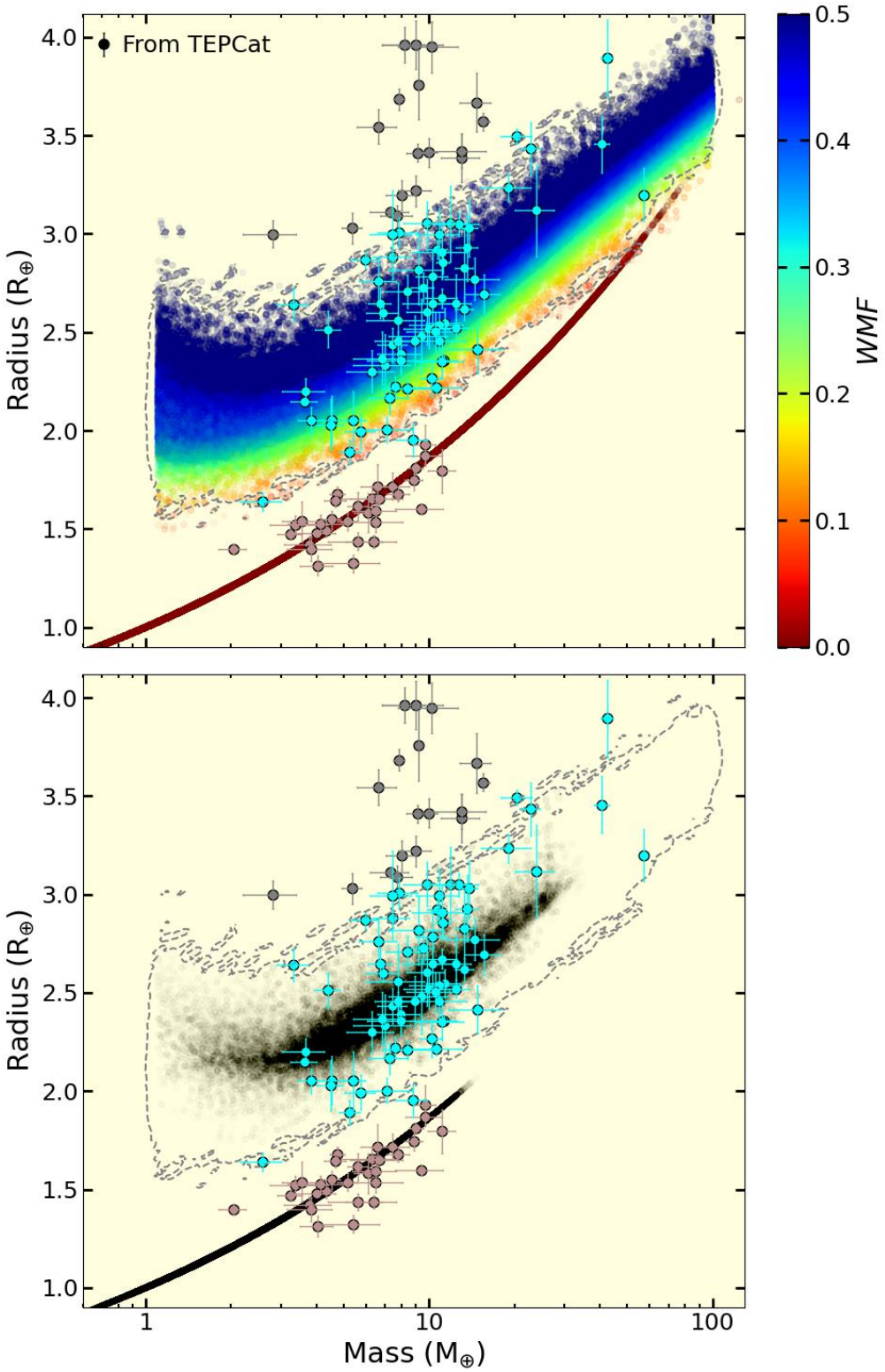

Figure 7. Mass–radius distribution of planets from Model I compared with the observed planets around G-type host stars from the TEPCat catalog Southworth (2011). Top: Weighted mass–radius distribution of simulated rocky planets and water worlds, shown with a color map of water mass fraction (WMF). Bottom: The same weighted distribution shown without the WMF color map, with weight contours highlighted to illustrate regions of higher model likelihood and observational completeness. The model broadly agrees with the observed distribution for planets smaller than $\sim 3\ R_\oplus$ in regions of high weights but fails to explain planets larger than $\sim 3\ R_\oplus$, suggesting the presence of a H/He-rich population, comprising $\sim 15\%$ of the full TEPCat sample and $\sim 20\%$ of the TEPCat sub-Neptunes.

rates for $m \sin i > 3\ M_\oplus$ based on independent radial velocity (RV) observations. Although their mass histogram uses broad bins and reflects minimum rather than true masses, this comparison serves to contrast the mass patterns inferred from a model-based analysis of Kepler planets with those derived from independent RV surveys. We also compare our results with the bias-uncorrected mass distribution from the TEPCat catalog for G-type host stars (described in §3.4) to assess whether the model-predicted underlying mass distribution is reflected in the current sample of planets with precise mass and radius measurements. This planet sample is corrected only for Kepler detection completeness based on planet size and orbital period.

All three mass distributions—Model I, TEPCat, and Howard et al. (2010)—are shown in Figure 8 using the same bins adopted by Howard et al. (2010). While the overall shapes of the mass (or minimum mass) distributions are broadly consistent, the occurrence rates per star inferred from TEPCat and RV observations are significantly lower than those predicted by our model—by nearly 50% in the 3–10 $M_\oplus$ range. This discrepancy suggests that RV surveys may underrepresent the Kepler population of 1–4 $R_\oplus$ planets.

A complementary avenue for identifying the origin of the radius valley is an age-based demographic study to determine whether this feature is primordial or the result of evolutionary processes. Recent studies using K2 and TESS data (Christiansen et al. 2023; Fernandes et al. 2025) have attempted to estimate occurrence rates as a function of stellar age. These results indicate a decline in the overall occurrence rate of small exoplanets with increasing stellar age, suggesting the influence of evolutionary processes such as atmospheric loss. However, these estimates are limited by the small number of data points with large uncertainties in stellar ages, and planetary radii binned over broad intervals, which prevent clear distinctions between super-Earths and sub-Neptunes across age. Similarly, Rogers (2025) conducts a simulation-driven investigation by generating synthetic TESS-biased young sub-Neptune populations under the 'gas dwarf' and 'water world' hypotheses and comparing them to observed TESS planets around stars younger than 40 Myr. His results tentatively disfavor the 'water worlds' hypothesis, though, as he points out, the small TESS sample limits the statistical significance. Therefore, the current observations cannot rule out a primordial origin driven by a bulk compositional dichotomy. Progress will require future surveys of larger, more uniform samples of young and intermediate-age stars, allowing a more precise characterization of age-dependent demographic studies of the super-Earths and sub-Neptunes.

Finally, atmospheric spectroscopy provides an independent means of constraining the nature of sub-Neptunes on an individual-planet basis. Several sub-Neptunes with flat transmission spectra have nonetheless shown strong secondary-eclipse or phase-curve signals, suggesting high-metallicity, possibly steam-rich atmospheres. For instance, GJ 1214 b displays a featureless transmission spectrum, but JWST thermal-emission data reveal a steep slope consistent with a high mean molecular weight atmosphere, supporting a steam- or mixed-composition atmosphere over a purely H/He-rich envelope (Kempton et al. 2023; Gao et al. 2023). Conversely, a combined analysis of JWST and HST spectra of GJ 9827 d indicates absorption features of water and a metal-rich atmosphere, suggesting a steam atmosphere with a hydrosphere that may constitute up to $\sim 40\%$ of the planet's mass (Piaulet-Ghorayeb et al. 2024). Thus, coordinated transmission, emission, and phase-curve observations of sub-Neptunes can help disentangle clouds from intrinsically high-metallicity atmospheric compositions and potentially reveal whether these planets host steam- or mixed-composition atmospheres rather than purely H/He-rich envelopes.

Our analysis uses the static internal-structure model of Aguichine et al. (2021), which can potentially overinflate the steam envelopes, especially for low-mass planets, as it does not account for thermal and structural evolution. Updated models that include evolutionary effects are now available (Aguichine et al. 2025), but incorporating them is beyond the scope of the present work and will be explored in future studies.

5. CONCLUSION

We model the Kepler planet population with radii between 0.9–6 $R_\oplus$ and period (P) < 100 days as a mixture of rocky planets and water worlds using a Bayesian hierarchical framework. We find that migration-based formation scenarios are consistent with the observed transition from super-Earths to sub-Neptunes, likely reflecting a shift in dominant composition from rocky to water-rich planets across orbital periods. This offers an alternative explanation to the observed hot Neptune desert. The period-radius distribution of sub-Neptunes supports a water-rich composition, suggesting that the radius cliff is primarily shaped by the *waterfall*—a sharp drop in the occurrence of water/steam-rich planets. At the same time, the observed mass–radius distribution from the TEPCat catalog shows an increasing prevalence of H/He-rich envelopes beyond $\sim 3R_\oplus$. Based on a comparison between our model and planets with precise

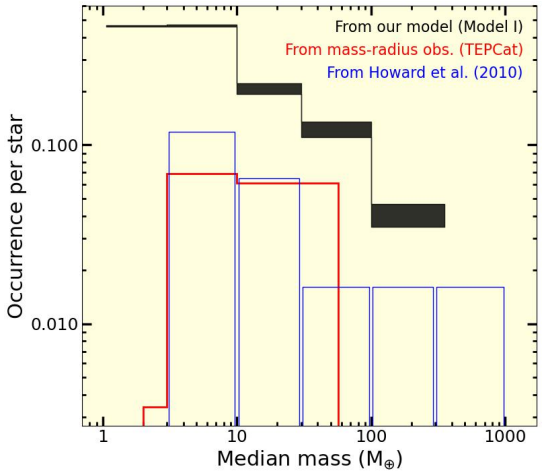

Figure 8. Comparison of the mass distribution from Model I with that from the TEPCat catalog (for G-type host stars) and the minimum mass ($m \sin i$) distribution from Howard et al. (2010), all shown in the same mass bins as used by Howard et al. (2010). The TEPCat distribution is corrected only for Kepler detection completeness based on planetary radius and orbital period. While Model I and the TEPCat distribution are restricted to small planets, the distribution from Howard et al. (2010) is not, since it is based on a pure RV sample with no measured radii. The high-mass tail from our model is found to be consistent with the observed distributions; because the biases preferentially enhance high-mass detections, this feature is unlikely to be spurious. However, the discrepancy in the occurrence rates between ~ 3 and $\sim 30 M_\oplus$ suggests that RV surveys likely underrepresent the Kepler population of small planets.

mass–radius measurements from the TEPCat catalog, we estimate that the small-planet population around G-type stars with orbital period < 100 days comprises approximately 25% rocky, 60% water-rich, and 15% H/He-rich planets. Our results also indicate that current RV surveys may underrepresent the planet population in terms of mass, and highlight the need for future missions with precise mass measurements and well-characterized completeness in the Kepler size range.

ACKNOWLEDGMENTS

We are thankful to the anonymous reviewer for the critical reading of the manuscript and for providing useful suggestions. A.C. thanks Steve Bryson and Matthias He for useful discussions and insightful feedback on this work. This research was supported by an appointment to the NASA Postdoctoral Program at NASA Ames Research Center, administered by Oak Ridge Associated Universities under contract with NASA. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of NASA or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. G.D. M. acknowledges support from FONDECYT project 1252141 and the ANID BASAL project FB210003.

REFERENCES

Aguichine, A., Batalha, N., Fortney, J. J., et al. 2025, ApJ, 988, 186, doi: 10.3847/1538-4357/add935

Aguichine, A., Mousis, O., Deleuil, M., & Marcq, E. 2021, ApJ, 914, 84, doi: 10.3847/1538-4357/abfa99

Anders, E., & Grevesse, N. 1989, GeoCoA, 53, 197, doi: 10.1016/0016-7037(89)90286-X

Bergsten, G. J., Pascucci, I., Mulders, G. D., Fernandes, R. B., & Koskinen, T. T. 2022, AJ, 164, 190, doi: 10.3847/1538-3881/ac8fea

Booth, R. A., Clarke, C. J., Madhusudhan, N., & Ilee, J. D. 2017, MNRAS, 469, 3994, doi: 10.1093/mnras/stx1103

Bryson, S., Coughlin, J., Batalha, N. M., et al. 2020, AJ, 159, 279, doi: 10.3847/1538-3881/ab8a30

Bryson, S., Kunimoto, M., Kopparapu, R. K., et al. 2021, AJ, 161, 36, doi: 10.3847/1538-3881/abc418

Burn, R., Mordasini, C., Mishra, L., et al. 2024, Nature Astronomy, 8, 463, doi: 10.1038/s41550-023-02183-7

Cassan, A., Kubas, D., Beaulieu, J. P., et al. 2012, Nature, 481, 167, doi: 10.1038/nature10684

Chakrabarty, A., & Mulders, G. D. 2024a, ApJ, 966, 185, doi: 10.3847/1538-4357/ad3802

—. 2024b, GPS: Genesis Population Synthesis, v1.0, Data Observatory, doi: 10.57716/1b63-nw69

Chen, J., & Kipping, D. 2017, ApJ, 834, 17, doi: 10.3847/1538-4357/834/1/17

Christiansen, J. L., Zink, J. K., Hardegree-Ullman, K. K., et al. 2023, AJ, 166, 248, doi: 10.3847/1538-3881/acf9f9

Dattilo, A., & Batalha, N. M. 2024, AJ, 167, 288, doi: 10.3847/1538-3881/ad434c

Dattilo, A., Batalha, N. M., & Bryson, S. 2023, AJ, 166, 122, doi: 10.3847/1538-3881/acebc8

Dominguez, G. 2020, MNRAS, 493, 661, doi: 10.1093/mnras/staa168

Fernandes, R. B., Bergsten, G. J., Mulders, G. D., et al. 2025, AJ, 169, 208, doi: 10.3847/1538-3881/adb97e

Fulton, B. J., & Petigura, E. A. 2018, AJ, 156, 264, doi: 10.3847/1538-3881/aae828

Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, AJ, 154, 109, doi: 10.3847/1538-3881/aa80eb
Fulton, B. J., Rosenthal, L. J., Hirsch, L. A., et al. 2021, ApJS, 255, 14, doi: 10.3847/1538-4365/abfcc1
Gao, P., Piette, A. A. A., Steinrueck, M. E., et al. 2023, ApJ, 951, 96, doi: 10.3847/1538-4357/acd16f
Genda, H., & Abe, Y. 2003, Icarus, 164, 149, doi: 10.1016/S0019-1035(03)00101-5
Ginzburg, S., Schlichting, H. E., & Sari, R. 2016, ApJ, 825, 29, doi: 10.3847/0004-637X/825/1/29
—. 2018, MNRAS, 476, 759, doi: 10.1093/mnras/sty290
Gupta, A., & Schlichting, H. E. 2019, MNRAS, 487, 24, doi: 10.1093/mnras/stz1230
Ho, C. S. K., & Van Eylen, V. 2023, MNRAS, 519, 4056, doi: 10.1093/mnras/stac3802
Howard, A. W., Marcy, G. W., Johnson, J. A., et al. 2010, Science, 330, 653, doi: 10.1126/science.1194854
Ikoma, M., & Hori, Y. 2012, ApJ, 753, 66, doi: 10.1088/0004-637X/753/1/66
Izidoro, A., Ogihara, M., Raymond, S. N., et al. 2017, MNRAS, 470, 1750, doi: 10.1093/mnras/stx1232
Izidoro, A., Schlichting, H. E., Isella, A., et al. 2022, ApJL, 939, L19, doi: 10.3847/2041-8213/ac990d
Johnson, J. A., Petigura, E. A., Fulton, B. J., et al. 2017, AJ, 154, 108, doi: 10.3847/1538-3881/aa80e7
Kempton, E. M. R., Zhang, M., Bean, J. L., et al. 2023, Nature, 620, 67, doi: 10.1038/s41586-023-06159-5
Kite, E. S., Fegley, Jr., B., Schaefer, L., & Ford, E. B. 2019, ApJL, 887, L33, doi: 10.3847/2041-8213/ab59d9
Liu, B., Ormel, C. W., & Lin, D. N. C. 2017, A&A, 601, A15, doi: 10.1051/0004-6361/201630017
Lodders, K. 2003, ApJ, 591, 1220, doi: 10.1086/375492
Marboeuf, U., Thiabaud, A., Alibert, Y., Cabral, N., & Benz, W. 2014, A&A, 570, A35, doi: 10.1051/0004-6361/201322207
Marcy, G. W., Isaacson, H., Howard, A. W., et al. 2014, ApJS, 210, 20, doi: 10.1088/0067-0049/210/2/20
Mazeh, T., Holczer, T., & Faigler, S. 2016, A&A, 589, A75, doi: 10.1051/0004-6361/201528065
Mordasini, C., Alibert, Y., & Benz, W. 2009, A&A, 501, 1139, doi: 10.1051/0004-6361/200810301
Mulders, G. D., Ciesla, F. J., Min, M., & Pascucci, I. 2015, ApJ, 807, 9, doi: 10.1088/0004-637X/807/1/9
Mulders, G. D., O'Brien, D. P., Ciesla, F. J., Apai, D., & Pascucci, I. 2020, ApJ, 897, 72, doi: 10.3847/1538-4357/ab9806
Mulders, G. D., Pascucci, I., Apai, D., Frasca, A., & Molenda-Żakowicz, J. 2016, AJ, 152, 187, doi: 10.3847/0004-6256/152/6/187
Ogihara, M., Morbidelli, A., & Guillot, T. 2015, A&A, 578, A36, doi: 10.1051/0004-6361/201525884
Owen, J. E., & Campos Estrada, B. 2020, MNRAS, 491, 5287, doi: 10.1093/mnras/stz3435
Owen, J. E., & Wu, Y. 2016, ApJ, 817, 107, doi: 10.3847/0004-637X/817/2/107
—. 2017, ApJ, 847, 29, doi: 10.3847/1538-4357/aa890a
Pascucci, I., Mulders, G. D., & Lopez, E. 2019, ApJL, 883, L15, doi: 10.3847/2041-8213/ab3dac
Petigura, E. A., Howard, A. W., Marcy, G. W., et al. 2017, AJ, 154, 107, doi: 10.3847/1538-3881/aa80de
Petigura, E. A., Marcy, G. W., Winn, J. N., et al. 2018, AJ, 155, 89, doi: 10.3847/1538-3881/aaa54c
Piaulet-Ghorayeb, C., Benneke, B., Radica, M., et al. 2024, ApJL, 974, L10, doi: 10.3847/2041-8213/ad6f00
Rogers, J. G. 2025, MNRAS, 539, 2230, doi: 10.1093/mnras/staf628
Rogers, J. G., & Owen, J. E. 2021, MNRAS, 503, 1526, doi: 10.1093/mnras/stab529
Southworth, J. 2011, MNRAS, 417, 2166, doi: 10.1111/j.1365-2966.2011.19399.x
Turbet, M., Bolmont, E., Ehrenreich, D., et al. 2020, A&A, 638, A41, doi: 10.1051/0004-6361/201937151
Van Eylen, V., Albrecht, S., Huang, X., et al. 2019, AJ, 157, 61, doi: 10.3847/1538-3881/aaf22f
Vehtari, A., Gelman, A., & Gabry, J. 2017, Statistics and Computing, 27, 1413, doi: 10.1007/s11222-016-9696-4
Venturini, J., Guilera, O. M., Haldemann, J., Ronco, M. P., & Mordasini, C. 2020, A&A, 643, L1, doi: 10.1051/0004-6361/202039141
Watanabe, S. 2010, in Journal of Physics Conference Series, Vol. 233, Journal of Physics Conference Series (IOP), 012014, doi: 10.1088/1742-6596/233/1/012014
Wolfgang, A., Rogers, L. A., & Ford, E. B. 2016, ApJ, 825, 19, doi: 10.3847/0004-637X/825/1/19
Wu, Y. 2019, ApJ, 874, 91, doi: 10.3847/1538-4357/ab06f8
Zeng, L., Jacobsen, S. B., Sasselov, D. D., et al. 2019, Proceedings of the National Academy of Science, 116, 9723, doi: 10.1073/pnas.1812905116

APPENDIX

A. CORNER PLOT FOR MODEL I

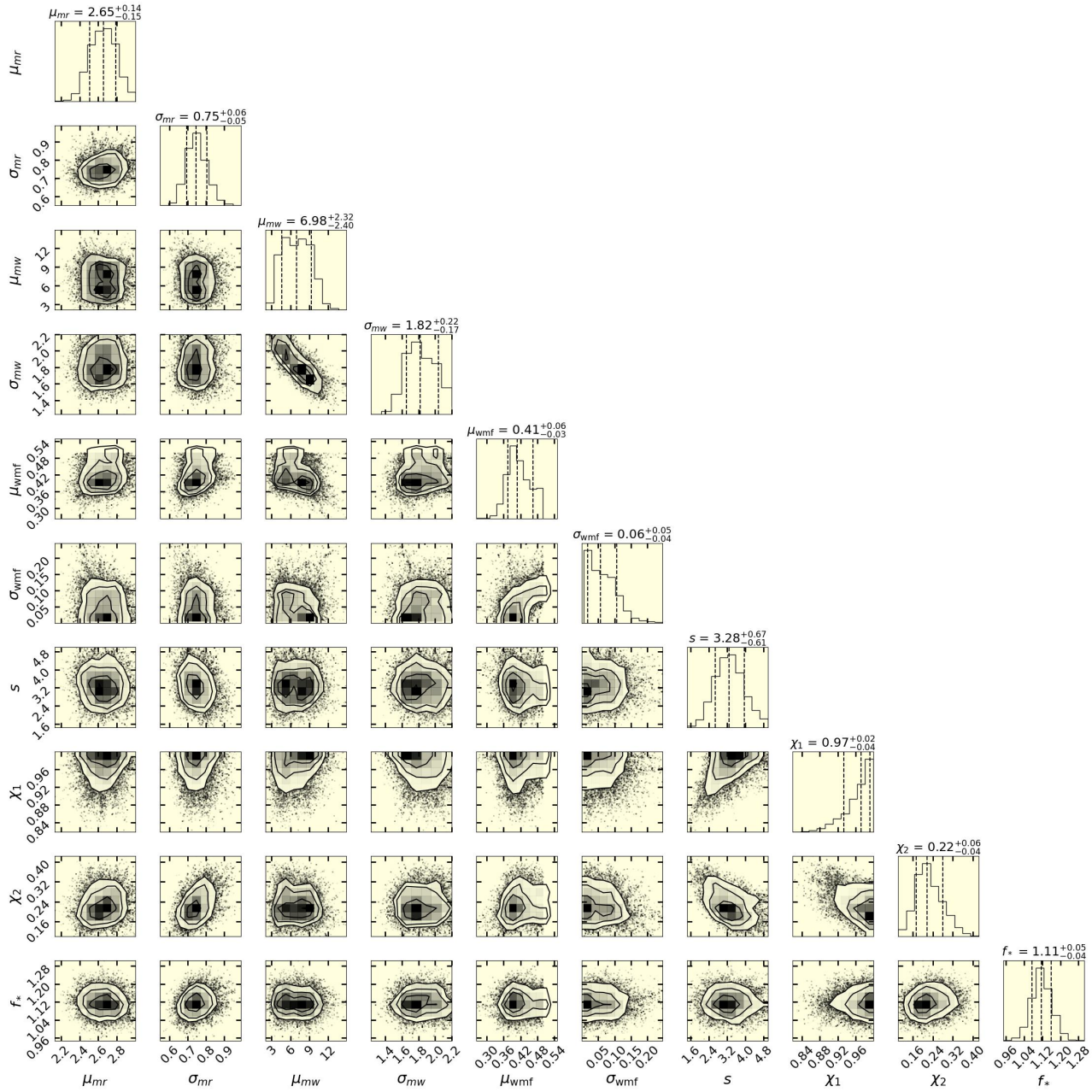

Figure A1. Corner plot showing the posterior distributions of the free parameters of Model I computed from the MCMC posterior sample. The quoted values represent the means and corresponding 1σ uncertainties.