A Bayesian Approach to the Understanding of Exoplanet Populations and the Origin of Life

Jingjing Chen

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018

ABSTRACT

A Bayesian Approach to the Understanding of Exoplanet Populations and the Origin of Life

Jingjing Chen

The study of extrasolar planets, or exoplanets for short, has developed rapidly over the last decade. While we have spent much effort building both ground-based and space telescopes to search for exoplanets, it is even more important that we use the observational data wisely to understand them. Exoplanets are of great interest to both astronomers and the general public because they have shown varieties of characteristics that we couldn't have anticipated from planets within our Solar System. To properly analyze the exoplanet populations, we need the tools of statistics. Therefore, in Chapter 1, I describe the science background as well as the statistical methods which will be applied in this thesis. In Chapter 2, I discuss how to train a hierarchical Bayesian model in detail to fit the relationship between masses and radii of exoplanets and categorize exoplanets based on that. A natural application that comes with the model is to use it for future observations of mass/radius and predict the other measurement. Thus I will show two application cases in Chapter 3. Composition of an exoplanet is also very much constrained by its mass and radius. I will show an easy way to constrain the composition of exoplanets in Chapter 4 and discuss how more complicated methods can be applied in future works.

Of even greater interest is whether there is life elsewhere in the Universe. Although

the future discovery of extraterrestrial life might be totally a fluke, a clear sketched plan always gives us some directions. Research in this area is still very preliminary. Fortunately, besides directly searching for extraterrestrial life, we can also apply statistical reasoning to first estimate the rate of abiogenesis, which will give us some clue on the question of whether there is extraterrestrial life in a probabilistic way. In Chapter 5, I will discuss how different methods can constrain the abiogenesis rate in an informatics perspective.

Finally I will give a brief summary in Chapter 6.

- Probabilistic forecasting of the masses and radii of other worlds, **J. Chen** & D. Kipping, The Astrophysical Journal 834 (1), 17

- Forecasted masses for 7000 *Kepler* Objects of Interest, **J. Chen** & D. Kipping, Monthly Notices of the Royal Astronomical Society 460 (3), 3335-3344

- Forecasting the detectability of known radial velocity planets with the upcoming CHEOPS mission, J. S. Yi, **J. Chen**, D. Kipping, Monthly Notices of the Royal Astronomical Society 475(3), 3090-3097

- A HARDCORE model for constraining an exoplanet's core size, G. Suissa, **J. Chen**, D. Kipping, Monthly Notices of the Royal Astronomical Society 476 (2), 2613-2620

- On the rate of abiogenesis from a Bayesian informatics perspective, **J. Chen**, D. Kipping, arXiv preprint arXiv:1806.08033

Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

Five years ago, in the late summer of 2013, I arrived in New York and started my journey of PhD studies.

I need to first thank the Astronomy department of Columbia University, especially the admission committee board, for giving me the chance of becoming a graduate student here. What I have experienced in my graduate study is very different from what I used to. It opened up my eyes.

I am most grateful to my advisor, Prof. David Kipping, who has so many brilliant ideas, encourages debate, listens to my point of view, and always puts my success in the center our relation. I couldn't have finished my PhD studies so fruitfully without his help.

I offer my special thanks to my family, my mother, father, my boyfriend, my cats, and my dog. You always treasure me, adore me, and put everything aside when I need you. You are the most important ones in my life.

I also thank my first and second year research advisors – Prof. David Schiminovich and Prof. Greg Bryan, my thesis committee and my thesis panel, every other professor who had helped me or taught me, the department secretaries Millie and Ayoune, and my folk graduate students. I have learned from and been supported by everyone of you and I truly appreciated it.

2018, New York City

Chapter 1

Introduction

1.1 Population Properties of Exoplanets and Prediction of Missing Features

The study of exoplanets, planets outside the Solar System, is currently one of the most active research areas in astronomy. In 1992, the very first exoplanetary system was found around a millisecond pulsar, PSR1257+12, using precise timing measurements (Wolszczan & Frail 1992). Over the past decade, astronomers have developed a diverse array of detection methods such as transits, radial velocity, micro-lensing etc. By May 2018, more than 3,700 confirmed exoplanets are cataloged (Exo 2018), and the number is still increasing. The great advance in exoplanet discoveries is largely due to survey projects such as ground-based HARPS (High Accuracy Radial Velocity Planet Searcher), HAT (HATNet and HATSouth Projects), OGLE (Optical Gravitational Lensing Experiment), space mission COROT (Convection, Rotation and planetary Transits), *Kepler*, K2, and there

are more to come in the near future such as TESS (Transiting Exoplanet Survey Satellite), CHEOPS (Characterizing ExOPlanets Satellite), and JWST (James Webb Space Telescope). Thanks to these efforts in searching for and gathering information of exoplanets, we are now able to perform statistical analyses on exoplanet populations and gain insights of their population properties.

In recent years, along with the rise of data science, statistical methods and models have become more and more widely used in astronomy. Besides some basic statistical techniques like least-square regressions, astronomers have delved deeper into realms such as Bayesian models, survival analysis, and machine learning. A Bayesian approach has become widely adopted in exoplanetary research as it easily incorporates updated information from previously conducted measurements, and gives posterior distribution estimates for the model parameters as opposed to point estimates by frequentist methods. Hierarchical Bayesian models (HBM), by adding more complexities to the basic Bayesian models, have been applied to many problems, inferring, for example, the eccentricty distribution of exoplanets (Hogg et al. 2010), the obliquities of *Kepler* stars (Morton & Winn 2014), the abundance of Earth analogs (Foreman-Mackey et al. 2014), the albedos of super-Earths (Demory 2014), the mass and radius relationship (Wolfgang et al. 2016) etc.

In this thesis, I will focus on using the Bayesian approach to build models to understand properties of exoplanets as an ensemble and use the models to make predictions on unobserved features. By using the Bayesian methods, Chapter 2 describes how I train a probabilistic model to fit a hierarchical relationship between mass and radius features of exoplanets. This model not only tells us the underlying population properties of exoplan-

ets, but also can be used to predict an unknown feature provided measurements of the related features. The prediction can then be used to estimate the range of the unknown feature and help with follow-up observations.

1.2   Extraterrestrial Life

One of the reasons behind the fever of searching for exoplanets is to answer the long-standing question on our mind – are we alone?

Actual detections of extraterrestrial life might be the only way that can answer this question with certainty. Many life explorers had been launched within the Solar System, especially on Mars, such as NASA's Curiosity rover. And there are more to come in the next decade, such as NASA's Mars 2020 rover and ESA's ExoMars rover. These previous life detection missions have been fruitful in finding life related evidence such as methane (Krasnopolsky et al. 2004; Atreya et al. 2007), organic matter (Grotzinger 2014; Eigenbrode et al. 2018), and water (Orosei et al. 2018). But direct evidence is still lacking.

Outside the Solar System, people have researched on a wide range of topics from theoretically how life originates and sustains (Pepe et al. 2011; Quintana et al. 2014; Khodachenko et al. 2007; Lammer et al. 2007), to observationally finding the signatures of life (Segura et al. 2005). For example, assuming that life is produced photochemically with UV light, Rimmer et al. (2018) calculates where abiogenesis zone lies and where it overlaps with the habitable zone.

In this thesis, I will show how one can use statistical methods to estimate the rate of abiogenesis, or how often life spontaneously originates from the environment, on Earth-

like planets. If the abiogenesis rate turns out to be high, it is more likely that extraterrestrial life exist. However, at the current stage, we have very limited information on the abiogenesis rate. First, abiogenesis events greatly depend on the specific environment and the evolution history of the given planet. Besides, since the Earth is the only known example with life, we have very little data upon which to condition our inference. Thus, the result of previous work showed that any estimate on the abiogenesis rate can be dramatically affected by assumptions (Spiegel & Turner 2012). In other words, additional information is required for meaningful estimates. In this thesis, I speculate three scenarios, inspired by possible results from other areas of research, which might provide some additional information to help better constrain the abiogenesis rate under the Bayesian statistic framework. I then show how these scenarios affect the final inferred distribution of abiogenesis rates.

1.3 The Bayesian Method

In the previous two sections, I briefly introduced the science background of the thesis. In this section, I will mainly talk about the methodology background.

The Bayesian framework has three elements: pre-existing information, which is called prior; new information, which is called data; and result, which is called posterior. For a parameter of interest, Bayesian inference uses the newly observed data to update the prior belief of the parameter's distribution and results in the posterior distribution of the parameter. For example, in weather forecast, tomorrow's probability of rain can be interpreted as a summary of the rain history updated with currently observed cloud

pattern in a Bayesian framework .

In this section, I will briefly introduce the Bayesian methods used throughout the thesis. For a more detailed explanation, the readers should refer to some standard textbooks in Bayesian statistics, such as Bayesian Data Analysis by Andrew Gelman (Gelman 2004).

1.3.1 Bayes' Rule

The idea of Bayesian analysis is to update the uncertainty in parameters of interest from their previous estimates conditioned on newly obtained data. The Bayes' rule can be described mathematically as the following equation,

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)},$$

where θ is the parameter of interest and y is the observation data.

The left hand side represents the posterior distribution of θ, which is the probability distribution of parameter θ conditioned on observation y. On the right hand side, $P(y)$ can be considered as a constant because y is observed and has no randomness. $P(\theta)$ is called prior distribution of parameter θ because it represents our assumption on θ prior to the observation of y. $P(y|\theta)$ means the probability of obtaining data y given θ, which is also known as the likelihood when considered as a function of θ. The specific forms of $P(\theta)$ and $P(y|\theta)$ depend on the mathematical model we want to use and the details of the problem.

1.3.2 Metropolis Algorithm

Most of the time in real-world applications, posteriors cannot be easily calculated with an analytical form. In those cases, we need simulation methods to derive an empirical distribution of the posterior. The Metropolis algorithm via random walk is one of the most widely used simulation methods. The algorithm generates a chain of samples step by step, which converges to the true distribution (Roberts et al. 1997). The algorithm operates as follows (Gelman 2004) to find the posterior distribution of θ.

1. Initialization: sample a random starting point θ^0 from the prior distribution of θ.

2. Jumping: for each step t ($n = 1, 2, ...$), sample a proposed θ^* from a symmetric jumping distribution $J(\theta|\theta^{t-1})$. By symmetric, it means that the probability of θ jumping from θ^* to θ^{t-1} is equal to the probability of jumping from θ^{t-1} to θ^*.

3. Acceptance/Rejection: accept the proposed θ^* with probability $min(r, 1)$, where

$$r = \frac{P(\theta^*|y)}{P(\theta^{t-1}|y)}$$

The algorithm will end up with a chain of θ. The chain has two different phases delineated by a "burn-in" point. Before burn-in, the chain is going in one direction in general to search for the area where the final samples will lie in. After burn-in, the chain becomes steady and converges to the posterior distribution. In practice, the initial samples of θ before burn-in will be thrown away.

1.3.3   Hierarchical Bayesian Model

Hierarchical models add another layer of parameters called hyper-parameters, that control the distribution of the original parameters, which are called local-parameters in a hierarchical model. Hierarchical Bayesian Models (HBM in short) are usually used in the analysis of a population of objects. In application, hyper-parameters are usually used to describe population properties while local-parameters are used to describe each single object. The goal of using hierarchical models is to fit both hyper and local parameters in a single framework. A general model can be described as the following.

$$P(\alpha, \theta | y) = \frac{P(y | \alpha, \theta) P(\alpha, \theta)}{P(y)}$$

where α, θ, y represents hyper-parameters, local-parameters and data respectively.

Because the observation y is only directly related to θ, $P(y | \alpha, \theta)$ can be written as $P(y | \theta)$. $P(\alpha, \theta)$ can be written as $P(\alpha) P(\theta | \alpha)$. So to build a hierarchical model we need to specify three things, the prior distribution of hyper parameters, the distribution of local-parameters conditioned on hypers, and the distribution of data conditioned on locals.

Graphical models are usually used to help explain the hierarchy in practice. Throughout this thesis, as is shown in Figure 1.1, hyper-parameters will be represented in yellow ovals, local-parameters will be in white ovals, and data will be in grey ovals. Within the box, there will be N samples of data and their corresponding local parameters.
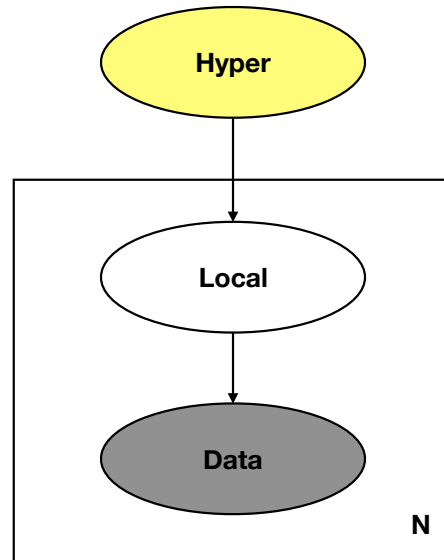
Figure 1.1 Graphical representation of hierarchical Bayesian models.

1.3.4 Information Gain

In Bayesian inference, the posterior distribution is considered to be the updated beliefs of a parameter with data from prior beliefs. Therefore we can ask the question, how much has the belief changed from the previous one? In information theory, the characteristics of a distribution are quantified through the entropy of the distribution, or the average amount of information carried in the distribution. The entropy of a continuous random

variable X with distribution $p(x)$ is defined as

$$H(X) = -\int p(x) log p(x) dx$$

Relative entropy, also known as the Kullback-Leibler divergence (KLD for short), between distributions P and Q is defined as

$$D_{KL}(P\|Q) = \int p(x) log \frac{p(x)}{q(x)} dx$$

As entropy measures the amount of information, relative entropy measures the information gain. Relative entropy can be applied to compare the posterior distribution with the prior distribution and quantify the impact of observation which plays in between.

Chapter 2

Forecasting the Masses and Radii of Other Worlds

Mass and radius are two of the most fundamental properties of an astronomical object. Increasingly, new planet discoveries are being announced with a measurement of one of these terms, but not both. This has led to a growing need to forecast the missing quantity using the other, especially when predicting the detectability of certain follow-up observations. We present an unbiased forecasting model built upon a probabilistic mass-radius relation conditioned on a sample of 316 well-constrained objects. Our publicly available code, `Forecaster`, accounts for **observational errors, hyper-parameter uncertainties** and the intrinsic dispersions observed in the calibration sample. By conditioning our model upon a sample spanning dwarf planets to late-type stars, `Forecaster` can predict the

---

[0]This chapter is a reproduction of a paper that has been published by The Astrophysical Journal. It can be found at `http://iopscience.iop.org/article/10.3847/1538-4357/834/1/17/meta`. The article has been reformatted for this section.

mass (or radius) from the radius (or mass) for objects covering nine orders-of-magnitude in mass. Classification is naturally performed by our model, which uses four classes we label as Terran worlds, Neptunian worlds, Jovian worlds and stars. Our classification identifies dwarf planets as merely low-mass Terrans (like the Earth), and brown dwarfs as merely high-mass Jovians (like Jupiter). We detect a transition in the mass-radius relation at $2.0^{+0.7}_{-0.6}\,M_{\oplus}$, which we associate with the divide between solid, Terran worlds and Neptunian worlds. This independent analysis adds further weight to the emerging consensus that rocky Super-Earths represent a narrower region of parameter space than originally thought. Effectively, then, the Earth is the Super-Earth we have been looking for.

2.1 Introduction

Over the last two decades, astronomers have discovered thousands of extrasolar worlds (see exoplanets.org; Han et al. 2014), filling in the parameter space from Moon-sized planets (e.g. Barclay et al. 2013) to brown dwarfs many times more massive than Jupiter (e.g. Deleuil et al. 2008). Over 98% of these detections have come from radial velocity, microlensing or transit surveys, yet each of these methods only directly measures the mass (M) *or* radius (R) of a planet, not both[1].

This leads to the common situation where it is necessary to forecast what the missing quantity is based on the other. A typical case would be when one needs to predict the detectability of a potentially observable effect for a resource-intensive, time-competitive observing facility, which in some way depends upon the missing quantity. For exam-

---

[1] Except for the rare cases of systems displaying invertible transit timing variations.

ple, the *TESS* mission (Ricker et al. 2014) will soon start detecting hundreds, possibly thousands, of nearby transiting planets for which the radius, but not the mass, will be measured. Planets with radii consistent with Super-Earths will be of great interest for follow-ups and so radial velocity facilities will need to forecast the detectability, which is proportional to the planet mass, for each case. Vice versa, the *CHEOPS* mission (Broeg et al. 2013) will try to detect the transits of planets discovered with radial velocities, necessitating a forecast of the radius based upon the mass.

In those two examples, the objective was to forecast the missing quantity in order to predict the feasibility of actually measuring it. However, the value of forecasting the mass/radius for the purposes of predicting detectability extends beyond this. As another example, exoplanet transit spectroscopy is expected to be a major function of the upcoming JWST mission (Seager et al. 2009). At the first-order level, the detectability of an exoplanet atmosphere is proportional to the scale height, H, which in turn is proportional to $1/g \propto R^2/M$. Given the limited supply of cryogen onboard *JWST*, discoveries of future Earth-analog candidates may be found with insufficient time to reasonably schedule a radial velocity campaign first (if even detectable at all). Therefore, there will likely be a critical need to accurately forecast the scale height of new planet discoveries from just either the mass or (more likely) the radius.

Forecasting the mass/radius of an object, based upon the other quantity is most obviously performed using a mass-radius (MR) relation. Such relations are known to display sharp changes at specific locations, such as the transition from brown dwarfs to hydrogen burning stars (e.g. see Hatzes & Rauer 2015). These transition points can

be thought of as bounding a set of classes of astronomical objects, where the classes are categorized using the features of the inferred MR relation. In this case then, it is apparent that inference of the MR relation enables both classification and forecasting.

Classification is more than a taxonomical enterprise, it can have dramatic implications in astronomy. Perhaps the most famous example of classification in astronomy is the Hertzsprung-Russell (HR) diagram (Hertzsprung 1909; Russell 1914) for luminosity versus effective temperature, which revealed the distinct regimes of stellar evolutions. A common concern in classification is that the very large number of possible features against which to frame the problem can be overwhelming. Mass and radius, though, are not random or arbitrary choices for framing such a problem. Rather, they are two of the most fundamental quantities describing any object in the cosmos and indeed represent two of the seven base quantities in the International System of Units (SI).

The value of classification extends beyond guiding physical understanding, it even affects the design of future instrumentation. As an example, the boundary between terrestrial planets and Neptune-like planets represents a truncation of the largest allowed habitable Earth-like body. The location of this boundary strongly affects estimates of the occurrence rate of Earth-like planets ($\eta_\oplus$) and thus in-turn the design requirements of future missions needed to characterize such planets (Dalcanton et al. 2015). To illustrate this, using the occurrence rate posteriors of Foreman-Mackey et al. (2014), $\eta_\oplus$ decreases by 42% when altering the definition of Earth-analogs from $R < 2.0\,R_\oplus$ to $R < 1.5\,R_\oplus$. In order to maintain the same exoEarth yield for the proposed *HDST* mission, this change corresponds to a 27% increase in the required mirror diameter (using yield equation in

§3.5.4 of Dalcanton et al. 2015).

We therefore argue that both forecasting and classification using the masses and radii of astronomical bodies will, at the very least, be of great utility for present/future missions and may also provide meaningful insights to guide our interpretation of these objects. Accordingly, the primary objective of this work is to build a statistically rigorous and empirically calibrated model

▶ to forecast the mass/radius of an astronomical object based upon a measurement of the other, and

▶ for the classification of astronomical bodies based upon their observed masses and/or radii.

The layout of this paper is as follows. In Section 2.2, we outline our model for the MR relation, which is used for forecasting and classification. In Section 2.3, we describe the regression algorithm used to conduct Bayesian parameter estimations of our model parameters. The results, in terms of both classification and forecasting are discussed separately in Sections 2.4 & 2.5. We summarize the main findings of our work in Section 2.6.

2.2 Model

2.2.1 Choosing a Model

We begin by describing the rationale behind the model used in this work. As discussed in Section 2.1 (and demonstrated later in Section 2.3), the two primary goals of this paper are both achievable through the use of a MR relation and this defines the approach in this work. Broadly speaking, such a relation can be cast as either a parametric (e.g. a polynomial) or non-parametric model (e.g. a nearest neighbor algorithm).

Parametric models, in particular power-laws, have long been popular for modeling the MR relation with many examples even in the recent literature (e.g. Valencia et al. 2006; Weiss et al. 2013; Hatzes & Rauer 2015; Wolfgang et al. 2016; Zeng et al. 2016). In our case, we note that such models are more straightforward for hierarchical Bayesian modeling (which we argue to be necessary later), since they allow for a simple prescription of the Bayesian framework. Moreover, based on those earlier cited works, power-laws ostensibly do an excellent job of describing the data and the greater flexibility afforded by non-parametric methods is not necessary. Accordingly, we adopt the power-law prescription in this work.

As noted earlier, the use of power-laws to describe the MR relation is common in the literature. However, many of the assumptions and model details in these previous implementations would make forecasts based upon these relations problematic. We identify three key aspects of the model proposed in this work which differentiate our work from previous studies.

[1] Largest data range: Inferences of the MR relation often censor the available data to a specific subset of parameter space (for example Wolfgang et al. 2016 consider the $R < 8\,R_\oplus$ exoplanets). Whilst it is inevitable that certain subjective choices will be made by those analyzing the MR relation, a more physically-motivated choice for the parameter limits can be established. Ideally, this range should be as large as possible such that forecasting is unlikely to encounter the extrema, leading to truncation errors. A natural lower bound is an object with sufficient mass to achieve hydrostatic equilibrium leading to a nearly spherical shape and thus a well-defined radius (a planemo), which would encompass dwarf planets. As an upper bound, late-type stars take longer than a Hubble time to leave the main-sequence and should exhibit a relatively tight trend between mass and radius.

[2] Fitted transitions: As a by-product of using such a wide mass range, several transitional regions are traversed where the MR relation exhibits sharp changes. For example, the onset of hydrogen burning leads to a dramatic change in the MR relation versus brown dwarfs (Hatzes & Rauer 2015). In previous works, such transitional points are often held as fixed, assumed locations (e.g. Weiss & Marcy 2014 assume a physically motivated, but not freely inferred, break at $1.5\,R_\oplus$). In contrast, we here seek to make a more agnostic, data-driven inference without imposing any assumed transition points from theory or previous data-driven inferences. In this way, the uncertainty in these transitions is propagated into the inference of all other parameters defining our model, leading to more robust uncertainty estimates for both forecasting and classification. Accordingly, in this work, the MR relation is described by a broken power-law with freely fitted transition

points (in addition to the other parameters).

[3] Probabilistic modeling: Whilst mass can be considered to be the primary influence on the size of an object, many second-order terms will also play a role. As an example, rocky planets of the same mass but different core mass fractions will exhibit distinct radii (Zeng et al. 2016). When viewed in the MR plane then, a particular choice of mass will not correspond to a single radius value. Rather, a distribution of radii is expected, as a consequence of the numerous hidden second-order effects influencing the size. Statistically speaking then, the MR relation is expected to be *probabilistic*, rather than deterministic. A probabilistic model fundamentally relaxes the assumption that the underlying model (in our case a broken power-law) is the "correct" or "true" description of the data, allowing an approximate model to absorb some (although it can never be all) of the error caused by model misspecifications (in our case via an intrinsic dispersion). Naturally, the closer one's underlying model is to the truth, the smaller this probabilistic dispersion need to be, and in the ultimate limit of a perfect model the probabilistic model tends towards a deterministic one. Since we do not make the claim that a broken power-law is the true description of the MR relation, the probabilistic model is essential for reliable forecasting, as it enables predictions in spite of the fact our model is understood to not represent the truth.

Whilst each of these three key features has been applied to MR relations in some form independently, a novel quality of our methodology is to adopt all three. For example, Wolfgang et al. (2016) inferred a probabilistic power-law conditioned on the masses and radii of 90 exoplanets with radii below $8\,R_\oplus$. This range crosses the expected divide

17

between solid planets and those with significant gaseous envelopes at 1.5-2.0 $R_\oplus$ (Lopez & Fortney 2014) and so the authors tried truncating the data at 1.6 $R_\oplus$ as an alternative model. In this work, we argue that the transitional points can actually be treated as free parameters in the model, enabling us to infer (rather than assume) their locations and test theoretical predictions. Additionally, the data need not be censored at $< 4\,R_\oplus$ and the wider range makes a forecasting model less susceptible to truncation issues at the extrema (we point out that Wolfgang et al. (2016) did not set out to develop a forecasting model explicitly, and thus this is not a criticism of their work, but rather just an example of how our work differs from previous studies).

2.2.2 Data Selection

Having broadly established the motivation (see Section 2.1) and the requirements (see Section 2.2.1) for our model, we will use the rest of Section 2.2 to provide a more detailed account of our methodology. To begin, we first define our basic criteria for a data point (a mass and radius measurement) to be included in what follows. Since our work focuses on the MR relation, all included objects must fundamentally have a well-defined mass and radius. Whilst the former is universally true, the latter requires that the object have a nearly spherical shape. Low mass objects, for example the comet 67P/Churyumov-Gerasimenko, may not have sufficient self-gravity to overcome rigid body forces and be assumed as in hydrostatic equilibrium shape (i.e. nearly spherical). The corresponding threshold mass limit should lie somewhere between the most massive body which is known to not be in hydrostatic equilibrium (Iapetus; 1.8×10^{21} kg; Sheppard 2016) and the least massive body

confirmed to be in hydrostatic equilibrium (Rhea; 2.3×10^{21} kg; Sheppard 2016). This leads us to adopt a boundary condition of $M > 2 \times 10^{21}$ kg for all objects considered in this work.

As for the upper limit, we choose the maximum mass to be that of a star that must still lie on the main-sequence within a Hubble time. The lifetime of a star is dependent upon its mass and luminosity, to first-order. Given that the Sun will spend 10 Gyr on the main-sequence and $L \propto M^{7/2}$, the lifetime $\tau \simeq (M/M_\odot)^{-5/2} 10$ Gyr. This results in an upper limit of $M < 0.87\, M_\odot$ (1.7×10^{30} kg) for $\tau = H_0^{-1}$ Gyr (where we set $H_0 = 69.7$ km/s; Planck Collaboration et al. 2014). Therefore, between our lower and upper limits, there is a difference of nine orders-of-magnitude in mass and three orders-of-magnitude in radius.

We performed a literature search for all objects within this range with a mass and radius measurement available. For Solar System moons, we used The Giant Planet Satellite and Moon Page (Sheppard 2016) which is curated by Scott Sheppard (Sheppard & Jewitt 2003; Sheppard et al. 2005, 2006) and for the planets we used the NASA Planetary Fact Sheet (Williams 2016). For extrasolar planets, we used the TEPCat catalog of "well-studied transiting planets", curated by John Southworth (Southworth 2008, 2009, 2010, 2011, 2012). Brown dwarfs and low-mass stars were drawn from a variety of sources, which we list (along with all other objects used in this work) in Table 2.1.

In order to later fit these data sources to an MR model, it is necessary to define a likelihood function of each datum. We later (see §2.2.9) make the assumption that for a quoted mass (or radius) measurement of $M = (a \pm b)$, one can reasonably approximate $M \sim \mathcal{N}(a, b)$. This assumption is a poor one for low signal-to-noise data, especially for upper limit constraints only, where M (or R) is more likely to follow an asymmetric profile

centered near zero. Without knowledge of the correct likelihood function, we argue that such data should be best excluded in what follows.

For this reason, we apply a 3σ cut to both mass $((M/\Delta M) > 3)$ and radius $((R/\Delta R) > 3)$. In what follows, we assume that both the mass and radius measurements follow normal distributions, which are symmetric. For those data which have substantially asymmetric errors $(\Delta_+ \neq \Delta_-)$ then, we only use cases where the errors differ by less than 10% (i.e. $(|\Delta_+ - \Delta_-|)/(\frac{1}{2}(\Delta_+ + \Delta_-)) \leq 0.1$). Together, these cuts remove 16% of the initial data, which, as discussed later in §2.3.3, do not bias (or even noticeably influence) our final results.. Next, we take the average of both errors $\frac{1}{2}(\Delta_+ + \Delta_-)$ as the standard deviations of the normal distributions. In the end, we have 316 of objects in total which are listed in Table 2.1.

The data spans a diverse range of environments, with a variety of orbital periods, insolations, metallicities, etc. Since these terms are not used in our analysis, the results presented here should be thought of as a MR relation marginalized over all of these other terms. Once again, we stress that the effects of these terms are naturally absorbed by the probabilistic framework of our model, meaning that forecasts may be made about any new data, provided that it can be considered a representative of the data used for our analysis.

Table 2.1: Masses and radii used for this study.

Name	Mass	Radius	Reference
NGC6791 KR V20	$(0.827 \pm 0.004)\, M_\odot$	$(0.768 \pm 0.006)\, R_\odot$	Torres et al. (2010)
HD 124784	$(0.854 \pm 0.003)\, M_\odot$	$(0.830 \pm 0.004)\, R_\odot$	Torres et al. (2010)
Parenago 1478	$(0.727 \pm 0.010)\, M_\odot$	$(1.063 \pm 0.011)\, R_\odot$	Torres et al. (2010)
HD 7700	$(0.764 \pm 0.004)\, M_\odot$	$(0.835 \pm 0.018)\, R_\odot$	Torres et al. (2010)
BD+34 4217	$(0.814 \pm 0.013)\, M_\odot$	$(0.838 \pm 0.011)\, R_\odot$	Torres et al. (2010)
TYC 3629-740-1	$(0.869 \pm 0.004)\, M_\odot$	$(0.964 \pm 0.004)\, R_\odot$	Torres et al. (2010)
GU Boo A	$(0.610 \pm 0.006)\, M_\odot$	$(0.627 \pm 0.016)\, R_\odot$	Torres et al. (2010)
GU Boo B	$(0.600 \pm 0.006)\, M_\odot$	$(0.624 \pm 0.016)\, R_\odot$	Torres et al. (2010)
YY Gem A	$(0.599 \pm 0.005)\, M_\odot$	$(0.619 \pm 0.006)\, R_\odot$	Torres et al. (2010)
YY Gem B	$(0.599 \pm 0.005)\, M_\odot$	$(0.619 \pm 0.006)\, R_\odot$	Torres et al. (2010)
CU Cnc A	$(0.435 \pm 0.001)\, M_\odot$	$(0.432 \pm 0.006)\, R_\odot$	Torres et al. (2010)
CU Cnc B	$(0.399 \pm 0.001)\, M_\odot$	$(0.392 \pm 0.009)\, R_\odot$	Torres et al. (2010)
CM Dra A	$(0.231 \pm 0.001)\, M_\odot$	$(0.253 \pm 0.002)\, R_\odot$	Torres et al. (2010)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
CM Dra B	$(0.214 \pm 0.001)\, M_\odot$	$(0.240 \pm 0.002)\, R_\odot$	Torres et al. (2010)
MOTESS-GNAT 78457 A	$(0.527 \pm 0.002)\, M_\odot$	$(0.505 \pm 0.011)\, R_\odot$	Kraus et al. (2011)
MOTESS-GNAT 78457 B	$(0.491 \pm 0.001)\, M_\odot$	$(0.471 \pm 0.011)\, R_\odot$	Kraus et al. (2011)
MOTESS-GNAT 116309 A	$(0.567 \pm 0.002)\, M_\odot$	$(0.552 \pm 0.014)\, R_\odot$	Kraus et al. (2011)
MOTESS-GNAT 116309 B	$(0.532 \pm 0.002)\, M_\odot$	$(0.532 \pm 0.009)\, R_\odot$	Kraus et al. (2011)
MOTESS-GNAT 506664 A	$(0.584 \pm 0.002)\, M_\odot$	$(0.560 \pm 0.004)\, R_\odot$	Kraus et al. (2011)
MOTESS-GNAT 506664 B	$(0.544 \pm 0.002)\, M_\odot$	$(0.513 \pm 0.008)\, R_\odot$	Kraus et al. (2011)
MOTESS-GNAT 646680 A	$(0.499 \pm 0.002)\, M_\odot$	$(0.457 \pm 0.007)\, R_\odot$	Kraus et al. (2011)
MOTESS-GNAT 646680 B	$(0.443 \pm 0.002)\, M_\odot$	$(0.427 \pm 0.006)\, R_\odot$	Kraus et al. (2011)
MOTESS-GNAT 1819499 A	$(0.557 \pm 0.001)\, M_\odot$	$(0.569 \pm 0.023)\, R_\odot$	Kraus et al. (2011)
MOTESS-GNAT 1819499 B	$(0.535 \pm 0.001)\, M_\odot$	$(0.500 \pm 0.014)\, R_\odot$	Kraus et al. (2011)
MOTESS-GNAT 2056316 A	$(0.469 \pm 0.002)\, M_\odot$	$(0.441 \pm 0.003)\, R_\odot$	Kraus et al. (2011)
MOTESS-GNAT 2056316 B	$(0.382 \pm 0.001)\, M_\odot$	$(0.374 \pm 0.003)\, R_\odot$	Kraus et al. (2011)
NSVS 11868841 A	$(0.870 \pm 0.074)\, M_\odot$	$(0.983 \pm 0.030)\, R_\odot$	Çakirli et al. (2010)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
NSVS 11868841 B	$(0.607 \pm 0.053)\ M_\odot$	$(0.901 \pm 0.026)\ R_\odot$	Çakirli et al. (2010)
KOI-686b	$(103.4 \pm 4.8)\ M_J$	$(1.216 \pm 0.037)\ R_J$	Díaz et al. (2014)
KOI-189b	$(78.0 \pm 3.4)\ M_J$	$(0.998 \pm 0.023)\ R_J$	Díaz et al. (2014)
OGLE-TR-123 B	$(0.085 \pm 0.011)\ M_\odot$	$(0.133 \pm 0.009), R_\odot$	Pont et al. (2006)
GJ 570 A	$(0.802 \pm 0.040)\ M_\odot$	$(0.739 \pm 0.019)\ R_\odot$	Demory et al. (2009)
GJ 845	$(0.762 \pm 0.038)\ M_\odot$	$(0.732 \pm 0.006)\ R_\odot$	Demory et al. (2009)
GJ 879	$(0.725 \pm 0.036)\ M_\odot$	$(0.629 \pm 0.051)\ R_\odot$	Demory et al. (2009)
GJ 887	$(0.503 \pm 0.025)\ M_\odot$	$(0.459 \pm 0.011)\ R_\odot$	Demory et al. (2009)
GJ 551	$(0.118 \pm 0.012)\ M_\odot$	$(0.141 \pm 0.007)\ R_\odot$	Boyajian et al. (2012)
SDSS0857+03 B	$(0.090 \pm 0.010)\ M_\odot$	$(0.110 \pm 0.004)\ R_\odot$	Parsons et al. (2012b)
NN Ser B	$(0.111 \pm 0.004)\ M_\odot$	$(0.149 \pm 0.002)\ R_\odot$	Parsons et al. (2010)
GK Vir B	$(0.116 \pm 0.003)\ M_\odot$	$(0.155 \pm 0.003)\ R_\odot$	Parsons et al. (2012a)
OGLE-TR-106 B	$(0.116 \pm 0.021)\ M_\odot$	$(0.181 \pm 0.013)\ R_\odot$	Pont et al. (2005)
HAT-TR-205-013 B	$(0.124 \pm 0.010)\ M_\odot$	$(0.167 \pm 0.006)\ R_\odot$	Beatty et al. (2007)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
SDSS 0138-00 B	$(0.132 \pm 0.003)\,M_\odot$	$(0.165 \pm 0.001)\,R_\odot$	Parsons et al. (2012a)
GJ 699	$(0.146 \pm 0.015)\,M_\odot$	$(0.187 \pm 0.001)\,R_\odot$	Boyajian et al. (2012)
SDSS 1210+33 B	$(0.158 \pm 0.006)\,M_\odot$	$(0.200 \pm 0.004)\,R_\odot$	Pyrzas et al. (2012)
SDSS 1548+40 B	$(0.173 \pm 0.027)\,M_\odot$	$(0.181 \pm 0.015)\,R_\odot$	Pyrzas et al. (2009)
RR Cae B	$(0.183 \pm 0.013)\,M_\odot$	$(0.209 \pm 0.014)\,R_\odot$	Maxted et al. (2007)
2MASS 0446+19 B	$(0.190 \pm 0.020)\,M_\odot$	$(0.210 \pm 0.010)\,R_\odot$	Hebb et al. (2006)
HATS551-019 B	$(0.17 \pm 0.01)\,M_\odot$	$(0.18 \pm 0.01)\,R_\odot$	Zhou et al. (2014a)
Moon	$0.0123\,M_\oplus$	$0.272\,R_\oplus$	Sheppard (2016)
Io	$0.0150\,M_\oplus$	$0.285\,R_\oplus$	Sheppard (2016)
Europa	$0.00804\,M_\oplus$	$0.245\,R_\oplus$	Sheppard (2016)
Ganymede	$0.0248\,M_\oplus$	$0.413\,R_\oplus$	Sheppard (2016)
Callisto	$0.0180\,M_\oplus$	$0.378\,R_\oplus$	Sheppard (2016)
Rhea	$0.000386\,M_\oplus$	$0.120\,R_\oplus$	Sheppard (2016)
Titan	$0.0225\,M_\oplus$	$0.404\,R_\oplus$	Sheppard (2016)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
Titania	$0.000590\,M_\oplus$	$0.124\,R_\oplus$	Sheppard (2016)
Oberon	$0.000505\,M_\oplus$	$0.119\,R_\oplus$	Sheppard (2016)
Triton	$0.00358\,M_\oplus$	$0.212\,R_\oplus$	Sheppard (2016)
Eris	$0.00278\,M_\oplus$	$0.182\,R_\oplus$	Sheppard (2016)
Mercury	$0.0553\,M_\oplus$	$0.383\,R_\oplus$	Williams (2016)
Venus	$0.815\,M_\oplus$	$0.949\,R_\oplus$	Williams (2016)
Earth	$M_\oplus$	$R_\oplus$	Williams (2016)
Mars	$0.107\,M_\oplus$	$0.532\,R_\oplus$	Williams (2016)
Jupiter	$317.8\,M_\oplus$	$11.21\,R_\oplus$	Williams (2016)
Saturn	$95.2\,M_\oplus$	$9.45\,R_\oplus$	Williams (2016)
Uranus	$14.5\,M_\oplus$	$4.01\,R_\oplus$	Williams (2016)
Neptune	$17.1\,M_\oplus$	$3.88\,R_\oplus$	Williams (2016)
Pluto	$0.00218\,M_\oplus$	$0.186\,R_\oplus$	Williams (2016)
55-Cnc e	$(0.0254 \pm 0.001)\,M_J$	$(0.1713 \pm 0.0071)\,R_J$	Demory et al. (2016)

Continued on next page

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
CoRoT-01 b	$(1.03 \pm 0.1)\,M_J$	$(1.551 \pm 0.064)\,R_J$	Southworth (2011)
CoRoT-02 b	$(3.57 \pm 0.15)\,M_J$	$(1.46 \pm 0.031)\,R_J$	Southworth (2012)
CoRoT-03 b	$(21.96 \pm 0.703)\,M_J$	$(1.037 \pm 0.069)\,R_J$	Southworth (2011)
CoRoT-06 b	$(2.96 \pm 0.34)\,M_J$	$(1.185 \pm 0.041)\,R_J$	Southworth (2011)
CoRoT-07 b	$(0.0181 \pm 0.0027)\,M_J$	$(0.1414 \pm 0.0057)\,R_J$	Barros et al. (2014)
CoRoT-08 b	$(0.216 \pm 0.036)\,M_J$	$(0.712 \pm 0.083)\,R_J$	Southworth (2011)
CoRoT-09 b	$(0.826 \pm 0.083)\,M_J$	$(1.037 \pm 0.082)\,R_J$	Southworth (2011)
CoRoT-10 b	$(2.78 \pm 0.14)\,M_J$	$(0.941 \pm 0.085)\,R_J$	Southworth (2011)
CoRoT-11 b	$(2.34 \pm 0.39)\,M_J$	$(1.426 \pm 0.057)\,R_J$	Southworth (2011)
CoRoT-12 b	$(0.887 \pm 0.078)\,M_J$	$(1.35 \pm 0.075)\,R_J$	Southworth (2011)
CoRoT-13 b	$(1.312 \pm 0.096)\,M_J$	$(1.252 \pm 0.076)\,R_J$	Southworth (2011)
CoRoT-14 b	$(7.67 \pm 0.49)\,M_J$	$(1.018 \pm 0.079)\,R_J$	Southworth (2011)
CoRoT-18 b	$(3.27 \pm 0.17)\,M_J$	$(1.251 \pm 0.083)\,R_J$	Southworth (2012)
CoRoT-20 b	$(5.06 \pm 0.36)\,M_J$	$(1.16 \pm 0.26)\,R_J$	Southworth (2012)

Continued on next page

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
CoRoT-21 b	$(2.26 \pm 0.33)\,M_J$	$(1.27 \pm 0.14)\,R_J$	Pätzold et al. (2012)
CoRoT-26 b	$(0.52 \pm 0.05)\,M_J$	$(1.26 \pm 0.13)\,R_J$	Almenara et al. (2013)
CoRoT-27 b	$(10.39 \pm 0.55)\,M_J$	$(1.007 \pm 0.044)\,R_J$	Parviainen et al. (2014)
CoRoT-28 b	$(0.484 \pm 0.087)\,M_J$	$(0.955 \pm 0.066)\,R_J$	Cabrera et al. (2015)
CoRoT-29 b	$(0.85 \pm 0.2)\,M_J$	$(0.9 \pm 0.16)\,R_J$	Cabrera et al. (2015)
EPIC-203771098 c	$(0.085 \pm 0.022)\,M_J$	$(0.698 \pm 0.064)\,R_J$	Petigura et al. (2015)
EPIC-204129699 b	$(1.774 \pm 0.079)\,M_J$	$(1.06 \pm 0.35)\,R_J$	Grziwa et al. (2015)
EPIC-204221263 b	$(0.038 \pm 0.009)\,M_J$	$(0.138 \pm 0.014)\,R_J$	Sinukoff et al. (2015)
GJ-0436 b	$(0.0799 \pm 0.0066)\,M_J$	$(0.366 \pm 0.014)\,R_J$	Lanotte et al. (2014)
GJ-1214 b	$(0.0197 \pm 0.0027)\,M_J$	$(0.254 \pm 0.018)\,R_J$	Harpsøe et al. (2013)
GJ-3470 b	$(0.0432 \pm 0.0051)\,M_J$	$(0.346 \pm 0.029)\,R_J$	Biddle et al. (2014)
HAT-P-01 b	$(0.525 \pm 0.019)\,M_J$	$(1.319 \pm 0.019)\,R_J$	Nikolov et al. (2014)
HAT-P-02 b	$(8.74 \pm 0.27)\,M_J$	$(1.19 \pm 0.12)\,R_J$	Southworth (2010)
HAT-P-03 b	$(0.584 \pm 0.027)\,M_J$	$(0.947 \pm 0.03)\,R_J$	Southworth (2012)

Continued on next page

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
HAT-P-05 b	$(1.06 \pm 0.11)\,M_J$	$(1.252 \pm 0.043)\,R_J$	Southworth et al. (2012c)
HAT-P-06 b	$(1.063 \pm 0.057)\,M_J$	$(1.395 \pm 0.081)\,R_J$	Southworth (2012)
HAT-P-07 b	$(1.87 \pm 0.03)\,M_J$	$(1.526 \pm 0.008)\,R_J$	Benomar et al. (2014)
HAT-P-08 b	$(1.275 \pm 0.053)\,M_J$	$(1.321 \pm 0.04)\,R_J$	Mancini et al. (2013c)
HAT-P-09 b	$(0.778 \pm 0.083)\,M_J$	$(1.38 \pm 0.1)\,R_J$	Southworth (2012)
HAT-P-11 b	$(0.084 \pm 0.0068)\,M_J$	$(0.3966 \pm 0.0094)\,R_J$	Southworth (2011)
HAT-P-12 b	$(0.21 \pm 0.012)\,M_J$	$(0.936 \pm 0.012)\,R_J$	Lee et al. (2012)
HAT-P-13 b	$(0.906 \pm 0.03)\,M_J$	$(1.487 \pm 0.041)\,R_J$	Southworth et al. (2012a)
HAT-P-14 b	$(2.271 \pm 0.083)\,M_J$	$(1.219 \pm 0.059)\,R_J$	Southworth (2012)
HAT-P-15 b	$(1.946 \pm 0.066)\,M_J$	$(1.072 \pm 0.043)\,R_J$	Kovács et al. (2010)
HAT-P-16 b	$(4.193 \pm 0.128)\,M_J$	$(1.19 \pm 0.037)\,R_J$	Ciceri et al. (2013)
HAT-P-17 b	$(0.534 \pm 0.018)\,M_J$	$(1.01 \pm 0.029)\,R_J$	Howard et al. (2012)
HAT-P-18 b	$(0.196 \pm 0.008)\,M_J$	$(0.947 \pm 0.044)\,R_J$	Esposito et al. (2014)
HAT-P-19 b	$(0.292 \pm 0.018)\,M_J$	$(1.132 \pm 0.072)\,R_J$	Hartman et al. (2011a)

Continued on next page

28

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
HAT-P-20 b	$(7.246 \pm 0.187)\,M_J$	$(0.867 \pm 0.033)\,R_J$	Bakos et al. (2011)
HAT-P-21 b	$(4.063 \pm 0.161)\,M_J$	$(1.024 \pm 0.092)\,R_J$	Bakos et al. (2011)
HAT-P-22 b	$(2.147 \pm 0.061)\,M_J$	$(1.08 \pm 0.058)\,R_J$	Bakos et al. (2011)
HAT-P-23 b	$(2.07 \pm 0.12)\,M_J$	$(1.224 \pm 0.037)\,R_J$	Ciceri et al. (2015b)
HAT-P-30 b	$(0.711 \pm 0.028)\,M_J$	$(1.34 \pm 0.065)\,R_J$	Johnson et al. (2011)
HAT-P-32 b	$(0.86 \pm 0.164)\,M_J$	$(1.789 \pm 0.025)\,R_J$	Hartman et al. (2011b)
HAT-P-33 b	$(0.762 \pm 0.101)\,M_J$	$(1.686 \pm 0.045)\,R_J$	Hartman et al. (2011b)
HAT-P-35 b	$(1.054 \pm 0.033)\,M_J$	$(1.332 \pm 0.098)\,R_J$	Bakos et al. (2012)
HAT-P-36 b	$(1.852 \pm 0.095)\,M_J$	$(1.304 \pm 0.025)\,R_J$	Mancini et al. (2015a)
HAT-P-37 b	$(1.169 \pm 0.103)\,M_J$	$(1.178 \pm 0.077)\,R_J$	Bakos et al. (2012)
HAT-P-40 b	$(0.615 \pm 0.038)\,M_J$	$(1.73 \pm 0.062)\,R_J$	Hartman et al. (2012)
HAT-P-42 b	$(1.044 \pm 0.083)\,M_J$	$(1.28 \pm 0.153)\,R_J$	Boisse et al. (2013)
HAT-P-50 b	$(1.35 \pm 0.073)\,M_J$	$(1.288 \pm 0.064)\,R_J$	Hartman et al. (2015b)
HAT-P-51 b	$(0.309 \pm 0.018)\,M_J$	$(1.293 \pm 0.054)\,R_J$	Hartman et al. (2015b)

Continued on next page

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
HAT-P-52 b	$(0.818 \pm 0.029)\,M_J$	$(1.009 \pm 0.072)\,R_J$	Hartman et al. (2015b)
HAT-P-53 b	$(1.484 \pm 0.056)\,M_J$	$(1.318 \pm 0.091)\,R_J$	Hartman et al. (2015b)
HAT-P-54 b	$(0.76 \pm 0.032)\,M_J$	$(0.944 \pm 0.028)\,R_J$	Bakos et al. (2015)
HAT-P-55 b	$(0.582 \pm 0.056)\,M_J$	$(1.182 \pm 0.055)\,R_J$	Juncher et al. (2015)
HAT-P-56 b	$(2.18 \pm 0.25)\,M_J$	$(1.466 \pm 0.04)\,R_J$	Huang et al. (2015)
HATS-02 b	$(1.345 \pm 0.15)\,M_J$	$(1.168 \pm 0.03)\,R_J$	Mohler-Fischer et al. (2013)
HATS-03 b	$(1.071 \pm 0.136)\,M_J$	$(1.381 \pm 0.035)\,R_J$	Bayliss et al. (2013)
HATS-04 b	$(1.323 \pm 0.028)\,M_J$	$(1.02 \pm 0.037)\,R_J$	Jordán et al. (2014)
HATS-05 b	$(0.237 \pm 0.012)\,M_J$	$(0.912 \pm 0.025)\,R_J$	Zhou et al. (2014b)
HATS-06 b	$(0.319 \pm 0.07)\,M_J$	$(0.998 \pm 0.019)\,R_J$	Hartman et al. (2015a)
HATS-09 b	$(0.837 \pm 0.029)\,M_J$	$(1.065 \pm 0.098)\,R_J$	Brahm et al. (2015b)
HATS-13 b	$(0.543 \pm 0.072)\,M_J$	$(1.212 \pm 0.035)\,R_J$	Mancini et al. (2015b)
HATS-15 b	$(2.17 \pm 0.15)\,M_J$	$(1.105 \pm 0.04)\,R_J$	Ciceri et al. (2015a)
HATS-16 b	$(3.27 \pm 0.19)\,M_J$	$(1.30 \pm 0.15)\,R_J$	Ciceri et al. (2015a)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
HATS-17 b	$(1.338 \pm 0.065)\,M_J$	$(0.777 \pm 0.056)\,R_J$	Brahm et al. (2015a)
HD-017156 b	$(3.262 \pm 0.113)\,M_J$	$(1.065 \pm 0.035)\,R_J$	Southworth (2011)
HD-080606 b	$(4.114 \pm 0.155)\,M_J$	$(1.003 \pm 0.027)\,R_J$	Southworth (2011)
HD-097658 b	$(0.0238 \pm 0.0026)\,M_J$	$(0.2005 \pm 0.0087)\,R_J$	Van Grootel et al. (2014)
HD-149026 b	$(0.368 \pm 0.014)\,M_J$	$(0.813 \pm 0.027)\,R_J$	Carter et al. (2009)
HD-189733 b	$(1.15 \pm 0.039)\,M_J$	$(1.151 \pm 0.038)\,R_J$	Southworth (2010)
HD-209458 b	$(0.714 \pm 0.017)\,M_J$	$(1.380 \pm 0.017)\,R_J$	Southworth (2010)
HD-219134 b	$(0.0136 \pm 0.0015)\,M_J$	$(0.1433 \pm 0.0077)\,R_J$	Motalebi et al. (2015)
K2-02 b	$(0.037 \pm 0.004)\,M_J$	$(0.226 \pm 0.016)\,R_J$	Vanderburg et al. (2015)
K2-19 b	$(0.138 \pm 0.038)\,M_J$	$(0.666 \pm 0.068)\,R_J$	Barros et al. (2015)
KELT-03 b	$(1.477 \pm 0.066)\,M_J$	$(1.345 \pm 0.072)\,R_J$	Pepper et al. (2013)
KELT-04 b	$(0.902 \pm 0.06)\,M_J$	$(1.699 \pm 0.046)\,R_J$	Eastman et al. (2015)
KELT-07 b	$(1.28 \pm 0.18)\,M_J$	$(1.533 \pm 0.047)\,R_J$	Bieryla et al. (2015)
KELT-15 b	$(1.196 \pm 0.072)\,M_J$	$(1.52 \pm 0.12)\,R_J$	Rodriguez et al. (2015)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
Kepler-07 b	$(0.453 \pm 0.068)\,M_J$	$(1.649 \pm 0.038)\,R_J$	Southworth (2012)
Kepler-08 b	$(0.59 \pm 0.12)\,M_J$	$(1.381 \pm 0.037)\,R_J$	Southworth (2011)
Kepler-09 b	$(0.142 \pm 0.005)\,M_J$	$(0.990 \pm 0.009)\,R_J$	Dreizler & Ofir (2014)
Kepler-09 c	$(0.098 \pm 0.003)\,M_J$	$(0.955 \pm 0.009)\,R_J$	Dreizler & Ofir (2014)
Kepler-14 b	$(7.68 \pm 0.38)\,M_J$	$(1.126 \pm 0.049)\,R_J$	Southworth (2012)
Kepler-15 b	$(0.696 \pm 0.099)\,M_J$	$(1.289 \pm 0.054)\,R_J$	Southworth (2012)
Kepler-18 c	$(0.054 \pm 0.006)\,M_J$	$(0.49 \pm 0.023)\,R_J$	Cochran et al. (2011)
Kepler-18 d	$(0.0516 \pm 0.0044)\,M_J$	$(0.623 \pm 0.029)\,R_J$	Cochran et al. (2011)
Kepler-25 c	$(0.077 \pm 0.018)\,M_J$	$(0.464 \pm 0.008)\,R_J$	Marcy et al. (2014)
Kepler-26 b	$(0.0161 \pm 0.002)\,M_J$	$(0.248 \pm 0.01)\,R_J$	Jontof-Hutter et al. (2015)
Kepler-26 c	$(0.0195 \pm 0.0021)\,M_J$	$(0.243 \pm 0.011)\,R_J$	Jontof-Hutter et al. (2015)
Kepler-29 b	$(0.0142 \pm 0.0046)\,M_J$	$(0.299 \pm 0.02)\,R_J$	Jontof-Hutter et al. (2015)
Kepler-29 c	$(0.0126 \pm 0.0041)\,M_J$	$(0.28 \pm 0.018)\,R_J$	Jontof-Hutter et al. (2015)
Kepler-30 b	$(0.0355 \pm 0.0044)\,M_J$	$(0.35 \pm 0.02)\,R_J$	Sanchis-Ojeda et al. (2012)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
Kepler-30 d	$(0.0727 \pm 0.0085)\,M_J$	$(0.79 \pm 0.04)\,R_J$	Sanchis-Ojeda et al. (2012)
Kepler-34 b	$(0.22 \pm 0.011)\,M_J$	$(0.764 \pm 0.045)\,R_J$	Welsh et al. (2012)
Kepler-35 b	$(0.127 \pm 0.02)\,M_J$	$(0.728 \pm 0.014)\,R_J$	Welsh et al. (2012)
Kepler-39 b	$(19.1 \pm 1.)\,M_J$	$(1.11 \pm 0.03)\,R_J$	Bonomo et al. (2015)
Kepler-40 b	$(2.16 \pm 0.43)\,M_J$	$(1.44 \pm 0.12)\,R_J$	Southworth (2012)
Kepler-41 b	$(0.56 \pm 0.08)\,M_J$	$(1.29 \pm 0.02)\,R_J$	Bonomo et al. (2015)
Kepler-43 b	$(3.09 \pm 0.21)\,M_J$	$(1.115 \pm 0.041)\,R_J$	Bonomo et al. (2015)
Kepler-44 b	$(1. \pm 0.1)\,M_J$	$(1.09 \pm 0.07)\,R_J$	Bonomo et al. (2015)
Kepler-45 b	$(0.5 \pm 0.06)\,M_J$	$(0.999 \pm 0.069)\,R_J$	Southworth (2012)
Kepler-48 c	$(0.046 \pm 0.007)\,M_J$	$(0.242 \pm 0.012)\,R_J$	Marcy et al. (2014)
Kepler-51 d	$(0.0239 \pm 0.0035)\,M_J$	$(0.865 \pm 0.045)\,R_J$	Masuda (2014)
Kepler-56 b	$(0.069 \pm 0.012)\,M_J$	$(0.581 \pm 0.025)\,R_J$	Huber et al. (2013)
Kepler-56 c	$(0.569 \pm 0.066)\,M_J$	$(0.874 \pm 0.041)\,R_J$	Huber et al. (2013)
Kepler-60 c	$(0.0121 \pm 0.0026)\,M_J$	$(0.17 \pm 0.013)\,R_J$	Jontof-Hutter et al. (2015)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
Kepler-74 b	$(0.63 \pm 0.12)\,M_J$	$(0.96 \pm 0.02)\,R_J$	Bonomo et al. (2015)
Kepler-75 b	$(10.1 \pm 0.4)\,M_J$	$(1.05 \pm 0.03)\,R_J$	Bonomo et al. (2015)
Kepler-76 b	$(2.18 \pm 0.42)\,M_J$	$(1.25 \pm 0.08)\,R_J$	Faigler & Mazeh (2015)
Kepler-77 b	$(0.43 \pm 0.032)\,M_J$	$(0.96 \pm 0.016)\,R_J$	Gandolfi et al. (2013)
Kepler-78 b	$(0.0059 \pm 0.0008)\,M_J$	$(0.107 \pm 0.008)\,R_J$	Grunblatt et al. (2015)
Kepler-79 e	$(0.0129 \pm 0.0038)\,M_J$	$(0.311 \pm 0.012)\,R_J$	Jontof-Hutter et al. (2014)
Kepler-87 b	$(1.02 \pm 0.028)\,M_J$	$(1.203 \pm 0.049)\,R_J$	Ofir et al. (2014)
Kepler-87 c	$(0.0201 \pm 0.0025)\,M_J$	$(0.548 \pm 0.026)\,R_J$	Ofir et al. (2014)
Kepler-88 b	$(0.0274 \pm 0.0079)\,M_J$	$(0.337 \pm 0.035)\,R_J$	Nesvorný et al. (2013)
Kepler-89 d	$(0.333 \pm 0.035)\,M_J$	$(1.005 \pm 0.095)\,R_J$	Weiss et al. (2013)
Kepler-93 b	$(0.0126 \pm 0.0021)\,M_J$	$(0.1319 \pm 0.0017)\,R_J$	Dressing et al. (2015)
Kepler-94 b	$(0.034 \pm 0.004)\,M_J$	$(0.313 \pm 0.013)\,R_J$	Marcy et al. (2014)
Kepler-95 b	$(0.041 \pm 0.009)\,M_J$	$(0.305 \pm 0.008)\,R_J$	Marcy et al. (2014)
Kepler-99 b	$(0.019 \pm 0.004)\,M_J$	$(0.132 \pm 0.007)\,R_J$	Marcy et al. (2014)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
Kepler-101 b	$(0.161 \pm 0.016)\,M_J$	$(0.515 \pm 0.076)\,R_J$	Bonomo et al. (2014)
Kepler-102 e	$(0.028 \pm 0.006)\,M_J$	$(0.198 \pm 0.006)\,R_J$	Marcy et al. (2014)
Kepler-105 c	$(0.0145 \pm 0.0029)\,M_J$	$(0.117 \pm 0.006)\,R_J$	Kostov et al. (2015)
Kepler-106 c	$(0.033 \pm 0.01)\,M_J$	$(0.223 \pm 0.029)\,R_J$	Marcy et al. (2014)
Kepler-117 c	$(1.84 \pm 0.18)\,M_J$	$(1.101 \pm 0.035)\,R_J$	Bruno et al. (2015)
Kepler-131 b	$(0.051 \pm 0.011)\,M_J$	$(0.215 \pm 0.018)\,R_J$	Marcy et al. (2014)
Kepler-289 c	$(0.013 \pm 0.003)\,M_J$	$(0.239 \pm 0.015)\,R_J$	Schmitt et al. (2014)
Kepler-289 d	$(0.415 \pm 0.053)\,M_J$	$(1.034 \pm 0.017)\,R_J$	Schmitt et al. (2014)
Kepler-307 b	$(0.0234 \pm 0.0029)\,M_J$	$(0.217 \pm 0.008)\,R_J$	Jontof-Hutter et al. (2015)
Kepler-406 b	$(0.02 \pm 0.004)\,M_J$	$(0.128 \pm 0.003)\,R_J$	Marcy et al. (2014)
Kepler-412 b	$(0.939 \pm 0.085)\,M_J$	$(1.325 \pm 0.043)\,R_J$	Deleuil et al. (2014)
Kepler-420 b	$(1.45 \pm 0.35)\,M_J$	$(0.94 \pm 0.12)\,R_J$	Santerne et al. (2014)
Kepler-422 b	$(0.43 \pm 0.13)\,M_J$	$(1.15 \pm 0.11)\,R_J$	Endl et al. (2014)
Kepler-423 b	$(0.595 \pm 0.081)\,M_J$	$(1.192 \pm 0.052)\,R_J$	Gandolfi et al. (2015)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
Kepler-433 b	$(2.82 \pm 0.52) M_J$	$(1.45 \pm 0.16) R_J$	Almenara et al. (2015)
Kepler-435 b	$(0.84 \pm 0.15) M_J$	$(1.99 \pm 0.18) R_J$	Almenara et al. (2015)
Kepler-447 b	$(1.37 \pm 0.16) M_J$	$(1.65 \pm 0.2) R_J$	Lillo-Box et al. (2015)
Kepler-454 b	$(0.0214 \pm 0.0044) M_J$	$(0.211 \pm 0.012) R_J$	Gettel et al. (2016)
KOI-188 b	$(0.25 \pm 0.08) M_J$	$(0.978 \pm 0.022) R_J$	Hébrard et al. (2014)
KOI-192 b	$(0.29 \pm 0.09) M_J$	$(1.23 \pm 0.21) R_J$	Hébrard et al. (2014)
KOI-195 b	$(0.34 \pm 0.08) M_J$	$(1.09 \pm 0.03) R_J$	Hébrard et al. (2014)
KOI-372 b	$(3.25 \pm 0.2) M_J$	$(0.882 \pm 0.088) R_J$	Mancini et al. (2015c)
KOI-830 b	$(1.27 \pm 0.19) M_J$	$(1.08 \pm 0.03) R_J$	Hébrard et al. (2014)
KOI-1474 b	$(2.6 \pm 0.3) M_J$	$(0.96 \pm 0.12) R_J$	Dawson et al. (2014)
LHS-6343 b	$(62.1 \pm 1.2) M_J$	$(0.783 \pm 0.011) R_J$	Montet et al. (2015)
OGLE-TR-010 b	$(0.68 \pm 0.15) M_J$	$(1.72 \pm 0.11) R_J$	Southworth (2010)
OGLE-TR-056 b	$(1.41 \pm 0.17) M_J$	$(1.734 \pm 0.058) R_J$	Southworth (2012)
OGLE-TR-111 b	$(0.55 \pm 0.1) M_J$	$(1.011 \pm 0.038) R_J$	Southworth (2012)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
OGLE-TR-113 b	$(1.23 \pm 0.2)\,M_J$	$(1.088 \pm 0.054)\,R_J$	Southworth (2012)
OGLE-TR-132 b	$(1.17 \pm 0.15)\,M_J$	$(1.229 \pm 0.075)\,R_J$	Southworth (2012)
OGLE-TR-182 b	$(1.06 \pm 0.15)\,M_J$	$(1.47 \pm 0.14)\,R_J$	Southworth (2010)
Qatar-1 b	$(1.294 \pm 0.052)\,M_J$	$(1.143 \pm 0.026)\,R_J$	Collins et al. (2015)
Qatar-2 b	$(2.494 \pm 0.054)\,M_J$	$(1.254 \pm 0.013)\,R_J$	Mancini et al. (2014c)
TrES-1 b	$(0.761 \pm 0.051)\,M_J$	$(1.099 \pm 0.035)\,R_J$	Southworth (2010)
TrES-2 b	$(1.206 \pm 0.049)\,M_J$	$(1.193 \pm 0.023)\,R_J$	Southworth (2011)
TrES-3 b	$(1.899 \pm 0.062)\,M_J$	$(1.31 \pm 0.019)\,R_J$	Southworth (2011)
TrES-5 b	$(1.79 \pm 0.068)\,M_J$	$(1.194 \pm 0.015)\,R_J$	Barstow et al. (2015)
WASP-02 b	$(0.88 \pm 0.038)\,M_J$	$(1.063 \pm 0.028)\,R_J$	Southworth (2012)
WASP-04 b	$(1.249 \pm 0.052)\,M_J$	$(1.364 \pm 0.028)\,R_J$	Southworth (2012)
WASP-05 b	$(1.595 \pm 0.052)\,M_J$	$(1.175 \pm 0.055)\,R_J$	Southworth (2012)
WASP-06 b	$(0.485 \pm 0.028)\,M_J$	$(1.23 \pm 0.037)\,R_J$	Tregloan-Reed et al. (2015)
WASP-07 b	$(0.98 \pm 0.13)\,M_J$	$(1.374 \pm 0.094)\,R_J$	Southworth (2012)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
WASP-11 b	$(0.492 \pm 0.024) M_J$	$(0.99 \pm 0.023) R_J$	Mancini et al. (2015a)
WASP-13 b	$(0.512 \pm 0.06) M_J$	$(1.528 \pm 0.084) R_J$	Southworth (2012)
WASP-15 b	$(0.592 \pm 0.019) M_J$	$(1.408 \pm 0.046) R_J$	Southworth et al. (2013)
WASP-16 b	$(0.832 \pm 0.038) M_J$	$(1.218 \pm 0.04) R_J$	Southworth et al. (2013)
WASP-17 b	$(0.477 \pm 0.033) M_J$	$(1.932 \pm 0.053) R_J$	Southworth et al. (2012b)
WASP-18 b	$(10.52 \pm 0.32) M_J$	$(1.204 \pm 0.028) R_J$	Maxted et al. (2013b)
WASP-19 b	$(1.139 \pm 0.036) M_J$	$(1.41 \pm 0.021) R_J$	Mancini et al. (2013a)
WASP-20 b	$(0.311 \pm 0.017) M_J$	$(1.462 \pm 0.059) R_J$	Anderson et al. (2015a)
WASP-21 b	$(0.276 \pm 0.019) M_J$	$(1.162 \pm 0.054) R_J$	Ciceri et al. (2013)
WASP-24 b	$(1.109 \pm 0.054) M_J$	$(1.303 \pm 0.047) R_J$	Southworth et al. (2014)
WASP-25 b	$(0.598 \pm 0.046) M_J$	$(1.247 \pm 0.032) R_J$	Southworth et al. (2014)
WASP-26 b	$(1.02 \pm 0.033) M_J$	$(1.216 \pm 0.047) R_J$	Southworth et al. (2014)
WASP-28 b	$(0.907 \pm 0.043) M_J$	$(1.213 \pm 0.042) R_J$	Anderson et al. (2015a)
WASP-29 b	$(0.244 \pm 0.02) M_J$	$(0.776 \pm 0.043) R_J$	Gibson et al. (2013)

Continued on next page

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
WASP-31 b	$(0.478 \pm 0.029)\,M_J$	$(1.549 \pm 0.05)\,R_J$	Anderson et al. (2011)
WASP-32 b	$(3.6 \pm 0.07)\,M_J$	$(1.18 \pm 0.07)\,R_J$	Maxted et al. (2010)
WASP-35 b	$(0.72 \pm 0.06)\,M_J$	$(1.32 \pm 0.05)\,R_J$	Enoch et al. (2011)
WASP-36 b	$(2.303 \pm 0.068)\,M_J$	$(1.281 \pm 0.029)\,R_J$	Smith et al. (2012)
WASP-38 b	$(2.691 \pm 0.058)\,M_J$	$(1.094 \pm 0.029)\,R_J$	Barros et al. (2011)
WASP-39 b	$(0.28 \pm 0.03)\,M_J$	$(1.27 \pm 0.04)\,R_J$	Faedi et al. (2011)
WASP-41 b	$(0.977 \pm 0.026)\,M_J$	$(1.178 \pm 0.018)\,R_J$	Southworth et al. (2015c)
WASP-42 b	$(0.527 \pm 0.028)\,M_J$	$(1.122 \pm 0.039)\,R_J$	Southworth et al. (2015c)
WASP-43 b	$(2.034 \pm 0.052)\,M_J$	$(1.036 \pm 0.019)\,R_J$	Gillon et al. (2012)
WASP-44 b	$(0.869 \pm 0.081)\,M_J$	$(1.002 \pm 0.038)\,R_J$	Mancini et al. (2013b)
WASP-45 b	$(1.002 \pm 0.062)\,M_J$	$(0.992 \pm 0.038)\,R_J$	Ciceri et al. (2016)
WASP-46 b	$(1.91 \pm 0.13)\,M_J$	$(1.174 \pm 0.037)\,R_J$	Ciceri et al. (2016)
WASP-47 b	$(1.13 \pm 0.06)\,M_J$	$(1.134 \pm 0.039)\,R_J$	Becker et al. (2015)
WASP-47 c	$(0.038 \pm 0.012)\,M_J$	$(0.1621 \pm 0.0058)\,R_J$	Dai et al. (2015)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
WASP-48 b	$(0.907 \pm 0.085) \, M_J$	$(1.396 \pm 0.051) \, R_J$	Ciceri et al. (2015b)
WASP-49 b	$(0.378 \pm 0.027) \, M_J$	$(1.115 \pm 0.047) \, R_J$	Lendl et al. (2012)
WASP-50 b	$(1.437 \pm 0.068) \, M_J$	$(1.138 \pm 0.026) \, R_J$	Tregloan-Reed & Southworth (2013)
WASP-52 b	$(0.46 \pm 0.02) \, M_J$	$(1.166 \pm 0.088) \, R_J$	Swift et al. (2015)
WASP-54 b	$(0.636 \pm 0.025) \, M_J$	$(1.653 \pm 0.09) \, R_J$	Faedi et al. (2013)
WASP-56 b	$(0.571 \pm 0.035) \, M_J$	$(1.092 \pm 0.035) \, R_J$	Faedi et al. (2013)
WASP-57 b	$(0.644 \pm 0.062) \, M_J$	$(1.05 \pm 0.053) \, R_J$	Southworth et al. (2015b)
WASP-58 b	$(0.89 \pm 0.07) \, M_J$	$(1.37 \pm 0.2) \, R_J$	Hébrard et al. (2013)
WASP-59 b	$(0.863 \pm 0.045) \, M_J$	$(0.775 \pm 0.068) \, R_J$	Hébrard et al. (2013)
WASP-60 b	$(0.514 \pm 0.034) \, M_J$	$(0.86 \pm 0.12) \, R_J$	Hébrard et al. (2013)
WASP-61 b	$(2.06 \pm 0.17) \, M_J$	$(1.24 \pm 0.03) \, R_J$	Hellier et al. (2012)
WASP-62 b	$(0.57 \pm 0.04) \, M_J$	$(1.39 \pm 0.06) \, R_J$	Hellier et al. (2012)
WASP-64 b	$(1.271 \pm 0.068) \, M_J$	$(1.271 \pm 0.039) \, R_J$	Gillon et al. (2013)
WASP-65 b	$(1.55 \pm 0.16) \, M_J$	$(1.112 \pm 0.059) \, R_J$	Gómez Maqueo Chew et al. (2013)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
WASP-66 b	$(2.32 \pm 0.13)\,M_J$	$(1.39 \pm 0.09)\,R_J$	Hellier et al. (2012)
WASP-67 b	$(0.406 \pm 0.035)\,M_J$	$(1.091 \pm 0.046)\,R_J$	Mancini et al. (2014a)
WASP-69 b	$(0.26 \pm 0.017)\,M_J$	$(1.057 \pm 0.047)\,R_J$	Anderson et al. (2014b)
WASP-71 b	$(2.242 \pm 0.08)\,M_J$	$(1.46 \pm 0.13)\,R_J$	Smith et al. (2013)
WASP-72 b	$(1.461 \pm 0.059)\,M_J$	$(1.27 \pm 0.2)\,R_J$	Gillon et al. (2013)
WASP-74 b	$(0.95 \pm 0.06)\,M_J$	$(1.56 \pm 0.06)\,R_J$	Hellier et al. (2015)
WASP-75 b	$(1.07 \pm 0.05)\,M_J$	$(1.27 \pm 0.048)\,R_J$	Gómez Maqueo Chew et al. (2013)
WASP-77 b	$(1.76 \pm 0.06)\,M_J$	$(1.21 \pm 0.02)\,R_J$	Maxted et al. (2013a)
WASP-78 b	$(0.89 \pm 0.08)\,M_J$	$(1.7 \pm 0.11)\,R_J$	Smalley et al. (2012)
WASP-79 b	$(0.9 \pm 0.08)\,M_J$	$(2.09 \pm 0.14)\,R_J$	Smalley et al. (2012)
WASP-80 b	$(0.562 \pm 0.027)\,M_J$	$(0.986 \pm 0.022)\,R_J$	Mancini et al. (2014b)
WASP-84 b	$(0.687 \pm 0.033)\,M_J$	$(0.976 \pm 0.025)\,R_J$	Anderson et al. (2015b)
WASP-85 b	$(1.265 \pm 0.062)\,M_J$	$(1.24 \pm 0.03)\,R_J$	Močnik et al. (2015)
WASP-87 b	$(2.18 \pm 0.15)\,M_J$	$(1.385 \pm 0.06)\,R_J$	Anderson et al. (2014a)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
WASP-89 b	$(5.9 \pm 0.4)\,M_J$	$(1.04 \pm 0.04)\,R_J$	Hellier et al. (2015)
WASP-90 b	$(0.63 \pm 0.07)\,M_J$	$(1.63 \pm 0.09)\,R_J$	West et al. (2016)
WASP-96 b	$(0.48 \pm 0.03)\,M_J$	$(1.2 \pm 0.06)\,R_J$	Hellier et al. (2014)
WASP-97 b	$(1.32 \pm 0.05)\,M_J$	$(1.13 \pm 0.06)\,R_J$	Hellier et al. (2014)
WASP-98 b	$(0.83 \pm 0.07)\,M_J$	$(1.1 \pm 0.04)\,R_J$	Hellier et al. (2014)
WASP-100 b	$(2.03 \pm 0.12)\,M_J$	$(1.69 \pm 0.29)\,R_J$	Hellier et al. (2014)
WASP-101 b	$(0.5 \pm 0.04)\,M_J$	$(1.41 \pm 0.05)\,R_J$	Hellier et al. (2014)
WASP-103 b	$(1.47 \pm 0.11)\,M_J$	$(1.554 \pm 0.045)\,R_J$	Southworth et al. (2015a)
WASP-104 b	$(1.272 \pm 0.047)\,M_J$	$(1.137 \pm 0.037)\,R_J$	Smith et al. (2014)
WASP-108 b	$(0.892 \pm 0.055)\,M_J$	$(1.284 \pm 0.047)\,R_J$	Anderson et al. (2014a)
WASP-109 b	$(0.91 \pm 0.13)\,M_J$	$(1.443 \pm 0.053)\,R_J$	Anderson et al. (2014a)
WASP-110 b	$(0.51 \pm 0.064)\,M_J$	$(1.238 \pm 0.056)\,R_J$	Anderson et al. (2014a)
WASP-111 b	$(1.83 \pm 0.15)\,M_J$	$(1.443 \pm 0.094)\,R_J$	Anderson et al. (2014a)
WASP-112 b	$(0.88 \pm 0.12)\,M_J$	$(1.191 \pm 0.049)\,R_J$	Anderson et al. (2014a)

Table 2.1 – continued from previous page

Name	Mass	Radius	Reference
WASP-120 b	$(5.01 \pm 0.26)\,M_J$	$(1.515 \pm 0.083)\,R_J$	Turner et al. (2015)
WASP-121 b	$(1.183 \pm 0.064)\,M_J$	$(1.865 \pm 0.044)\,R_J$	Delrez et al. (2015)
WASP-122 b	$(1.372 \pm 0.072)\,M_J$	$(1.792 \pm 0.069)\,R_J$	Turner et al. (2015)
WASP-123 b	$(0.92 \pm 0.05)\,M_J$	$(1.327 \pm 0.074)\,R_J$	Turner et al. (2015)
WASP-135 b	$(1.9 \pm 0.08)\,M_J$	$(1.3 \pm 0.09)\,R_J$	Spake et al. (2015)
WTS-2 b	$(1.12 \pm 0.16)\,M_J$	$(1.363 \pm 0.061)\,R_J$	Birkby et al. (2014)
XO-1 b	$(0.924 \pm 0.077)\,M_J$	$(1.206 \pm 0.041)\,R_J$	Southworth (2010)
XO-2 b	$(0.597 \pm 0.021)\,M_J$	$(1.019 \pm 0.031)\,R_J$	Damasso et al. (2015)
XO-3 b	$(11.83 \pm 0.38)\,M_J$	$(1.248 \pm 0.049)\,R_J$	Southworth (2010)
XO-5 b	$(1.19 \pm 0.031)\,M_J$	$(1.142 \pm 0.034)\,R_J$	Smith (2015)

2.2.3   Probabilistic Broken Power-Law

We elect to model the MR relation with a probabilistic broken power-law, for the reasons described in §2.2.1. By probabilistic, we mean that this model includes intrinsic dispersion in the MR relation to account for additional variance beyond that of the formal measurement uncertainties. This dispersion represents the variance observed in nature itself around our broken power-law model. To put this in context, a deterministic MR power-law would be described via

$$\frac{R}{R_\oplus} = C\left(\frac{M}{M_\oplus}\right)^S,$$

(2.1)

where R & M are the mass and radius of the object respectively and C & S are the parameters describing the power-law. However, it is easy to conceive of two objects with the exact same mass but different compositions, thereby leading to different radii. For this reason, we argue that a deterministic model provides an unrealistic description of the MR relation. In the probabilistic model, for any given mass there is a corresponding distribution of radius. In this work, we assume a normal distribution in the logarithm of radius. The mean of the distribution takes the result of the deterministic model, and the standard deviation is the intrinsic dispersion, a new free parameter.

A power-law relation can be converted to a linear relation by taking logarithm on both axes. In practice, we take the logarithm base ten of both mass and radius in Earth units, and use a linear relation to fit them. In what follows, we will use M, R to represent mass and radius, and $\mathcal{M}$, $\mathcal{R}$ to represent $\log_{10}(M/M_\oplus)$ and $\log_{10}(R/R_\oplus)$. The power-law relation turns into

$$\mathcal{R} = C + \mathcal{M} \times S, \tag{2.2}$$

where $\mathcal{R} = \log_{10}(R/R_\oplus)$, $\mathcal{M} = \log_{10}(M/M_\oplus)$, and $C = \log_{10} C$. In what follows, we will use $\mathcal{N}(\mu, \sigma)$ as the normal distribution, where μ is the mean and σ is the standard deviation. The corresponding probabilistic relation in log scale becomes

$$\mathcal{R} \sim \mathcal{N}(\mu = C + \mathcal{M} \times S, \sigma = \sigma_\mathcal{R}) \tag{2.3}$$

On a logarithmic scale, the data still approximately follow normal distributions, because the logarithm of a normal distribution is approximately a normal distribution when the standard deviation is small relative to the mean, which is true here since we made a 3σ cut in both mass and radius. The original data, $M \sim \mathcal{N}(M_t, \Delta M)$, will turn into $\mathcal{M}_{ob} \sim \mathcal{N}(\mathcal{M}_t, \Delta\mathcal{M}_{ob})$, where $\mathcal{M}_t = \log_{10}(M_t/M_\oplus)$ and $\Delta\mathcal{M}_{ob} = \log_{10}(e)(\Delta M/M)$.

We consider it more reasonable to assume that the intrinsic dispersion in radius will be a fractional dispersion, rather than an absolute dispersion. For example, the dispersion of Earth-radius planets might be $\mathcal{O}[0.1\,R_\oplus]$ but for stars it should surely be much larger in an absolute sense. Since a fractional dispersion on a linear scale corresponds to an absolute dispersion on logarithmic scale, this assumption is naturally accounted for by our model. To implement the probabilistic model, we employ a hierarchical Bayesian model, or HBM for short.

2.2.4 Hierarchical Bayesian Modeling

The difference between an HBM and the more familiar Bayesian method is that HBMs have two sets of parameters; a layer of hyper parameters, Θ_{hyper}, on top of the local parameters, Θ_{local} (see Hogg et al. 2010 for a pedagogical explanation). The local parameters usually describe the properties of each individual datum, whilst the hypers describe the overall ensemble properties. For example, in this work, the local parameters are the true $\log_{10}(M/M_\oplus)$, $\log_{10}(R/R_\oplus)$ (or $\mathcal{M}_t$, $\mathcal{R}_t$) of all the objects, and the hyper parameters, Θ_{hyper}, are those that represent the broken power-law. This hierarchical structure is illustrated in Figure 2.1, which may be compared to the analogous graphical model shown in Figure 1 of Wolfgang et al. (2016).

Some of the first applications of this method are Loredo & Wasserman (1995), Graziani & Lamb (1996), and Hogg et al. (2010) (in exoplanets research).

For the local parameters, we define a mass, $\mathcal{M}_t$, and radius, $\mathcal{R}_t$, term for each object, giving 632 local variables. In practice, the $\mathcal{R}_t$ local parameters are related to the $\mathcal{M}_t$ term through the broken power-law and each realization of the hyper parameters. In total then, our model includes 632 local parameters and a compact set of hyper parameters, as described later in the MCMC subsection.

2.2.5 Continuous Broken Power-Law Model

Plotting the masses and radii on a log-log scale, (as shown later in Figure 2.3), it is clear that single, continuous power-law is unable to provide a reasonable description of the data. For example, one might reasonably expect that the Neptune-like planets follow a
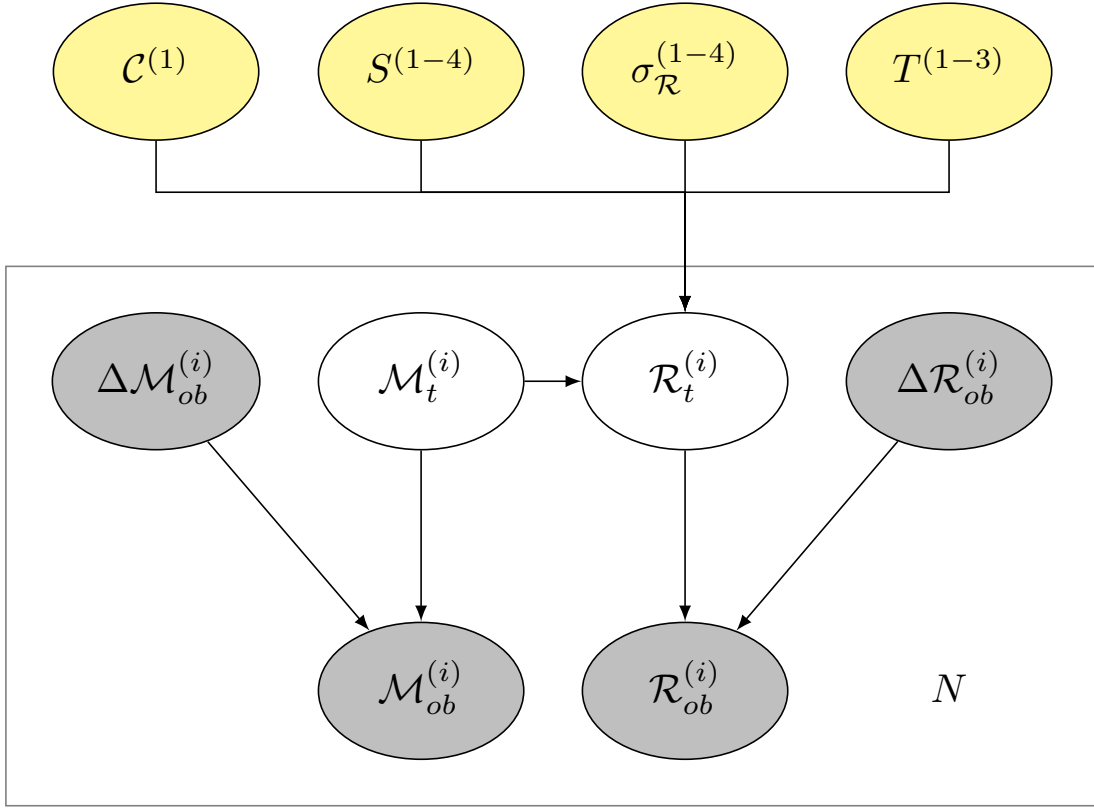
Figure 2.1 Graphical model of the HBM used to infer the probabilistic MR relation in this work. Yellow ovals represent hyper-parameters, white represent the true local parameters and gray represent data inputs. All objects on the plate have N members.

different MR relation from the terrestrial planets, since the voluminous gaseous envelopes of the former dominate their radii (Lopez & Fortney 2014). This therefore argues in favor of using a segmented (or broken) power-law.

At least three fundamentally distinct regimes are expected using some simple physical insights; a segment for terrestrial planets, gas giants, and stars. Indeed, the MR data clearly shows distinct changes in the power-index, corresponding to the transition points between each segment. However, a visual inspection also reveals a turn-over in the MR

relation at around a Saturn-mass. Therefore, from one roughly Saturn-mass to the on-set of stars, there is a strong case for a fourth segment which we consequently include in our model. Later, in Section 2.3.2, we perform a model comparison of a three- versus four-segment model to validate that the four-segment broken power-law is strongly favored.

Our favored model consists of 12 free hyper parameters; 1 offset ($C^{(1)}$), 4 slopes ($S^{(1-4)}$), 4 intrinsic dispersions ($\sigma_{\mathcal{R}}^{(1-4)}$), and 3 transition points ($T^{(1-3)}$). Critically then, we actually fit for the locations of transition points and include an independent intrinsic dispersion for each segment (making our model probabilistic). Also note that the "slopes" in log-log space are the power-law indices in linear space. The hyper parameter vector is therefore

$$
\begin{aligned}
\Theta_{\text{hyper}} = \{ & S^{(1)}, S^{(2)}, S^{(3)}, S^{(4)}, \\
& \sigma_{\mathcal{R}}^{(1)}, \sigma_{\mathcal{R}}^{(2)}, \sigma_{\mathcal{R}}^{(3)}, \sigma_{\mathcal{R}}^{(4)}, \\
& T^{(1)}, T^{(2)}, T^{(3)}, C^{(1)} \}.
\end{aligned}
\tag{2.4}
$$

There is only one free parameter for the offset since we impose the condition that each segment of the power-law is connected, i.e. a continuous broken power-law. By requiring that two segments meet at the transition point between them, we can derive the offsets for the rest of the segments. At each transition point $T^{(j)}$,

$$
C^{(j)} + S^{(j)} \times T^{(j)} = C^{(j+1)} + S^{(j+1)} \times T^{(j)} \quad \text{for } j = 1, 2, 3.
\tag{2.5}
$$

We can now iteratively derive the other offsets as,

$$C^{(j+1)} = C^{(j)} + (S^{(j)} - S^{(j+1)}) \times T^{(j)} \quad \text{for j} = 1, 2, 3. \tag{2.6}$$

2.2.6 Hyper Priors

The hyper priors, or the priors on the hyper-parameters, are selected to be sufficiently broad to allow an extensive exploration of parameter space and to be identical for each segment. Uniform priors are used for the location parameters, namely the offset, C, and transition points, T. For scale parameters, namely the intrinsic dispersion $\sigma_{\mathcal{R}}$, we adopt log-uniform priors.

For the slope parameters, we don't want to constrain them in a specific range, so we use the normal distribution with a large variance. This leads to a prior distribution which is approximately uniform in any small region yet loosely constrains the MCMC walkers to the relevant scale of the data. A detailed list of the priors is provided in Table 2.2.

Table 2.2: Description and posterior of hyper parameters. The prior distributions of the hyper parameters are $C^{(1)} \sim \mathcal{U}(-1,1)$; $S^{(1-4)} \sim \mathcal{N}(0,5)$; $\log_{10}\left[\sigma_{\mathcal{R}}^{(1-4)}\right] \sim \mathcal{U}(-3,2)$; $T^{(1-3)} \sim \mathcal{U}(-4,6)$.

Θ_{hyper}	Description	Credible Interval
$10^C R_\oplus$	Power-law constant for the Terran (T-class) worlds MR relation	$1.008^{+0.046}_{-0.045} R_\oplus$
$S^{(1)}$	Power-law index of Terran worlds; $R \propto M^S$	$0.2790^{+0.0092}_{-0.0094}$
$S^{(2)}$	Power-law index of Neptunian worlds; $R \propto M^S$	$0.589^{+0.044}_{-0.031}$
$S^{(3)}$	Power-law index of Jovian worlds; $R \propto M^S$	$-0.044^{+0.017}_{-0.019}$
$S^{(4)}$	Power-law index of Stellar worlds; $R \propto M^S$	$0.881^{+0.025}_{-0.024}$
$\sigma_{\mathcal{R}}^{(1)}$	Fractional dispersion of radius for the Terran MR relation	$4.03^{+0.94}_{-0.64}$ %
$\sigma_{\mathcal{R}}^{(2)}$	Fractional dispersion of radius for the Neptunian MR relation	$14.6^{+1.7}_{-1.3}$ %
$\sigma_{\mathcal{R}}^{(3)}$	Fractional dispersion of radius for the Jovian MR relation	$7.37^{+0.46}_{-0.45}$ %
$\sigma_{\mathcal{R}}^{(4)}$	Fractional dispersion of radius for the Stellar MR relation	$4.43^{+0.64}_{-0.47}$ %
$10^{T(1)} M_\oplus$	Terran-to-Neptunian transition point	$2.04^{+0.66}_{-0.59} M_\oplus$

Continued on next page

50

Table 2.2 – continued from previous page

Θ_{hyper}	Description	Credible Interval
$10^{T(2)} M_{\oplus}$	Neptunian-to-Jovian class transition point	$0.414^{+0.057}_{-0.065} M_J$
$10^{T(3)} M_{\oplus}$	Jovian-to-Stellar class transition point	$0.0800^{+0.0081}_{-0.0072} M_{\odot}$

51

2.2.7 Two Different Categories of Local Parameters

The local parameters in our model are formally $\mathcal{M}_t$ and $\mathcal{R}_t$, although in practice $\mathcal{R}_t$ doesn't need to be fitted explicitly since it is derived from the realization of the broken power-law (as described in more detail later).

Even for $\mathcal{M}_t$ though, there are two categories that we must distinguish between. Objects within the Solar System tend to have very precise measurements of their fundamental properties such that their formal uncertainties are negligible relative to the uncertainties encountered for extrasolar objects, for which we must account for the measurement uncertainties in our model.

For objects with negligible error, we simply fix $\mathcal{M}_t = \mathcal{M}_{ob}$ and $\mathcal{R}_t = \mathcal{R}_{ob}$, since $\Delta\mathcal{M}_{ob}, \Delta\mathcal{R}_{ob} \propto \frac{\Delta M}{M}, \frac{\Delta R}{R} = 0$. For objects in the second category, $\mathcal{M}_t$ are set to be independently uniformly distributed in $[-4, 6]$. Throughout the paper, we will use $\mathcal{U}(a, b)$ to denote a uniform distribution, where a and b are the lower and upper bounds of the distribution. So

$$\mathcal{M}_t^{(i)} \sim \mathcal{U}(-4, 6) \quad \text{for i} = 1, 2, ..., 316. \tag{2.7}$$

2.2.8 Inverse Sampling

We use the inverse sampling method to sample the parameters $\mathcal{M}_t$ and Θ_{hyper}. By inverse sampling, we mean that the walkers directly sample in the probability space, rather than the parameter space itself. By directly walking in the prior probability space with Gaussian function as our proposal distribution, inverse sampling is more efficient than walking in

the parameter space plus likelihood penalization (see Devroye 1986 further details on the inverse sampling method).

For each jump in the MCMC chain, we sample a probability, p, for each parameter with $\mathcal{U}(0,1)$. We then determine this parameter's cumulative distribution from its prior probability distribution. With p and the cumulative distribution, we can then calculate the corresponding sample value of the parameter.

The equations of the prior distributions of $\mathcal{M}_t$ and Θ_{hyper} are already shown in Table 2.2 and Equation (2.7). With inverse sampling, the effects of the priors have already been accounted for, meaning that we do not need to add the prior probabilities of parameters into the total log-likelihood function.

2.2.9 Total Log-Likelihood

As discussed above, since $\mathcal{M}_t$ and Θ_{hyper} are drawn with inverse sampling, there is no need to add corresponding penalty terms to the log-likelihood function. The total log-likelihood is now based on how we sample $\mathcal{R}_t$ from $\mathcal{M}_t$ and Θ_{hyper}, and the relations between $\mathcal{M}_t$, $\mathcal{R}_t$, and data. The relations are given by

$$\mathcal{M}_{ob}^{(i)} \sim \mathcal{N}\left(\mathcal{M}_t^{(i)}, \Delta\mathcal{M}_{ob}^{(i)}\right). \tag{2.8}$$

When $\Delta\mathcal{M}_{ob}^{(i)} = 0$, the above equation can be interpreted as $\mathcal{M}_{ob}^{(i)} = \mathcal{M}_t^{(i)}$, which corresponds to the case where measurement errors are zero. This is also true for $\mathcal{R}_{ob}^{(i)}$, such that

53

$$\mathcal{R}_{ob}^{(i)} \sim \mathcal{N}\big(\mathcal{R}_{t}^{(i)}, \Delta\mathcal{R}_{ob}^{(i)}\big), \tag{2.9}$$

and

$$\mathcal{R}_{t}^{(i)} \sim \mathcal{N}\big(f(\mathcal{M}_{t}^{(i)}, \boldsymbol{\Theta}_{\text{hyper}}), \sigma'_{\mathcal{R}}\big), \tag{2.10}$$

where we define

$$
\big(f(\mathcal{M}_{t}^{(i)}, \boldsymbol{\Theta}_{\text{hyper}}), \sigma'_{\mathcal{R}}\big) =
\begin{cases}
\big(C^{(1)} + \mathcal{M}_{t}^{(i)} S^{(1)}, \sigma_{\mathcal{R}}^{(1)}\big) & \mathcal{M}_{t}^{(i)} \leq T^{(1)} \\[2mm]
\big(C^{(2)} + \mathcal{M}_{t}^{(i)} S^{(2)}, \sigma_{\mathcal{R}}^{(2)}\big) & T^{(1)} < \mathcal{M}_{t}^{(i)} \leq T^{(2)} \\[2mm]
\big(C^{(3)} + \mathcal{M}_{t}^{(i)} S^{(3)}, \sigma_{\mathcal{R}}^{(3)}\big) & T^{(2)} < \mathcal{M}_{t}^{(i)} \leq T^{(3)} \\[2mm]
\big(C^{(4)} + \mathcal{M}_{t}^{(i)} S^{(4)}, \sigma_{\mathcal{R}}^{(4)}\big) & T^{(3)} < \mathcal{M}_{t}^{(i)}
\end{cases}
. \tag{2.11}
$$

Combining Equation (2.9) and (2.10), we have

$$\mathcal{R}_{ob}^{(i)} \sim \mathcal{N}\Big(f(\mathcal{M}_{t}^{(i)}, \boldsymbol{\Theta}_{\text{hyper}}), \sqrt{(\Delta\mathcal{R}_{ob}^{(i)})^2 + (\sigma'_{\mathcal{R}})^2}\Big). \tag{2.12}$$

Equation (2.12) shows that if we have already sampled $\mathcal{M}_t$ and $\boldsymbol{\Theta}_{\text{hyper}}$, we don't need to sample $\mathcal{R}_t$ anymore since $\mathcal{R}_{ob}$ can be directly related to $\mathcal{M}_t$ and $\boldsymbol{\Theta}_{\text{hyper}}$. From Equation (2.8) and (2.12), we can see that the total log-likelihood of the model is

$$-2\log\mathcal{L} = \sum_{i=1}^{N}\left(\frac{\mathcal{M}_{ob}^{(i)} - \mathcal{M}_{t}^{(i)}}{\Delta\mathcal{M}_{ob}^{(i)}}\right)^2 + \sum_{i=1}^{N}(\Delta\mathcal{M}_{ob}^{(i)})^2 +$$

$$\sum_{i=1}^{N}\frac{\left(\mathcal{R}_{ob}^{(i)} - f(\mathcal{M}_{t}^{(i)}, \boldsymbol{\Theta}_{\text{hyper}})\right)^2}{\left(\Delta\mathcal{R}_{ob}^{(i)}\right)^2 + \left(\sigma_{\mathcal{R}}'\right)^2} +$$

$$\sum_{i=1}^{N}\log\left[\left(\Delta\mathcal{R}_{ob}^{(i)}\right)^2 + \left(\sigma_{\mathcal{R}}'\right)^2\right]. \tag{2.13}$$

Note that in the above, we assume mass and radius have no covariance, which is almost always true given the independent methods of their measurement.

2.3   Analysis

2.3.1   Parameter Inference with Markov Chain Monte Carlo

We used the Markov Chain Monte Carlo (MCMC) method with the Metropolis algorithm (Metropolis et al. 1953) to explore the parameter space and infer the posterior distributions for both the hyper and local parameters. The Metropolis algorithm uses jumping walkers, proceeding by accepting or rejecting each jump by comparing its likelihood with the likelihood of the previous step. Since we have 12 hyper parameters and 316 data points (corresponding to 316 $\mathcal{M}_t$), the walker jumps in a probability hyper cube of (12+316) dimensions.

We begin by running 5 independent initial chains for 500,000 **accepted** steps each, seeding the parameters $T^{(1-3)}$ from 0.5, 2, and 4 with Gaussian distributions of sigma one

(but keeping all other terms seeded randomly from the hyper priors).

We identify the burn-in point by eye, searching for the instant where the local variance in the log-likelihood (with respect to the chain step) stabilizes to a relatively small scatter in comparison to the initial steps. This burn-in point tended to occur after $\simeq 200,000$ accepted steps, largely driven by the fact that both the hyper and the local parameters were not seeded from local minimums (with the exception of $T^{(1-3)}$) and therefore required a substantial number of steps to converge.

Combining these initial chains, we chose 10 different realizations which have the highest log-likelihoods but also not too close to each other. We then start 10 new independent chains, where each chain is seeded from one of the top 200 log-like solutions found from the stacked initial chains. We run each of these 10 chains for 10^7 trials with acceptance rates $\sim 5\%$ (i.e. 500,000 accepted steps each) and find, as expected, that each chain is burnt-in right from the beginning.

To check for adequate mixing, we calculated the effective length, defined as the length of the chain divided by the correlation length, where the correlation length is defined as $\ell_{corr} = \min_{lag}\{|\text{AutoCorrelation(chain, lag)}| < 0.5\}$ (Tegmark et al. 2004). We find that the sum of the effective lengths exceeds 2000 (i.e. is $\gg 1$), indicating good mixing. We also verified that the Gelman-Rubin statistic (Gelman & Rubin 1992) dropped below 1.1 (it was 1.02), indicating that the chains had converged. Finally, we thinned the 10 chains by a factor of 100, and stacked them together, which gives a combined chain of length of 10^6. The hyper-parameter posteriors, available at this URL, are shown as a triangle plot in Figure 2.2. We list the median and corresponding 68.3% credible interval of each hyper

parameter posterior distribution in Table 2.2. Our model, evaluated at the spatial median of the hyper parameters, is shown in Figure 2.3 compared to the data upon which it was conditioned. The spatial median simply finds the sample from the joint posteriors which minimizes the Euclidean distance to all other samples.

2.3.2 Model Comparison

The model with four segments was at first selected by visual inspection of the data. Two of the three transition points, $T(1)$ and $T(3)$, occur at locations which can be associated with physically well-motivated boundaries (planets accreting volatile envelopes, Rogers 2015, and hydrogen burning, Dieterich et al. 2014), whereas the $T(2)$ transition is not as well studied.

In order to demonstrate that this model is statistically favored over the three-segment model, we repeated all of the fits for a simpler three-segment model. We seed the remaining two transition points from the approximate locations of $T(1)$ and $T(3)$ found from the four-segment model fit. We label this model as $\mathcal{H}_3$ and the four-segment model used earlier as $\mathcal{H}_4$.

For this simpler model, $\mathcal{H}_3$ uses only two transition points which break the data into three different segments. To implement this model, the only difference is that the hyper parameters vector Θ'_{hyper} becomes

$$\Theta'_{\text{hyper}} = (C, S^{(1-3)}, \sigma_{\mathcal{R}}^{(1-3)}, T^{(1-2)}). \tag{2.14}$$

We find that the maximum log-likelihood of $\mathcal{H}_3$ is considerably less than that of $\mathcal{H}_4$,
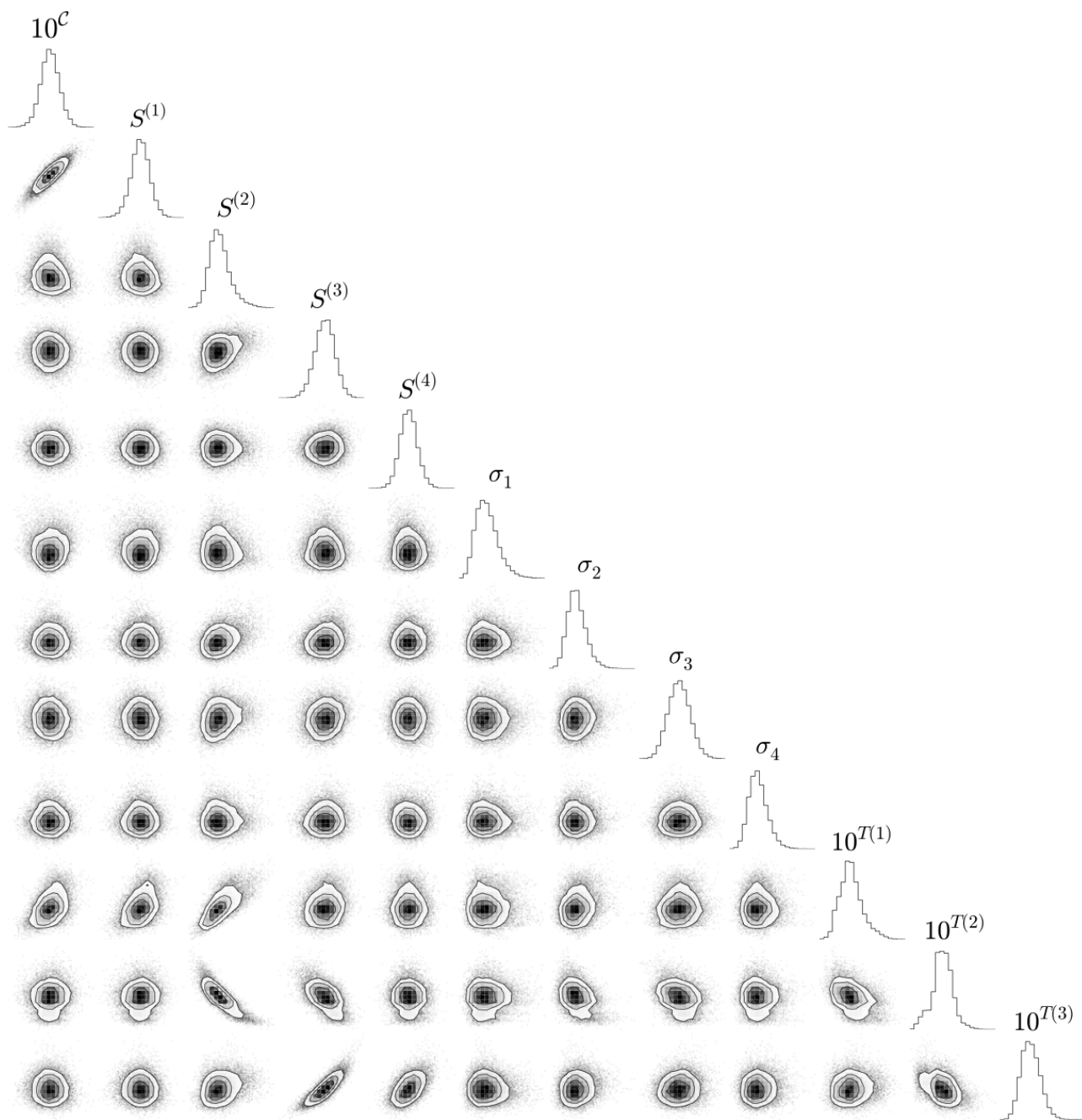
Figure 2.2 Triangle plot of the hyper-parameter joint posterior distribution (generated using corner.py). Contours denote the 0.5, 1.0, 1.5 and 2.0 σ credible intervals.
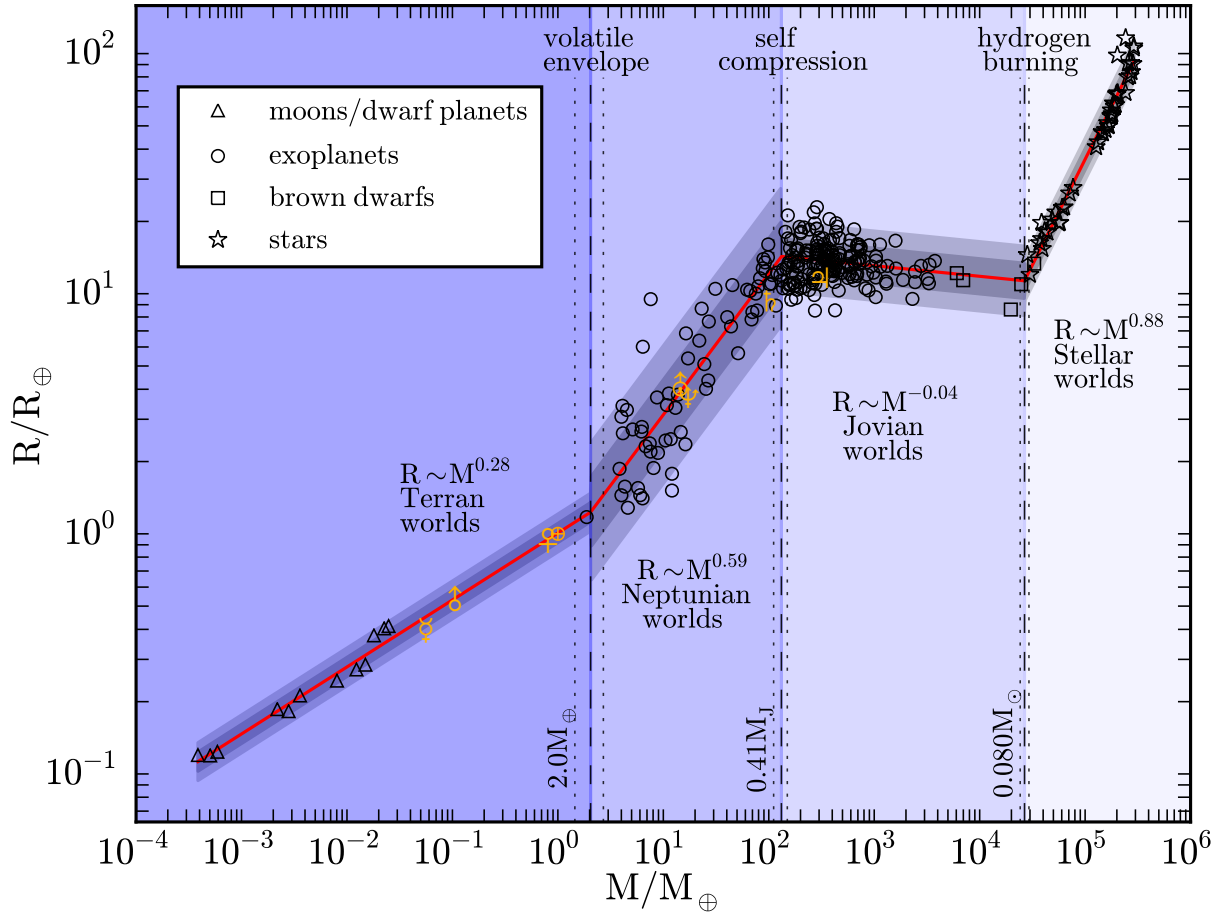
Figure 2.3 The mass-radius relation from dwarf planets to late-type stars. Points represent the 316 data against which our model is conditioned, with the data key in the top-left. Although we do not plot the error bars, both radius and mass uncertainties are accounted for. The red line shows the mean of our probabilistic model and the surrounding light and dark gray regions represent the associated 68% and 95% credible intervals, respectively. The plotted model corresponds to the spatial median of our hyper parameter posterior samples.

less by 34.16 (corresponding to $\Delta\chi^2 = 68.3$ for 316 data points) at the gain of just three fewer free parameters. The marginal likelihood cannot be easily computed in the very high dimensional parameter space of our problem, and the Bayesian and Akaike information criteria (BIC, Schwarz 1978 & AIC, Akaike 1974) are also both invalid for such high dimensionality. Instead, we used the Deviance information criterion (DIC, Spiegelhalter et al. 2002), a hierarchical modeling generalization of the AIC, to compare the two models. When comparing two models with the DIC, the smaller value is understood to be the preferred model. We find that $DIC(\mathcal{H}_4) = -665.5$ and $DIC(\mathcal{H}_3) = -333.5$, indicating a strong preference for model $\mathcal{H}_4$.

2.3.3 The Effect of our Data Cuts

As discussed earlier in §2.2.2, our data cuts removed 16% of the initial data considered. Since these points are low SNR data, they, by definition, have weak effects on the likelihood function. As is evident in Figure 2.3, there is an abundance of precise data constraining the slope parameters in each segment and none of the segments can be described as residing in a poorly constrained region. Given that the transition points are defined as the intercepts of the slopes, they too are well constrained by virtue of the construction of our model. Critically, then, a paucity of data at the actual transition point locations (as is true for $T(1)$) has little influence on our inference of their locations. In order for the results of this work to be significantly affected by the exclusion of these low SNR data, these points would have to have modified the inference of the slope parameters.

To demonstrate this effect is negligible, we consider the Neptunian segment in iso-

lation, since it strongly affects the critical transition $T(1)$ and features the largest fraction of excluded points (24%). Since the excluded data were due to lossy mass measurements, we ignore the radius errors and perform a simple weighted linear least squares regression with and without the excluded data, where we approximate the observations to be normally distributed. We find that the slope parameter, $S(2)$, changes from 0.782 ± 0.058 to 0.784 ± 0.050 by re-introducing the excluded data, illustrating the negligible impact of these data.

2.3.4 Injection/Recovery Tests

In order to verify the robustness of our algorithm, we created ten fake data sets and blindly ran our algorithm again on each. The data sets are generated by making random, fair draws from our joint posteriors (both the local and the hyper parameters), ensuring that each draw is from an different effective chain. We then re-ran our original algorithm as before, except that the number of steps in the final chain is reduced by a factor of ten for computational expedience.

We computed the one and two-sigma credible intervals on each hyper-parameter and compare them to the injected truth in Figure 2.4. As evident from this figure, we are able to easily recover all of the inputs to within the expected range, validating the robustness of the main results presented in this work.
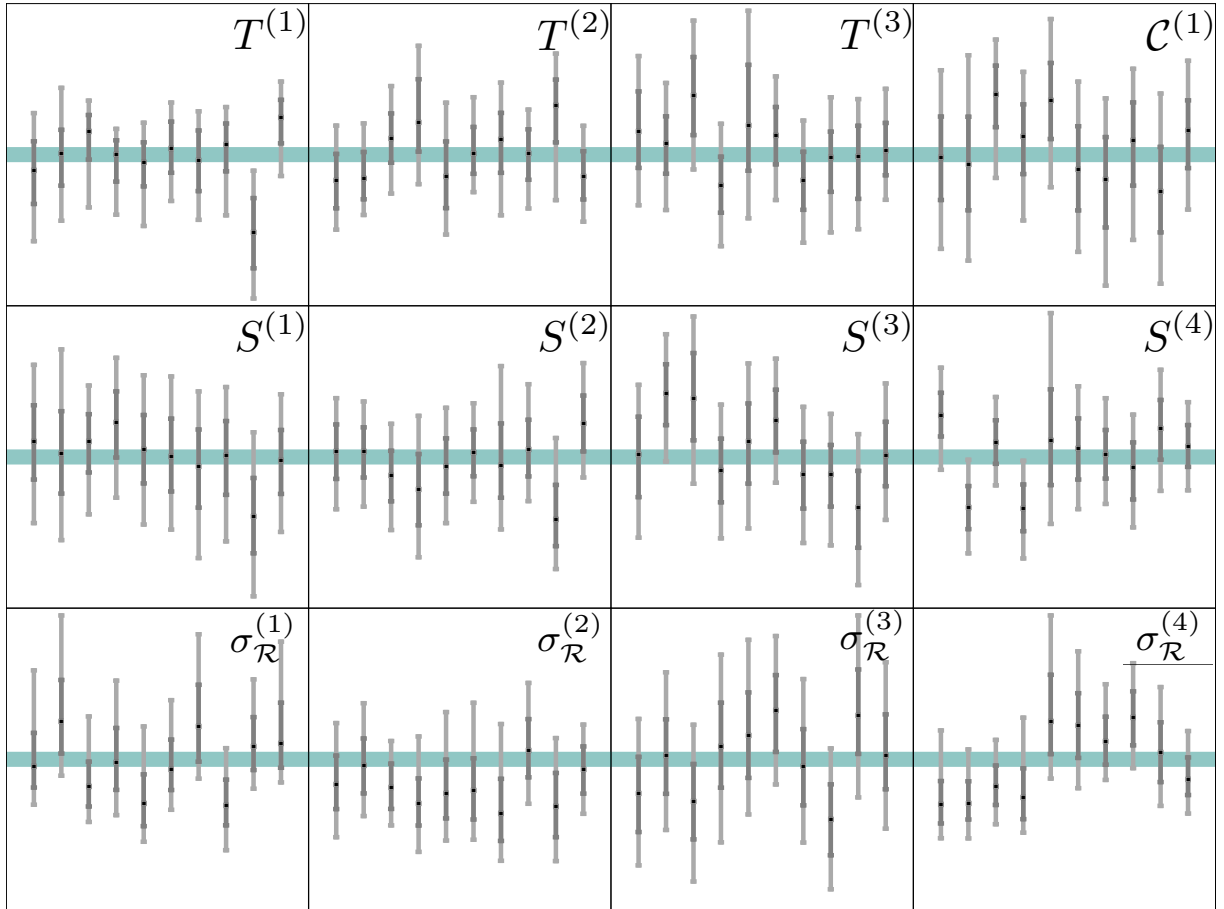
Figure 2.4 Each sub-panel shows the residuals of a hyper-parameter in our model, as computed between ten injected truths and the corresponding recovered values. The black square denotes the recovered posterior median and the dark & light gray bars denote the 1 & 2 σ credible intervals. The green horizontal bar marks the zero-point expected for a perfect recovery.

2.4 Classification

2.4.1 Classification with an MR relation

A unique aspect of this work was to use freely fitted transitional points in our MR relation. As discussed earlier, these transitional points essentially classify the data between distinct categories, where the class boundaries occur in mass and are defined using the feature of $d\mathcal{R}/dM$. Such classes are evident even from visual inspection of the MR data (see Figure 2.3), but our Bayesian inference of a self-consistent probabilistic broken power-law provides statistically rigorous estimates of these class boundaries. In what follows, we discuss the implications of the inferred locations of the class boundaries ($T^{(1)}$, $T^{(2)}$ and $T^{(3)}$).

2.4.2 Naming the Classes

Rather than refer to each class as segments 1, 2, 3 and 4, we here define a name for each class to facilitate a more physically intuitive discussion of the observed properties. A naming scheme based on the physical processes is appealing but ultimately disingenuous since our model is deliberately chosen to be a data-driven inference, free of physical assumptions about the mechanics and evolution sculpting these worlds. We consider it more appropriate, then, to name each class based upon a typical and well-known member.

For segment 2, Neptune and Uranus are typical members and are of course very similar to one another in basic properties. We therefore consider this class a sub-sample of Neptune-like worlds, or "Neptunian" worlds more succinctly. Similarly, we identify Jupiter as a typical member of segment 3, unlike Saturn which lies close to a transitional

point. Accordingly, we define this sub-sample to be representative of Jupiter-like worlds, or "Jovian" worlds.

For the hydrogen-burning late-type stars of segment 4, these objects can be already classified by their spectral types spanning M, K, and late-type G dwarfs. Rather than refer to them as M/K/late-G class stars, we simply label them as stars for the sake of this work and for consistency with the "worlds" taxonomy dub them "Stellar" worlds.

Finally, we turn to segment 1 which is comprised largely of Solar System members and thus all of which are relatively well-known. The objects span from dwarf planets to terrestrial planets, from silicate worlds to icy worlds, making naming this broad class quite challenging. Additionally, calling this class Earth-like worlds would be confusing given the usual association of this phrase with habitable Earth-analogs. For consistency with the naming scheme used thus far, we decided that dubbing these objects as "Terran" worlds to be the most appropriate.

2.4.3 $T^{(1)}$: The Terran-Neptunian Worlds Divide

From masses of $\sim 10^{-4} \, M_\oplus$ to a couple of Earth masses, we find that a continuous power-law of $R \sim M^{0.279 \pm 0.009}$ provides an excellent description of these Terran worlds. No break is observed between "dwarf planets" and "planets". If the Terrans display a constant mean density, we would expect $R \sim M^{1/3}$, and so the slightly depressed measured index indicates modest compression with increasing mass ($\rho \sim M^{0.16 \pm 0.03}$). Our result is in close agreement with theoretical models, which typically predict $R \sim M^{0.27}$ for Earth-like compositions (e.g. see Valencia et al. 2006).

We find the first transition to be located at $(2.0 \pm 0.7)\, M_\oplus$, defining the transition from Terrans to Neptunians. After this point, the density trend reverses with $\bar{\rho} \sim M^{-0.77 \pm 0.13}$, indicating the accretion of substantial volatile gas envelopes. This transition is not only evident in the power-law index, but also in the intrinsic dispersion, which increases by a factor of (3.6 ± 0.9) from Terrans to Neptunians. This transition point is of major interest to the community, since it caps the possibilities of rocky, habitable Super-Earth planets, with implications for future mission designs (e.g. see Dalcanton et al. 2015).

Our result is compatible with independent empirical and theoretical estimates of this transition. Starting with the former, we compare our result to Rogers (2015), who sought the transition in radius rather than mass. This was achieved by identifying radii which exceed that of a solid planet, utilizing a principle first proposed by Kipping et al. (2013). Assuming an Earth-like compositional model, the radius threshold was inferred to be $1.48^{+0.08}_{-0.04}\, R_\oplus$ (Rogers 2015). Our result may be converted to a radius by using our derived relation. However, since our model imposes intrinsic radius dispersion (i.e. the probabilistic nature of our model), the uncertainty in radius is somewhat inflated by this process. Nevertheless, we may convert our mass posterior samples to fair radii realizations using our `Forecaster` public code (described later in Section 2.5). Accordingly, we find that the transition occurs at $1.23^{+0.44}_{-0.22}\, R_\oplus$, which is fully compatible with Rogers (2015).

A comparison to theory comes from Lopez & Fortney (2014), who scale down compositional models of gaseous planets to investigate the minimum size of a H/He rich sub-Neptune. From this theoretical exercise, the authors estimate that $1.5\, R_\oplus$ is the minimum radius of a H/He rich sub-Neptune, which is also compatible with our measurement.

Therefore, despite the fact that we do not impose any physical model (unlike Lopez & Fortney 2014 & Rogers 2015), our broken power-law model recovers the transition from Terrans to Neptunians.

2.4.4 $T^{(3)}$: The Jovian-Stellar Worlds Divide

Another well-understood transition is recovered by our model at $(0.080 \pm 0.008) \, M_\odot$, which we interpret as the onset of hydrogen burning. Like the Terran-Neptunian worlds transition, we may compare this transition to other estimates of the critical boundary. In the recent work of Dieterich et al. (2014), the authors performed a detailed observational campaign around this boundary. Inspecting the T_{eff}-R plane, the authors identified a minimum at $\simeq 0.086 \, R_\odot$, which corresponds to $\simeq 0.072 \, M_\odot$, with the 5 Gyr isochrones[2] of Baraffe et al. (1998). Based on this, we conclude that the result is fully compatible with our own prediction.

From stellar modeling, estimates of the minimum mass for hydrogen-burning range from $0.07 \, M_\odot$ to $0.09 \, M_\odot$ (Burrows et al. 1993, 1997; Baraffe et al. 1998; Chabrier et al. 2000; Baraffe et al. 2003; Saumon & Marley 2008). Therefore, both independent observational studies and theoretical estimates are consistent with our broken power-law estimate.

2.4.5 $T^{(2)}$: The Neptunian-Jovian Worlds Divide

We find strong evidence for a transition in our broken power-law at $(0.41 \pm 0.07) \, M_J$, corresponding to the transition between Neptunians and Jovians. Whilst this transition

---

[2]Although the point slightly precedes the first point in the Baraffe et al. 1998 grid, requiring a small linear extrapolation to compute.

has been treated as an assumed, fixed point in previous works (e.g. Weiss et al. 2013 adopt a fixed transition at $150\,M_\oplus$, or $0.47\,M_J$), our work appears to be the first instance of a **data-driven inference** of this transition.

A plausible physical interpretation of this boundary is that a Neptunian world rapidly grows in radius as more mass is added, depositing more gaseous envelope to its outer layer. Eventually, the object's mass is sufficient for gravitational self-compression to start reversing the growth, leading into a Jovian world. The existence of such a transition is not unexpected, but our model allows for an actual measurement of its location.

We infer the significance of this transition to be high at nearly $10\,\sigma$ (see §2.3.2), motivating us to propose that this transition is physically real and that a class of Jovians is taxonomically rigorous in the mass-radius plane. A defining feature of the Jovian worlds is that the MR power-index is close to zero (-0.04 ± 0.02), with radius being nearly degenerate with respect to mass.

We find that brown dwarfs are absorbed into this class, displaying no obvious transition (also see Figure 2.3) at $\sim 13\,M_J$, as was also argued by Hatzes & Rauer (2015). When viewed in terms of mass and radius then, brown dwarfs are merely high-mass members of a continuum of Jovians and more closely resemble "planets" than "stars".

The fact that the Neptunian-to-Jovian transition occurs at around one Saturn mass is generally incompatible with theoretical predictions of a H/He rich planet, such as Saturn. Calculations by Zapolsky & Salpeter (1969) predict that a cold sphere of H/He is expected to reach a maximum size somewhere between $1.2\,M_J$ to $3.3\,M_J$. The suite of models produced by Fortney et al. (2007) for H/He rich giant planets, for various insolations and

metallicities, peak at masses between $\sim 2.5\,M_J$ to $\sim 5.2\,M_J$. Nevertheless, Jupiter and Saturn have similar radii (within 20%) of one another despite a factor of three difference in mass, crudely indicating that Jovians commence at a mass less than or equal to that of Saturn.

2.5   Forecasting

2.5.1   `Forecaster`: An Open-Source Package

Using our probabilistic model for MR relation inferred in this work, it is possible to now achieve our primary objective: to forecast the mass (or radius) of an object given the radius (or mass). Crucially, our forecasting model can not only propagate measurement uncertainty on the inputs (easily achieved using Monte Carlo draws), but also the uncertainty in the model itself thanks to the probabilistic nature of our model. Thus, even for an input with perfect measurement error (i.e. none), our forecasting model will still return a probability distribution for the forecasted quantity, due to (i) our measurement uncertainty in the hyper-parameters describing the model; and (ii) the intrinsic variability seen in nature itself around the imposed model.

To enable the community to make use of this model, we have written a `Python` package, `Forecaster` [3] (MIT license), which allows a user to input a mass (or radius) posterior distribution and return a radius (or mass) forecasted distribution. Alternatively, one can simply input a mean and standard deviation of mass (or radius), and the package

[3]https://github.com/chenjj2/forecaster

will return a forecasted mean and standard deviation of the radius (or mass). This code works for any object with mass in the range of $[3 \times 10^{-4} \ M_\oplus, 3 \times 10^5 \ M_\oplus(0.87 \ M_\odot)]$, or radius in the range of $[0.1 \ R_\oplus, 100 \ R_\oplus(9 \ R_J)]$.

We present the details of how we use the MR relation we obtained to forecast one quantity from the other below.

2.5.2 Forecasting Radius

Predicting radius given mass is straightforward from our model. If the input is the mean and standard deviation of mass, `Forecaster` will first generate a vector of masses, $\{M^{(i)}, i = 1, 2, ..., n\}$, following a normal distribution truncated within the mass range. Otherwise, the code will accept the input mass posterior distribution as $\{M^{(i)}, i = 1, 2, ..., n\}$. `Forecaster` will then randomly choose n realizations of the hyper parameters from the hyper posteriors derived in this work. A radius will be drawn for each $M^{(i)}$ with each set of hyper parameters $\Theta_{\mathrm{hyper}}^{(i)}$, as

$$R^{(i)} \sim \mathcal{N}(f(M^{(i)}, \Theta_{\mathrm{hyper}}^{(i)}), \sigma_{\mathcal{R}}^{(i)}). \tag{2.15}$$

The output in this case is a vector of radii $\{R^{(i)}, i = 1, 2, ..., n\}$. It is worth pointing out that since our model uses a Gaussian distribution, it is possible that the predicted radius for a given mass turns out to be so small that no current physical composition model can explain. However we choose not to truncate the prediction with any theoretical model and let our code users to choose what's suitable for them.

2.5.3 Forecasting Mass

Mass cannot be directly sampled given $\{R^{(i)}, i = 1, 2, ..., n\}$ with our model. To sample mass, Forecaster first creates a grid of masses as $\{M^{(j)}_{grid}, j = 1, 2, ..., m\}$ in the whole mass range of our model. Similarly, then we randomly chose n sets of hyper parameters from the hyper posteriors of our model. For each radius $R^{(i)}$, Forecaster calculates the probability $\{p^{(j)}_{grid}, j = 1, 2, ..., m\}$ of $R^{(i)}$ given $M^{(j)}$ with $\Theta^{(i)}_{hyper}$. Finally, Forecaster samples $M^{(i)}$ from $\{M^{(j)}_{grid}, j = 1, 2, ..., m\}$ with $\{p^{(j)}_{grid}, j = 1, 2, ..., m\}$. The output in this case is a vector of masses $\{M^{(i)}, i = 1, 2, ..., n\}$.

2.5.4 Examples: Kepler-186f and Kepler-452b

To give illustrative examples of Forecaster, we here forecast the masses of arguably the two most Earth-like planets discovered by *Kepler*, Kepler-186f and Kepler-452b.

Kepler-186f was discovered by Quintana et al. (2014), reported to be $(1.11 \pm 0.14) R_\oplus$ and receiving $32^{+6}_{-4}\%$ the insolation received by the Earth. A re-analysis by Torres et al. (2015) refined the radius to $(1.17 \pm 0.08) R_\oplus$ and we use the radius posterior samples from that work as our input to Forecaster. As shown in Figure 2.5, we predict a mass of $1.74^{+1.31}_{-0.60} M_\oplus$, with 59% of the samples lying within the Terrans category. Therefore, in agreement with the discovery paper of Quintana et al. (2014), we also predict that Kepler-186f is most likely a rocky planet.

Kepler-452b was discovered by Jenkins et al. (2015) and was found to have a very similar insolation to that of the Earth, differing by a factor of just $1.10^{+0.29}_{-0.22}$. The reported radius of $1.63^{+0.23}_{-0.20} R_\oplus$ means that Kepler-452b would be unlikely to be rocky using the
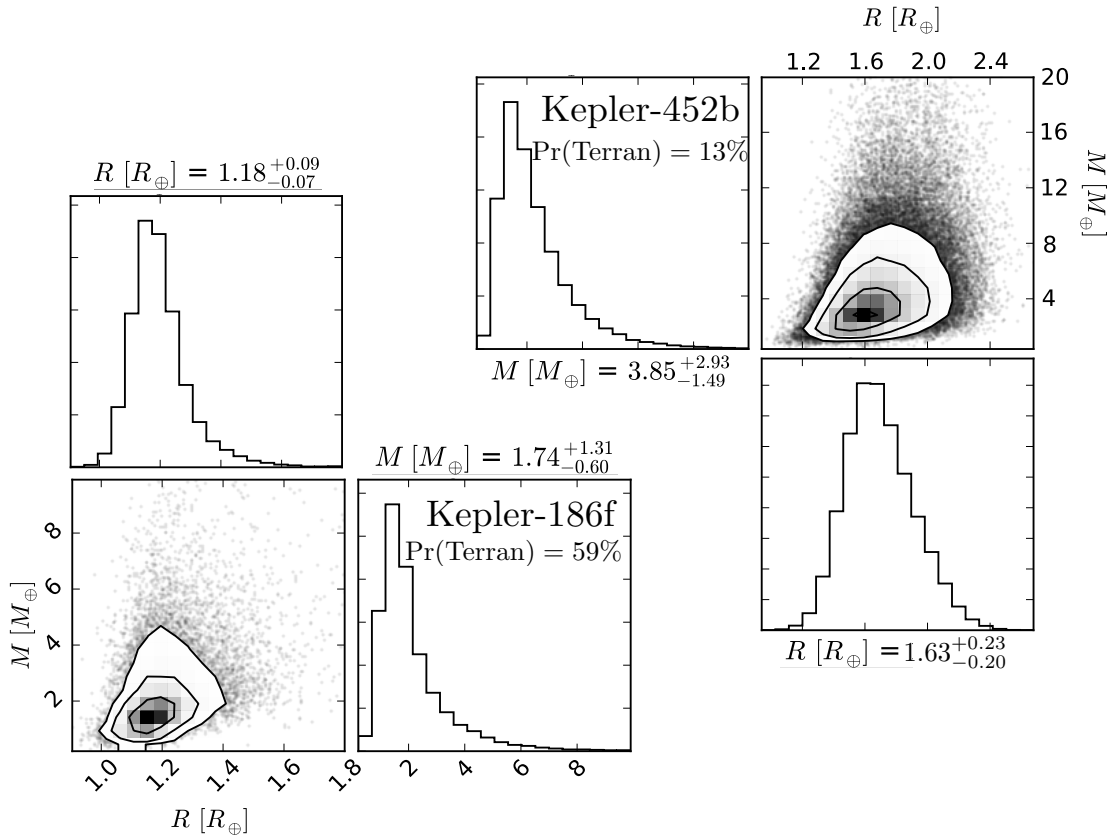
Figure 2.5 Posterior distributions of the radius (measured) and mass (forecasted) of two habitable-zone small planets, with predictions produced by `Forecaster` (triangle plots generated using corner.py).

definition resulting from the analysis of Rogers (2015). Using the reported radius with `Forecaster` predicts that $M = 3.9^{+2.9}_{-1.5} M_\oplus$, with only 13% of samples lying within the Terran worlds (see Figure 2.5). Therefore, in contrast to the discovery paper of Jenkins et al. (2015), we predict that Kepler-452b is unlikely to be a rocky planet.

2.6 Discussion

In this work, we have developed a new package, called `Forecaster`, to predict the mass (or radius) of an object based upon the radius (or mass). Our code uses a new probabilistic mass-radius relation which has been conditioned upon the masses and radii of 316 objects spanning from dwarf planets to late-type stars. Aside from enabling forecasting, this exercise naturally performs classification of the observed population, since we fit for the transitional points. Since the observed population has been classified in this way, future objects can also be probabilistically classified too, which is another feature of `Forecaster`.

As discussed in Section 2.1, expected applications may include a newly discovered transiting planet candidate could have its mass forecasted in order to estimate the detectability with radial velocities. Vice versa, a newly discovered planet found via radial velocities may be considered for transit follow-up and our code can predict the detectability given the present constraints. Another example might be to forecast the scale height of a small planet found by *TESS* for atmospheric follow-up with *JWST*, where `Forecaster` would also calculate the probability of the object being a Terran world.

The classification aspect of our work, which is essentially a free by-product of our approach, provides some interesting insights:

▶ There is no discernible change in the MR relation from Jupiter to brown dwarfs. Brown dwarfs are merely high-mass planets, when classified using their size and mass.

▶ There is no discernible change in the MR relation from dwarf planets to the Earth.

Dwarf planets are merely low-mass planets, when classified using their size and mass.

▶ The transition from Neptunians to Jovians occurs at $0.414^{-0.057}_{-0.065}$ M_J, meaning that Saturn is close to being the largest occurring Neptunian world. This is the first empirical inference of this divide.

▶ The transition from Terrans to Jovians occurs at $2.04^{+0.66}_{-0.59}$ $M_\oplus$, meaning that the Earth is close to being the largest occurring solid world. Rocky "Super-Earths", then, can be argued to be a fictional category.

This latter point may seem remarkable given that "Super-Earths" have become part of the astronomical lexicon. The large number of 2-10 $M_\oplus$ planets discovered is often cited as evidence that Super-Earths are very common and thus Solar System's makeup is unusual (Haghighipour 2013). However, if the boundary between Terran and Neptunian worlds is shifted down to $2\,M_\oplus$, the Solar System is no longer unusual. Indeed, by our definition, three of the eight Solar System planets are Neptunian worlds, which are the most common type of planet around other Sun-like stars (Foreman-Mackey et al. 2014).

As shown earlier, whilst our value is lower than previous estimates, it is fully compatible with previous estimates from both theory (e.g. see Lopez & Fortney 2014) and independent population studies (e.g. see Rogers 2015). The uncertainty on our inference of this key transition is large ($\sim 33\%$) due to the paucity of objects with $> 3\,\sigma$ precision masses and radii in the Earth-mass regime. Future works could hopefully improve the precision to $\sim 10\%$ by using a larger sample size, which will inevitably be found in the coming years, or by extending our method to include lossier measurements and upper

limits by directly re-fitting the original observations (which was beyond the scope of this work). In any case, these divides are unlikely sharp, with counter-examples such as the $M_\oplus$-mass Neptunian world KOI-314c (Kipping et al. 2015).

A wholly independent line of thinking can also be shown to support the provocative hypothesis that the divide between Terran and Neptunian worlds is much lower than the canonical $10\,M_\oplus$ limit. Recently, Simpson (2016) made a Bayesian argument using the population bias to infer that inhabited, Terran worlds should have radii of $R < 1.2\,R_\oplus$ to 95% confidence. Assuming an Earth-like core-mass fraction, this limit corresponds to $2.0\,M_\oplus$ (Zeng et al. 2016). This is also compatible with our determination and again argues for effectively a paucity of Super-Earths. It may be, then, that the Earth is the Super-Earth we have been looking for all along.

Chapter 3

Application of `Forecaster`

In the previous chapter, I used the hierarchical Bayesian method to fit the masses and radii of exoplanets and built a predictive model called `Forecaster`. In this chapter, I will show two cases of the application of `Forecaster`. From Section 3.1 to Section 3.4, I use `Forecaster` to predict the masses of 7000 exoplanet candidates which have their radii measured with the transit survey *Kepler*. From Section 3.5 to Section 3.7, I briefly discuss the vice versa application, predicting radii from measured masses.

3.1 Forecasted masses for 7000 *Kepler* Objects of Interest

Recent transit surveys have discovered thousands of planetary candidates with directly measured radii, but only a small fraction have measured masses. Planetary mass is

---

[0]This chapter is a reproduction of two papers published by The Monthly Notices of the Road Astronomical Society, which can be found at `https://academic.oup.com/mnras/article/473/2/2753/4209246` and `https://academic.oup.com/mnras/article/475/3/3090/4803965`. The articles have been reformatted for this section.

crucial in assessing the feasibility of numerous observational signatures, such as radial velocities (RVs), atmospheres, moons and rings. In the absence of a direct measurement, a data-driven, probabilistic forecast enables observational planning. So here we compute posterior distributions for the forecasted masses of approximately seven thousand *Kepler* Objects of Interest (KOIs).

We find that mass forecasts are unlikely to improve through more precise planetary radii, with the error budget presently dominated by the intrinsic model uncertainty. Our forecasts identify a couple of dozens KOIs near the Terran-Neptunian divide with particularly large RV semi-amplitudes which could be promising targets to follow up, especially in the near-IR. With several more transit surveys planned in the near-future, the need to quickly forecast observational signatures is likely to grow and the work here provides a template example of such calculations.

3.2 Forecasting KOI Masses

3.2.1 Data Requirements

The principal objective of this work is to compute self-consistent and homogeneous a-posteriori distributions for the predicted mass of each KOI. The predicted mass of each KOI will be solely determined by its radius and the empirical, probabilistic forecasting model of Chen & Kipping (2017). Because of this conditional relationship, masses and radii will certainly be covariant, along with any other derived terms based on these quantities. We therefore aim to derive the joint posteriors for all the parameters of interest, which

will encode any resulting covariances.

To accomplish this goal, we first require posterior distributions for the KOI radii. The transit light curve of each KOI enables a measurement of the planet-to-star radius ratio, p, with the precision depending upon photometric quality, number and duration of observed transit events and modest degeneracies with other covariant terms describing the transit shape (Carter et al. 2008). With the quantity p in-hand, it may be combined with the stellar radius, $R_\star$, to infer the true planetary size, R_P. Therefore, to make progress we need homogenous posterior probability distributions for 1) basic transit parameters of each KOI 2) fundamental stellar properties of each KOI.

3.2.2 Transit Posteriors

Basic transit parameters of each KOI are provided in the NASA Exoplanet Archive (NEA, Akeson et al. 2013), but these are summary statistics rather than full posterior distributions, as required for this work. Fortunately, Rowe & Thompson (2015) provide posterior distributions for almost every KOI, with 100,000 samples for each obtained via a Markov Chain Monte Carlo (MCMC) regression of a Mandel & Agol (2002) light curve model to the *Kepler* photometric time series (we direct the reader to Rowe & Thompson 2015 for details). We downloaded all available posteriors and found 7106 object files.

Except for two KOIs, all of these objects appear in the currently listed NEA database (comprising of 9564 KOIs), with the exceptions being KOI-1168.02 and KOI-1611.02. We deleted these two objects from our sample in what follows, giving us 7104 KOIs. The other 2460 KOIs were not modeled by Rowe & Thompson (2015) and thus are not considered

further in what follows.

3.2.3 Stellar Posteriors

For stellar properties, we again note that summary statistics are available on NEA but posteriors are not directly available. As one of our criteria is a homogenous set of posteriors, and we wish to calculate masses for as many KOIs as possible, inferences for particular subsets of the KOI database are not as useful for this work. Instead, we used the publicly available posteriors from Mathur et al. (2017) who used information such as colors, spectroscopy, and asteroseismology to fit Dartmouth isochrone models (Dotter et al. 2008) for each *Kepler* star, giving 40,000 posterior samples per star.

We attempted to download stellar posteriors for all 7104 KOIs in the Rowe & Thompson (2015) database, but found that posteriors were missing for 93 KOIs spanning 87 stars. These KOIs are flagged with a "1" in Table 3.1 and were not considered further for analysis.

3.2.4 KOI Radii Posteriors

We next combined these distributions together to generate fair realizations of the KOI radii. We do this by consecutively stepping through each row of the Mathur et al. (2017) samples and drawing a random row from the corresponding Rowe & Thompson (2015) posterior samples for p. This is possible because the two posteriors are completely independent and share no covariance.

This process results in 40,000 fair realizations for the radius of each KOI, where we

Table 3.1 Data flags assigned to 7104 KOIs considered in this analysis. "0" denotes no problems, "1" denotes that stellar posteriors were missing and "2" denotes transit posteriors were missing. Only a portion of the table is shown here, the full version is available on GitHub repository chenjj2/forecasts.

KOI	data flag
0001.01	0
0002.01	0
0003.01	0
0004.01	0
0005.01	0
0005.02	2
0006.01	0
0007.01	0
0008.01	0
0009.01	0
0010.01	0
⋮	⋮

report the radii in units of Earth radii ($R_\oplus = 6378.1\,\mathrm{km}$), which are made available in the public repository at this URL (https://github.com/chenjj2/forecasts).

During this process, we found 38 KOIs that could not locate the corresponding MCMC file for the Rowe & Thompson (2015) transit parameters. It is unclear why these were missing but given their relativity small number, we simply flagged them with a "2" in Table 3.1 and did not consider them further for analysis. At this point, we are left with 6973 KOIs for which we have been able to derive radius posteriors, of which 2283 are dispositioned as "CONFIRMED" on NEA, 1665 are "CANDIDATE", and 3025 are "FALSE POSITIVE" (these dispositions are also provided in Table 3.2).

3.2.5 Predicting Masses with `Forecaster`

The next step is to take the radius posterior of each KOI and, row-by-row, predict a corresponding mass with `Forecaster`. We direct the reader to Chen & Kipping (2017) for a full description of `Forecaster`. We also direct the reader to Wolfgang et al. (2016) for their prior and alternative discussion of probabilistic mass-radius relations.

`Forecaster` is fundamentally probabilistic, by which we mean that the model includes intrinsic dispersion in the mass-radius relation to account for additional variance beyond measurement uncertainties. This dispersion represents the variance observed in nature itself.

Another characteristic of `Forecaster` is that it was only trained within a specific (albeit broad) mass range, from $3 \times 10^{-4}\ M_\oplus$ to $3 \times 10^{5}\ M_\oplus$, corresponding to dwarf planets to late-type stars. However, some extreme KOIs have radii which fall outside of the expected corresponding radius range. To enable a homogeneous analysis, we simply extend the first and last part of the broken power-law relation, so that it could cover a semi-infinite interval. In practice, this is only necessary for very large KOIs exceeding a Solar radius, for which the KOI cannot be a planet in any case, and thus these extrapolations simply highlight the unphysical nature of these rare cases.

As discussed earlier, the model was applied to each radius posterior sample for each KOI (278.92 million individual forecasts). It is important to stress that the probabilistic nature of `Forecaster` means that re-running the same script again on the same KOI would lead to a slightly different set of posterior samples for the planetary mass, although they would still of course be fair and representative samples.

The 68.3% credible intervals of the forecasted masses are depicted in upper panel of Figure 3.1 and listed along with the 95.5% credible intervals in Table 3.2.

To perform the analysis, we used the function "Rpost2M" in `Forecaster`. To be clearer, a summary of the steps taken in the function is listed as below. 1. generates a grid of mass in the allowed mass range; 2. calculates the probabilities of the measured radius given each mass in the grid; 3. uses the above probabilities as weights to redraw mass. For more details, we direct the reader to the original `Forecaster` paper (Chen & Kipping 2017).
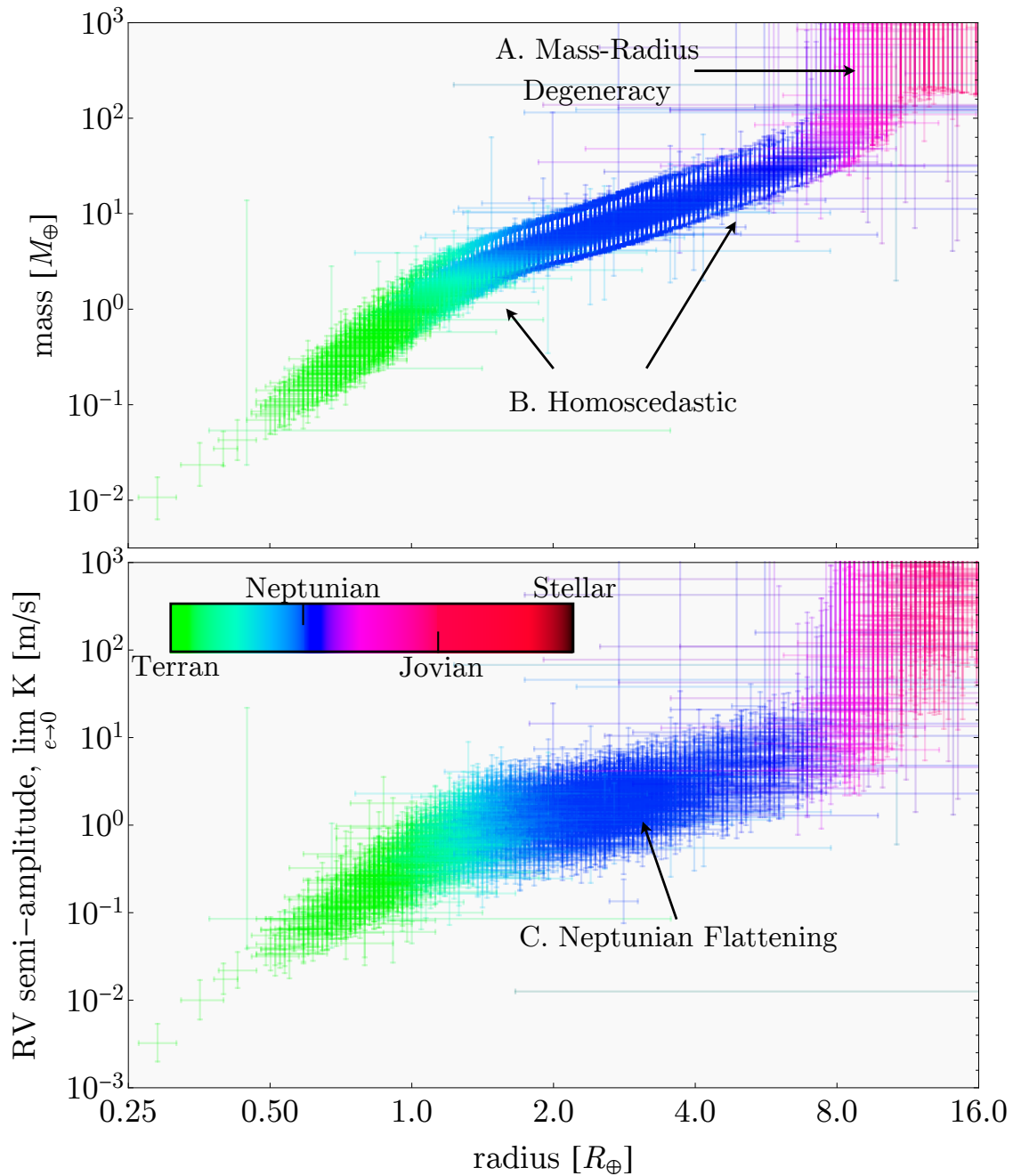
Figure 3.1 Forecasted masses (top) and radial velocity semi-amplitudes (bottom) for a circular orbit, as a function of the observed KOI radius. We here only show objects with a NEA disposition of being either a candidate or validated planet. Error bars depict the 68.3% credible intervals.

Table 3.2 Final predicted properties for 6973 KOIs using Forecaster. Quoted values are the $[-2, -1, 0, +1, +2]\sigma$ credible intervals. First letter of the flag column gives modal planet classification ("T"=Terran, "N"=Neptunian, "J"=Jovian & "S"=Stellar). Second letter of the flag column denotes NEA KOI disposition where "f" is a false-positive, "c" is a candidate and "v" is validated/confirmed. Only a portion of the table is shown here, the full version is available on GitHub repository chenj2/forecasts.

KOI	flag	$\log_{10}(R_P\,[R_\oplus])$	$\log_{10}(M_P\,[M_\oplus])$	$\log_{10}(K\,[\mathrm{m\,s^{-1}}])$	
0001.01	Jv	[1.08,1.10,1.11,1.13,1.15]	[1.83,2.33,3.20,4.14,4.53]	[1.51,2.02,2.88,3.82,4.19]	...
0002.01	Jv	[1.16,1.18,1.20,1.22,1.24]	[1.86,2.23,2.95,4.51,4.67]	[1.44,1.82,2.54,4.08,4.22]	...
0003.01	Nv	[0.66,0.67,0.69,0.70,0.71]	[0.81,1.07,1.32,1.58,1.83]	[0.45,0.71,0.96,1.21,1.46]	...
0004.01	Jc	[0.85,0.92,1.00,1.08,1.17]	[1.34,1.72,3.04,4.13,4.46]	[0.90,1.27,2.55,3.64,3.95]	...
0005.01	Nc	[0.83,0.85,0.87,0.89,0.91]	[1.13,1.40,1.67,2.00,4.28]	[0.64,0.91,1.18,1.51,3.77]	...
0006.01	Sf	[1.21,1.38,1.60,1.91,2.10]	[2.10,4.77,5.05,5.39,5.48]	[1.83,4.46,4.69,4.96,5.03]	...
0007.01	Nv	[0.58,0.60,0.62,0.65,0.67]	[0.70,0.96,1.22,1.47,1.73]	[0.31,0.57,0.82,1.07,1.32]	...
0008.01	Nf	[0.19,0.22,0.26,0.32,0.38]	[0.26,0.44,0.66,0.90,1.17]	[0.06,0.24,0.45,0.70,0.97]	...
0009.01	Jf	[0.99,1.06,1.15,1.27,1.37]	[1.77,2.26,3.27,4.56,4.80]	[1.32,1.80,2.83,4.08,4.24]	...
0010.01	Jv	[1.11,1.14,1.17,1.20,1.23]	[1.85,2.27,3.03,4.23,4.63]	[1.43,1.85,2.61,3.81,4.17]	...
0011.01	Sf	[0.77,0.93,1.18,1.51,1.76]	[1.32,1.86,3.78,4.95,5.24]	[0.97,1.51,3.43,4.52,4.75]	...
0012.01	Jv	[1.09,1.13,1.17,1.21,1.27]	[1.86,2.27,3.04,4.31,4.67]	[1.14,1.56,2.34,3.59,3.91]	...
0013.01	Jv	[1.05,1.12,1.21,1.32,1.50]	[1.87,2.29,3.35,4.71,4.95]	[1.42,1.85,2.93,4.21,4.36]	...
0014.01	Nf	[0.55,0.62,0.69,0.80,0.90]	[0.77,1.05,1.35,1.65,2.04]	[0.29,0.57,0.85,1.14,1.52]	...
0015.01	Sf	[1.45,1.70,2.02,2.28,2.46]	[2.01,4.96,5.40,5.48,5.48]	[1.42,4.40,4.75,4.83,4.86]	...
0016.01	Sf	[0.79,1.01,1.64,1.94,2.13]	[1.36,2.49,5.09,5.42,5.48]	[1.06,2.15,4.70,4.97,5.02]	...
0017.01	Jv	[1.08,1.10,1.13,1.15,1.19]	[1.82,2.32,3.14,4.14,4.56]	[1.44,1.93,2.76,3.75,4.14]	...
0018.01	Jv	[1.12,1.14,1.18,1.21,1.24]	[1.86,2.27,3.01,4.29,4.64]	[1.41,1.81,2.56,3.83,4.15]	...
0019.01	Sf	[0.96,1.12,1.32,1.45,1.56]	[1.77,2.39,4.67,4.88,5.03]	[1.55,2.16,4.40,4.59,4.70]	...
0020.01	Jv	[1.11,1.16,1.23,1.30,1.38]	[1.86,2.25,3.25,4.68,4.82]	[1.43,1.82,2.82,4.20,4.31]	...
...	...	...	...	...	...

3.3 Implication & Limitations

3.3.1 Densities, Gravities and RV Amplitudes

With a joint posterior for both the stellar and the planetary fundamental parameters in hand, we can compute other derived quantities of interest, such as surface gravity, bulk density, and radial velocity (RV) amplitude (these are given in Table 3.2). Once again, we stress that these values should be treated as forecasts and not measurements.

We have computed forecasted masses for all KOIs where available, irrespective of whether the KOI is a known false-positive or confirmed planet. False-positives frequently have extreme radii associated with them, giving rise to anomalous derived parameters. For example, in the case of KOI-5385.01, a known false-positive, we obtain a radius of $R_P = 1010^{+450}_{-500} R_\odot$ (for comparison NEA report $R_P = 62^{+20}_{-5} R_\odot$), implying that the Rowe & Thompson (2015) MCMC chains ultimately diverged to a very high p. We caution that this effect also appears for some KOIs which are not dispositioned as false-positives. For example in the case of KOI-3891.01, which is listed as a candidate on NEA, we obtain $R_P = 126^{+132}_{-89} R_\odot$ (NEA also report an anomalously large radius of $R_P = 0.81^{+0.28}_{-0.52} R_\odot$). Despite this, these cases are straight-forward to identify and essentially represent cases where the MCMC diverged, indicating poor quality light curves or false-positives.

The density forecasts are computed assuming spherical shapes for the planets and are listed in Table 3.2. Density forecasts are particularly useful in cases where one wishes to predict the Roche radii around KOIs for estimating potential ring radii (Zuluaga et al. 2015).

The surface gravity forecasts are also made assuming spherical planets. These estimates are particularly important for climate and atmospheric modeling (Heng & Vogt 2011), as well as predicting the scale height of exoplanetary atmospheres (Seager & Sasselov 2000) when planning observational campaigns.

Finally, the radial velocity amplitudes are computed assuming $i \simeq \pi/2$, which is appropriate for transiting planets, and additionally under the explicit assumption of a circular orbit, such that

$$\lim_{e \to 0} K = \left(\frac{2\pi G}{P} \right)^{1/3} \frac{M_P}{(M_\star + M_P)^{2/3}}. \tag{3.1}$$

We do not assume $M_P \ll M_\star$ and ensure that the uncertainties in the stellar mass are propagated correctly into the calculation of K, although we assume negligible uncertainty in orbital period, P. Circularity is assumed to provide a tight fiducial value rather than marginalizing over poorly constrained eccentricity distributions for these worlds. These predictions should help observers decide which KOIs have potentially detectable signatures from the ground.

We highlight that our predicted masses are calculated homogeneously for all KOIs, regardless to whether the system is known to exhibit strong transit timing variations (TTV) or not. This is relevant since planetary masses detected through the TTV method, rather than the RV method, are subject to distinct selection effects leading to the possibility of ostensibly distinct mass distributions Steffen (2016). Further, the original Chen & Kipping (2017) calibration of Forecaster is dominated by planets with masses measured through

RVs. As this issue continues to be investigated, it will be interesting to test if subsequent TTV detected masses are offset from the predictions in this work.

3.3.2   Observed Patterns

From studying the results, we highlight three noticeable patterns, which are shown in Figure 3.1. Observation A: the masses (and indeed RV predictions) follow a relatively tight relation up to $\simeq 10\,R_\oplus$ after which point the uncertainties greatly increase. Naively, one might assume that the uncertainties should be smaller here, since larger planets give rise to deeper transits and thus we acquire a higher relative signal-to-noise. However, at around a Jupiter-radius, degeneracy pressure leads to an almost flat mass-radius relation all the way until deuterium burning starts in the stars, which occurs at $M = 0.008^{+0.0081}_{-0.0072}\,M_\odot$ (Chen & Kipping 2017). As a result of this degeneracy, mass predictions end up spanning almost the entire Jovian range, leading to much larger credible intervals.

Observation B: we note that the radius-mass diagram in Figure 3.1 shows that the credible interval of the Neptunians and even many of the Terrans have approximately the same width (in logarithmic space), i.e. the logarithmic variance is homoscedastic. Ultimately, these uncertainties are a combination of the measurement error in the observed radii and the intrinsic dispersion in the mass-radius relation. Accordingly, the fact that the uncertainties appear approximately homoscedastic indicates that they are dominated by the intrinsic dispersion term, rather than the measurement errors on R. This, in turn, implies that it will not be possible to forecast noticeably reduced credible intervals in the future by using more precise planetary radii (for example by using more precise

stellar radii e.g. Johnson et al. 2017). Therefore, the only way to *forecast* improved credible intervals would be to train a revised `Forecaster` model, perhaps using additional variables treated as latent in the current version (e.g. insolation). Until that happens though, the forecasted masses for $R \lesssim 10\,R_\oplus$ presented here can be treated as the most precise possible regardless to observational errors.

Observation C: we observe a flattening in the gradient of the predicted RV amplitude as a function of planetary radius for the Neptunian planets (see Figure 3.1). This is somewhat surprising because the RV amplitude is proportional to planetary mass, yet the radius-mass diagram shows a steeper dependency in the Neptunian range. The implication from this is that large Neptunians detected by *Kepler* are predicted to have almost the same RV amplitudes as small, detected Neptunians. Given that $M_P \ll M_\star$ in this regime, and that we've assumed circular orbits, there are only two ways to explain this: i) the larger Neptunians tend to have longer orbital periods, ii) the larger Neptunians tend to have higher mass host stars (or a combination of the two effects). Both effects are also detection biases for *Kepler* (Kipping & Sandford 2016), since long-period planets transit less frequently and high-mass stars tend to be larger, giving rise to smaller transit depths. To investigate which effect dominates, we plot the orbital periods and host star masses as a function of planetary radii for the Neptunian planets in Figure 3.2.

Figure 3.2 reveals that there is indeed an apparent trend between the planetary radius and the orbital period for the *Kepler* Neptunians, which is most easily explained as being due to detection biases. We find no such trend for stellar mass objects, although *Kepler* had a strong selection bias towards Sun-like stars and thus the sample is limited for low-mass
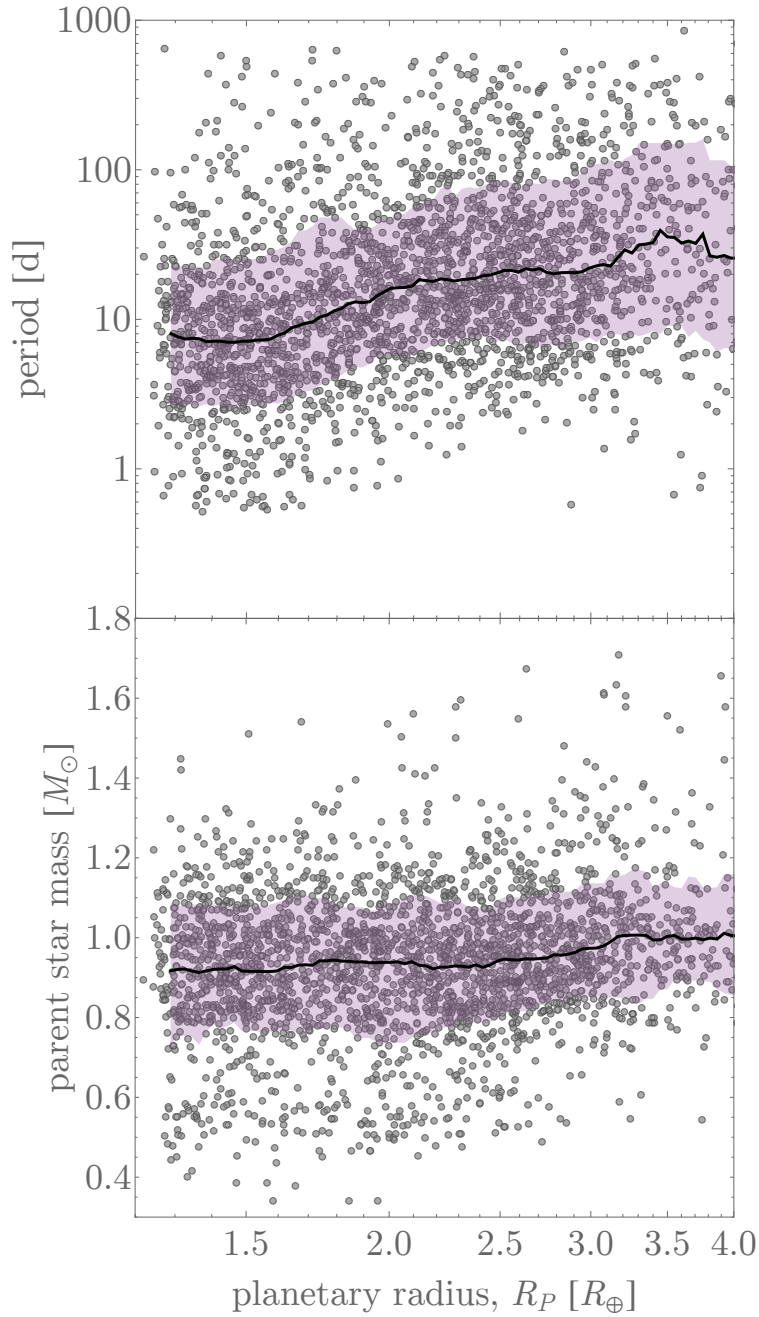
Figure 3.2 Orbital periods (top) and host star masses (bottom) as a function of planetary radius for the 2936 KOIs with modal class probability of being Neptunian and not dispositioned as a false-positive on NEA. The trend between period and radius is most easily explained as being due to detection bias, which in turn gives rise to the RV plateau seen for the Neptunians in Figure 3.1. Median (black line) and 68.3% credible interval (purple) shown using a top-hat smoothing kernel of 0.14 dex.

stars.

3.3.3 Promising Small Planets for RV follow-up

We here demonstrate perhaps the most useful application of this work, identifying promising small planets for RV follow-up. In this work, we have applied `Forecaster` to *Kepler* planets but this should be treated as a demonstration of what could be done with the upcoming *TESS* survey too (Ricker et al. 2015). In particular, the narrow-but-deep nature of *Kepler* combined with the fact that RV amplitudes are enhanced for low-mass stars means that the most favorable targets (in terms of absolute RV amplitude) will be around faint stars. The all-sky nature of *TESS* will lead to brighter targets and here `Forecaster` can clearly play a direct role in identifying promising follow-up targets.

We list 28 promising targets in Table 3.3, where we have down-selected on KOIs not dispositioned as false-positives on NEA, with radii between one half to four times that of the Earth and with unusually high RV amplitude predictions. For the latter criterium, we specifically used $\hat{K} - K_{\mathrm{med}} > 1.64(\hat{K} - K_{\mathrm{lower}})$, where $\hat{K}$ is the median forecast for K, K_{lower} is the 15.9% lower quantile of the forecasted K distribution and K_{med} is the running median of the median forecasts of K with the same window size as used before (note that $1.64\,\sigma = 90\%$). These KOIs are visualized versus the ensemble of small KOIs in Figure 3.3.

As can be seen from Table 3.3, this sample of 28 KOIs all have *Kepler* bandpass magnitudes $K_P > 14$. The $1.35\,R_{\oplus}$ planetary candidate KOI-2119.01 is brightest at $K_P = 14.1$, for which we predict a $\sim 2\,\mathrm{m/s}$ amplitude. This sample of targets also highlights the great potential of near-infrared spectrographs (e.g. Cersullo et al. 2017), where the apparent

Table 3.3 List of planetary candidates with maximum likelihood radii between 0.5 to 4.0 Earth radii and unusually high forecasted RV semi-amplitudes.

KOI	$R_P\,[R_\oplus]$	$K\,[\mathrm{m/s}]$	$M_\star\,[M_\odot]$	$P_P\,[\mathrm{days}]$	K_P
0596.01	$1.29^{+0.12}_{-0.14}$	$1.82^{+1.42}_{-0.62}$	$0.485^{+0.039}_{-0.045}$	1.6827	14.818
0739.01	$1.45^{+0.10}_{-0.10}$	$2.57^{+1.90}_{-0.87}$	$0.533^{+0.033}_{-0.037}$	1.2871	15.488
0936.02	$1.29^{+0.12}_{-0.14}$	$2.29^{+1.78}_{-0.78}$	$0.479^{+0.039}_{-0.049}$	0.8930	15.073
0952.05	$1.38^{+0.13}_{-0.12}$	$2.88^{+2.13}_{-1.02}$	$0.506^{+0.037}_{-0.041}$	0.7430	15.801
0961.01	$0.81^{+0.16}_{-0.12}$	$0.98^{+0.93}_{-0.45}$	$0.144^{+0.030}_{-0.022}$	1.2138	15.920
0961.02	$0.87^{+0.15}_{-0.13}$	$1.78^{+1.77}_{-0.80}$	$0.144^{+0.030}_{-0.022}$	0.4533	15.920
1202.01	$1.23^{+0.09}_{-0.08}$	$1.74^{+1.28}_{-0.56}$	$0.614^{+0.030}_{-0.029}$	0.9283	15.854
1300.01	$1.12^{+0.08}_{-0.10}$	$1.58^{+1.17}_{-0.54}$	$0.540^{+0.034}_{-0.034}$	0.6313	14.285
1367.01	$1.55^{+0.11}_{-0.07}$	$2.95^{+2.06}_{-1.05}$	$0.840^{+0.037}_{-0.039}$	0.5679	15.055
1880.01	$1.35^{+0.20}_{-0.12}$	$2.29^{+1.78}_{-0.81}$	$0.553^{+0.030}_{-0.031}$	1.1512	14.440
2119.01	$1.35^{+0.10}_{-0.06}$	$2.14^{+1.66}_{-0.73}$	$0.849^{+0.039}_{-0.038}$	0.5710	14.098
2250.02	$1.66^{+0.12}_{-0.15}$	$3.24^{+2.39}_{-1.15}$	$0.812^{+0.055}_{-0.059}$	0.6263	15.622
2347.01	$1.00^{+0.07}_{-0.07}$	$1.05^{+0.81}_{-0.36}$	$0.580^{+0.029}_{-0.030}$	0.5880	14.934
2393.02	$1.23^{+0.09}_{-0.08}$	$1.58^{+1.17}_{-0.51}$	$0.790^{+0.031}_{-0.036}$	0.7667	14.903
2409.01	$1.55^{+0.19}_{-0.14}$	$3.09^{+2.28}_{-1.10}$	$0.764^{+0.080}_{-0.065}$	0.5774	14.859
2480.01	$1.35^{+0.16}_{-0.12}$	$2.63^{+2.16}_{-0.89}$	$0.558^{+0.050}_{-0.048}$	0.6668	15.745
2493.01	$1.55^{+0.19}_{-0.14}$	$2.75^{+2.14}_{-0.98}$	$0.844^{+0.066}_{-0.068}$	0.6631	15.304
2699.01	$1.62^{+0.20}_{-0.14}$	$3.16^{+2.33}_{-1.17}$	$0.853^{+0.065}_{-0.077}$	0.5689	15.230
2704.01	$2.40^{+0.62}_{-0.58}$	$7.41^{+6.08}_{-3.15}$	$0.189^{+0.068}_{-0.050}$	4.8712	17.475
2704.02	$1.55^{+0.45}_{-0.37}$	$4.07^{+3.69}_{-1.67}$	$0.189^{+0.068}_{-0.050}$	2.9842	17.475
2735.01	$1.48^{+0.18}_{-0.13}$	$2.69^{+2.09}_{-0.95}$	$0.864^{+0.054}_{-0.063}$	0.5588	15.600
2783.01	$0.65^{+0.05}_{-0.06}$	$0.23^{+0.14}_{-0.08}$	$0.518^{+0.030}_{-0.032}$	0.5269	14.694
2793.02	$1.48^{+0.14}_{-0.16}$	$2.63^{+2.05}_{-0.93}$	$0.456^{+0.043}_{-0.046}$	1.7668	16.283
2817.01	$1.55^{+0.19}_{-0.17}$	$2.82^{+2.19}_{-1.04}$	$0.805^{+0.076}_{-0.074}$	0.6340	15.760
2842.01	$1.95^{+0.24}_{-0.21}$	$5.50^{+4.05}_{-2.18}$	$0.343^{+0.044}_{-0.047}$	1.5654	16.257
2842.03	$1.58^{+0.19}_{-0.17}$	$3.09^{+2.28}_{-1.10}$	$0.343^{+0.044}_{-0.047}$	3.0362	16.257
3119.01	$1.20^{+0.18}_{-0.16}$	$2.14^{+1.75}_{-0.82}$	$0.249^{+0.045}_{-0.036}$	2.1844	16.946
4002.01	$1.38^{+0.13}_{-0.09}$	$2.24^{+1.74}_{-0.76}$	$0.883^{+0.044}_{-0.042}$	0.5242	15.040
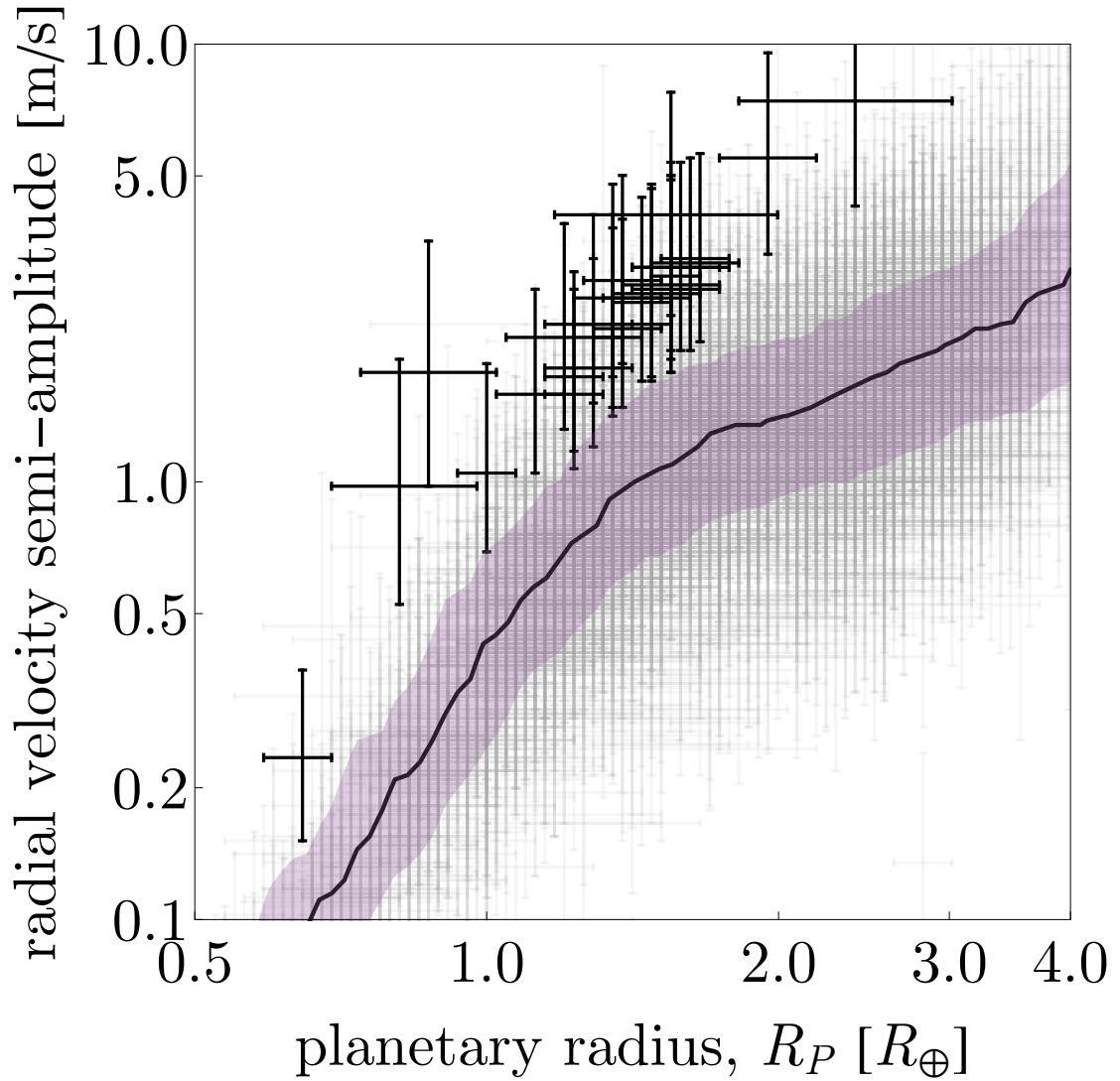
Figure 3.3 Forecasted radial velocity amplitudes for small planets. The colored region contains the 68.3% credible interval of maximum likelihood forecasted K values from a moving window. Points plotted use 68.3% uncertainties and the black points are those more than $1.64\,\sigma$ (90%) above the moving median (black).

magnitude is significantly brighter.

Despite their faintness, the very short-period nature of these objects (as evident from Table 3.3) enables repeated monitoring and potentially easier disentanglement of spurious signals originating from the stars (Pepe et al. 2013; Howard et al. 2013). Thus, despite the challenges, these targets may be worth pursuing for *Kepler* and indeed one can expect this approach to yield far more suitable targets in the *TESS*-era.

3.4   Discussion

In this work, we have used a data-driven and probabilistic mass-radius relation to forecast the masses of approximately seven thousands KOIs with careful attention to correctly propagate parameter covariances and uncertainties (see Wolfgang et al. 2016 and Chen & Kipping 2017). We report 1 and 2σ credible intervals for each KOI in Table 2, including derived parameters such as density, surface gravity and radial velocity semi-amplitude. Full joint posterior samples are made publicly available at this URL(https://github.com/chenjj2/forecasts). We stress that these results should be treated as informed, probabilistic, and model-conditioned *predictions*, but not *measurements*.

`Forecaster` has already seen numerous applications (e.g. see Foreman-Mackey et al. 2016; Obermeier et al. 2016; Dressing et al. 2017; Rodriguez et al. 2017) to individual systems or small subsets, but we believe this work to be the first application to a large (several hundred) ensemble of planets ensuring homogeneous methodology. This work serves as a demonstration of ensemble predictions for planetary masses based on transit-derived radii. In the near future, one can expect the number of planetary candidates

discovered through transits to continue to rapidly grow, exacerbating the demands on follow-up resources, for which planetary mass is frequently a key driver for various observational signatures (e.g. radial velocities, atmospheric scale heights, ring/moon stabilities). For example, *TESS* is expected to discover $\sim 5,000$ transiting planets (Bouma et al. 2017), LSST a further $\sim O[10^3]$-$O[10^4]$ (Lund et al. 2015) and $\sim O[10^4]$ from PLATO (Rauer et al. 2014). In choosing the subset of targets for follow-up, forecasted masses will be beneficial, and thus the analysis provided in this work can be considered as a template for these future surveys too.

As an example, we have shown how this approach can quickly identify small KOIs for which the radial velocity amplitudes are unusually high, increasing the chances for detections (see Section 3.3.3). The combination of the forecasted mass, the low parent star mass, and the short orbital period maximizes K. However, the deep-sky nature of *Kepler* and the bias towards low-mass stars mean that these targets are faint - a problem likely to be resolved with the all-sky surveys of *TESS* and PLATO. Of course, realistic target selections will depend upon more than just signal amplitude, likely factoring in expected noise, planet properties, and other program-specific questions of interest (e.g. see Kipping & Lam 2017 for an example of target selections based on forecasted multiplicity).

Future improvements to the forecasts made in this work are unlikely to result from simply obtaining more precise planetary radii, since the predicted mass uncertainties are observed to be dominated by the intrinsic dispersion inherent to the model itself (see Section 3.3.2). Future work could attempt to build an updated probabilistic model using more data and likely a second dependent variable, such as insolation. Until that time, the

forecasts made here are likely to be the most credible estimates available without direct observations.

3.5 The detectability of known radial velocity planets with the upcoming CHEOPS mission

The *Characterizing Exoplanets Satellite* (CHEOPS) mission is planned for launch next year with a major objective being to search for transits of known RV planets, particularly those orbiting bright stars. Using an empirically calibrated probabilistic mass-radius relation, `Forecaster`, we predict a catalog of 468 planets discovered via radial velocities. Our work aims to assist the CHEOPS team in scheduling efforts and highlights the great value of quantifiable, statistically robust estimates for upcoming exoplanetary missions.

The purpose of this and the following sections are to exemplify the application of `Forecaster` in predicting the radii of planets. As the code execution was performed by Joo Sung Yi, a high school student who I co-advised during the summer of 2017, I won't go into too many details of the methodology and results.

3.6 Forecasting Radii from Masses

3.6.1 Probabilistic predictions with `Forecaster`

A basic requirement of this work is to predict the transit depth, and thus the radius, of each of the few hundred exoplanets discovered through the radial velocity method. A major

challenge facing any attempt to convert masses to radii, or vice versa, is that the mass-radius relation of exoplanets displays intrinsic spread. This spread, likely due to intrinsic variations in composition, chemistry, environment, age, and formation mechanism, means that simple deterministic models are unable to reliably predict the range of plausible radii expected for any given mass.

One solution to this problem is to model the exoplanetary masses and radii with a probabilistic relation. This treats each mass as corresponding to a probability distribution of radius, rather than a single, deterministic estimate. Chen & Kipping (2017) inferred such a relation for masses spanning nine orders-of-magnitude, both providing a comprehensive scale for inversion, as well as calibrating the relation as precisely as possible by utilizing the full dynamic range of data available. An additional key quality of the `Forecaster` model is that the relationship is trained on the actual measurement posteriors of each training sample via a hierarchical Bayesian model, meaning that the measurement errors of these data are propagated into the `Forecaster` predictions.

For these reasons, `Forecaster` is a natural tool for predicting the radii of radial velocity planets.

3.6.2 Accounting for measurement uncertainties

Whilst `Forecaster` accounts for the intrinsic dispersion displayed by nature in the mass-radius relation, as well as the measurement uncertainties of its training set, a third source of error exists that requires accounting for on our end in this work. Specifically, for each RV planet, there exists often sizable uncertainty in the mass of each body.

To account for this, we technically require the a-posteriori probability distribution of each planet's mass, m. For n samples randomly drawn from such a mass posterior, $\Pr(m)$, a robust forecast of the planetary radius can be made by proceeding through the samples, row-by-row, and executing `Forecaster` at each line. The compiled list of predicted radii represents the covariant posterior distribution for the forecasted radius. A similar process is utilized in the reverse-direction application described in Chen & Kipping (2018).

A practical challenge to implement the scheme above is that posterior distributions are rarely made available in papers announcing or studying RV planets[1]. Fortunately, of all the parameters describing the RV model, the RV semi-amplitude (K) posterior rarely exhibits multi-modality or extreme covariance (Ford 2006) and can often, in practice, be reasonably approximated as Gaussian (e.g. see Tuomi 2012 and Hou et al. 2014). By extension, since $K \propto m$ for $m \ll M_\star$, we will assume that the planetary mass can be described as Gaussian, such that $m \sim \mathcal{N}[\mu_m, \sigma_m]$ in what follows in order to make progress.

3.6.3 Approximate form for the posterior distributions

Naively using Gaussians for m can be problematic though for two reasons. First, Gaussians have non-zero probability density at negative values and thus negative masses will occasionally be sampled from a Gaussian distribution. This can be solved by performing a truncation at zero for the Gaussian distribution, preventing the distribution from drawing negative samples. Second, Gaussians are perfectly symmetric yet literature quoted credible intervals for m may include asymmetric uncertainties, e.g. $m = (\mu_m)^{+\sigma_{m+}}_{-\sigma_{m-}}$. But

[1]This is in contrast to our earlier work predicting masses from radii in Chen & Kipping (2018), where transit-derived posteriors are often available.

in practice, we note that none of the reported planetary masses used for this work is asymmetric.

3.6.4 Treating the minimum mass as being equal to the true mass

We assume that the minimum mass derived from radial velocity measurements equals the true mass of the exoplanet, in other words $\sin i \simeq 1$. Whilst one might naively assume $\sin i$ to be, a-priori, isotropically distributed, our work primarily concerns itself with transiting planets. Within the range of inclination angles which lead to transits, we can safely assume $\sin i \simeq 1$ to be a reasonable approximation.

3.7 An Overview of Results

We predicted the posterior distribution for radius as described earlier. To illustrate a specific result, we show a corner plot of the posteriors produced in this work for the planet HD 20794c in Figure 3.4. We mark the median, as well as the 68.3% and 95.5% credible intervals.

Predictions in this work are validated by extracting the RV planets in our sample for which the planetary radius has been directly measured via a transit detection. We compare the predictions versus the observations, which reveals that our predictions are fully compatible with the observed values. Only one of our predicted radii falls outside of the 1σ credible interval (HD 149026b), yet this object falls within the 2σ prediction. This is not statistically surprising, since amongst a sample of nine objects, it should not be
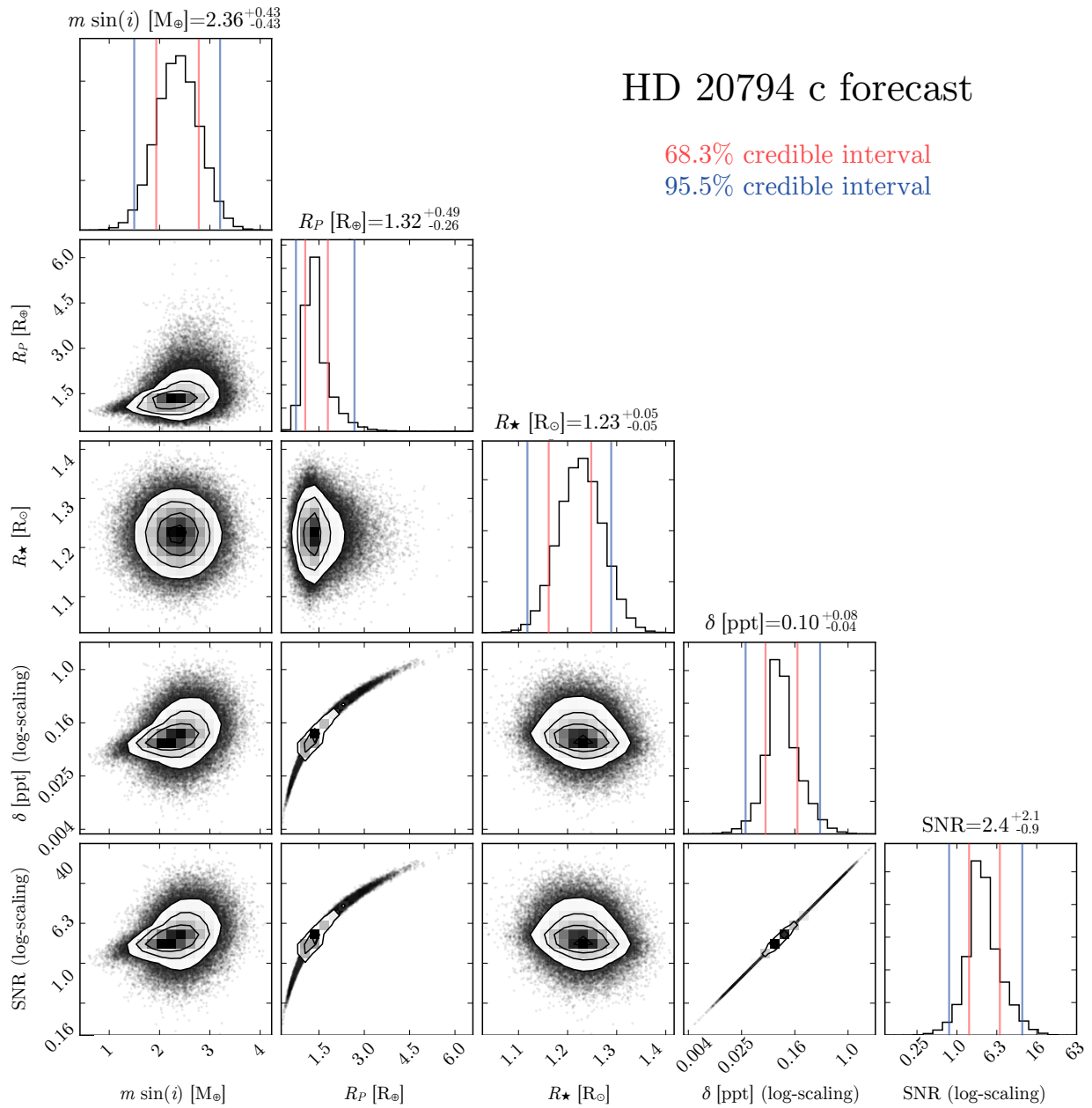
Figure 3.4 Corner plot of our resulting posteriors for HD 20794c an illustrative example. Covariances are evident between the parameters, including between mass and radius.

generally expected that all nine fall within the 68.3% probability interval.

Chapter 4

Constraining an Exoplanet's

Composition

The interior structure of an exoplanet cannot be directly measured with today's observational techniques. However the interior structure plays an important role in determining the evolution and the environment of an exoplanet, such as predicting the habitability of Earth analogs. With measurements of only two bulk parameters: mass and radius, inference of the interior structure with any multi-layer model suffers from a fundamental degeneracy. In this work, we show that although the problem is indeed degenerate, there exist two boundary conditions that enable one to infer the minimum and maximum core radius fraction, CRF_{min} & CRF_{max}. These hold true even for planets with light volatile envelopes, but require the planet to be fully differentiated and that layers denser than

iron are forbidden. In the end, we also propose using the hierarchical Bayesian model to derive the typical fraction of each layer given a sample of similar exoplanets.

4.1 Introduction

Despite recent strides in our ability to characterize exoplanets (Kaltenegger 2017), knowledge regarding the internal structure of distant worlds remains almost entirely lacking (Spiegel et al. 2014; Baraffe et al. 2014). Unlike the search for exoplanetary atmospheres (Burrows 2014), moons (Kipping 2014), or tomography (McTier & Kipping 2018), our remote observations do not have direct access to what we seek to infer - the planet's interior. The habitability of an Earth-like planet, in particular via the likelihood of plate tectonics, is likely to be strongly influenced by the internal structure (Noack et al. 2014) and thus the community is strongly motivated to infer what lies beneath, as part of the broader goal of understanding our own planet's uniqueness.

In general, the only information we have about an exoplanet, which is directly affected by internal structure, is the bulk mass and radius of the planet[1]. Aside from this, we highlight that there are some special cases where additional information about the planetary interior can become available. For example, Kaltenegger et al. (2010) argue that volcanism and planetary outgassing could be detectable using atmospheric characterization techniques. Certain dynamical configurations of planetary systems, such as tidal fixed points (Batygin et al. 2009), can also enable inference of the planetary tidal properties, which in turn constrain internal compositions (see also Kramm et al. 2012). Finally, direct

[1]Quantities such as bulk density and surface gravity are of course derivative of mass and radius

measurements of oblateness may also provide constraints on internal structures (Seager & Hui 2002; Carter & Winn 2010; Zhu et al. 2014).

Whilst there is some hope of identifying outgassing of exoplanets, providing clues to the mantle composition (Kaltenegger et al. 2010), and measuring tidal dissipation constants in special cases (Batygin et al. 2009), full structure inference will likely be limited to indirect methods based on theoretical models. In this approach, one takes the basic observables we do have access to, in particular planetary mass and radius, and compares them to theoretical models in an effort to find families of compatible solutions. Since theoretical models depend on not only masses and radii, but also factors such as chemical compositions (Seager et al. 2007), ultraviolet environments (Lopez & Fortney 2013; Batygin & Stevenson 2013), and ages (Fortney & Nettelmann 2010), the problem is degenerate, in a general sense.

Since we do not have direct access to the interior of exoplanets, the interior structure is generally modeled by assuming several key chemical constituents. In the case of solid exoplanets, extrapolation from the Solar System implies that they should be comprised of three primary chemical ingredients, namely water, H_2O, enstatite, $MgSiO_3$, and iron, Fe. If we assume that the planet is not young and has thus become fully differentiated, the equations of state of these three layers can be solved to provide theoretical estimates of the mass and radius of solid bodies (Zeng & Sasselov 2013). A fourth layer describing a light volatile envelope can be placed on top to capture the behavior of mini-Neptunes, where the light envelope is assumed to have negligible relative mass and thus only affects the bulk radius but not the mass (Kipping et al. 2013; Wolfgang et al. 2016).

These four constituents can be combined in multiple ways to re-create the same combination of mass and radius. Even in the absence of a volatile envelope this degeneracy persists, leading to the common use of ternary diagrams to illustrate their symplectic yet degenerate loci of solutions (e.g. see Seager et al. 2007). This degeneracy is a major barrier to inferring the unique solution for the planetary interior, leading authors to either switch out to simpler and non-degenerate two-layer models (e.g. Zeng et al. 2016) or add a chemical proxy from the parent star (e.g. Dorn et al. 2017) to break the degeneracy. Whilst these are both certainly promising avenues for tackling interior inference, in this work we focus on a third approach based on boundary conditions.

The possibility of exploiting boundary conditions was first highlighted in Kipping et al. (2013), where the authors focused on the concept of "minimum atmospheric height". The method works by first predicting the maximum allowed radius of a planet without any extended envelope given its measured mass. This atmosphere-less planet is typically assumed to be a pure water/icy body, for which detailed models are widely available. If the observed radius exceeds this maximum limit, some finite volume of atmosphere must sit on top of the planetary interior, and the difference in radii represents the "minimum atmospheric height" (MAH). The approach therefore formally describes a key boundary condition of a general four-layer exoplanet.

In this chapter, we explore the other extreme, asking the question under what conditions would the observed mass and radius definitively prove some finite iron-core to exist, and what is the minimum radius fraction that the core must comprise? Going further, we argue that the maximum core radius fraction is another boundary condition in the

103

problem and thus can be derived to provide a complete bounding box of a planet's core size in a general four-layer model framework.

We introduce the concept of the minimum core size in Section 4.2, as well as our fast parametric model to interpolate the Zeng & Sasselov (2013) grid models. Section 4.3 discusses our approach to inverting the relation to solve for core radius fraction directly, as well as the much more straightforward method for inferring the maximum core size. Finally, we discuss applying hierarchical models on the composition fractions of exoplanets in Section 4.4.

4.2   Boundary Conditions on the Core

4.2.1   Outline and Assumptions

Exoplanets are generally expected to display a diverse range of physical characteristics, owing to their presumably distinct formation mechanisms, chemical environments, and histories, as three examples. Even if the mass and radius of an exoplanet were known to infinite precision, and the body was known to be definitively solid, these two observed parameters are insufficient to provide a unique solution for the relative fractions of water, silicate, and iron typically assumed to represent the major constituents of solid planets (Kipping et al. 2013). In other words, mass and radius alone cannot confidently reveal an exoplanet's CRF or CMF (core radius fraction or core mass fraction, respectively).

As a concrete example, a planet composed of water and iron can have the same mass and radius as an iron-silicon planet, and thus have very different CRFs (as illustrated in
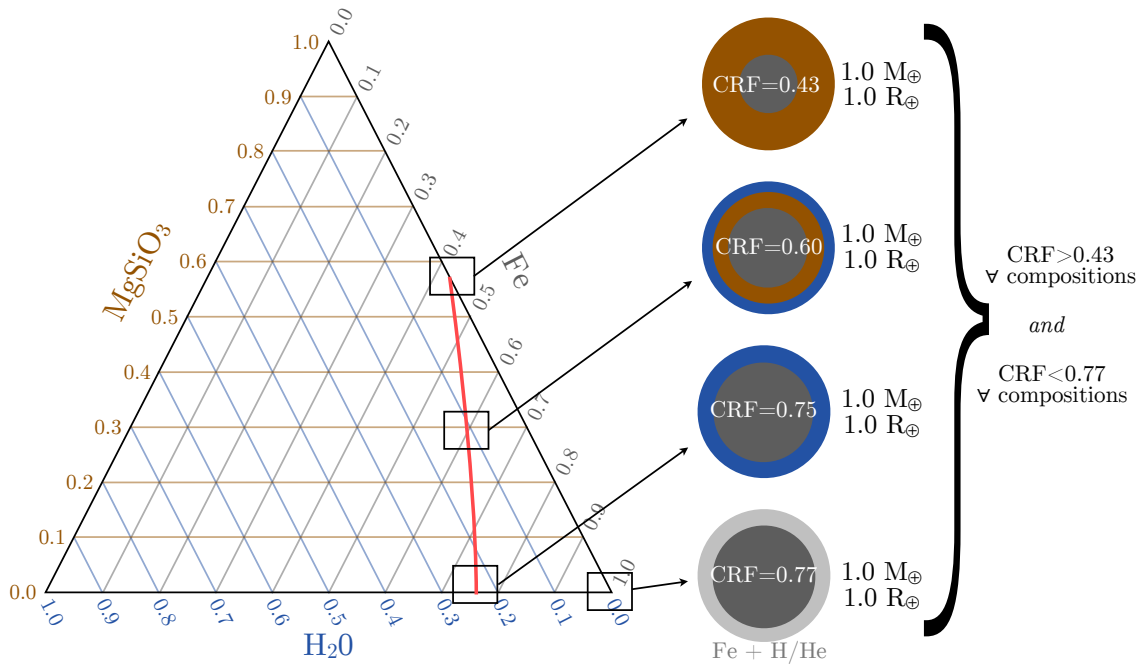
Figure 4.1 Ternary diagram of a three-layer interior structure model for a solid planet. All points along the red line lead to a $M = 1\,M_\odot$ and $R = 1\,R_\odot$ planet. Although indistinguishable from each other with current observations, all points satisfy having a core-radius fraction exceeding 43%, a boundary condition we exploit in this work. The largest iron core size allowed is depicted by the lowest sphere, where the volatile envelope contributes negligible mass.

Figure 4.1).

As touched on in Section 4.1, four-layer theoretical models of solid planets are degenerate for a single mass-radius observation. However, there exists a boundary condition when the composition is pure silicate and iron. At this point, the CRF takes the smallest value out of all possible models, since the second-layer (the mantle) is now as heavy as it can be, being pure silicate (the second densest material). Therefore, for any given mass-radius pair, we can solve for the corresponding CRF of a pure silicate-iron model (which is not a degenerate problem) and define that this CRF must equal the minimum

CRF, CRF_{min}, allowed across all models.

Similarly, another boundary condition we can exploit is to consider the maximum allowed core size. As depicted in Figure 4.1, this occurs when all of the mass is located within a pure iron core, padded by a second layer of a light volatile envelope. The core can't possibly exceed this fractional size else the mass would be incompatible with the observed value.

Note that although we refer to CRF rather than CMF here in what follows, once armed with a CRF, the mass and radius, it is easy to convert back to CMF. Note too that here and throughout what follows, we refer to the CRF strictly in terms of an iron core. Although technically we acknowledge that a water-silicate body could be described as having a finite sized silicate core, that core would not qualify as being a "core" in this work.

Before continuing, we highlight some key assumptions of our model, for the sake of transparency:

- The planet is fully differentiated and is not recently formed.

- The outer volatile envelope has insufficient mass and thus insufficient gravitational pressure to significantly affect the equation-of-state of the inner layers.

- The densest core permitted is that of iron i.e. heavy-element (e.g. uranium) cores are forbidden.

- We have accurate models for a planet's mass and radius given a particular compositional mix.

We stress that what has been described thus far includes the possibility of a light volatile envelope, and is not limited to some special case of two- or three-layer conditions, as discussed earlier.

Under these assumptions, the limiting core radii fractions should be determinable, although violating any of the assumptions listed above would invalidate our argument. An obvious one is that the theoretical models used are invalid or inaccurate, for example because their assumed equations of state are wrong. In this work, we primarily use the Zeng & Sasselov (2013) model but we point out that should a user believe an alternative model to be superior, it is straight-forward to reproduce the methods described in this paper using the model of their preference. The actual existence of a boundary condition remains true.

Another more serious flaw would be if the planetary body in question has a significant mineral fraction based on some heavier element than iron, for example a uranium core. Such a body could feasibly have a significantly smaller core than that derived using our approach. If evidence for such cores emerges in the coming years, we advice against users employing the model described in this work.

The remaining two assumptions, a differentiated, non-young planet and a light volatile envelope, mean that young systems are not suitable and gas giants would not be either. In general then, the model described is expected to be valid for most planets smaller than mini-Neptunes.

4.2.2 A parametric interpolative model for CRF_{min}

For any combination of mass and radius, we need to be able to predict what the corresponding CRF would be for a silicate-iron two-layer model, in order to determine CRF_{min}. We use the models of Zeng & Sasselov (2013), which are made available as a regular grid of theoretical points. Whilst we could simply perform a nearest neighbor look-up, this is unsatisfactory since our precision will be limited to the grid spacing and resulting posteriors would be rasterized to the same grid resolution. Instead, we seek a means to perform an interpolation of the grid.

The first successful literature interpolation of the Zeng & Sasselov (2013) models comes from Kipping et al. (2013), who found that for a specific fixed CRF, each of the various two-layer models of Zeng & Sasselov (2013) is very well-approximated by a seventh-order polynomial of radius with respect to the logarithm of mass.

In this work we used a similar approach by training a suite of seventh-order polynomials with varying CRFs assuming the two-layer iron-silicate models of Zeng & Sasselov (2013), and validated the model with all of the original training data Zeng & Sasselov (2013) but omitting a random single datum each time, serving as a hold-out validation point. The model, `hardCORE`, was made publicly available at this URL (https://github.com/gsuissa/hardCORE).

4.2.3 A parametric model for CRF_{max}

Determining CRF_{max} is far more straight-forward than CRF_{min}. One may simply take the 100% iron mass-radius models, in our case from Zeng & Sasselov (2013), and directly

compute the expected radius of a pure iron planet given an observed mass, $R_{\text{iron}}(M_{\text{obs}})$.

4.3 Solving for the CRF Limits

4.3.1 From forward- to inverse-modeling

Thus far we have described a method to solve for CRF_{max} but not for CRF_{min}. The silicate-iron interpolative model is a forward model in which we begin with knowledge of both the planet's mass and the minimum core radius fraction to compute the corresponding radius. In practice, however, we are interested in the inverse model, where we wish to determine CRF_{min} from the mass and radius.

The nested coefficient structure makes the problem non-linear with respect to CRF_{min}, yet it is one dimensional and found to be unimodal in practice. For these reasons, it is amenable to a large number of possible optimization algorithms, but in what follows we adopted Newton's method, since we are able to directly differentiate our functions thanks to their parametric nature.

Specifically, in our implementation we minimize the following cost function, J, with respect to one degree of freedom, the CRF:

$$J = (R_{\text{Fe}-\text{Si}}(\text{CRF}; M_{\text{obs}}) - R_{\text{obs}})^2, \tag{4.1}$$

where M_{obs} and R_{obs} are the observed mass and radius of the planet, and $R_{\text{Fe}-\text{Si}}(\text{CRF}; M)$

is the radius of a two-layer iron + silicate model with core radius fraction CRF and mass M. The latter function is determined using the smooth parametric interpolation model described in Section 4.2.2. Our inversion algorithm, starts at an initial guess of $CRF_{min} = 0.5$ and then iterates by computing

$$CRF_{i+1} = CRF_i - \frac{J(CRF_i)}{[dJ/dCRF](CRF_i)}. \tag{4.2}$$

To improve speed and stability, we impose a check as to whether R_{obs} is below that of a pure iron planet of mass M_{obs}, in which case we fix $CRF_{min} = 1$, or if the radius exceeds that of an pure silicate planet of mass M_{obs} (where again we use the interpolative model of Kipping et al. 2013), in which case we fix $CRF_{min} = 0$.

4.3.2 The Earth as an example

Let us use the Earth itself as an example of our method. We took a $1\,M_\oplus$ and $1\,R_\oplus$ planet and used the methods described earlier to solve for CRF_{min} and CRF_{max}, giving $CRF_{min} = 0.43$ and $CRF_{max} = 0.77$.

In reality, the Earth is not perfectly described by the Zeng & Sasselov (2013) model and the core in particular is only $\sim$ 80% iron, with nickel and other heavy elements comprising the rest. The mantle-core boundary occurs at a radius of 3480 km relative to the Earth's mean radius of 6371 km, meaning that its actual CRF = 0.55. Accordingly, our CRF bounds correctly bracket the true solution, as expected.

We go further by treating these limits as being bounds on a prior distribution for CRF. Adopting the least informative continuous distribution for a parameter constrained by only two limits corresponds to a uniform distribution. Sampling from this distribution yields a marginalized CRF of $CRF_{marg} = 0.600 \pm 0.098$, which is again fully compatible with the true value.

To test our inversions in a probabilistic sense, we decided to create a mock posterior distribution of an Earth-like planet where $M \sim \mathcal{N}[1.0M_\oplus, 0.01M_\oplus]$ and $R \sim \mathcal{N}[1.0R_\oplus, 0.01R_\oplus]$ (we also apply a truncation to the distributions at zero to prevent negative masses/radii). This is clearly an optimistic assumption but a more detailed investigation of sensitivity for different relative errors is tackled later in Section 4.3.3. Generating 10^5 samples, we inverted each sample as described earlier to produce a posterior for CRF_{min} and CRF_{max}. Our experiment returns near-Gaussian distributions for both terms with the mean and standard deviation given by $CRF_{min} = (0.43 \pm 0.04)$ and $CRF_{max} = (0.7716 \pm 0.0080)$. This establishes that the inversions are stable against perturbations around physical solutions.

As a brief aside, we argued earlier in Section 4.2.1 that the principle of exploiting the boundary condition of theoretical models to infer CRF_{min} does not explicitly require solid planets and works for mini-Neptunes too. To demonstrate this point with a specific example, let us return to the earlier thought experiment of the Earth as a gaseous planet consisting of a solid iron core surrounded by a light H/He envelope. We consider that the mass and radius of the planet remain the same as the real Earth, and that the envelope significantly influences the radius but has negligible mass. Using the 100% iron-model of Zeng & Sasselov (2013) and the corresponding 7th order polynomial interpolation of

Kipping et al. (2013), we estimate that a $1\,M_\oplus$ iron core would have a radius of $0.77\,R_\oplus$ and therefore the remaining $0.23\,R_\oplus$ is given by the light, H/He envelope as depicted earlier in Figure 4.1. Accordingly, such a body would have a core radius fraction exceeding our inferred *minimum* value of (which was $\mathrm{CRF} > 0.43$), **which is expected and self-consistent with our definition of** $\mathrm{CRF_{min}}$**.**

We highlight that the counter-example described above is highly unphysical though; $1\,R_\oplus$ planets are not expected to retain significant volatile envelopes in mature systems, both from a theoretical perspective (Lopez & Fortney 2014) and an observational one (Rogers 2015; Chen & Kipping 2017; Fulton et al. 2017). Planets exceeding $\sim 1.5\,R_\oplus$, on the other hand, are certainly at risk of retaining large envelopes (e.g. see Kipping et al. 2015) and we generally advise against using our model for such large planets which cannot be reasonably assumed to be solid.

4.3.3 Sensitivity analysis for an Earth

A basic and important question to ask is what kind of precisions on a planet's mass and radius lead to meaningful constraints on $\mathrm{CRF_{min}}$. In other words, what is the correspondence we might expect between $\{(\Delta M/M), (\Delta R/R)\}$ and $(\Delta \mathrm{CRF_{min}}/\mathrm{CRF_{min}})$. This is key for designing future observational surveys, where primary science objectives may center around inferring internal compositions. To investigate this, we repeated the retrieval experiment described in Section 4.3.2, but varied the fractional error on mass and radius away from the fixed 1% value previously assumed.

In total, we generated $81^2 = 6561$ experiments, where for each one we generated a new

mock posterior of 10^5 mass-radius samples, which was then converted into a posterior of CRF_{min} and CRF_{max}. For each experiment, we record the standard deviation of the resulting posteriors as ΔCRF_{min} and ΔCRF_{max}. The errors on the mass and radius were independently varied with a fractional error given by 10^x, where x was varied across a regular grid from -4 to 0 in steps of size 0.05, giving 81 grid points in each dimension, and thus 6561 across both. In all experiments, the underlying mass and radius posteriors are generated assuming a mean of $\mu_M = 1\,M_\oplus$ and $\mu_R = 1\,R_\oplus$.
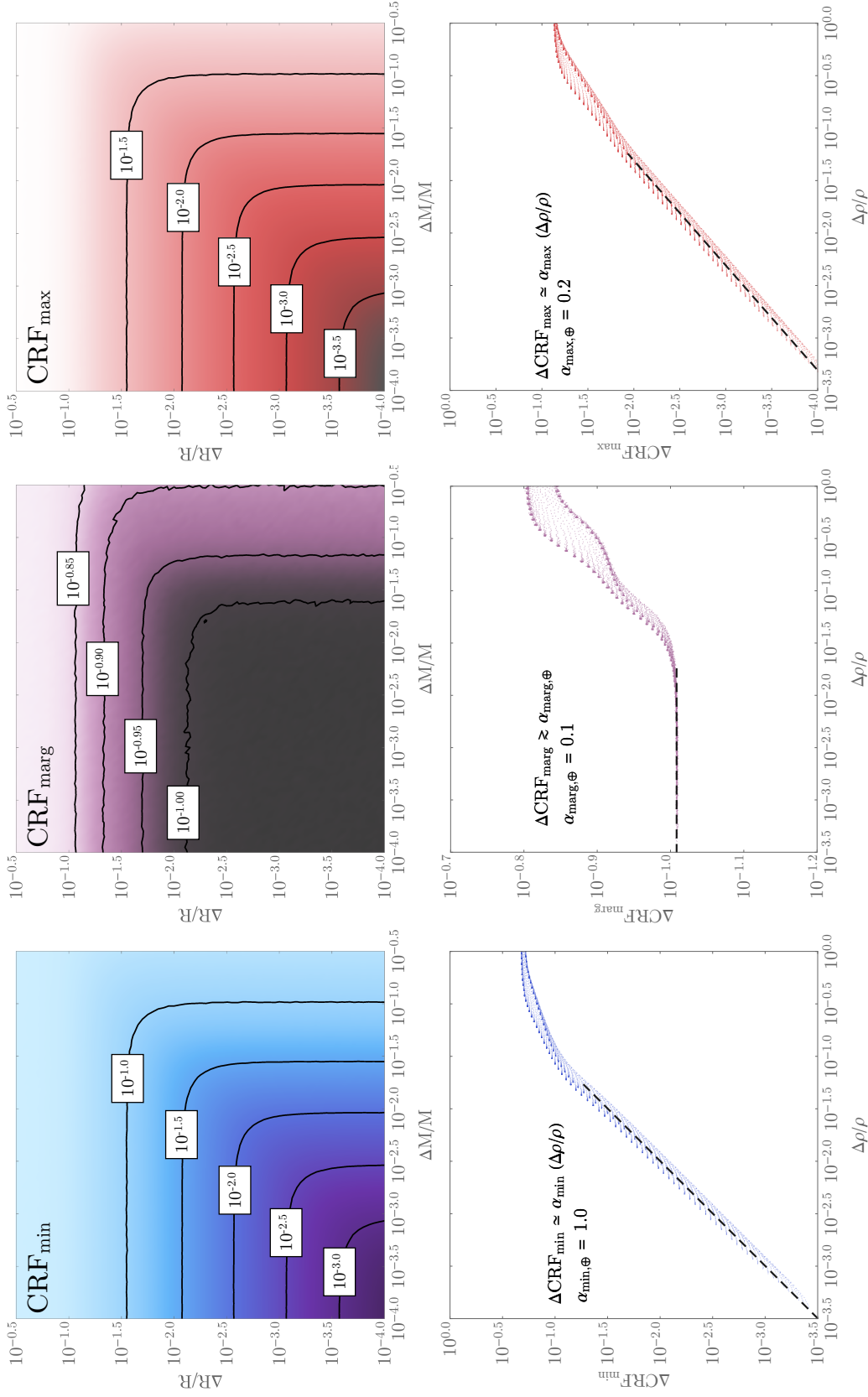
Figure 4.2 Upper: Contour plots depicting the a-posteriori standard deviation on the minimum (left), marginalized (center) and maximum (right) CRF, as a function of the fractional errors on mass and radius. Lower: Reparameterization of the above plots by combining the mass and radius axes into a single density term, demonstrating a strong dependency in all cases.

Figure 4.2 displays the results of this effort for each combination of mass and radius error. Given the shape of darker areas of the color plot, one can see that radius is the dominant constraint, and that for the same fractional error on mass and radius, it is the radius term which mostly strongly constrains CRF_{min}. For example, we find that in order to obtain a precision of 10% on CRF_{min}, we require a measurement on the mass better than 11% and a measurement on the radius better than 3%.

The ratio of these two numbers, close to three-to-one, led us to hypothesize that density is the underlying driving term. This can be seen by calculating error on density for independent mass and radius via

$$\frac{\Delta\rho}{\rho} = \frac{\Delta M}{M} + 3\frac{\Delta R}{R}. \tag{4.3}$$

This is verified in the lower panels of Figure 4.2, where we find that although density doesn't perfectly capture the dependency, it describes the vast majority of the variance. For precise densities ($\lesssim 1\%$), the dependency is strictly linear where we give the coefficients in the panels. This linear dependency breaks down as the errors grow, likely as a result of the truncated normals used to generate the masses and radii, becoming increasingly skewed and the finite support interval (zero to unity) of the CRF itself causing a saturation effect.

The marginalized CRF behaves quite different to the other two. The upper-central panel of Figure 4.2 alone looks fairly consistent with the previous inflated errors. This is to be expected by the very act of marginalization. However, the bottom-central panel does

not exhibit a simple linear dependency, even at precise densities. In contrast, at precise densities the marginalized CRF appears to saturate to ~10%. This implies that no better than 10% precision can ever be obtained on the CRF using just mass and radius alone.

4.3.4 Generalized sensitivity analysis

Thus far, we have assumed a $M = 1\,M_\oplus$, $R = 1\,R_\oplus$ planet. In order to generalize the scalings found, we decided to vary these inputs and repeat the entire process described above. We varied the mass from 1 to 10 Earth masses logarithmically and the CRF from 0.2 to 0.8 uniformly, exploring over 1000 different realizations. For the CRF_{max} term, we find that

$$\Delta\text{CRF}_{\text{max}} \simeq \alpha_{\text{max}}\left(\frac{\Delta\rho}{\rho}\right) \tag{4.4}$$

provides an excellent fit across all simulations, where the best-fitting value of the coefficient term ranged in $0.187 < \alpha_{\text{max}} < 0.237$. The relationship is sufficiently tight that it is reasonable to simply adopt $\alpha_{\text{max}} \simeq 0.2$ as a general rule of thumb. Repeating for the minimum limit on the CRF we find that the function

$$\Delta\text{CRF}_{\text{min}} \simeq \alpha_{\text{min}}\left(\frac{\Delta\rho}{\rho}\right) \tag{4.5}$$

again provides excellent fits, but now the coefficient term varies dramatically in the range of $0.6 < \alpha_{\text{min}} < 4$ across all simulations, or logarithmically in the range of $-0.3 <$

$\log_{10} \alpha_{\min} < 0.66$. The behavior of $\alpha_{\max}$ appears to display a peculiar and periodic dependency with respect to the dependent variables, CRF and mass. We were unable to identify a simple physically motivated relation after substituting for terms such as density and surface gravity, but were able to capture the most of the variance using an empirical approximate formula given by

$$\log_{10} \alpha_{\min} \simeq -0.340985 + 0.0766358 \exp \beta_{\min}, \tag{4.6}$$

where

$$\beta_{\min} \simeq \lfloor 4.934(\text{CRF} - \tfrac{8}{3}\log_{10} M - \tfrac{2}{3}\mathcal{R}[1.34 - 4\log_{10} M]\rfloor$$

$$- 3.28933\mathcal{R}[4(\log_{10} M - 0.335)] - 4.934\text{CRF}$$

$$+ 13.1573 \log_{10} M. \tag{4.7}$$

where $\mathcal{R}$ is a rounding function. We find that the above function has a robust (via the median absolute deviation) estimate of the RMS in the residuals of $\Delta(\log_{10} \alpha_{\min}) = 0.058$.

The marginal CRF plateau, which was 10% in the case of the Earth, also varies in a nontrivial manner, ranging from 2.8% to 15.7% across our suite of simulations. Fortunately, the location of the plateau appears to be directly related to $\alpha_{\min}$, and thus we can use our earlier empirical function to predict this term too with

$$\Delta\text{CRF}_{\text{marg}} \simeq \alpha_{\text{marg}} \tag{4.8}$$

where

$$\log_{10} \alpha_{\text{marg}} \simeq -0.9819 - 0.00583 \exp(-13.6 \log_{10} \alpha_{\text{min}})$$

$$+ 0.321 \log_{10} \alpha_{\text{marg}}. \tag{4.9}$$

We briefly comment that this saturation-behavior can be understood as follows. With imprecise data, the posteriors on the minimum and maximum limits will be broad, and so sampling a point between them will yield an even broader distribution. As the data become more precise, the posterior distributions for CRF_{min} and CRF_{max} converge towards sharp delta functions, but these two limits have no reason or expectation to be on top of one another. Accordingly, when we draw samples between them uniformly, we will still get a broad distribution, albeit not as broad as that obtained when CRF_{min} and CRF_{max} were also broad.

4.4 Discussion

As we have mentioned above, any interior model with more than two layers will suffer from fundamental degeneracies and so cannot be solved using only mass and radius data

because the number of constraints is smaller than the number of degrees of freedom. However, with an ensemble of similar exoplanets, which we assume to have similar interior structures, the degeneracy problem can be partially overcome with a hierarchical Bayesian model as we can use multiple masses and radii measurements to constrain the 'typical' composition fractions for a certain type of exoplanets. In what follows, we propose a simple hierarchical model to derive the composition fractions.

For simplicity, we can assume a three-layer model comprising of an iron core, a silicate mantle, and a water envelope. A more sophisticated interior model can be easily generalized with the same framework. Let's say the radius fraction of the three layers are x_i, x_s, and x_w. As they always sum up to one, we can assume them to follow the Dirichlet distribution $Dir(a_1, a_2, a_3)$. This is the hyper layer of the model. Given the three fractions and the true radius of an exoplanet, we can derive the true mass with a interior structure model, such as that of Zeng & Sasselov (2013). Then we can assume that the mass and radius measurements follow some distribution, which in my cases is well-represented by a normal distribution, related to the planet's true mass and radius. This is the local layer of the model. The full model is shown in Figure 4.3.

We didn't have time to actually run the model as explained above. But we expect that this method should lead to some improvements over deriving boundary conditions only. As is pointed out in Dorn et al. (2017), additional constraints such as relative abundance derived from the host star must be applied to overcome the inherent degeneracy of deriving compositions from bulk mass and radius of exoplanets. Here the additional constraints comes from stacking the masses and radii of multiple planets, instead of from
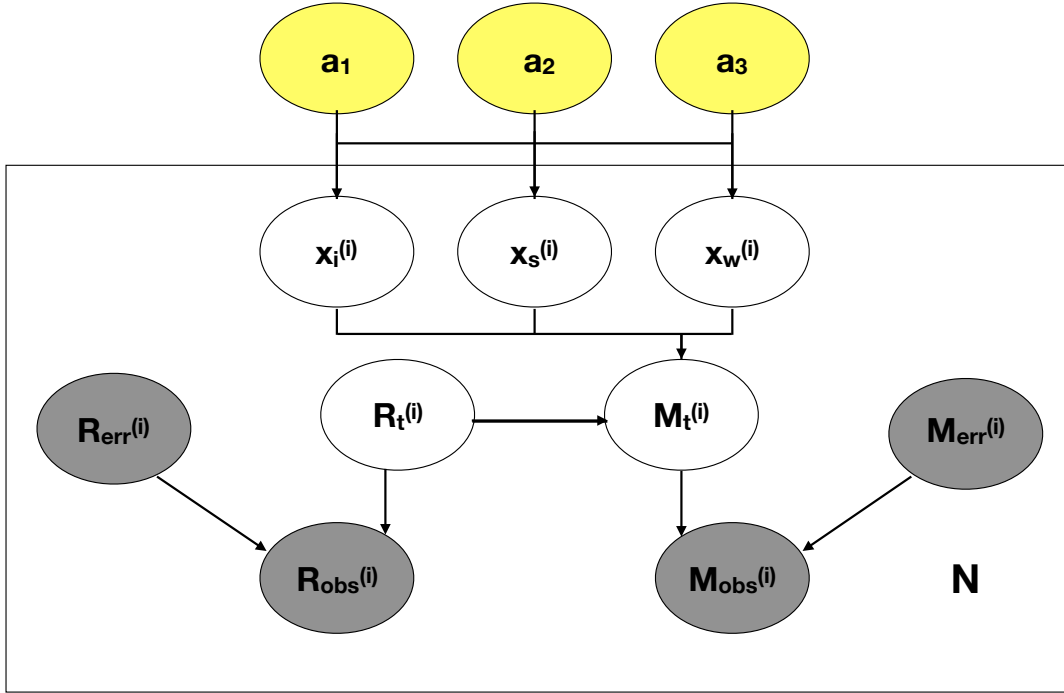
Figure 4.3 Proposed graphical model of the HBM used to infer the composition fractions. Yellow ovals represent hyper-parameters, white represent the true local parameters, and gray represent data inputs. All objects on the plate have N members.

a different dimension of the same individual planet. This model, on the other hand, also has its problems. First, there will be individual difference between the composition fractions of any two planets even if they belong to the same category. But this model mainly considers a typical interior structure. Even though the local parameters will represent each individual exoplanet's composition fractions, they are derived based on the assumption that all the samples share similar interior structures. If an exoplanet of extreme interior structure exist, this model might not be able to catch it. In this case, adding information

from another perspective, such as the relative abundance, and making inference on an individual exoplanet should be more helpful. Second, this model requires a spectrum of either the mass and or the radius measurements. Consider the extreme case of multiple planets having the exact same masses and radii, which means there is only one unique mass and radius data point. We still don't know if a possible composition combination is more probable than another one even though there is more information in hand. For these reasons, the above would be another substantial research project that we did not have time to execute but believe it would be a valuable investigation to add to the literature for future work.

Chapter 5

On the Rate of Abiogenesis from a Bayesian Informatics Perspective

The abiogenesis rate (λ) defines how often life spontaneously originates from the environment. This number plays a crucial role in estimating if extraterrestrial life exists and how many inhabited exoplanets there are. However, since the Earth is the only place we know life to exist, previous Bayesian formalisms for the posterior distribution of λ have demonstrated that the posterior distribution is dominated by the prior distribution. Instead of inferring the true λ posterior, we here compute the relative change in λ when new information is provided. By calculating the information gains, we compare three experimental pressures on λ: 1) evidence for an earlier start to life, t_{obs}; 2) constraints on spontaneous abiogenesis from the lab; and 3) an exoplanet survey for bio-signatures.

5.1 Introduction

The number of communicative civilizations in our galaxy is often framed in terms of the famous Drake Equation, described by seven parameters which can be considered to guide the observational inputs needed to make an inference. When first proposed by Frank Drake over half a century ago, all of the terms were essentially unknown (Drake 1977). But today observational constraints exist for the first three terms: the rate of star formation (Robitaille & Whitney 2010), the fraction of stars with planets (Cassan et al. 2012), and the average number of habitable planets per star (Foreman-Mackey et al. 2014; Dressing & Charbonneau 2015). It is therefore timely to consider what type of observations might enable us to best constrain the next term, f_l, the fraction of habitable planets which develop life.

At the time of writing, interstellar panspermia is considered to be an unlikely pathway for planets to develop life (Wesson 2010), which leaves the most likely pathway as abiogenesis: the spontaneous process by which inert matter becomes living things.

The rate of abiogenesis, λ, is surely functionally dependent upon the very specific geophysical, chemical, and environmental conditions of a given world (Scharf & Cronin 2016). For example, it would seem manifestly wrong to assume that frigid Pluto should have the same rate of abiogenesis events as the early Earth. Consequently, one might reasonably ask whether knowledge of the Earth's rate of abiogenesis has any useful merit for predicting the prevalence of life elsewhere? Whilst this is certainly a reasonable and valid concern, we are frankly left with no choice but to operate under the assumption that the Earth is not an outlier but a sample drawn from a distribution if we wish to

123

make any meaningful progress from a statistical perspective. This concern can at least be somewhat mitigated by assuming that the Earth's rate of abiogenesis is a representative sample drawn for a particular subset of planet types - namely "Earth-like" planets. The precise definition of what constitutes a member of this subset is an in-depth discussion in itself (Tasker et al. 2017), but let us continue under the assumption that such a definition can be expressed, whatever it may be.

There are no other known examples of life beyond the Earth. Further, all known extant and extinct forms of life are generally thought to share a universal common ancestry (UCA), an idea which was first suggested by Darwin (1968) and today forms a central pillar of modern evolutionary theory (Sober 2008). We are therefore faced with the challenge of trying to infer the rate of abiogenesis from a single datum. Further, we cannot exclude the possibility that multiple abiogenesis events occurred or are indeed in the process of occurring, merely that at least one event transpired which led to a planet-dominating tree of life. A statistic of one is difficult to work with at the best of times, but the situation is exacerbated by the anthropic bias that we would not be here had at least one event not been successful.

A critical piece of observational evidence guiding these discussions is that life appears to have emerged on the Earth over 3.5 billion years ago (Schopf et al. 2007), and thus relatively early.

A significant advancement in the statistical treatment was presented in Spiegel & Turner (2012), who cast the problem in a Bayesian framework. As we show later in a re-derivation of their formalism, this naturally accounts for the selection bias imposed by

the fact that our existence is conditioned upon a success. A key conclusion of Spiegel & Turner (2012) was that the probability distribution of λ conditioned upon the observation of life's early emergence, known as the posterior distribution, is highly sensitive to what prior assumptions (including tacit assumptions) one imposes about how λ is distributed, known as the prior distribution.

In general, the functional form and the corresponding shape parameters describing the prior probability distribution for λ are completely unknown. Even changing between different but ostensibly reasonable uninformative priors (e.g. see Lacki 2016) leads to dramatic changes in the posterior (Spiegel & Turner 2012), given the current data in hand.

Rather than trying to estimate the absolute posterior of λ, which may be simply infeasible with present information, a more achievable goal is to compare how the posterior would be expected to change if additional information were obtained - a distinct approach. By quantifying how much we learn from various hypothetical experiments, the goal here is to guide future experimenters to the strategies that are most likely to constrain λ, and critically *in what way* does each experiment constrain it. Accordingly, in this work, we consider three types of experiments that might be expected to constrain λ:

- An earlier estimate for the time at which life first appeared on the Earth, t_{obs}.

- An upper limit on the rate of abiogenesis, λ_{max} (e.g. from a series of Miller & Urey (1959) type laboratory experiments).

- The unambigious detection or null-detection of life amongst a subset of N Earth-like exoplanets/exomoons.

Observational evidence for an earlier t_{obs} may come from numerous different sources but the details are unimportant for this work - we only assume that the evidence is unambiguous. This means that we do not invoke a probabilistic model where the data itself are uncertain, since the probability distribution describing that uncertainty would be itself unknown.

In a similar vein, an upper limit on λ from laboratory experiments is not a soft probabilistic threshold but assumed to be a hard cut off, since again the softness shape would add additional unknowns into our model. Further, we do not consider the more informative case of an actual instance of spontaneous lab-based abiogenesis in this work. If reproducible, such a detection would be so constraining that it would make the results of our paper moot, and is therefore not worth including here.

Finally, extraterrestrial examples of life are formally (but not necessarily strictly) limited to extrasolar worlds (i.e. not worlds such as Mars), in order to greatly reduce the false-alarm probability that the observed life is in fact related to us via a prior panspermia event. We also assume that it is possible to collect observations that can categorically determine whether an extrasolar body has or does not have recognizable life. In this way, one might limit our definition of λ to be only applicable to events which ultimately lead to recognizable life. For example, it is possible that a body is inhabited by a life form which produces almost no bio-signatures, making remote detections effectively impossible and our λ would not include these more elusive organisms. A final caveat is that the planets/moons in this sample should be not only Earth-like now but Earth-like for geological timescales (we formally adopt 4 Gyr as a representative habitable age).

In each of these three experiments, we can compare how the posterior distribution of λ is expected to change, both qualitatively and quantitatively. Before doing so, it is necessary to first formally introduce our Bayesian framework in Section 5.2. Following this, we describe the results of the three experiments in Section 5.3. Finally, Section 5.4 summarizes our findings and describes the main conclusions of this work, as well as listing the limitations of this exercise.

5.2 Bayesian Model

5.2.1 A uniform rate model for abiogenesis

We follow the prescription of Spiegel & Turner (2012) in adopting the simple yet reasonable assumption that abiogenesis events occur at a uniform rate over time, thus following a Poisson process. Let us denote the Poisson shape parameter as λ, which describes the mean number of abiogenesis events, N, occurring in a fixed time interval, which we set as one Gyr. For example, a rate of $\lambda = 4$ means that, on average, four abiogenesis events occur every Gyr. The probability distribution for the number of events that actually occur is described by a Poisson distribution, with a shape parameter λ; i.e. $N \sim \Pr(\lambda)$. Over a time interval of t then, the expected value for the number of abiogenesis events would be λt. Consequently, the probability of N abiogenesis events having transpired during a time interval t is

$$\Pr(N|\lambda, t) = e^{-\lambda t} \frac{(\lambda t)^N}{N!}. \tag{5.1}$$

5.2.2 Time for life to emerge as an exponential distribution

As discussed in Section 5.1, we do not know how many abiogenesis events have occurred on Earth to date, only that at least one must have occurred over a time interval t, where t represents the interval from now back to when the Earth first became capable of supporting life. The probability of at least one abiogenesis event occurring in this time is unity minus the probability of zero event occurring; or equivalently the probability of life emerging on a planet within a time t is

$$\Pr(N \geq 1|\lambda, t) = 1 - \Pr(N = 0|\lambda, t),$$

$$= 1 - e^{-\lambda t}, \tag{5.2}$$

which is the cumulative density function (CDF) of an exponential distribution. When we consider that this CDF describes the probability that life arose by a time t, it therefore follows that the probability distribution for the time at which life first arose, which we can denote as t_{life}, must be an exponential distribution i.e. $t_{\text{life}} \sim \mathcal{E}(\lambda)$.

5.2.3 Accounting for selection bias

For the moment, let us treat $t_{\rm life}$ as a completely random variable drawn from $\mathcal{E}(\lambda)$ with no other constraints whatsoever on its value (i.e. we will not yet include the paleotological information regarding the earliest evidence for life). Let us consider setting λ to be some low rate, such as $0.01\,\rm Gyr^{-1}$. In such a case, the vast majority of random draws from the the probability distribution describing $t_{\rm life}$ will exceed $\sim 4.5\,\rm Gyr$. Even if the Earth can be considered to be habitable immediately after the hypothesized Moon-forming impact $4.47\,\rm Gyr$ ago (Bottke et al. 2015), clearly draws exceeding this age are incompatible with humanity's existence, else life should not have arisen yet. Accordingly, one might justifiably include this constraint into our statistical treatment by setting the maximum limit on $t_{\rm life}$ to be less than $\sim 4.5\,\rm Gyr$.

More generally, one might reasonably argue that $4.5\,\rm Gyr$ is too optimistic, and that following the Moon-forming impact, it would have taken some time for conditions to be suitable for life - for example the crust probably did not condense until $4.4\,\rm Gyr$ ago (Valley et al. 2014). On the other hand, one might argue that even if conditions were immediately suitable for life, $4.5\,\rm Gyr$ is still too generous a maximum limit, since had life begun just $0.5\,\rm Gyr$ ago, there would have been insufficient time for intelligent observers, such as ourselves, to have evolved. We may absorb these arguments into a single term, τ, which describes the latest time for which life could have arisen on the Earth and yet still be compatible with both the habitable history of our planet, and the time it would take intelligent observers such as ourselves to evolve. The selection bias introduced by τ can be formally encoded into our model by applying a truncation to the PDF for $t_{\rm life}$, such that

$$\Pr(t_{\text{life}}; \tau) = \frac{\lambda e^{-\lambda t_{\text{life}}}}{1 - e^{-\lambda \tau}} \mathbf{1}(t_{\text{life}} \geq \tau). \tag{5.3}$$

5.2.4 Accounting for observation time

Consider that we have some measurement for the existence of the earliest life on Earth which occurs at a time $t = t_{\text{obs}}$. We set $t = 0$ to be the time at which the planet became habitable and thus t_{obs} is perhaps better thought of as the difference between these two times. We set the reference time this way in order to remove a degree of freedom from our model, in contrast to Spiegel & Turner (2012) who leaves this as an extra unknown. With our reference time, τ is directly interpreted as the minimum time it takes for life to evolve from whatever biological entity emerged from the abiogenesis event to an intelligent observer.

The measurement of t_{obs} is our most constraining datum but the emergence of life must in fact predate this time, such that $t_{\text{life}} \leq t_{\text{obs}}$. This constraint can be encoded in our probability framework by evaluating the probability of life emerging before a time t_{obs}:

$$\Pr(t_{\text{life}} \leq t_{\text{obs}}; \tau) = \int_{t_{\text{life}}=0}^{t_{\text{life}}=t_{\text{obs}}} \frac{\lambda e^{-\lambda t_{\text{life}}}}{1 - e^{-\lambda \tau}} \, \mathrm{d}t_{\text{life}}, \tag{5.4}$$

which, after simplification, gives

$$\Pr(t_{\text{life}} \leq t_{\text{obs}}; \tau) = \frac{1 - e^{-\lambda t_{\text{obs}}}}{1 - e^{-\lambda \tau}}. \tag{5.5}$$

5.2.5 Likelihood function for λ

Consider that we want to infer λ, conditioned upon a measurement of t_{obs}, in other words we wish to infer

$$\underbrace{\Pr(\lambda|t_{\text{obs}})}_{\text{posterior}} = \frac{\overbrace{\Pr(t_{\text{obs}}|\lambda)}^{\text{likelihood}} \overbrace{\Pr(\lambda)}^{\text{prior}}}{\underbrace{\Pr(t_{\text{obs}})}_{\text{evidence}}}. \tag{5.6}$$

The likelihood function is given by Equation 5.5, or re-writing in a conditional form:

$$\Pr(t_{\text{obs}}|\lambda) = \frac{1 - e^{-\lambda t_{\text{obs}}}}{1 - e^{-\lambda \tau}} \mathbf{1}(t_{\text{obs}} \geq \tau), \tag{5.7}$$

which is the same result obtained by Spiegel & Turner (2012). This is not surprising given that we begin from the same basic assumption of a uniform rate model for abiogenesis, but we hope our independent take on the derivation more clearly explains it's origins. It is worth noting that the likelihood function has a maximum at $\lambda \to \infty$ and displays asymptotic behavior towards that limit.

5.2.6 Multiplanet likelihood function for λ

Thus far, we have presented a Bayesian model for inferring the λ conditioned upon evidence for life on the Earth by a time t_{obs}. An extension to the above, which was not derived in Spiegel & Turner (2012), is to consider surveying N Earth-like worlds for life and detecting M samples.

If we surveyed planets and moons in the Solar System, there is a plausible causal connection between the bodies via panspermia (Wesson 2010), which would significantly complicate the analysis. Instead, we focus on exoplanets where it can be reasonably assumed that each exoplanetary *system* is an independent sample. Generally, in what follows, we refer to each exoplanet system as simply a planet for brevity, but this can include systems of planets and moons. The other condition is that the surveyed worlds belong to a subset of worlds which share similar properties to the Earth i.e. Earth-like.

One difference from our earlier framework is that each exoplanet will typically have a fairly weak age constraint and an even worse constraint on t_{obs}. This simplifies our analysis though, since we can approximately assume that all planets in the sample share the same age, and life simply arose before that age at any time. We set this fiducial age to be 4 Gyr, comparable to when the Earth became habitable (Valley et al. 2014). We make the further assumption that λ is a common value to all planets surveyed. As discussed in Section 5.1, an assumption of this type is fundamentally necessary to make further progress, but can be considered to be justifiable when the survey sample belong to a class of Earth-like worlds.

The probability that any one of these planets is observed to be a positive detection is

given by $\Pr(t_{obs}|\lambda)$ i.e. the likelihood function defined for the single planet case. Indeed, we can consider the probability of a detection for each planet to be treated as a Bernoulli experiment with $p = \Pr(t_{obs}|\lambda)$. The probability of obtaining M successes from a total of N Bernoulli experiments describes a Binomial distribution and thus we may write that

$$\Pr(N, M|p[t_{obs}, \tau, \lambda]) = (1 - p)^{N-M} p^M \binom{N}{M}. \tag{5.8}$$

In the limit of $N = 1$ and $M = 1$, we get back to the same result for the single-planet case, as expected:

$$\lim_{N \to 1} \Pr(N, M|p[t_{obs}, \tau, \lambda]) = \begin{cases} p & \text{if } M = 1, \\ 1 - p & \text{if } M = 0. \end{cases} \tag{5.9}$$

5.2.7 Choosing priors for λ and τ

A key result of Spiegel & Turner (2012) is that changing the prior distribution on λ has a major effect on its posterior distribution. As discussed in Section 5.1, this makes the goal of an absolute determination of the posterior somewhat unachievable, unless we are confident about our choice of the prior distribution. However, this work is chiefly concerned with how the posterior changes when new information is acquired, seeking the relative differences between hypothetical posteriors rather than the absolute truth. For this reason, the choice of prior is less crucial than before and we elect to adopt a simple objective prior in the form of a log uniform distribution:

$$\Pr(\lambda) = \frac{1}{\lambda} \frac{1}{\log \lambda_{\max} - \log \lambda_{\min}}. \tag{5.10}$$

We highlight that this was one of the candidate priors considered by Spiegel & Turner (2012) too. For $\lambda_{\min}$, we fix it at $10^{-3}\,\text{Gyr}^{-1}$ in the majority of simulations that follow. In contrast, Spiegel & Turner (2012) considered three different values of $10^{-3}\,\text{Gyr}^{-1}$, $10^{-11}\,\text{Gyr}^{-1}$, and $10^{-22}\,\text{Gyr}^{-1}$, which roughly correspond to life occurring once per 200 stars, once in our galaxy, and once in the observable Universe (assuming one Earth-like planet per star). We will return to these other choices later in Section 5.3. The effect of $\lambda_{\max}$ will be investigated in detail in one of our three hypothetical experiments and so we also discuss it later.

The final term we require a prior distribution for is τ. The largest allowed value for this term is well-defined as being the age of the Earth, 4.5 Gyr. The most conservative estimate for the first life on Earth is ~3.5 Gyr ago (Schopf et al. 2007) and thus we might reasonably posit $(3.5 - \sqrt{3.5})$ Gyr as lower estimate for how rapidly life could have evolved to an intelligent observer, giving 1.6 Gyr. We round this down to 1.5 Gyr in what follows giving us a full 3 Gyr plausible range for τ. The least informative prior is a uniform distribution given by

$$\Pr(\tau) = U[1.5, 4.5], \tag{5.11}$$

and this is the distribution adopted in what follows.

5.2.8   Sampling method

From a sampling perspective, our model has relatively compact dimensionality and has only a single data point. In such a case, we are able to employ a highly efficient sampling algorithm known as the "bootstrap filter", which is a class of "particle filter" (Künsch et al. 2013).

To briefly describe the algorithm, first N sets of hyper-parameters $(\Theta_H^1, \Theta_H^2, ..., \Theta_H^N)$ are drawn according to the hyper prior distribution. N sets of local parameters are then also drawn given the hyper-parameters $(\Theta_L^i, \; i = 1, .., N)$. Next, we calculate the likelihood of our data given the parameters $(L^i, \; i = 1, .., N)$. Finally, we resample parameters (Θ_H, Θ_L) with their corresponding likelihoods L as weights.

In order for the final sampled parameters to be well spread over the parameter space, the parameter space needs to be well explored in the first step. This indicates that the method would not work for high dimensional problems, since the required number of samples would be intractable in such a parameter space. In such cases, other sampling methods, like Markov Chain Monte Carlo (Metropolis et al. 1953), would be more suitable. But clearly for a simple model like the one we have, the bootstrap filter sampling method is well-suited and highly efficient.

5.2.9   Using information gain to compare different posteriors

The goal of the paper is to compare how the λ posterior distribution changes with different experimental setups. To evaluate the difference in a quantitative way, we choose to use the Kullback-Leibler divergence (KLD), which is also known as relative entropy (Kullback &

Leibler 1951). It quantifies the entropy change from probability distribution Q to P, where a result of zero implies no difference and non-zero (but always positive) values imply a finite difference, and is given by

$$\text{KLD}(\mathcal{P}\|Q) = \int \mathcal{P}(x) \log \frac{\mathcal{P}(x)}{Q(x)} \, dx, \tag{5.12}$$

where the integrand limits cover the complete supported domain of the functions P and Q. We highlight that the popular Kolmogorov Smirnov (Kolmogorov 1933; Smirnov 1948) or Anderson-Darling (Anderson & Darling 1952) tests are primarily designed for cases when two distributions show significant departures from each other and thus would not be directly applicable here. A test statistic, such as the Kolmogorov-Smirnov distance, could be utilized to quantify differences although we note that this metric only codifies the maximum difference in the CDF between two distributions, and does not fully account for the ensemble of differences occurring across the parameter space. For these reasons, we ultimately concluded that the KLD would be a well-suited tool for quantifying the differences observed in our hypothetical experiments.

Computationally we need to use the discretized version of the equation to calculate the information gain. For this work, we use an R package `entropy` for all the KLD calculations. However, as the results of the paper show, when we update the posterior with different experiments, thus changing the posterior in different ways, a number is not enough to describe the results.

5.3 Results

5.3.1 Experiment 1: Reducing t_{obs}

5.3.1.1 Overview

With our model and objective established, we now describe the results from our hypothetical experiments, starting with reducing t_{obs}. We therefore consider here, what effect would an earlier estimate for the first life on Earth have on our knowledge concerning the rate of abiogenesis events, λ?

The earliest undisputed evidence for life on Earth comes from ~3465 Myr Archean deposits in the Apex Basalt of Western Australia, containing morphotype units which are concluded to be microfossils in Schopf et al. (2007). Given that the Earth formed (4.54 ± 0.05) Gyr ago (Dalrymple 2001), the maximum plausible value we can assign to t_{obs} would be $\simeq 1$ Gyr.

In contrast, the very earliest claim for the first evidence of life extends as far back as 4280 Myr ago (Dodd et al. 2017), from putative fossilized microorganisms in ferruginous sedimentary rocks from the Nuvvuagittuq belt in Quebec, Canada. The study of ancient zircons in Jack Hills, Western Australia, indicates the presence of oceans on the Earth as far back as (4408 ± 8) Myr (Wilde et al. 2001), and so one might argue from these studies that t_{obs} could be as short as ~ 100 Myr.

As evident from the cited literature, this is an active and rapidly developing field and thus it is quite likely that further revisions to t_{obs} may occur in the near future. If the age is revised down by another factor of ten, though, how much does this really teach us

about λ? The intuitive temptation is to assign earlier start dates with evidence that life is not fussy and can start quite easily i.e. a high λ (Allwood 2016). As discussed earlier, this question can be more readily addressed in a Bayesian informatics framework such as the one presented here.

To investigate this, we computed the posterior distributions for several different hypothetical values of $t_{\rm obs}$ log-uniformly spaced between 10^0 Gyr (corresponding to the modern conservative limit) down to 10^{-3} Gyr (a deliberately highly optimistic choice). In each simulation, we fix $\lambda_{\rm min} = 10^{-3}$ Gyr^{-1}, as discussed in Section 5.2.7, but explore four different candidate values for $\lambda_{\rm max}$ of 10^0 Gyr^{-1}, 10^{+1} Gyr^{-1}, 10^{+2} Gyr^{-1}, and 10^{+3} Gyr^{-1}.

5.3.1.2 Qualitative impacts on the λ posterior

In the right panel of Figure 5.1, we present the posterior distributions of λ for all four particular $t_{\rm obs}$ choices when $\lambda_{\rm max} = 10^{+2}$ Gyr^{-1}, to illustrate the qualitative differences. Broadly speaking, it can be seen that as $t_{\rm obs}$ becomes smaller, a higher value of the abiogenesis rate is favored. In all cases, the posterior is a monotonic function consistent with a lower limit constraint on λ rather than a peaked, quasi-Gaussian measurement. This immediately reveals that measurements of $t_{\rm obs}$ alone can only hope to ever place probabilistic lower limits on λ, but never produce what might be considered as a direct measurement.

When $t_{\rm obs} = 10^0$ Gyr (line A), which corresponds to the conservative limit, the posterior stays almost flat after a rise at around $\lambda = 10^0$ Gyr^{-1} (i.e. one event per Gyr). This can be understood by the observation that we can't distinguish between, for example, 10 or 100 events per Gyr, as they are both sufficient to have $t_{\rm obs} = 10^0$ Gyr. Accordingly, the
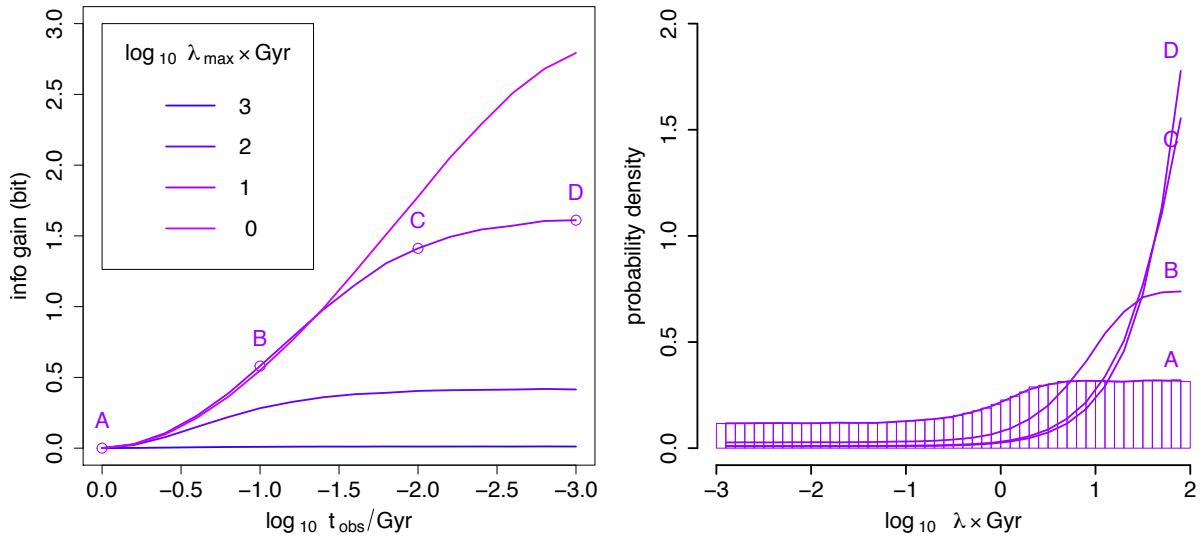
Figure 5.1 The left panel shows information gain calculated by KLD as t_{obs} gets smaller. The right panel gives the detailed posterior distribution of λ for fixed λ_{max} of $10^2\,\text{Gyr}^{-1}$. Line A, B, C, D represents t_{obs} of $10^0\,\text{Gyr}$, $10^{-1}\,\text{Gyr}$, $10^{-2}\,\text{Gyr}$ and $10^{-3}\,\text{Gyr}$ respectively.

posterior largely follows the prior. More generally, one expects an inflection point in the posterior to occur at $\lambda = \lambda_{\text{inflection}} \simeq t_{\text{obs}}^{-1}\,\text{Gyr}^{-1}$.

As t_{obs} gets larger, a higher rate becomes clearly more favorable. The observed posterior morphologies are also responsive to the fact that we set λ_{max} at $10^2\,\text{Gyr}^{-1}$. If say, it is set as high as $10^5\,\text{Gyr}^{-1}$, we should expect all four lines to reach a plateau (which we indeed verified). Accordingly, one can conclude that the actual posterior distribution is highly sensitive to whatever value one adopts for the prior's upper limit, λ_{max}, reinforcing the conclusions of Spiegel & Turner 2012.

5.3.1.3   Quantitative impacts on the λ posterior

The left panel of Figure 5.1 shows the information gain obtained from revising t_{obs} down. Specifically, we set Q in Equation 5.12 to be the posterior distribution of λ when $t_{\text{obs}} = 10^0$ Gyr, corresponding to our current observational constraint. We do not use the prior on λ, since our "starting point" in terms of current knowledge includes a measurement of t_{obs}, and that should be incorporated in our comparison. Accordingly, P in Equation 5.12 is now set to be equal to the derived posteriors at other choices of t_{obs}. As a result of this setup, the information gain between P and Q when $t_{\text{obs}} = 1$ Gyr is, by definition, zero.

The four lines shown broadly reproduce expectation. The information gain increases, resembling a non-linear logistic function, as t_{obs} is revised down. Clearly the choice of λ_{max} again has a significant impact and specifically seems to control a saturation limit in the information gain. This is most clearly seen for the curve describing $\lambda_{\text{max}} = 10^1$ Gyr^{-1}, where there is negligible information gain once t_{obs} drops below $10^{-1.5}$ Gyr. Indeed, if $\lambda_{\text{max}} = 10^0$ Gyr^{-1}, even our current conservative limit on t_{obs} lives on this saturated plateau. In such a circumstance, there is essentially no value in paleontologists continuing to try to revise t_{obs} further back (if their goal is to constrain λ).

The reason behind the saturation can be explained by careful examination of the right panel plot. In the case of line A, the first abiogenesis event happened at $t_{\text{obs}} = 1$ Gyr. Since there is an inflection point in the posterior at $\lambda_{\text{inflection}} = t_{\text{obs}}^{-1}$ Gyr^{-1} (as established earlier), the posterior experiences a steep rise at around once per Gyr. Likewise, line B/C/D should rise at around 10^1 Gyr^{-1}, 10^2 Gyr^{-1}, and 10^3 Gyr^{-1} respectively. However, for the case of line D, since the upper limit is set at $\lambda_{\text{max}} = 10^2$ Gyr^{-1}, many of its morphological changes

are truncated. This upper limit constraint forces line D to perform similarly like line C. If we translate that to information gain, we can see that point D doesn't increase very much from point C, and the whole line will gradually saturate.

5.3.2   Experiment 2: Reducing $\lambda_{\max}$

5.3.2.1   Overview

The second type of experimental pressure we consider on λ is one driven by lab-based experiments. In particular, we consider the thought experiment where a large suite of containers are constructed, within each exists a representative environment of an Earth-like planet, and simply count how often life spontaneously emerges from these containers. This is not meant to be a practical experimental setup but rather a toy example of how one might construct a series of experiments to constrain λ in the lab.

In principle, one or more of these experiments might successfully spawn a new form of life. If this result were reproducible and verifiable, it would provide such dramatic insights into abiogenesis that comparing it to the other hypothetical experiments in this work is somewhat of a moot point. Instead, we take the pessimistic angle that the experiments do not successfully produce a single abiogenesis event (which is of course consistent with current experiments). In such a case, one could reasonably infer an upper limit on λ from the null results and thus we envisage that these experiments return a single new datum for our setup - $\lambda_{\max}$.

It is therefore instructive to compare how obtaining even tighter limits on $\lambda_{\max}$ affects the posterior on λ without changing t_{obs}. To do so, we varied $\lambda_{\max}$ log-uniformly across

the same range as used in Section 5.3.1, from $10^3\,\mathrm{Gyr}^{-1}$ to $10^0\,\mathrm{Gyr}^{-1}$, but now at a much finer resolution. In each simulation we derive the resulting λ posterior and repeat the entire excercise for four different choices of t_{obs}, namely $10^0\,\mathrm{Gyr}$, $10^{-1}\,\mathrm{Gyr}$, $10^{-2}\,\mathrm{Gyr}$, and $10^{-3}\,\mathrm{Gyr}$. Thus, our parameter range directly mirrors the range considered in Section 5.3.1. As before, λ_{min} is fixed at $10^{-3}\,\mathrm{Gyr}^{-1}$.

5.3.2.2 Qualitative impacts on the λ posterior

In the right panel of Figure 5.2, we show four examples of how the λ posterior changes for different λ_{max} values, keeping a fixed $t_{\mathrm{obs}} = 10^{-2}\,\mathrm{Gyr}$. It is clear that the main difference between the updated posteriors (line b, c, d) and the original posterior (histogram/line a) is that the posterior becomes truncated at smaller λ values, corresponding directly to λ_{max}. As a result, each line shows a plateau, which is eroded by the truncation of the sharper λ_{max} constraints, for similar reasons discussed in Section 5.3.1.

5.3.2.3 Quantitative impacts on the λ posterior

The left-panel of Figure 5.2 shows the information gain by varying λ_{max} from $10^3\,\mathrm{Gyr}^{-1}$ to $10^0\,\mathrm{Gyr}^{-1}$, with each of the four lines showing a different assumed t_{obs} value. In comparison to Figure 5.1, the information gains are, in general, higher in this second experiment. This can be understood to be resulting from the truncation effect, which significantly elevates the density of lower λ values in order to maintain normalization. Both experiment 1 and 2 can be seen to provide merely lower limits on λ, rather than strong measurements, since they are both ultimately conditioned upon the same datum. Nevertheless, in a

142
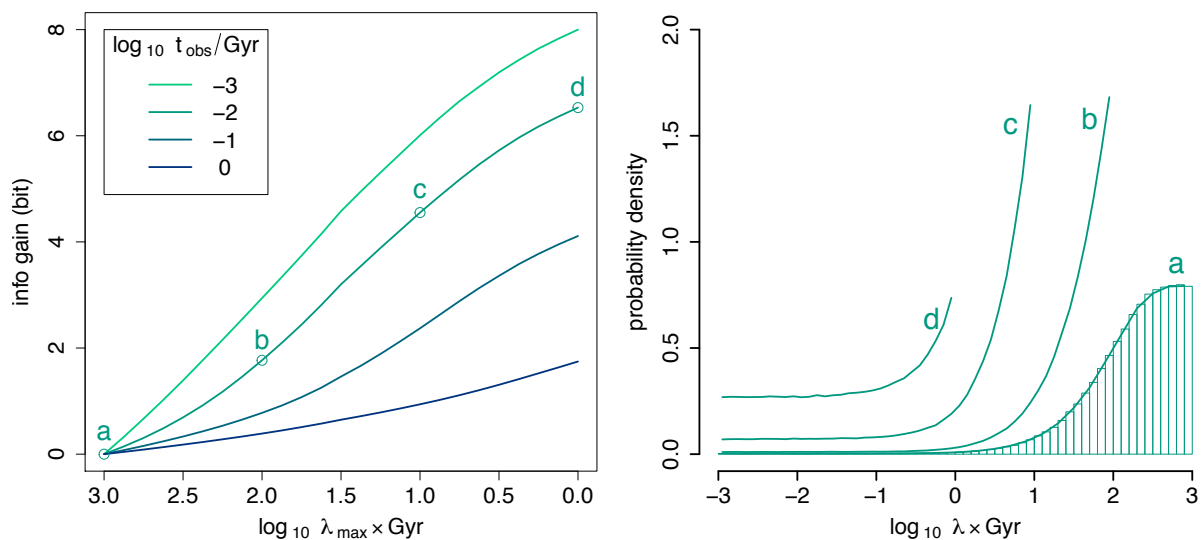
Figure 5.2 The left panel shows information gain calculated by KLD as λ_{max} gets smaller. The right panel gives the detailed posterior distribution of λ for fixed t_{obs} of 10^{-1} Gyr. Line a, b, c, d represents λ_{max} of $10^3\,\mathrm{Gyr}^{-1}$, $10^2\,\mathrm{Gyr}^{-1}$, $10^1\,\mathrm{Gyr}^{-1}$ and $10^0\,\mathrm{Gyr}^{-1}$ respectively.

quantitative sense, our analysis indicates that there is a greater potential for information gain in experiment 2 within the parameter range considered.

This conclusion is reinforced by the fact that none of our simulations in experiment 2 led to a negligible information gain, which was observed in certain cases of experiment 1. Specifically, once λ_{max} can be constrained to less than or equal to $\sim 10^0$-$10^1\,\mathrm{Gyr}^{-1}$, there is very little information gain from revising t_{obs}. In contrast, even the most conservative limit on $t_{obs} = 10^0$ Gyr allows for sizable information gains by revising λ_{max}. Furthermore, the remote possibility of an outright successful abiogenesis event in the lab makes a compelling case that this enterprise is more likely to teach us about abiogenesis than experiment 1.

Although we only show lines spanning up to $\lambda_{\text{max}} = 10^3\,\text{Gyr}^{-1}$, it is worthwhile to consider the behavior at more extreme choices. As can be seen in Figure 5.2, line "a" saturates but it is only the line to do so. This is because in this case λ_{max} (which equals $10^3\,\text{Gyr}$) substantially exceeds $\lambda_{\text{inflection}}$ (which equals $10^1\,\text{Gyr}$), allowing for the saturation behavior to take place. We might therefore hypothesize that if we had chosen a far higher choice of λ_{max}, say $10^{10}\,\text{Gyr}^{-1}$, the posterior would broadly look similar to line "a" except that the truncation would occur much later. Accordingly, we might hypothesize that if we compared the information gain from $\lambda_{\text{max}} = 10^{10}\,\text{Gyr}^{-1}$ to $\lambda_{\text{max}} = 10^9\,\text{Gyr}^{-1}$, there would be very little information gain (compared to, say, going from $10^2\,\text{Gyr}^{-1}$ to $10^1\,\text{Gyr}^{-1}$) since in both cases $\lambda_{\text{max}} \gg \lambda_{\text{inflection}}$. If true, this would mean that if we extended the left panel plot in Figure 5.2 far to the left (i.e. much higher λ_{max}), the curves would be very flat until we start to encroach upon values in the domain of $\lambda_{\text{inflection}}$. To verify this hypothesis, we repeated the experiments up to $\lambda_{\text{max}} = 10^{10}\,\text{Gyr}^{-1}$ and indeed verified that the information gain curves are approximately flat when $\lambda_{\text{max}} \gg \lambda_{\text{inflection}}$.

5.3.3 Future Evidence from Exoplanets

5.3.3.1 Overview

The third experiment we consider is a future astronomical telescope capable of discovering unambiguously whether an observed exoplanet hosts life or not. As with the previous experiments, the specific details of how this is achieved is not important for the following discussion, although a plausible strategy would be to seek atmospheric bio-signatures (Léger et al. 1996). We note that surveying a large number of Earth-like worlds is beyond

the abilities of existing facilities (Seager 2014), but it is not unreasonable to suppose that it should be plausible in the future (Rauscher et al. 2015). If the goal of such an enterprise is to quantify our uniqueness and thus the rate at which life springs forth on Earth-like worlds, the thought experiment described here provides a direct evaluation of how informative such an effort should be expected to be.

Naturally, prior to having conducted this experiment in reality, the number of abiogenesis detections, M, amongst a sample of N exoplanets is unknown, yet the ratio M/N will clearly strongly affect the resulting λ posterior. To account for this, we therefore have two control variables in this experiment, rather than one: the success rate, M/N, and the survey size, N.

5.3.3.2 Yield expectations

It is instructive to first pose the question, what kind of ratio values do we actually expect, based on current information? A ratio which agrees with our naive expectation (whatever that may be) would lead to only a small change in the λ posterior and thus a small degree of information gain. In contrast, a ratio M/N resulting from this hypothetical survey that is in tension with our prior expectation would dramatically change the λ posterior, and thus lead to a large information gain. These considerations provide some initial insights as to why particular values of M/N may not necessarily lead to significant gains in our knowledge of λ. It is therefore interesting and somewhat counter-intuitive to note that a survey of N exoplanets for life may not necessarily lead to any substantial gains in our knowledge about the rate of abiogenesis, depending on what ratio of success is observed.

From the above arguments, it is clear that an observed M/N close to our prior expectation on M/N should represent the minimum in the possible information gain. But what exactly is our prior expectation on M/N? Optimists would say that life starts everywhere and thus we expect $M/N \sim 1$ (Lineweaver & Davis 2002). If so, then detecting a high success rate in an exoplanet survey would actually teach us very little. Like dropping a coin and seeing it fall to the ground under gravity, results which match expectation generally don't teach us as much as if the coin had travelled upwards. In contrast, a pessimistic might say they are convinced that life is rare and thus any detection elsewhere would be highly surprising, much like the coin traveling upwards, thereby teaching us a great deal.

We now turn to estimating what our a-priori expected M/N should be. This is of course closely related to our current posterior for λ. However, as we have argued earlier and indeed as concluded by Spiegel & Turner (2012), an absolute inference of λ is not possible unless we know what the correct prior should be. Although we have investigated the effect of varying $\lambda_{\max}$ and this may be constrainable via experiments (see Section 5.3.2), it is unclear how one should assign $\lambda_{\min}$ at this time (of course another problem is the shape of the prior itself, but we leave that aside at this time).

Using the fiducial choice of $\lambda_{\min} = 10^{-3}\,\mathrm{Gyr}^{-1}$, which has been used in both Sections 5.3.1 & 5.3.2, our expectation is that M/N is generally quite high. We demonstrate this in the left panel of Figure 5.3, where we compute the average number of detections expected using our earlier λ posteriors on $5\,\mathrm{Gyr}$ old Earth-like planets (by Monte Carlo experiments drawing Bernoulli trials using the probability defined in Equation 5.2). For almost any choice of t_{obs} or $\lambda_{\max}$ (within the ranges used throughout this paper), one

146

can see that M/N is expected to be high. Accordingly, if we set $\lambda_{\min} = 10^{-3}\,\text{Gyr}^{-1}$ as before, experiments where we set $M/N \sim 1$ will tend to yield minimal gains in information content on λ. We highlight this point carefully due to its somewhat counter-intuitive consequences.
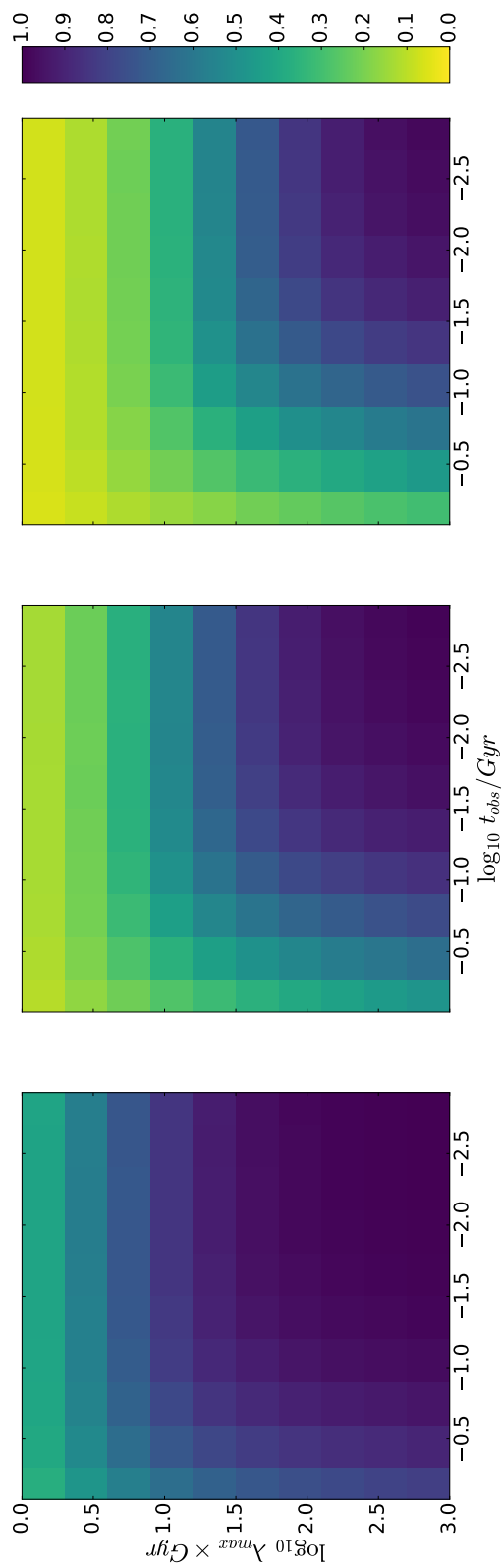
Figure 5.3 The three plots show the predicted ratio of planets with lives given the situation on the Earth. From left to right panel, $\lambda_{\min} = 10^{-3}$ Gyr^{-1}, 10^{-11} Gyr^{-1}, and 10^{-22} Gyr^{-1}. In each panel, we vary t_{obs} (x-axis) and $\lambda_{\max}$ (y-axis).

As Figure 5.3 makes clear, changing $\lambda_{\min}$ leads to a significant decrease in the expected success yield. For these reasons, the experiment 3 information gains presented here are highly sensitive to the assumed values of $\lambda_{\min}$. We note that this was not the case in experiments 1 and 2, where we verified that varying $\lambda_{\min}$ to $10^{-11}\,\mathrm{Gyr}^{-1}$ or $10^{-22}\,\mathrm{Gyr}^{-1}$ did not significantly impact the shape of the posteriors and only slightly affected the scaling of the information gain plots. For this reason, in what follows, we limit our discussion to being largely a qualitative one.

5.3.3.3 Qualitative impacts on the λ posterior and information gain

In the right panel of Figure 5.4, the dark green lines show four example posteriors for λ at $N = 10$, 50, 200, and 1000. The dark green lines all assume $M/N = 30\%$. If 30% of the planets harbor life, this implies a λ of $0.3/5\,\mathrm{Gyr}^{-1}$, and this is indeed where the posteriors all peak. The peakiness of the posterior naturally increases as it becomes conditioned upon larger samples of data.

The yellow lines show the case of no detections, i.e. $M = 0$, which strongly opposes our prior expectation of a high rate. However, because these are essentially null detections, the posterior peaks at $\lambda_{\min}$ and does not converge to a single quasi-Gaussian shape. For this reason, the posterior does certainly lead to a large information gain but not maximal.

On the other hand, if $M = N$, shown by the dark blue lines, the information gain is minimal. As explained in detail earlier, this results from our high prior expectation of λ anyway, conditioned upon the Earth's early start of life. Adding more data only reinforces this belief, leading to minor gains in information content. Further, as with the $M = 0$ case,
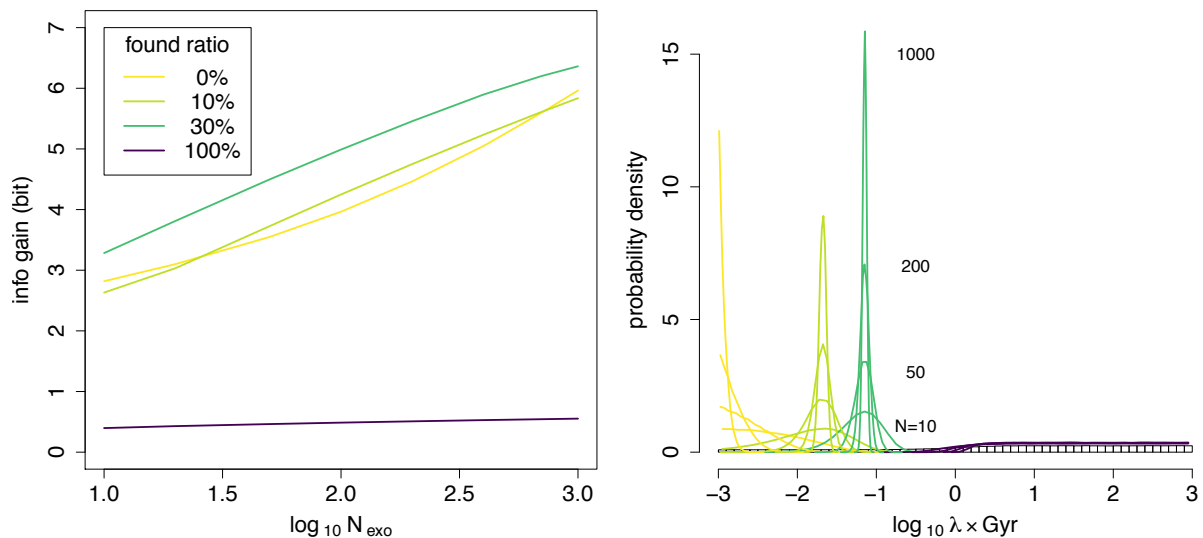
Figure 5.4 The left panel shows information gain calculated as we increase the number of observed exoplanets. The right panel gives the detailed posterior distribution of λ for different ratios.

the posterior peaks at a prior bound, and thus does not represent a converged, constrained datum.

The argument made above does not include effects regarding the sample inclusion though. For example, if all Earth-like planets surveyed show evidence for life, there would be an opportunity to learn about abiogenesis in a different way by gently expanding the sample to sub-optimal worlds and observing when the success rate decreases. This is not formally encoded within our model and represents just one of the ways that a large sample of exo-life detections would lead to large gains in the understanding of abiogenesis, even though not significantly improving our inference of the abiogenesis rate of Earth-like worlds.

5.4 Discussion

In this work we have sought to understand how three different experimental approaches would be expected to inform our knowledge of the rate of abiogenesis, λ, on Earth-like worlds. Our approach adopts the Bayesian formalism of Spiegel & Turner (2012) as a means for deriving the probability distribution of λ conditioned upon some observation of the earliest evidence for life (t_{obs}) and some choice for the prior. As discussed in Spiegel & Turner (2012), the resulting posteriors are highly influenced by the choice of prior, and thus an absolute inference of the posterior is somewhat unachievable. Instead, our paper focuses on the relative gain in information acquired as new experiment evidence is introduced into the problem across a range of plausible input parameters.

Our Bayesian informatics approach seeks to understand exactly how the λ posterior changes in response to new experimental constraints, and whether there are certain experiments or parameter regimes which can be concluded as either negligibly informative, or vice versa, greatly informative.

Our work focuses on three thought experiments, which could be plausibly conducted at present or in the near future: i) paleontological evidence for an earlier start of life; ii) Miller-Urey experiments seeking to create a laboratory abiogenesis event; and iii) a survey of exoplanets for bio-signatures. To quantify the information gained from each experiment, we employ the Kullback-Leibler divergence, or relative entropy, to calculate the difference between the original and the new λ posterior. Additionally, we have performed detailed analyses of the resulting posteriors in an attempt to understand how their morphologies are sculpted by new constraints.

Without knowledge of the correct limits on the prior (or indeed the shape of the prior), it is not possible to unambiguously claim that any of these experiments will always be superior/inferior to the others. Despite this, there are general trends which emerge from our thought experiments. These are briefly summarized in Figure 5.5, and we urge the reader to explore the more detailed results of each found within this paper.

	paleontology (decreases t_{obs})	biology (decreases λ_{max})	astronomy (new samples with lossy t_{obs})
posterior distribution	lower limit on λ only, λ posterior inflects at $\lambda_{inflection}=t_{obs}^{-1}$ Gyr^{-1}, followed by saturation until λ_{max}	λ posterior same as experiment 1 except becomes increasingly truncated with experimental pressure	λ posterior can be constrained from both sides with the mode occurring at $[(M/N)/\tau]$ Gyr^{-1} and a width scaling inversely with N
information gain	saturation in gains once $\lambda_{max}<\lambda_{inflection}$, which occurs for $t_{obs}<\lambda_{max}^{-1}$; for high λ_{max} this can lead to negligible information gain for almost any reasonable choice of t_{obs}	little information gain if $\lambda_{max}\gg\lambda_{inflection}=t_{obs}^{-1}$; otherwise large gains possible	if λ_{min} is large (comparable to $\lambda_{inflection}$), then very little gain when (M/N) is high; otherwise large gains expected
other potential not calculated in this work	definitive evidence for two (or more) independent abiogenesis events on Earth would greatly inform λ	if a reproducible experiment could produce abiogenesis, information gain regarding the origin of life would be very large	aside from the peaked posterior, detections on non-Earth-like planets would yield rich and orthogonal constraints

Figure 5.5 Summary chart of some key general results found in this work.

We highlight a couple of important general trends. First, it is non-intuitive that an exoplanet survey detecting many instances of life can be highly uninformative in certain regimes. This is because for certain choices of the prior, the early start of life on the Earth leads to an expectation that all Earth-like planets are inhabited, thus new detections don't significantly affect the posterior. As highlighted earlier though, clearly such a result could be greatly informative if the sample is extended beyond Earth-like worlds to include more exotic locations, echoing the conclusion of Lenardic & Seales (2018) but for different reasons.

Second, we have ignored the possibility of a successful laboratory abiogenesis in this work and instead assumed they only yield null-results and thus an upper limit on λ. This is motivated by the assumption that the ability to create life in the lab, in a reproducible experiment, would provide an enormous amount of information that would make direct comparisons to the other experiments somewhat meaningless. This is related to the possibility of paleontology/paleongenetics finding unambiguous evidence for a second abiogenesis at early times, which we also did not explicitly consider.

A general trend we highlight is that unless abiogenesis occurs in the lab, the biological and paleontological experiments are conditioned upon the same datum - the earliest evidence for life on Earth - and thus both yield only lower limits on λ. In contrast, exoplanet surveys will, in general, yield peaked posteriors constrained from both sides i.e. a measurement rather than a limit. It is important to emphasize this difference as this is not fully captured by simply comparing the relative entropies between posteriors.

A final point we emphasize is that paleontology in particular has regimes in which

154

essentially no information is gained. For reasons discussed in detail earlier, this occurs when $t_{\mathrm{obs}} \ll \lambda_{\max}^{-1}$. Since there is no agreed value for $\lambda_{\max}$, we are unable to infer what the corresponding time really is, but it is important to note that even a very early start of life can, for some choices of $\lambda_{\max}$, make us no more wiser about abiogenesis.

Being the first effort at a Bayesian informatics analysis of abiogenesis, we acknowledge that there are outstanding issues that we haven't been able to address in detail here. First, we did not conduct a full exploration of the effect of changing $\lambda_{\min}$, largely since we couldn't conceive of a direct empirical way of constraining it. Another issue is that we have assumed λ to follow a universal distribution for all exoplanets and at all times, while of course we should expect that it will change with different environments. This was largely done as a simplifying assumption, following Spiegel & Turner (2012) and is partly mitigated by our assumption that only "Earth-like" worlds are included in an exoplanet survey. This problem might be better solved in the future with hierarchical models, which were not explored here. Finally, we have assumed that experiment 3 is actually feasible - which remains somewhat unclear. Specifically, we assume that life can be unambiguously detected or ruled out on a planet from remote observations. A suggestion for future work would be to change this hard binary flag to a softer probability, again implying a hierarchical model.

Together, we generally find that all of the experiments can certainly yield constraints on λ and in non-overlapping ways. A lab abiogenesis event would be such an informative experiment that even if it could be argued that the information gains from null results are negligible, there is a strong case to conduct the experiment regardless. Similarly, early

starts of life, whilst they may not formally improve λ, do clearly provide information about the conditions upon which life began, and this is not formally encoded in our model. If an actual measurement of λ is desired, rather than a limit, we would argue that the exoplanet survey is the most direct way to infer this. Moreover, the ability to expand to non-Earth-like worlds can probe different conditions and thus offer an orthogonal type of information. Together then, all three experiments deserve our attention and resources in our quest to answer one of the modern science's greatest questions.

Chapter 6

Conclusion

I will give a summary of each piece of work in this thesis, as well as provide some directions for future improvements. I will also discuss the difference between Bayesian approaches and frequentist approaches, and explain why the Bayesian method is adopted in this work.

6.1   Overview of Exoplanets' Probabilistic Mass and Radius Relation

I used the hierarchical Bayesian framework to build a probabilistic power-law relation between the masses and radii of a wide range of objects from dwarf planets to dwarf stars. The power-law, broken into four segments to capture the different relationships of different types of astronomical objects, performs classification naturally by fitting the transition points between different segments.

157

After the publication of this work, we have also seen results coming from other groups showing agreement with our work. Precise radius measurements based on the California-*Kepler* Survey (Fulton et al. 2017), Gaia parallaxes (Fulton & Petigura 2018), and more precise stellar parameters derived from asteroseismology (Van Eylen et al. 2018) showed a gap in the radius distribution and suggested that planets with radius smaller than 1.5 $R_\oplus$ have rocky cores and belong to a different category from planets with radius larger than 2.0 $R_\oplus$, which is consistent with the Terran Neptunian divide at 2.0 ± 0.7 $M_\oplus$, corresponding to $1.23^{+0.44}_{-0.22}$ $R_\oplus$ in our work.

The fitted mass-radius relation can also be used as a forecasting tool to predict mass with radius measurement or vice versa, for objects within the sample range. With the pre-trained relationship, I predict masses for 7000 KOIs with their radii measurements. Such prediction is particularly important because *Kepler* has delivered so many planetary candidates that it is often impractical to schedule follow-up for each object. Finite resources demand prioritization, and one obvious criterium for ranking the objects is whether an observation signature is even expected to be detectable. Moreover, the detectability of an exoplanet, such as radial velocity semi-amplitude, is strongly constrained by its mass. Accordingly, we performed a similar analysis for CHEOPS.

But there are still possible improvements for this mass-radius relation. First, more data and finer errors that can be provided by future observations may reveal additional categories of planets. Also it would be better to take the posteriors of masses and radii measurements as data instead of using point estimates and assuming Gaussian distributions for measurement errors, since distributions are only approximations. Thirdly,

adding another dimension of data, such as metallicity, may result in smaller intrinsic dispersion estimates, and lead to more precise predictions.

6.2   Overview of Exoplanets' Composition

The interior structure of an exoplanet is hidden from direct view yet likely plays a crucial role in influencing the habitability of Earth analogs. Inferences of the interior structure are impeded by a fundamental degeneracy that exists in any model comprising of more than two layers and observations constraining just two bulk parameters: mass and radius. However there exist two boundary conditions that enable one to infer the minimum and maximum core radius fraction, CRF_{min} & CRF_{max}.

Besides inferring the boundary conditions, I also proposed using the hierarchical Bayesian framework to derive the typical composition fractions for an ensemble of planets. I gave an example of such model in the discussion part of that chapter. This framework can be applied with even more sophisticated interior structure models such as a four-layer model. In this framework, we can simultaneously derive the local parameters, which represent the composition fractions of each single planet.

6.3   Overview of Abiogenesis Rate

Life appears to have emerged relatively quickly on the Earth, a fact sometimes used to justify a high rate of spontaneous abiogenesis (λ) among Earth-like worlds. Conditioned upon a single datum - the time of earliest evidence for life (t_{obs}) - previous Bayesian

formalisms for the posterior distribution of λ have demonstrated how inferences are highly sensitive to the priors. Rather than attempting to infer the true λ posterior distribution, we compute the relative change in λ when new experimental/observational evidence is introduced. By simulating posterior distributions and the resulting information gains, we compare three experimental pressures on λ: 1) evidence for an earlier start of life, t_{obs}; 2) constraints on spontaneous abiogenesis from the lab; and 3) an exoplanet survey for bio-signatures.

First, we find that experiments 1 and 2 can only yield lower limits on λ, unlike 3. Second, evidence for an earlier start of life can yield negligible information on λ if $t_{\mathrm{obs}} \ll \lambda_{\max}^{-1}$. Vice versa, experiment 2 is uninformative when $\lambda_{\max} \gg t_{\mathrm{obs}}^{-1}$. Whilst experiment 3 appears to be the most direct means of measuring λ, we highlight that early starts inform us of the conditions of abiogenesis, and that lab experiments could succeed in building new life. Together then, the three experiments are complimentary and we encourage activity in all to solve this grand challenge.

6.4   On Bayesian V.S. Frequentist Methods

The Bayesian approach has been widely adopted in exoplanet studies and proven to be useful. Another approach that is often compared with the Bayesian approach is the frequentist approach. In practice, both can be used for parameter estimations and model comparisons. But these two approaches do stem from different philosophies. While frequentist takes probabilities as frequencies of an event, Bayesian takes probabilities as a degree of certainty about an event. This fundamental difference means that the model

used in this thesis may be confusing from a frequentist perspective.

For example, in building the relationship between masses and radii, we specify the local parameters to be the true mass and radius of each sample. But from a frequentist perspective, the true value should be a fixed number and it wouldn't appear to be anything else. In our hierarchical Bayesian model, the true values are indispensable parts of the model which connect between the hyper parameters and the measurement data, and make it possible for us to address the intrinsic dispersion of the relation and the measurement error separately in calculations.

The difference may seem even clearer in the context of inferring abiogenesis rate. As in here, we only have one data point, which is the time life first appeared on the Earth. In a frequentist perspective, making statistical inference with one data point might be ridiculous. Actually a frequentist approach requires that the number of parameters be less than the number of data points. However, in a Bayesian perspective, as we can always use some previous information or simply guess a prior distribution for the parameter, we can always derive a posterior distribution, although in this extreme case the posterior is simply dominated by the prior. While the prior distribution is a powerful tool of the Bayesian approach, it is also the reason many people doubt about Bayesian inferences, as almost always there will be some arbitrary choices involved.

Bibliography

2018, NASA Exoplanet Archive

Akaike, H. 1974, IEEE Transactions on Automatic Control, 19, 716

Akeson, R. L., Chen, X., Ciardi, D., Crane, M., Good, J., Harbut, M., Jackson, E., Kane, S. R., Laity, A. C., Leifer, S., Lynn, M., McElroy, D. L., Papin, M., Plavchan, P., Ramírez, S. V., Rey, R., von Braun, K., Wittman, M., Abajian, M., Ali, B., Beichman, C., Beekley, A., Berriman, G. B., Berukoff, S., Bryden, G., Chan, B., Groom, S., Lau, C., Payne, A. N., Regelson, M., Saucedo, M., Schmitz, M., Stauffer, J., Wyatt, P., & Zhang, A. 2013, PASP, 125, 989

Allwood, A. C. 2016, Nature, 537, 500

Almenara, J. M., Bouchy, F., Gaulme, P., Deleuil, M., Havel, M., Gandolfi, D., Deeg, H. J., Wuchterl, G., Guillot, T., Gardes, B., Pasternacki, T., Aigrain, S. a nd Alonso, R., Auvergne, M., Baglin, A., Bonomo, A. S., Bordé, P., Cabrera, J., Carpano, S., Cochran, W. D., Csizmadia, S., Damiani, C., Diaz, R. F., Dvorak, R., Endl, M., Erikson, A., Ferraz-Mello, S., Fridlund, M. a nd Hébrard, G., Gillon, M., Guenther, E., Hatzes, A. a nd Léger, A., Lammer, H., MacQueen, P. J., Mazeh, T. a nd Moutou, C., Ollivier, M., Ofir, A., Pätzold, M., Parviainen, H., Queloz, D., Rauer, H., Rouan, D., Santerne, A., Samuel, B., Schneider, J., Tal-Or, L., Tingley, B., & Weingrill, J. 2013, A&A, 555, A118

Almenara, J. M., Damiani, C., Bouchy, F., Havel, M. ., Bruno, G., Hébrard, G., Diaz, R. F., Deleuil, M. an d Barros, S. C. C., Boisse, I., Bonomo, A. S., Montagnier, G., & Santerne, A. 2015, A&A, 575, A71

Anderson, D. R., Brown, D. J. A., Collier Cameron, A., Delrez, L., Fumel, A., Gillon, M., Hellier, C., Jehin, E., Lendl, M., Maxted, P. F. L., Neveu-VanMalle, M., Pepe, F., Pollacco, D., Queloz, D., Rojo, P., Segransan, D., Serenelli, A. M., Smalley, B., Smith, A. M. S.,

Southworth, J., Triaud, A. H. M. J., Turner, O. D., Udry, S., & West, R. G. 2014a, ArXiv e-prints

Anderson, D. R., Collier Cameron, A., Delrez, L., Doyle, A. P., Faedi, F., Fumel, A., Gillon, M., Gómez Maqueo Chew, Y., Hellier, C., Jehin, E., Lendl, M., Maxted, P. F. L., Pepe, F., Pollacco, D., Queloz, D., Ségransan, D., Skillen, I., Smalley, B., Smith, A. M. S., Southworth, J., Triaud, A. H. M. J., Turner, O. D., Udry, S., & West, R. G. 2014b, MNRAS, 445, 1114

Anderson, D. R., Collier Cameron, A., Hellier, C., Lendl, M., Lister, T. A., Maxted, P. F. L., Queloz, D., Smalley, B., Smith, A. M. S., Triaud, A. H. M. J., Brown, D. J. A., Gillon, M., Neveu-VanMalle, M., Pepe, F., Pollacco, D., Ségransan, D., Udry, S., West, R. G., & Wheatley, P. J. 2015a, A&A, 575, A61

Anderson, D. R., Collier Cameron, A., Hellier, C., Lendl, M., Lister, T. A., Maxted, P. F. L., Queloz, D., Smalley, B., Smith, A. M. S., Triaud, A. H. M. J., West, R. G., Brown, D. J. A., Gillon, M., Pepe, F., Pollacco, D., Ségransan, D., Street, R. A., & Udry, S. 2011, A&A, 531, A60

Anderson, D. R., Triaud, A. H. M. J., Turner, O. D., Brown, D. J. A., Clark, B. J. M., Smalley, B., Collier Cameron, A., Doyle, A. P., Gillon, M., Hellier, C., Lovis, C., Maxted, P. F. L., Pollacco, D., Queloz, D., & Smith, A. M. S. 2015b, ApJL, 800, L9

Anderson, T. W. & Darling, D. A. 1952, The annals of mathematical statistics, 193

Atreya, S. K., Mahaffy, P. R., & Wong, A.-S. 2007, Planetary and Space Science, 55, 358

Bakos, G. Á., Hartman, J., Torres, G., Latham, D. W., Kovács, G., Noyes, R. W., Fischer, D. A., Johnson, J. A., Marcy, G. W., Howard, A. W., Kipping, D., Esquerdo, G. A., Shporer, A., Béky, B., Buchhave, L. A., Perumpilly, G., Everett, M., Sasselov, D. D., Stefanik, R. P., Lázár, J., Papp, I., & Sári, P. 2011, ApJ, 742, 116

Bakos, G. Á., Hartman, J. D., Bhatti, W., Bieryla, A., de Val-Borro, M., Latham, D. W., Buchhave, L. A., Csubry, Z., Penev, K., Kovács, G., Béky, B., Falco, E., Kovács, T., Howard, A. W., Johnson, J. A., Isaacson, H., Marcy, G. W., Torres, G., Noyes, R. W., Berlind, P., Calkins, M. L., Esquerdo, G. A., Lázár, J., Papp, I., & Sári, P. 2015, AJ, 149, 149

Bakos, G. Á., Hartman, J. D., Torres, G., Béky, B., Latham, D. W., Buchhave, L. A., Csubry, Z., Kovács, G., Bieryla, A., Quinn, S., Szklenár, T., Esquerdo, G. A., Shporer, A., Noyes,

R. W., Fischer, D. A., Johnson, J. A., Howard, A. W., Marcy, G. W., Sato, B., Penev, K., Everett, M., Sasselov, D. D., Fűrész, G., Stefanik, R. P., Lázár, J., Papp, I., & Sári, P. 2012, AJ, 144, 19

Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, A&A, 337, 403

Baraffe, I., Chabrier, G., Barman, T. S., Allard, F., & Hauschildt, P. H. 2003, A&A, 402, 701

Baraffe, I., Chabrier, G., Fortney, J., & Sotin, C. 2014, Protostars and Planets VI

Barclay, T., Rowe, J. F., Lissauer, J. J., Huber, D., Fressin, F., Howell, S. B., Bryson, S. T., Chaplin, W. J., Désert, J.-M., Lopez, E. D., Marcy, G. W., Mullally, F., Ragozzine, D., Torres, G., Adams, E. R. a nd Agol, E., Barrado, D., Basu, S., Bedding, T. R., Buchhave, L. A., Charbonneau, D., Christiansen, J. L., Christensen-Dalsgaard, J., Ciardi, D., Cochran, W. D., Dupree, A. K., Elsworth, Y., Everett, M., Fischer, D. A. ., Ford, E. B., Fortney, J. J., Geary, J. C., Haas, M. R., Handberg, R., Hekker, S., Henze, C. E., Horch, E., Howard, A. W., Hunter, R. C., Isaacson, H., Jenkins, J. M., Karoff, C., Kawaler, S. D., Kjeldsen, H., Klaus, T. C., Latham, D. W., Li, J., Lillo-Box, J., Lund, M. N., Lundkvist, M., Metcalfe, T. S., Miglio, A., Morris, R. L., Quintana, E. V., Stello, D., Smith, J. C., & Still, M. an d Thompson, S. E. 2013, Nature, 494, 452

Barros, S. C. C., Almenara, J. M., Deleuil, M., Diaz, R. F., Csizmadia, S., Cabrera, J., Chaintreuil, S. ., Collier Cameron, A., Hatzes, A., Haywood, R., Lanza, A. F., Aigrain, S., Alonso, R., Bordé, P. an d Bouchy, F., Deeg, H. J., Erikson, A., Fridlund, M., Grziwa, S., Gandolfi, D., Guillot, T., Guenther, E., Leger, A., Moutou, C., Ollivier, M., Pasternacki, T. an d Pätzold, M., Rauer, H., Rouan, D., Santerne, A., Schneider, J., & Wuchterl, G. 2014, A&A, 569, A74

Barros, S. C. C., Almenara, J. M., Demangeon, O., Tsantaki, M., Santerne, A., Armstrong, D. J., Barrado, D., Brown, D., Deleuil, M., Lillo-Box, J., Osborn, H., Pollacco, D., Abe, L., Andre, P., Bendjoya, P., Boisse, I., Bonomo, A. S., Bouchy, F., Bruno, G., Cerda, J. R., Courcol, B., Díaz, R. F. ., Hébrard, G., Kirk, J., Lachurié, J. C., Lam, K. W. F., Martinez, P., McCormac, J., Moutou, C. a nd Rajpurohit, A., Rivet, J.-P., Spake, J., Suarez, O., Toublanc, D., & Walker, S. R. 2015, MNRAS, 454, 4267

Barros, S. C. C., Faedi, F., Collier Cameron, A., Lister, T. A., McCormac, J., Pollacco, D., Simpson, E. K., Smalley, B., Street, R. A., Todd, I., Triaud, A. H. M. J., Boisse, I., Bouchy, F., Hébrard, G., Moutou, C., Pepe, F., Queloz, D., Santerne, A., Segransan, D., Udry, S., Bento, J., Butters, O. W., Enoch, B., Haswell, C. A., Hellier, C., Keenan, F. P., Miller,

G. R. M., Moulds, V., Norton, A. J., Parley, N., Skillen, I., Watson, C. A., West, R. G., & Wheatley, P. J. 2011, A&A, 525, A54

Barstow, J. K., Aigrain, S., Irwin, P. G. J., Kendrew, S., & Fletcher, L. N. 2015, MNRAS, 448, 2546

Batygin, K., Bodenheimer, P., & Laughlin, G. 2009, The Astrophysical Journal Letters, 704, L49

Batygin, K. & Stevenson, D. J. 2013, ApJL, 769, L9

Bayliss, D., Zhou, G., Penev, K., Bakos, G. Á., Hartman, J. D., Jordán, A., Mancini, L., Mohler-Fischer, M., Suc, V., Rabus, M., Béky, B. an d Csubry, Z., Buchhave, L., Henning, T., Nikolov, N., Csák, B., Brahm, R., Espinoza, N., Noyes, R. W., Schmidt, B., Conroy, P., Wright, D. J., Tinney, C. G. a nd Addison, B. C., Sackett, P. D., Sasselov, D. D., Lázár, J., Papp, I., & Sári, P. 2013, AJ, 146, 113

Beatty, T. G., Fernández, J. M., Latham, D. W., Bakos, G. Á., Kovács, G., Noyes, R. W., Stefanik, R. P., Torres, G., Everett, M. E., & Hergenrother, C. W. 2007, ApJ, 663, 573

Becker, J. C., Vanderburg, A., Adams, F. C., Rappaport, S. A., & Schwengeler, H. M. 2015, ApJL, 812, L18

Benomar, O., Masuda, K., Shibahashi, H., & Suto, Y. 2014, PASJ, 66, 94

Biddle, L. I., Pearson, K. A., Crossfield, I. J. M., Fulton, B. J., Ciceri, S., Eastman, J., Barman, T., Mann, A. W., Henry, G. W., Howard, A. W., Williamson, M. . H., Sinukoff, E., Dragomir, D., Vican, L., Mancini, L., Southworth, J., Greenberg, A., Turner, J. D., Thompson, R., Taylor, B. W., Levine, S. E., & Webber, M. W. 2014, MNRAS, 443, 1810

Bieryla, A., Collins, K., Beatty, T. G., Eastman, J., Siverd, R. J., Pepper, J., Gaudi, B. S., Stassun, K. G., Cañas, C., Latham, D. W., Buchhave, L. A., Sanchis-Ojeda, R., Winn, J. N., Jensen, E. L. N., Kielkopf, J. F., McLeod, K. K., Gregorio, J., Colón, K. D., Street, R., Ross, R., Penny, M., Mellon, S. N., Oberst, T. E., Fulton, B. J., Wang, J., Berlind, P., Calkins, M. L., Esquerdo, G. A., DePoy, D. L., Gould, A., Marshall, J., Pogge, R., Trueblood, M., & Trueblood, P. 2015, AJ, 150, 12

Birkby, J. L., Cappetta, M., Cruz, P., Koppenhoefer , J., Ivanyuk, O., Mustill, A. J., Hodgkin, S. T., Pinfield, D. J., Sipőcz, B., Kovács, G., Saglia, R., Pavlenko, Y., Barrado, D., Bayo, A., Campbell, D., Catalan, S., Fossati, L., Gálvez-Ortiz , M.-C., Kenworthy, M., Lillo-Box,

J., Martín, E. L., Mislis, D., de Mooij, E. J. W., Nefs, S. V., Snellen, I. A. G., Stoev, H., Zendejas, J., del Burgo, C., Barnes, J., Goulding, N., Haswell, C. A., Kuznetsov, M., Lodieu, N., Murgas, F., Palle, E., Solano, E., Steele, P., & Tata, R. 2014, MNRAS, 440, 1470

Boisse, I., Hartman, J. D., Bakos, G. Á., Penev, K., Csubry, Z., Béky, B., Latham, D. W., Bieryla, A., Torres, G., Kovács, G., Buchhave, L. A. ., Hansen, T., Everett, M., Esquerdo, G. A., Szklenár, T., Falco, E., Shporer, A., Fulton, B. J., Noyes, R. W., Stefanik, R. P., Lázár, J., Papp, I., & Sári, P. 2013, A&A, 558, A86

Bonomo, A. S., Sozzetti, A., Lovis, C., Malavolta, L., Rice, K., Buchhave, L. A., Sasselov, D., Cameron, A. C., Latham, D. W., Molinari, E., Pepe, F., Udry, S., Affer, L., Charbonneau, D., Cosentino, R., Dressing, C. D., Dumusque, X., Figueira, P., Fiorenzano, A. F. M., Gettel, S., Harutyunyan, A., Haywood, R. D., Horne, K., Lopez-Morales, M., Mayor, M., Micela, G., Motalebi, F., Nascimbeni, V., Phillips, D. F., Piotto, G., Pollacco, D., Queloz, D., Ségransan, D., Szentgyorgyi, A., & Watson, C. 2014, A&A, 572, A2

Bonomo, A. S., Sozzetti, A., Santerne, A., Deleuil, M., Almenara, J.-M., Bruno, G., Díaz, R. F., Hébrard, G., & Moutou, C. 2015, A&A, 575, A85

Bottke, W., Vokrouhlickỳ, D., Marchi, S., Swindle, T., Scott, E., Weirich, J., & Levison, H. 2015, Science, 348, 321

Bouma, L. G., Winn, J. N., Kosiarek, J., & McCullough, P. R. 2017, ArXiv e-prints

Boyajian, T. S., von Braun, K., van Belle, G., McAlister, H. A., ten Brummelaar, T. A., Kane, S. R., Muirhead, P. S., Jones, J., White, R., Schaefer, G., Ciardi, D., Henry, T., López-Morales, M., Ridgway, S., Gies, D., Jao, W.-C., Rojas-Ayala, B., Parks, J. R., Sturmann, L., Sturmann, J., Turner, N. H., Farrington, C., Goldfinger, P. J., & Berger, D. H. 2012, ApJ, 757, 112

Brahm, R., Jordán, A., Bakos, G. Á., Penev, K., Espinoza, N., Rabus, M., Hartman, J. D., Bayliss, D., Ciceri, S., Zhou, G., Mancini, L., Tan, T. G., de Val-Borro, M., Bhatti, W., Csubry, Z. an d Bento, J., Henning, T., Schmidt, B., Suc, V., Lázár, J., Papp, I., & Sári, P. 2015a, ArXiv e-prints

Brahm, R., Jordán, A., Hartman, J. D., Bakos, G. Á., Bayliss, D., Penev, K., Zhou, G., Ciceri, S., Rabus, M., Espinoza, N., Mancini, L., de Val-Borro, M., Bhatti, W., Sato, B., Tan, T. G., Csubry, Z., Buchhave, L., Henning, T., Schmidt, B., Suc, V., Noyes, R. W., Papp, I., Lázár, J., & Sári, P. 2015b, AJ, 150, 33

Broeg, C., Fortier, A., Ehrenreich, D., Alibert, Y., Baumjohann, W., Benz, W., Deleuil, M., Gillon, M., Ivanov, A., Liseau, R., Meyer, M., Oloffson, G., Pagano, I., Piotto, G., Pollacco, D., Queloz, D., Ragazzoni, R., Renotte, E., Steller, M., & Thomas, N. 2013, in European Physical Journal Web of Conferences, Vol. 47, European Physical Journal Web of Conferences, 03005

Bruno, G., Almenara, J.-M., Barros, S. C. C., Santerne, A., Diaz, R. F., Deleuil, M., Damiani, C., Bonomo, A. S., Boisse, I., Bouchy, F., Hébrard, G., & Montagnier, G. 2015, A&A, 573, A124

Burrows, A., Hubbard, W. B., Saumon, D., & Lunine, J. I. 1993, ApJ, 406, 158

Burrows, A., Marley, M., Hubbard, W. B., Lunine, J. I., Guillot, T., Saumon, D., Freedman, R., Sudarsky, D., & Sharp, C. 1997, ApJ, 491, 856

Burrows, A. S. 2014, Nature, 513, 345

Çakirli, Ö., Ibanoglu, C., & Dervisoglu, A. 2010, Rev. Mexicana Astron. Astrofis., 46, 363

Cabrera, J., Csizmadia, S., Montagnier, G., Fridlun d, M., Ammler-von Eiff, M., Chaintreuil, S., Damiani, C., Deleuil, M., Ferraz-Mello, S., Ferrigno, A., Gandolfi, D., Guillot, T., Guenther, E. W., Hatzes, A., Hébrard, G., Klagyivik, P., Parviainen, H., Pasternacki, T., Pätzold, M., Sebastian, D., Tadeu dos Santos, M., Wuchterl, G., Aigrain, S., Alonso, R., Almenara, J.-M., Armstrong, J. D., Auvergne, M., Baglin, A., Barge, P., Barros, S. C. C. a nd Bonomo, A. S., Bordé, P., Bouchy, F., Carpano, S. a nd Chaffey, C., Deeg, H. J., Díaz, R. F., Dvorak, R., Erikson, A., Grziwa, S., Korth, J., Lammer, H., Lindsay, C., Mazeh, T., Moutou, C., Ofir, A., Ollivier, M., Pallé, E., Rauer, H., Rouan, D., Samuel, B., Santerne, A., & Schneider, J. 2015, A&A, 579, A36

Carter, J. A. & Winn, J. N. 2010, The Astrophysical Journal, 709, 1219

Carter, J. A., Winn, J. N., Gilliland, R., & Holman, M. J. 2009, ApJ, 696, 241

Carter, J. A., Yee, J. C., Eastman, J., Gaudi, B. S., & Winn, J. N. 2008, ApJ, 689, 499

Cassan, A., Kubas, D., Beaulieu, J.-P., Dominik, M., Horne, K., Greenhill, J., Wambsganss, J., Menzies, J., Williams, A., Jørgensen, U. G., Udalski, A., Bennett, D. P., Albrow, M. D., Batista, V., Brillant, S., Caldwell, J. A. R., Cole, A., Coutures, C., Cook, K. H., Dieters, S., Dominis Prester, D., Donatowicz, J., Fouqué, P., Hill, K., Kains, N., Kane, S., Marquette, J.-B., Martin, R., Pollard, K. R., Sahu, K. C., Vinter, C., Warren, D., Watson, B., Zub,

M., Sumi, T., Szymański, M. K., Kubiak, M., Poleski, R., Soszynski, I., Ulaczyk, K., Pietrzyński, G., & Wyrzykowski, Ł. 2012, Nature, 481, 167

Cersullo, F., Wildi, F., Chazelas, B., & Pepe, F. 2017, A&A, 601, A102

Chabrier, G., Baraffe, I., Allard, F., & Hauschildt, P. 2000, ApJ, 542, 464

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

Chen, J. & Kipping, D. M. 2018, MNRAS, 473, 2753

Ciceri, S., Mancini, L., Henning, T., Bakos, G. 'A., Penev, K., Brahm, R., Zhou, G., Hartman, J. D., Bayliss, D., Jordán, A., Csubry, Z., de Val-Borro, M., Bhatti, W., Rabus, M., Espinoza, N., Suc, V., Schmidt, B., Noyes, R., Howard, A. W., Fulton, B. J. an d Isaacson, H., Marcy, G. W., Butler, R. P., Arriagada, P. ., Crane, J., Shectman, S., Thompson, I., Tan, T. G., Lázár, J., Papp, I., & Sari, P. 2015a, ArXiv e-prints

Ciceri, S., Mancini, L., Southworth, J., Bruni, I., Nikolov, N., D'Ago, G., Schröder, T., Bozza, V., Tregloan-Reed, J., & Henning, T. 2015b, A&A, 577, A54

Ciceri, S., Mancini, L., Southworth, J., Lendl, M., Tregloan-Reed, J., Brahm, R., Chen, G., D'Ago, G., Dominik, M., Figuera Jaimes, R., Galianni, P., Harpsøe, K., Hinse, T. C., Jørgensen, U. G., Juncher, D., Korhonen, H., Liebig, C., Rabus, M., Bonomo, A. S., Bott, K., Henning, T., Jordán, A. an d Sozzetti, A., Alsubai, K. A., Andersen, J. M., Bajek, D., Bozza, V., Bramich, D. M., Browne, P., Calchi Novati, S., Damerdji, Y., Diehl, C., Elyiv, A. a nd Giannini, E., Gu, S.-H., Hundertmark, M., Kains, N., Penny, M., Popovas, A., Rahvar, S., Scarpetta, G., Schmidt, R. W., Skottfelt, J., Snodgrass, C., Surdej, J., Vilela, C., Wang, X.-B., & Wertz, O. 2016, MNRAS, 456, 990

Ciceri, S., Mancini, L., Southworth, J., Nikolov, N., Bozza, V., Bruni, I., Calchi Novati, S., D'Ago, G., & Henning, T. 2013, A&A, 557, A30

Cochran, W. D., Fabrycky, D. C., Torres, G., Fressin, F., Désert, J.-M., Ragozzine, D., Sasselov, D., Fortney, J. J., Rowe, J. F., Brugamyer, E. . J., Bryson, S. T., Carter, J. A., Ciardi, D. R., Howell, S. B., Steffen, J. H., Borucki, W. J., Koch, D. G., Winn, J. N., Welsh, W. F., Uddin, K., Tenenbaum, P., Still, M., Seager, S., Quinn, S. N., Mullally, F., Miller, N., Marcy, G. W., MacQueen, P. J., Lucas, P., Lissauer, J. J., Latham, D. W., Knutson, H., Kinemuchi, K., Johnson, J. A., Jenkins, J. M., Isaacson, H., Howard, A., Horch, E., Holman, M. J., Henze, C. E., Haas, M. R., Gilliland, R. L., Gautier, III, T. N., Ford, E. B.,

Fischer, D. A., Everett, M., Endl, M., Demory, B.-O., Deming, D., Charbonneau, D., Caldwell, D., Buchhave, L., Brown, T. M., & Batalha, N. 2011, ApJS, 197, 7

Collins, K. A., Kielkopf, J. F., & Stassun, K. G. 2015, ArXiv e-prints

Dai, F., Winn, J. N., Arriagada, P., Butler, R. P., Crane, J. D., Johnson, J. A., Shectman, S. A., Teske, J. K., Thompson, I. B., Vanderburg, A., & Wittenmyer, R. A. 2015, ApJL, 813, L9

Dalcanton, J., Seager, S., Aigrain, S., Battel, S., Brandt, N., Conroy, C., Feinberg, L., Gezari, S., Guyon, O., Harris, W., Hirata, C., Mather, J., Postman, M., Redding, D., Schiminovich, D., Stahl, H. P., & Tumlinson, J. 2015, ArXiv e-prints

Dalrymple, G. B. 2001, Geological Society, London, Special Publications, 190, 205

Damasso, M., Biazzo, K., Bonomo, A. S., Desidera, S., Lanza, A. F., Nascimbeni, V., Esposito, M., Scandariato, G., Sozzetti, A., Cosentino, R., Gratton, R., Malavolta, L., Rainer, M., Gandolfi, D., Poretti, E., Zanmar Sanchez, R., Ribas, I., Santos, N., Affer, L., Andreuzzi, G., Barbieri, M., Bedin, L. R., Benatti, S., Bernagozzi, A., Bertolini, E., Bonavita, M., Borsa, F., Borsato, L., Boschin, W., Calcidese, P., Carbognani, A., Cenadelli, D., Christille, J. M., Claudi, R. U., Covino, E., Cunial, A., Giacobbe, P., Granata, V., Harutyunyan, A., Lattanzi, M. G., Leto, G., Libralato, M., Lodato, G. an d Lorenzi, V., Mancini, L., Martinez Fiorenzano, A. F., Marzari, F., Masiero, S., Micela, G., Molinari, E., Molinaro, M., Munari, U., Murabito, S., Pagano, I., Pedani, M., Piotto, G., Rosenberg, A., Silvotti, R., & Southworth, J. 2015, A&A, 575, A111

Darwin, C. 1968, London: Murray Google Scholar

Dawson, R. I., Johnson, J. A., Fabrycky, D. C., Foreman-Mackey, D., Murray-Clay, R. A., Buchhave, L. A., Cargile, P. A., Clubb, K. I., Fulton, B. J., Hebb, L., Howard, A. W., Huber, D., Shporer, A., & Valenti, J. A. 2014, ApJ, 791, 89

Deleuil, M., Almenara, J.-M., Santerne, A., Barros, S. C. C., Havel, M., Hébrard, G., Bonomo, A. S., Bouchy, F. a nd Bruno, G., Damiani, C., Díaz, R. F., Montagnier, G., & Moutou, C. 2014, A&A, 564, A56

Deleuil, M., Deeg, H. J., Alonso, R., Bouchy, F., Rouan, D., Auvergne, M., Baglin, A., Aigrain, S., Almenara, J. M., Barbieri, M., Barge, P., Bruntt, H., Bordé, P., Collier Cameron, A., Csizmadia, S., de La Reza, R., Dvorak, R., Erikson, A., Fridlund, M., Gandolfi, D., Gillon, M., Guenther, E., Guillot, T., Hatzes, A., Hébrard, G., Jorda, L., Lammer, H., Léger, A., Llebaria, A., Loeillet, B., Mayor, M., Mazeh, T., Moutou, C.,

Ollivier, M., Pätzold, M., Pont, F., Queloz, D., Rauer, H., Schneider, J., Shporer, A., Wuchterl, G., & Zucker, S. 2008, A&A, 491, 889

Delrez, L., Santerne, A., Almenara, J.-M., Anderson , D. R., Collier-Cameron, A., Díaz, R. F., Gillon, M., Hellier, C., Jehin, E., Lendl, M., Maxted, P. F. L., Neveu-VanMalle, M., Pepe, F., Pollacco, D., Queloz, D., Ségransan, D., Smalley, B., Smith, A. M. S., Triaud, A. H. M. J., Udry, S., Van Grootel, V., & West, R. G. 2015, ArXiv e-prints

Demory, B.-O. 2014, The Astrophysical Journal Letters, 789, L20

Demory, B.-O., Gillon, M., Madhusudhan, N., & Queloz, D. 2016, MNRAS, 455, 2018

Demory, B.-O., Ségransan, D., Forveille, T., Queloz, D., Beuzit, J.-L., Delfosse, X., di Folco, E., Kervella, P., Le Bouquin, J.-B., Perrier, C., Benisty, M., Duvert, G., Hofmann, K.-H., Lopez, B., & Petrov, R. 2009, A&A, 505, 205

Devroye, L. 1986, in Proceedings of the 18th Conference on Winter Simulation, WSC '86 (New York, NY, USA: ACM), 260–265

Díaz, R. F., Montagnier, G., Leconte, J., Bonomo, A. S., Deleuil, M., Almenara, J. M., Barros, S. C. C., Bouchy, F., Bruno, G., Damiani, C., Hébrard, G., Moutou, C., & Santerne, A. 2014, A&A, 572, A109

Dieterich, S. B., Henry, T. J., Jao, W.-C., Winters, J. G., Hosey, A. D., Riedel, A. R., & Subasavage, J. P. 2014, AJ, 147, 94

Dodd, M. S., Papineau, D., Grenne, T., Slack, J. F., Rittner, M., Pirajno, F., ONeil, J., & Little, C. T. 2017, Nature, 543, 60

Dorn, C., Venturini, J., Khan, A., Heng, K., Alibert, Y., Helled, R., Rivoldini, A., & Benz, W. 2017, Astronomy & Astrophysics, 597, A37

Dotter, A., Chaboyer, B., Jevremović, D., Kostov, V., Baron, E., & Ferguson, J. W. 2008, ApJS, 178, 89

Drake, F. D. 1977, in Focus on the Stars, ed. H. Messel & S. T. Butler, 257–287

Dreizler, S. & Ofir, A. 2014, ArXiv e-prints

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

Dressing, C. D., Charbonneau, D., Dumusque, X., Gettel, S., Pepe, F., Collier Cameron, A., Latham, D. W. ., Molinari, E., Udry, S., Affer, L., Bonomo, A. S., Buchhave, L. A., Cosentino, R., Figueira, P., Fiorenzano, A. F. M., Harutyunyan, A., Haywood, R. D., Johnson, J. A., Lopez-Morales, M., Lovis, C., Malavolta, L., Mayor, M., Micela, G., Motalebi, F., Nascimbeni, V., Phillips, D. F., Piotto, G., Pollacco, D., Queloz, D., Rice, K., Sasselov, D., Ségransan, D., Sozzetti, A., Szentgyorgyi, A., & Watson, C. 2015, ApJ, 800, 135

Dressing, C. D., Vanderburg, A., Schlieder, J. E., Crossfield, I. J. M., Knutson, H. A., Newton, E. R., Ciardi, D. R., Fulton, B. J., Gonzales, E. J., Howard, A. W., Isaacson, H., Livingston, J., Petigura, E. A., Sinukoff, E., Everett, M., Horch, E., & Howell, S. B. 2017, AJ, 154, 207

Eastman, J. D., Beatty, T. G., Siverd, R. J., Antognini, J. M. O., Penny, M. T., Gonzales, E. J., Crepp, J. R., Howard, A. W., Avril, R. L., Bieryla, A., Collins, K., Fulton, B. J., Ge, J., Gregorio, J., Ma, B., Mellon, S. N., Oberst, T. E., Wang, J., Gaudi, B. S., Pepper, J., Stassun, K. G., Buchhave, L. A., Jensen, E. L. N., Latham, D. W., Berlind, P., Calkins, M. L., Cargile, P. A., Colon, K. D., Dhital, S., Esquerdo, G. A., Johnson, J. A., Kielkopf, J. F., Manner, M., Mao, Q., McLeod, K. K., Penev, K., Stefanik, R. P., Street, R., Zambelli, R. an d DePoy, D. L., Gould, A., Marshall, J. L., Pogge, R. W., Trueblood, M., & Trueblood, P. 2015, ArXiv e-prints

Eigenbrode, J. L., Summons, R. E., Steele, A., Freissinet, C., Millan, M., Navarro-González, R., Sutter, B., McAdam, A. C., Franz, H. B., Glavin, D. P., et al. 2018, Science, 360, 1096

Endl, M., Caldwell, D. A., Barclay, T., Huber, D., Isaacson, H., Buchhave, L. A., Brugamyer, E., Robertson, P., Cochran, W. D., MacQueen, P. J., Havel, M., Lucas, P., Howell, S. B., Fischer, D., Quintana, E., & Ciardi, D. R. 2014, ApJ, 795, 151

Enoch, B., Anderson, D. R., Barros, S. C. C., Brown, D. J. A., Collier Cameron, A., Faedi, F., Gillon, M., Hébrard, G., Lister, T. A., Queloz, D., Santerne, A., Smalley, B., Street, R. A., Triaud, A. H. M. J., West, R. G., Bouchy, F., Bento, J., Butters, O., Fossati, L., Haswell, C. A., Hellier, C., Holmes, S. an d Jehin, E., Lendl, M., Maxted, P. F. L., McCormac, J. an d Miller, G. R. M., Moulds, V., Moutou, C., Norton, A. J., Parley, N., Pepe, F., Pollacco, D., Segransan, D., Simpson, E., Skillen, I., Smith, A. M. S., Udry, S., & Wheatley, P. J. 2011, AJ, 142, 86

Esposito, M., Covino, E., Mancini, L., Harutyunyan, A., Southworth, J., Biazzo, K., Gandolfi, D., Lanza, A. F., Barbieri, M., Bonomo, A. S., Borsa, F., Claudi, R., Cosentino, R.,

Desidera, S., Gratton, R., Pagano, I. an d Sozzetti, A., Boccato, C., Maggio, A., Micela, G., Molinari, E., Nascimbeni, V., Piotto, G., & Poretti, E. a nd Smareglia, R. 2014, A&A, 564, L13

Faedi, F., Barros, S. C. C., Anderson, D. R., Brown, D. J. A., Collier Cameron, A., Pollacco, D., Boisse, I., Hébrard, G., Lendl, M., Lister, T. A. a nd Smalley, B., Street, R. A., Triaud, A. H. M. J., Bento, J., Bouchy, F., Butters, O. W., Enoch, B., Haswell, C. A., Hellier, C., Keenan, F. P., Miller, G. R. M., Moulds, V., Moutou, C., Norton, A. J., Queloz, D., Santerne, A., Simpson, E. K., Skillen, I., Smith, A. M. S., Udry, S., Watson, C. A., West, R. G., & Wheatley, P. J. 2011, A&A, 531, A40

Faedi, F., Pollacco, D., Barros, S. C. C., Brown, D., Collier Cameron, A., Doyle, A. P., Enoch, R., Gillon, M., Gómez Maqueo Chew, Y., Hébrard, G., Lendl, M., Liebig, C., Smalley, B., Triaud, A. H. M. J., West, R. G., Wheatley, P. J., Alsubai, K. A., Anderson, D. R., Armstrong, D., Bento, J., Bochinski, J. ., Bouchy, F., Busuttil, R., Fossati, L., Fumel, A., Haswell, C. A., Hellier, C., Holmes, S., Jehin, E., Kolb, U., McCormac, J., Miller, G. R. M., Moutou, C. an d Norton, A. J., Parley, N., Queloz, D., Santerne, A., Skillen, I., Smith, A. M. S., Udry, S., & Watson, C. 2013, A&A, 551, A73

Faigler, S. & Mazeh, T. 2015, ApJ, 800, 73

Ford, E. B. 2006, ApJ, 642, 505

Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, The Astrophysical Journal, 795, 64

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

Foreman-Mackey, D., Morton, T. D., Hogg, D. W., Agol, E., & Schölkopf, B. 2016, AJ, 152, 206

Fortney, J. J., Marley, M. S., & Barnes, J. W. 2007, ApJ, 659, 1661

Fortney, J. J. & Nettelmann, N. 2010, Space Science Reviews, 152, 423

Fulton, B. J. & Petigura, E. A. 2018, arXiv preprint arXiv:1805.01453

Fulton, B. J., Petigura, E. A., Howard, A. W., Isaacson, H., Marcy, G. W., Cargile, P. A., Hebb, L., Weiss, L. M., Johnson, J. A., Morton, T. D., et al. 2017, The Astronomical Journal, 154, 109

Gandolfi, D., Parviainen, H., Deeg, H. J., Lanza, A. F., Fridlund, M., Prada Moroni, P. G., Alonso, R., Augusteijn, T., Cabrera, J., Evans, T., Geier, S., Hatzes, A. P., Holczer, T., Hoyer, S., Kangas, T., Mazeh, T., Pagano, I., Tal-Or, L., & Tingley, B. 2015, A&A, 576, A11

Gandolfi, D., Parviainen, H., Fridlund, M., Hatzes, A. P., Deeg, H. J., Frasca, A., Lanza, A. F., Prada Moroni, P. G., Tognelli, E., McQuillan, A., Aigrain, S., Alonso, R. an d Antoci, V., Cabrera, J., Carone, L., Csizmadia, S., Djupvik, A. A., Guenther, E. W., Jessen-Hansen, J., Ofir, A., & Telting, J. 2013, A&A, 557, A74

Gelman, A. 2004, Bayesian Data Analysis, Texts in statistical science (Chapman & Hall/CRC)

Gelman, A. & Rubin, D. 1992, Statistical Science, 7, 457, `http://www.stat.columbia.edu/~gelman/research/published/itsim.pdf`

Gettel, S., Charbonneau, D., Dressing, C. D., Buchhave, L. A., Dumusque, X., Vanderburg, A., Bonomo, A. S., Malavolta, L., Pepe, F., Collier Cameron, A., Latham, D. W., Udry, S., Marcy, G. W., Isaacson, H., Howard, A. W., Davies, G. R., Silva Aguirre, V., Kjeldsen, H., Bedding, T. R., Lopez, E., Affer, L., Cosentino, R., Figueira, P., Fiorenzano, A. F. M., Harutyunyan, A., Johnson, J. A., Lopez-Morales, M., Lovis, C., Mayor, M., Micela, G., Molinari, E., Motalebi, F., Phillips, D. F., Piotto, G., Queloz, D. a nd Rice, K., Sasselov, D., Ségransan, D., Sozzetti, A., Watson, C., Basu, S., Campante, T. L., Christensen-Dalsg aard, J., Kawaler, S. D., Metcalfe, T. S., Handberg, R., Lund, M. N., Lundkvist, M. S., Huber, D., & Chaplin, W. J. . 2016, ApJ, 816, 95

Gibson, N. P., Aigrain, S., Barstow, J. K., Evans, T. M., Fletcher, L. N., & Irwin, P. G. J. 2013, MNRAS, 428, 3680

Gillon, M., Anderson, D. R., Collier-Cameron, A., Doyle, A. P., Fumel, A., Hellier, C., Jehin, E., Lendl, M., Maxted, P. F. L., Montalbán, J., Pepe, F., Pollacco, D., Queloz, D., Ségransan, D. a nd Smith, A. M. S., Smalley, B., Southworth, J., Triaud, A. H. M. J., Udry, S., & West, R. G. 2013, A&A, 552, A82

Gillon, M., Triaud, A. H. M. J., Fortney, J. J., Demory, B.-O., Jehin, E., Lendl, M., Magain, P., Kabath, P., Queloz, D., Alonso, R., Anderson, D. R., Collier Cameron, A., Fumel, A., Hebb, L., Hellier, C. a nd Lanotte, A., Maxted, P. F. L., Mowlavi, N., & Smalley, B. 2012, A&A, 542, A4

Gómez Maqueo Chew, Y., Faedi, F., Pollacco, D., Brown, D. J. A., Doyle, A. P., Collier Cameron, A., Gillon, M., Lendl, M., Smalley, B., Triaud, A. H. M. J., West, R. G., Wheatley, P. J., Busuttil, R., Liebig, C., Anderson, D. R., Armstrong, D. J., Barros, S. C. C., Bento, J., Bochinski, J., Burwitz, V., Delrez, L., Enoch, B., Fumel, A., Haswell, C. A., Hébrard, G. a nd Hellier, C., Holmes, S., Jehin, E., Kolb, U., Maxted, P. F. L., McCormac, J., Miller, G. R. M., Norton, A. J., Pepe, F., Queloz, D., Rodríguez, J., Ségransan, D., Skillen, I., Stassun, K. G., Udry, S., & Watson, C. 2013, A&A, 559, A36

Graziani, C. & Lamb, D. Q. 1996, in American Institute of Physics Conference Series, Vol. 366, High Velocity Neutron Stars, ed. R. E. Rothschild & R. E. Lingenfelter, 196–200

Grotzinger, J. P. 2014, Habitability, taphonomy, and the search for organic carbon on Mars

Grunblatt, S. K., Howard, A. W., & Haywood, R. D. 2015, ApJ, 808, 127

Grziwa, S., Gandolfi, D., Csizmadia, S., Fridlund, M., Parviainen, H., Deeg, H. J., Cabrera, J., Djupvik, A. A. ., Albrecht, S., Palle, E. B., Pätzold, M., Béjar, V. J. S., Arranz, J. P., Eigmüller, P., Erikson, A., Fynbo, J. P. U., Guenther, E. W., Hatzes, A. P., Kiilerich, A., Korth, J., Kuutma, T., Montanés-Rodríguez, P., Nespral, D., Nowak, G. an d Rauer, H., Saario, J., Sebastian, D., & Slumstrup, D. 2015, ArXiv e-prints

Haghighipour, N. 2013, Annual Review of Earth and Planetary Sciences, 41, 469

Han, E., Wang, S. X., Wright, J. T., Feng, Y. K., Zhao, M., Fakhouri, O., Brown, J. I., & Hancock, C. 2014, PASP, 126, 827

Harpsøe, K. B. W., Hardis, S., Hinse, T. C., Jørgensen, U. G., Mancini, L., Southworth, J., Alsubai, K. A., Bozza, V., Browne, P., Burgdorf, M. J., Calchi Novati, S., Dodds, P., Dominik, M., Fang, X.-S., Finet, F., Gerner, T., Gu, S.-H., Hundertmark, M., Jessen-Hansen, J., Kains, N., Kerins, E., Kjeldsen, H., Liebig, C., Lund, M. N., Lundkvist, M., Mathiasen, M. a nd Nesvorný, D., Nikolov, N., Penny, M. T., Proft, S., Rahvar, S., Ricci, D., Sahu, K. C., Scarpetta, G., Schäfer, S., Schönebeck, F., Snodgrass, C., Skottfelt, J., Surdej, J., Tregloan-Reed, J., & Wertz, O. 2013, A&A, 549, A10

Hartman, J. D., Bakos, G. Á., Béky, B., Torres, G., Latham, D. W., Csubry, Z., Penev, K., Shporer, A., Fulton, B. J., Buchhave, L. A., Johnson, J. A., Howard, A. W., Marcy, G. W., Fischer, D. A., Kovács, G., Noyes, R. W., Esquerdo, G. A., Everett, M., Szklenár, T., Quinn, S. N., Bieryla, A., Knox, R. P., Hinz, P., Sasselov, D. D., Fűrész, G., Stefanik, R. P., Lázár, J., Papp, I., & Sári, P. 2012, AJ, 144, 139

174

Hartman, J. D., Bakos, G. Á., Sato, B., Torres, G., Noyes, R. W., Latham, D. W., Kovács, G., Fischer, D. A., Howard, A. W., Johnson, J. A., Marcy, G. W., Buchhave, L. A., Füresz, G., Perumpilly, G., Béky, B., Stefanik, R. P., Sasselov, D. D., Esquerdo, G. A., Everett, M., Csubry, Z., Lázár, J., Papp, I., & Sári, P. 2011a, ApJ, 726, 52

Hartman, J. D., Bakos, G. Á., Torres, G., Latham, D. W., Kovács, G., Béky, B., Quinn, S. N., Mazeh, T., Shporer, A., Marcy, G. W., Howard, A. W., Fischer, D. A., Johnson, J. A., Esquerdo, G. A., Noyes, R. W., Sasselov, D. D., Stefanik, R. P., Fernandez, J. M., Szklenár, T., Lázár, J., Papp, I., & Sári, P. 2011b, ApJ, 742, 59

Hartman, J. D., Bayliss, D., Brahm, R., Bakos, G. Á., Mancini, L., Jordán, A., Penev, K., Rabus, M., Zhou, G., Butler, R. P., Espinoza, N., de Val-Borro, M., Bhatti, W., Csubry, Z., Ciceri, S., Henning, T., Schmidt, B., Arriagada, P., Shectman, S., Crane, J., Thompson, I., Suc, V., Csák, B., Tan, T. G., Noyes, R. W., Lázár, J., Papp, I., & Sári, P. 2015a, AJ, 149, 166

Hartman, J. D., Bhatti, W., Bakos, G. Á., Bieryla, A., Kovács, G., Latham, D. W., Csubry, Z., de Val-Borro, M., Penev, K., Buchhave, L. A., Torres, G., Howard, A. W., Marcy, G. W., Johnson, J. A., Isaacson, H., Sato, B., Boisse, I., Falco, E., Everett, M. E., Szklenar, T., Fulton, B. J., Shporer, A., Kovács, T., Hansen, T., Béky, B. a nd Noyes, R. W., Lázár, J., Papp, I., & Sári, P. 2015b, AJ, 150, 168

Hatzes, A. P. & Rauer, H. 2015, ArXiv e-prints

Hebb, L., Wyse, R. F. G., Gilmore, G., & Holtzman, J. 2006, AJ, 131, 555

Hébrard, G., Collier Cameron, A., Brown, D. J. A. a nd Díaz, R. F., Faedi, F., Smalley, B., Anderson, D. . R., Armstrong, D., Barros, S. C. C., Bento, J., Bouchy, F., Doyle, A. P., Enoch, B., Gómez Maqueo Chew, Y., Hébrard, É. M., Hellier, C., Lendl, M., Lister, T. A., Maxted, P. F. L., McCormac, J., Moutou, C., Pollacco, D., Queloz, D., Santerne, A., Skillen, I., Southworth, J., Tregloan-Reed, J., Triaud, A. H. M. J., Udry, S., Vanhuysse, M., Watson, C. A., West, R. G., & Wheatley, P. J. 2013, A&A, 549, A134

Hébrard, G., Santerne, A., Montagnier, G., Bruno, G., Deleuil, M., Havel, M., Almenara, J.-M., Damiani, C., Barros, S. C. C., Bonomo, A. S., Bouchy, F., Díaz, R. F., & Moutou, C. 2014, A&A, 572, A93

Hellier, C., Anderson, D. R., Cameron, A. C., Delrez, L., Gillon, M., Jehin, E., Lendl, M., Maxted, P. F. L., Pepe, F., Pollacco, D., Queloz, D. an d Ségransan, D., Smalley, B., Smith, A. M. S., Southworth, J., Triaud, A. H. M. J., Udry, S., & West, R. G. 2014, MNRAS, 440, 1982

Hellier, C., Anderson, D. R., Collier Cameron, A., Delrez, L., Gillon, M., Jehin, E., Lendl, M., Maxted, P. F. L., Pepe, F., Pollacco, D., Queloz, D. an d Ségransan, D., Smalley, B., Smith, A. M. S., Southworth, J., Triaud, A. H. M. J., Turner, O. D., Udry, S., & West, R. G. 2015, AJ, 150, 18

Hellier, C., Anderson, D. R., Collier Cameron, A., Doyle, A. P., Fumel, A., Gillon, M., Jehin, E., Lendl, M., Maxted, P. F. L., Pepe, F., Pollacco, D., Queloz, D., Ségransan, D., Smalley, B., Smith, A. M. . S., Southworth, J., Triaud, A. H. M. J., Udry, S., & West, R. G. 2012, MNRAS, 426, 739

Heng, K. & Vogt, S. S. 2011, MNRAS, 415, 2145

Hertzsprung, E. 1909, Astronomische Nachrichten, 179, 373

Hogg, D. W., Myers, A. D., & Bovy, J. 2010, The Astrophysical Journal, 725, 2166

Hogg, D. W., Myers, A. D., & Bovy, J. 2010, ApJ, 725, 2166

Hou, F., Goodman, J., & Hogg, D. W. 2014, ArXiv e-prints

Howard, A. W., Bakos, G. Á., Hartman, J., Torres, G., Shporer, A., Mazeh, T., Kovács, G., Latham, D. W., Noyes, R. W., Fischer, D. A., Johnson, J. A., Marcy, G. W., Esquerdo, G. A., Béky, B., Butler, R. P., Sasselov, D. D., Stefanik, R. P., Perumpilly, G., Lázár, J., Papp, I., & Sári, P. 2012, ApJ, 749, 134

Howard, A. W., Sanchis-Ojeda, R., Marcy, G. W., Johnson, J. A., Winn, J. N., Isaacson, H., Fischer, D. A., Fulton, B. J., Sinukoff, E., & Fortney, J. J. 2013, Nature, 503, 381

Huang, C. X., Hartman, J. D., Bakos, G. Á., Penev, K., Bhatti, W., Bieryla, A., de Val-Borro, M. an d Latham, D. W., Buchhave, L. A., Csubry, Z., Kovács, G., Béky, B., Falco, E., Berlind, P., Calkins, M. L. an d Esquerdo, G. A., Lázár, J., Papp, I., & Sári, P. 2015, AJ, 150, 85

Huber, D., Carter, J. A., Barbieri, M., Miglio, A., Deck, K. M., Fabrycky, D. C., Montet, B. T., Buchhave, L. A., Chaplin, W. J., Hekker, S., Montalbán, J., Sanchis-Ojeda, R., Basu, S., Bedding, T. R., Campante, T. L., Christensen-Dalsgaard, J. a nd Elsworth, Y. P., Stello, D., Arentoft, T., Ford, E. B., Gilliland, R. L., Handberg, R., Howard, A. W., Isaacson, H., Johnson, J. A., Karoff, C., Kawaler, S. D. ., Kjeldsen, H., Latham, D. W., Lund, M. N., Lundkvist, M., Marcy, G. W., Metcalfe, T. S., Silva Aguirre, V., & Winn, J. N. 2013, Science, 342, 331

Jenkins, J. M., Twicken, J. D., Batalha, N. M., Caldwell, D. A., Cochran, W. D., Endl, M., Latham, D. W. ., Esquerdo, G. A., Seader, S., Bieryla, A., Petigura, E., Ciardi, D. R., Marcy, G. W., Isaacson, H., Huber, D. an d Rowe, J. F., Torres, G., Bryson, S. T., Buchhave, L. an d Ramirez, I., Wolfgang, A., Li, J., Campbell, J. R., Tenenbaum, P., Sanderfer, D., Henze, C. E., Catanzarite, J. H., Gilliland, R. L., & Borucki, W. J. 2015, AJ, 150, 56

Johnson, J. A., Petigura, E. A., Fulton, B. J., Marcy, G. W., Howard, A. W., Isaacson, H., Hebb, L., Cargile, P. A., Morton, T. D., Weiss, L. M., Winn, J. N., Rogers, L. A., Sinukoff, E., & Hirsch, L. A. 2017, AJ, 154, 108

Johnson, J. A., Winn, J. N., Bakos, G. Á., Hartman, J. D., Morton, T. D., Torres, G., Kovács, G., Latham, D. W., Noyes, R. W., Sato, B., Esquerdo, G. A., Fischer, D. A., Marcy, G. W., Howard, A. W., Buchhave, L. A., Főrész, G., Quinn, S. N., Béky, B., Sasselov, D. D., Stefanik, R. P., Lázár, J., Papp, I., & Sári, P. 2011, ApJ, 735, 24

Jontof-Hutter, D., Ford, E. B., Rowe, J. F., Lissauer, J. J., Fabrycky, D. C., Van Laerhoven, C., Agol, E., Deck, K. M., Holczer, T., & Mazeh, T. 2015, ArXiv e-prints

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

Jordán, A., Brahm, R., Bakos, G. Á., Bayliss, D., Penev, K., Hartman, J. D., Zhou, G., Mancini, L., Mohler-Fischer, M., Ciceri, S., Sato, B., Csubry, Z., Rabus, M., Suc, V., Espinoza, N., Bhatti, W., de Val-Borro, M., Buchhave, L. ., Csák, B., Henning, T., Schmidt, B., Tan, T. G., Noyes, R. W., Béky, B., Butler, R. P., Shectman, S., Crane, J., Thompson, I., Williams, A., Martin, R., Contreras, C., Lázár, J., Papp, I., & Sári, P. . 2014, AJ, 148, 29

Juncher, D., Buchhave, L. A., Hartman, J. D., Bakos, G. Á., Bieryla, A., Kovács, T., Boisse, I., Latham, D. W., Kovács, G., Bhatti, W. a nd Csubry, Z., Penev, K., de Val-Borro, M., Falco, E., Torres, G., Noyes, R. W., Lázár, J., Papp, I., & Sári, P. 2015, PASP, 127, 851

Kaltenegger, L. 2017, Annual Review of Astronomy and Astrophysics, 55, 433

Kaltenegger, L., Henning, W., & Sasselov, D. 2010, The Astronomical Journal, 140, 1370

Khodachenko, M. L., Ribas, I., Lammer, H., Grießmeier, J.-M., Leitner, M., Selsis, F., Eiroa, C., Hanslmeier, A., Biernat, H. K., Farrugia, C. J., et al. 2007, Astrobiology, 7, 167

Kipping, D. M. 2014, arXiv preprint arXiv:1405.1455

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

Kipping, D. M. & Sandford, E. 2016, MNRAS, 463, 1323

Kipping, D. M., Schmitt, A. R., Huang, X., Torres, G., Nesvorný, D., Buchhave, L. A., Hartman, J., & Bakos, G. Á. 2015, ApJ, 813, 14

Kipping, D. M., Spiegel, D. S., & Sasselov, D. D. 2013, MNRAS, 434, 1883

Kolmogorov, A. 1933, Inst. Ital. Attuari, Giorn., 4, 83

Kostov, V. B., Orosz, J. A., Welsh, W. F., Doyle, L. R., Fabrycky, D. C., Haghighipour, N., Quarles, B., Short, D. R., Cochran, W. D., Endl, M., Ford, E. B., Gregorio, J., Hinse, T. C., Isaacson, H., Jenkins, J. M. ., Jensen, E. L. N., Kull, I., Latham, D. W., Lissauer, J. J., Marcy, G. W., Mazeh, T., Muller, T. W. A., Pepper, J. a nd Quinn, S. N., Ragozzine, D., Shporer, A., Steffen, J. H. ., Torres, G., Windmiller, G., & Borucki, W. J. 2015, ArXiv e-prints

Kovács, G., Bakos, G. Á., Hartman, J. D., Torres, G., Noyes, R. W., Latham, D. W., Howard, A. W., Fischer, D. A., Johnson, J. A., Marcy, G. W., Isaacson, H., Sasselov, D. D., Stefanik, R. P., Esquerdo, G. A., Fernandez, J. M., Lázár, B. B. J. a nd Papp, I., & Sári, P. 2010, ApJ, 724, 866

Kramm, U., Nettelmann, N., Fortney, J., Neuhäuser, R., & Redmer, R. 2012, Astronomy & Astrophysics, 538, A146

Krasnopolsky, V. A., Maillard, J. P., & Owen, T. C. 2004, Icarus, 172, 537

Kraus, A. L., Tucker, R. A., Thompson, M. I., Craine, E. R., & Hillenbrand, L. A. 2011, ApJ, 728, 48

Kullback, S. & Leibler, R. A. 1951, The annals of mathematical statistics, 22, 79

Künsch, H. R. et al. 2013, Bernoulli, 19, 1391

Lacki, B. C. 2016, arXiv preprint arXiv:1609.05931

Lammer, H., Lichtenegger, H. I., Kulikov, Y. N., Grießmeier, J.-M., Terada, N., Erkaev, N. V., Biernat, H. K., Khodachenko, M. L., Ribas, I., Penz, T., et al. 2007, Astrobiology, 7, 185

Lanotte, A. A., Gillon, M., Demory, B.-O., Fortney, J. J., Astudillo, N., Bonfils, X., Magain, P., Delfosse, X. an d Forveille, T., Lovis, C., Mayor, M., Neves, V., Pepe, F., Queloz, D., Santos, N., & Udry, S. 2014, A&A, 572, A73

Lee, J. W., Youn, J.-H., Kim, S.-L., & Lee, C.-U. an d Hinse, T. C. 2012, AJ, 143, 95

Léger, A., Mariotti, J.-M., Mennesson, B., Ollivier, M., Puget, J., Rouan, D., & Schneider, J. 1996, Icarus, 123, 249

Lenardic, A. & Seales, J. 2018, arXiv preprint arXiv:1801.09146

Lendl, M., Anderson, D. R., Collier-Cameron, A., Doyle, A. P., Gillon, M., Hellier, C., Jehin, E., Lister, T. A., Maxted, P. F. L., Pepe, F., Pollacco, D., Queloz, D., Smalley, B., Ségransan, D., Smith, A. M. . S., Triaud, A. H. M. J., Udry, S., West, R. G., & Wheatley, P. . J. 2012, A&A, 544, A72

Lillo-Box, J., Barrado, D., Santos, N. C., Mancini, L., Figueira, P., Ciceri, S., & Henning, T. 2015, A&A, 577, A105

Lineweaver, C. H. & Davis, T. M. 2002, Astrobiology, 2, 293

Lopez, E. D. & Fortney, J. J. 2013, The Astrophysical Journal, 776, 2

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

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

Lund, M. B., Pepper, J., & Stassun, K. G. 2015, AJ, 149, 16

Mancini, L., Ciceri, S., Chen, G., Tregloan-Reed, J., Fortney, J. J., Southworth, J., Tan, T. G., Burgdorf, M. ., Calchi Novati, S., Dominik, M., Fang, X.-S., Finet, F., Gerner, T., Hardis, S., Hinse, T. C., Jørgensen, U. G., Liebig, C., Nikolov, N., Ricci, D., Schäfer, S., Schönebeck, F., Skottfelt, J., Wertz, O., Alsubai, K. A., Bozza, V., Browne, P., Dodds, P., Gu, S.-H., Harpsøe, K., Henning, T., Hundertmark, M., Jessen-Hansen, J., Kains, N., Kerins, E., Kjeldsen, H., Lund, M. N., Lundkvist, M., Madhusudhan, N., Mathiasen, M., Penny, M. T., Prof, S., Rahvar, S., Sahu, K., Scarpetta, G., Snodgrass, C., & Surdej, J. 2013a, MNRAS, 436, 2

Mancini, L., Esposito, M., Covino, E., Raia, G. an d Southworth, J., Tregloan-Reed, J., Biazzo, K., Bonomo, A. S., Desidera, S., Lanza, A. F., Maciejewski, G., Poretti, E., Sozzetti, A., Borsa, F., Bruni, I., Ciceri, S., Claudi, R., Cosentino, R., Gratton, R., Martinez Fiorenzano, A. F., Lodato, G., Lorenzi, V., Marzari, F., Murabito, S., Affer, L., Bignamini, A., Bedin, L. R., Boccato, C., Damasso, M., Henning, T., Maggio, A., Micela, G., Molinari, E., Pagano, I., Piotto, G., Rainer, M., Scandariato, G., Smareglia, R., & Zanmar Sanchez, R. 2015a, A&A, 579, A136

Mancini, L., Hartman, J. D., Penev, K., Bakos, G. Á., Brahm, R., Ciceri, S., Henning, T., Csubry, Z., Bayliss, D., Zhou, G., Rabus, M., de Val-Borro, M., Espinoza, N., Jordán, A., Suc, V., Bhatti, W., Schmidt, B., Sato, B., Tan, T. G., Wright, D. J., Tinney, C. G., Addison, B. C., Noyes, R. W., Lázár, J., Papp, I., & Sári, P. 2015b, A&A, 580, A63

Mancini, L., Lillo-Box, J., Southworth, J., Borsato , L., Gandolfi, D., Ciceri, S., Barrado, D., Brahm, R., & Henning, T. 2015c, ArXiv e-prints

Mancini, L., Nikolov, N., Southworth, J., Chen, G., Fortney, J. J., Tregloan-Reed, J., Ciceri, S., van Boekel, R., & Henning, T. 2013b, MNRAS, 430, 2932

Mancini, L., Southworth, J., Ciceri, S., Calchi Nov ati, S., Dominik, M., Henning, T., Jørgensen, U. G., Korhonen, H., Nikolov, N., Alsubai, K. A., Bozza, V. an d Bramich, D. M., D'Ago, G., Figuera Jaimes, R., Galianni, P., Gu, S.-H., Harpsøe, K., Hinse, T. C. a nd Hundertmark, M., Juncher, D., Kains, N., Popovas, A. an d Rabus, M., Rahvar, S., Skottfelt, J., Snodgrass, C., Street, R., Surdej, J., Tsapras, Y., Vilela, C., Wang, X.-B., & Wertz, O. 2014a, A&A, 568, A127

Mancini, L., Southworth, J., Ciceri, S., Dominik, M., Henning, T., Jørgensen, U. G., Lanza, A. F., Rabus, M., Snodgrass, C., Vilela, C., Alsubai, K. A. an d Bozza, V., Bramich, D. M., Calchi Novati, S., D'Ago, G., Figuera Jaimes, R., Galianni, P., Gu, S.-H., Harpsøe, K., Hinse, T., Hundertmark, M. a nd Juncher, D., Kains, N., Korhonen, H., Popovas, A., Rahvar, S., Skottfelt, J., Street, R., Surdej, J., Tsapras, Y., Wang, X.-B., & Wertz, O. 2014b, A&A, 562, A126

Mancini, L., Southworth, J., Ciceri, S., Fortney, J. J., Morley, C. V., Dittmann, J. A., Tregloan-Reed, J., Bruni, I., Barbieri, M., Evans, D. F., D'Ago, G., Nikolov, N., & Henning, T. 2013c, A&A, 551, A11

Mancini, L., Southworth, J., Ciceri, S., Tregloan-R eed, J., Crossfield, I., Nikolov, N., Bruni, I., & Zambelli, R. an d Henning, T. 2014c, MNRAS, 443, 2391

Mandel, K. & Agol, E. 2002, ApJL, 580, L171

Marcy, G. W., Isaacson, H., Howard, A. W., Rowe, J. . F., Jenkins, J. M., Bryson, S. T., Latham, D. W., Howell, S. B., Gautier, III, T. N., Batalha, N. M., Rogers, L., Ciardi, D., Fischer, D. A., Gilliland, R. L. ., Kjeldsen, H., Christensen-Dalsgaard, J., Huber, D., Chaplin, W. J., Basu, S., Buchhave, L. A., Quinn, S. N., Borucki, W. J., Koch, D. G., Hunter, R., Caldwell, D. A. ., Van Cleve, J., Kolbl, R., Weiss, L. M., Petigura, E. an

d Seager, S., Morton, T., Johnson, J. A., Ballard, S., Burke, C., Cochran, W. D., Endl, M., MacQueen, P., Everett, M. E., Lissauer, J. J., Ford, E. B., Torres, G., Fressin, F., Brown, T. M., Steffen, J. H. a nd Charbonneau, D., Basri, G. S., Sasselov, D. D., Winn, J., Sanchis-Ojeda, R., Christiansen, J., Adams, E., Henze, C., Dupree, A., Fabrycky, D. C., Fortney, J. J., Tarter, J., Holman, M. J., Tenenbaum, P. ., Shporer, A., Lucas, P. W., Welsh, W. F., Orosz, J. A. a nd Bedding, T. R., Campante, T. L., Davies, G. R., Elsworth, Y., Handberg, R., Hekker, S., Karoff, C., Kawaler, S. D., Lund, M. N., Lundkvist, M., Metcalfe, T. . S., Miglio, A., Silva Aguirre, V., Stello, D., White, T. R., Boss, A., Devore, E., Gould, A., Prsa, A., Agol, E., Barclay, T., Coughlin, J., Brugamyer, E., Mullally, F., Quintana, E. V., Still, M., Thompson, S. E., Morrison, D., Twicken, J. D., Désert, J.-M., Carter, J., Crepp, J. R., Hébrard, G., Santerne, A., Moutou, C., Sobeck, C., Hudgins, D., Haas, M. R., Robertson, P., Lillo-Box, J., & Barrado, D. 2014, ApJS, 210, 20

Masuda, K. 2014, ApJ, 783, 53

Mathur, S., Huber, D., Batalha, N. M., Ciardi, D. R., Bastien, F. A., Bieryla, A., Buchhave, L. A., Cochran, W. D., Endl, M., Esquerdo, G. A., Furlan, E., Howard, A., Howell, S. B., Isaacson, H., Latham, D. W., MacQueen, P. J., & Silva, D. R. 2017, ApJS, 229, 30

Maxted, P. F. L., Anderson, D. R., Collier Cameron, A., Doyle, A. P., Fumel, A., Gillon, M., Hellier, C., Jehin, E., Lendl, M., Pepe, F., Pollacco, D. L., Queloz, D., Ségransan, D., Smalley, B., Southworth, K., Smith, A. M. S., Triaud, A. H. M. J., Udry, S., & West, R. G. 2013a, PASP, 125, 48

Maxted, P. F. L., Anderson, D. R., Collier Cameron, A., Gillon, M., Hellier, C., Queloz, D., Smalley, B., Triaud, A. H. M. J., West, R. G., Enoch, R., Lister, T. A., Pepe, F., Pollacco, D. L., Ségransan, D., Skillen, I., & Udry, S. 2010, PASP, 122, 1465

Maxted, P. F. L., Anderson, D. R., Doyle, A. P., Gillon, M., Harrington, J., Iro, N., Jehin, E., Lafrenière, D., Smalley, B., & Southworth, J. 2013b, MNRAS, 428, 2645

Maxted, P. F. L., O'Donoghue, D., Morales-Rueda, L., Napiwotzki, R., & Smalley, B. 2007, MNRAS, 376, 919

McTier, M. A. & Kipping, D. M. 2018, Monthly Notices of the Royal Astronomical Society

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, J. Chem. Phys., 21, 1087

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, The journal of chemical physics, 21, 1087

Miller, S. L. & Urey, H. C. 1959, Science, 130, 245

Mohler-Fischer, M., Mancini, L., Hartman, J. D., Bakos, G. Á., Penev, K., Bayliss, D., Jordán, A. ., Csubry, Z., Zhou, G., Rabus, M., Nikolov, N., Brahm, R., Espinoza, N., Buchhave, L. A., Béky, B., Suc, V., Csák, B., Henning, T., Wright, D. J., Tinney, C. G., Addison, B. C., Schmidt, B., Noyes, R. W. ., Papp, I., Lázár, J., Sári, P., & Conroy, P. 2013, A&A, 558, A55

Montet, B. T., Johnson, J. A., Muirhead, P. S., Villar, A., Vassallo, C., Baranec, C., Law, N. M., Riddle, R., Marcy, G. W., Howard, A. W., & Isaacson, H. 2015, ApJ, 800, 134

Morton, T. D. & Winn, J. N. 2014, The Astrophysical Journal, 796, 47

Motalebi, F., Udry, S., Gillon, M., Lovis, C., Ségransan, D., Buchhave, L. A., Demory, B. O., Malavolta, L., Dressing, C. D., Sasselov, D., Rice, K., Charbonneau, D., Collier Cameron, A., Latham, D., Molinari, E., Pepe, F., Affer, L., Bonomo, A. S., Cosentino, R., Dumusque, X., Figueira, P. ., Fiorenzano, A. F. M., Gettel, S., Harutyunyan, A., Haywood, R. D., Johnson, J., Lopez, E., Lopez-Morales, M., Mayor, M., Micela, G., Mortier, A., Nascimbeni, V., Philips, D., Piotto, G., Pollacco, D., Queloz, D., Sozzetti, A., Vanderburg, A., & Watson, C. A. 2015, A&A, 584, A72

Močnik, T., Clark, B., Anderson, D. R., Hellier, C., & Brown, D. J. A. 2015, ArXiv e-prints

Nesvorný, D., Kipping, D., Terrell, D., Hartman , J., Bakos, G. Á., & Buchhave, L. A. 2013, ApJ, 777, 3

Nikolov, N., Sing, D. K., Pont, F., Burrows, A. S., Fortney, J. J., Ballester, G. E., Evans, T. M., Huitson, C. M., Wakeford, H. R., Wilson, P. A., Aigrain, S., Deming, D., Gibson, N. P., Henry, G. W. an d Knutson, H., Lecavelier des Etangs, A., Showman, A. P., Vidal-Madjar, A., & Zahnle, K. 2014, MNRAS, 437, 46

Noack, L., Godolt, M., von Paris, P., Plesa, A.-C., Stracke, B., Breuer, D., & Rauer, H. 2014, Planetary and Space Science, 98, 14

Obermeier, C., Henning, T., Schlieder, J. E., Crossfield, I. J. M., Petigura, E. A., Howard, A. W., Sinukoff, E., Isaacson, H., Ciardi, D. R., David, T. J., Hillenbrand, L. A., Beichman, C. A., Howell, S. B., Horch, E., Everett, M., Hirsch, L., Teske, J., Christiansen, J. L., Lépine, S., Aller, K. M., Liu, M. C., Saglia, R. P., Livingston, J., & Kluge, M. 2016, AJ, 152, 223

Ofir, A., Dreizler, S., Zechmeister, M., & Husser, T. .-O. 2014, A&A, 561, A103

Orosei, R., Lauro, S., Pettinelli, E., Cicchetti, A., Coradini, M., Cosciotti, B., Di Paolo, F., Flamini, E., Mattei, E., Pajola, M., et al. 2018, Science, 361, 490

Parsons, S. G., Gänsicke, B. T., Marsh, T. R., Bergeron, P., Copperwheat, C. M., Dhillon, V. S., Bento, J., Littlefair, S. P., & Schreiber, M. R. 2012a, MNRAS, 426, 1950

Parsons, S. G., Marsh, T. R., Copperwheat, C. M., Dhillon, V. S., Littlefair, S. P., Gänsicke, B. T., & Hickman, R. 2010, MNRAS, 402, 2591

Parsons, S. G., Marsh, T. R., Gänsicke, B. T., Dhillon, V. S., Copperwheat, C. M., Littlefair, S. P., Pyrzas, S., Drake, A. J., Koester, D., Schreiber, M. R., & Rebassa-Mansergas, A. 2012b, MNRAS, 419, 304

Parviainen, H., Gandolfi, D., Deleuil, M., Moutou, C., Deeg, H. J., Ferraz-Mello, S., Samuel, B., Csizmadia, S. ., Pasternacki, T., Wuchterl, G., Havel, M., Fridlund, M., Angus, R., Tingley, B., Grziwa, S., Korth, J., Aigrain, S., Almenara, J. M., Alonso, R., Baglin, A. an d Barros, S. C. C., Bordé, P., Bouchy, F., Cabrera, J., Díaz, R. F., Dvorak, R., Erikson, A., Guillot, T., Hatzes, A., Hébrard, G. a nd Mazeh, T., Montagnier, G., Ofir, A., Ollivier, M., Pätzold, M., Rauer, H., Rouan, D., Santerne, A., & Schneider, J. 2014, A&A, 562, A140

Pätzold, M., Endl, M., Csizmadia, S., Gandolfi, D., Jorda, L., Grziwa, S., Carone, L., Pasternacki, T., Aigrain, S., Almenara, J. M., Alonso, R., Auvergne, M., Baglin, A., Barge, P., Bonomo, A. S., Bordé, P., Bouchy, F., Cabrera, J., Cavarroc, C., Cochran, W. B. a nd Deleuil, M., Deeg, H. J., Díaz, R., Dvorak, R. a nd Erikson, A., Ferraz-Mello, S., Fridlund, M., Gillon, M., Guillot, T., Hatzes, A., Hébrard, G. an d Léger, A., Llebaria, A., Lammer, H., MacQueen, P. J. ., Mazeh, T., Moutou, C., Ofir, A., Ollivier, M., Parviainen, H., Queloz, D., Rauer, H., Rouan, D., Santerne, A., Schneider, J., Tingley, B., Weingrill, J., & Wuchterl, G. 2012, A&A, 545, A6

Pepe, F., Cameron, A. C., Latham, D. W., Molinari, E., Udry, S., Bonomo, A. S., Buchhave, L. A., Charbonneau, D., Cosentino, R., Dressing, C. D., Dumusque, X., Figueira, P., Fiorenzano, A. F. M., Gettel, S., Harutyunyan, A., Haywood, R. D., Horne, K., Lopez-Morales, M., Lovis, C., Malavolta, L., Mayor, M., Micela, G., Motalebi, F., Nascimbeni, V., Phillips, D., Piotto, G., Pollacco, D., Queloz, D., Rice, K., Sasselov, D., Ségransan, D., Sozzetti, A., Szentgyorgyi, A., & Watson, C. A. 2013, Nature, 503, 377

Pepe, F., Lovis, C., Segransan, D., Benz, W., Bouchy, F., Dumusque, X., Mayor, M., Queloz, D., Santos, N., & Udry, S. 2011, Astronomy & Astrophysics, 534, A58

Pepper, J., Siverd, R. J., Beatty, T. G., Gaudi, B. . S., Stassun, K. G., Eastman, J., Collins, K., Latham, D. W., Bieryla, A., Buchhave, L. A., Jensen, E. L. N., Manner, M., Penev, K., Crepp, J. R., Cargile, P. A., Dhital, S., Calkins, M. L., Esquerdo, G. A., Berlind, P., Fulton, B. J., Street, R., Ma, B., Ge, J., Wang, J., Mao, Q., Richert, A. J. W., Gould, A., DePoy, D. L., Kielkopf, J. F., Marshall, J. L., Pogge, R. W., Stefanik, R. P., Trueblood, M., & Trueblood, P. 2013, ApJ, 773, 64

Petigura, E. A., Howard, A. W., Lopez, E. D., Deck, K. M., Fulton, B. J., Crossfield, I. J. M., Ciardi, D. R., Chiang, E., Lee, E. J., Isaacson, H., Beichman, C. A., Hansen, B. M. S., Schlieder, J. E., & Sinukoff, E. 2015, ArXiv e-prints

Planck Collaboration, Ade, P. A. R., Aghanim, N., Alves, M. I. R., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Aussel, H., & et al. 2014, A&A, 571, A1

Pont, F., Bouchy, F., Melo, C., Santos, N. C., Mayor, M., Queloz, D., & Udry, S. 2005, A&A, 438, 1123

Pont, F., Moutou, C., Bouchy, F., Behrend, R., Mayor, M., Udry, S., Queloz, D., Santos, N., & Melo, C. 2006, A&A, 447, 1035

Pyrzas, S., Gänsicke, B. T., Brady, S., Parsons, S. G., Marsh, T. R., Koester, D., Breedt, E., Copperwheat, C. M., Nebot Gómez-Morán, A., Rebassa-Mansergas, A., Schreiber, M. R., & Zorotovic, M. 2012, MNRAS, 419, 817

Pyrzas, S., Gänsicke, B. T., Marsh, T. R., Aungwerojwit, A., Rebassa-Mansergas, A., Rodríguez-Gil, P., Southworth, J., Schreiber, M. R., Nebot Gomez-Moran, A., & Koester, D. 2009, MNRAS, 394, 978

Quintana, E. V., Barclay, T., Raymond, S. N., Rowe, J. F., Bolmont, E., Caldwell, D. A., Howell, S. B. ., Kane, S. R., Huber, D., Crepp, J. R., Lissauer, J. J. a nd Ciardi, D. R., Coughlin, J. L., Everett, M. E., Henze, C. E., Horch, E., Isaacson, H., Ford, E. B., Adams, F. C., Still, M., Hunter, R. C., Quarles, B., & Selsis, F. 2014, Science, 344, 277

Quintana, E. V., Barclay, T., Raymond, S. N., Rowe, J. F., Bolmont, E., Caldwell, D. A., Howell, S. B., Kane, S. R., Huber, D., Crepp, J. R., et al. 2014, Science, 344, 277

Rauer, H., Catala, C., Aerts, C., Appourchaux, T., Benz, W., Brandeker, A., Christensen-Dalsgaard, J., Deleuil, M., Gizon, L., Goupil, M.-J., Güdel, M., Janot-Pacheco, E., Mas-Hesse, M., Pagano, I., Piotto, G., Pollacco, D., Santos, Ċ., Smith, A., Suárez, J.-C., Szabó, R., Udry, S., Adibekyan, V., Alibert, Y., Almenara, J.-M., Amaro-Seoane, P., Eiff, M. A.-v., Asplund, M., Antonello, E., Barnes, S., Baudin, F., Belkacem, K., Bergemann, M., Bihain, G., Birch, A. C., Bonfils, X., Boisse, I., Bonomo, A. S., Borsa, F., Brandão, I. M., Brocato, E., Brun, S., Burleigh, M., Burston, R., Cabrera, J., Cassisi, S., Chaplin, W., Charpinet, S., Chiappini, C., Church, R. P., Csizmadia, S., Cunha, M., Damasso, M., Davies, M. B., Deeg, H. J., Díaz, R. F., Dreizler, S., Dreyer, C., Eggenberger, P., Ehrenreich, D., Eigmüller, P., Erikson, A., Farmer, R., Feltzing, S., de Oliveira Fialho, F., Figueira, P., Forveille, T., Fridlund, M., García, R. A., Giommi, P., Giuffrida, G., Godolt, M., Gomes da Silva, J., Granzer, T., Grenfell, J. L., Grotsch-Noels, A., Günther, E., Haswell, C. A., Hatzes, A. P., Hébrard, G., Hekker, S., Helled, R., Heng, K., Jenkins, J. M., Johansen, A., Khodachenko, M. L., Kislyakova, K. G., Kley, W., Kolb, U., Krivova, N., Kupka, F., Lammer, H., Lanza, A. F., Lebreton, Y., Magrin, D., Marcos-Arenal, P., Marrese, P. M., Marques, J. P., Martins, J., Mathis, S., Mathur, S., Messina, S., Miglio, A., Montalban, J., Montalto, M., Monteiro, M. J. P. F. G., Moradi, H., Moravveji, E., Mordasini, C., Morel, T., Mortier, A., Nascimbeni, V., Nelson, R. P., Nielsen, M. B., Noack, L., Norton, A. J., Ofir, A., Oshagh, M., Ouazzani, R.-M., Pápics, P., Parro, V. C., Petit, P., Plez, B., Poretti, E., Quirrenbach, A., Ragazzoni, R., Raimondo, G., Rainer, M., Reese, D. R., Redmer, R., Reffert, S., Rojas-Ayala, B., Roxburgh, I. W., Salmon, S., Santerne, A., Schneider, J., Schou, J., Schuh, S., Schunker, H., Silva-Valio, A., Silvotti, R., Skillen, I., Snellen, I., Sohl, F., Sousa, S. G., Sozzetti, A., Stello, D., Strassmeier, K. G., Švanda, M., Szabó, G. M., Tkachenko, A., Valencia, D., Van Grootel, V., Vauclair, S. D., Ventura, P., Wagner, F. W., Walton, N. A., Weingrill, J., Werner, S. C., Wheatley, P. J., & Zwintz, K. 2014, Experimental Astronomy, 38, 249

Rauscher, B. J., Bolcar, M. R., Clampin, M., Domagal-Goldman, S. D., McElwain, M. W., Moseley, S., Stahle, C., Stark, C. C., & Thronson, H. A. 2015, in UV/Optical/IR Space Telescopes and Instruments: Innovative Technologies and Concepts VII, Vol. 9602, International Society for Optics and Photonics, 96020D

Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. Á., Bean, J. L., Berta-Thompson, Z. K., Brown, T. M., Buchhave, L., Butler, N. R., Butler, R. P., Chaplin, W. J., Charbonneau, D., Christensen-Dalsgaard, J., Clampin, M., Deming, D., Doty, J., De Lee, N., Dressing, C., Dunham, E. W., Endl, M., Fressin, F., Ge, J., Henning, T., Holman, M. J., Howard, A. W., Ida, S., Jenkins, J., Jernigan, G., Johnson, J. A., Kaltenegger, L., Kawai,

N., Kjeldsen, H., Laughlin, G., Levine, A. M., Lin, D., Lissauer, J. J., MacQueen, P., Marcy, G., McCullough, P. R., Morton, T. D., Narita, N., Paegert, M., Palle, E., Pepe, F., Pepper, J., Quirrenbach, A., Rinehart, S. A., Sasselov, D., Sato, B., Seager, S., Sozzetti, A., Stassun, K. G., Sullivan, P., Szentgyorgyi, A., Torres, G., Udry, S., & Villasenor, J. 2014, in Proc. SPIE, Vol. 9143, Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave, 914320

Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. Á., Bean, J. L., Berta-Thompson, Z. K., Brown, T. M., Buchhave, L., Butler, N. R., Butler, R. P., Chaplin, W. J., Charbonneau, D., Christensen-Dalsgaard, J., Clampin, M., Deming, D., Doty, J., De Lee, N., Dressing, C., Dunham, E. W., Endl, M., Fressin, F., Ge, J., Henning, T., Holman, M. J., Howard, A. W., Ida, S., Jenkins, J. M., Jernigan, G., Johnson, J. A., Kaltenegger, L., Kawai, N., Kjeldsen, H., Laughlin, G., Levine, A. M., Lin, D., Lissauer, J. J., MacQueen, P., Marcy, G., McCullough, P. R., Morton, T. D., Narita, N., Paegert, M., Palle, E., Pepe, F., Pepper, J., Quirrenbach, A., Rinehart, S. A., Sasselov, D., Sato, B., Seager, S., Sozzetti, A., Stassun, K. G., Sullivan, P., Szentgyorgyi, A., Torres, G., Udry, S., & Villasenor, J. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

Rimmer, P. B., Xu, J., Thompson, S. J., Gillen, E., Sutherland, J. D., & Queloz, D. 2018, Science advances, 4, eaar3302

Roberts, G. O., Gelman, A., & Gilks, W. R. 1997, Ann. Appl. Probab., 7, 110

Robitaille, T. P. & Whitney, B. A. 2010, ApJL, 710, L11

Rodriguez, J. E., Colon, K. D., Stassun, K. G., Wright, D., Cargile, P. A., Bayliss, D., Pepper, J., Collins, K. A., Kuhn, R. B., Lund, M. B., Siverd, R. J., Zhou, G., Gaudi, B. S., Tinney, C. G., Penev, K., Tan, T. G., Stockdale, C., Curtis, I. A., James, D., Udry, S., Segransan, D., Bieryla, A., Latham, D. W., Beatty, T. G., Eastman, J. D., Myers, G., Bartz, J., Bento, J., Jensen, E. L. N., Oberst, T. E., & Stevens, D. J. 2015, ArXiv e-prints

Rodriguez, J. E., Zhou, G., Vanderburg, A., Eastman, J. D., Kreidberg, L., Cargile, P. A., Bieryla, A., Latham, D. W., Irwin, J., Mayo, A. W., Calkins, M. L., Esquerdo, G. A., & Mink, J. 2017, AJ, 153, 256

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

Rowe, J. F. & Thompson, S. E. 2015, ArXiv e-prints

Russell, H. N. 1914, Popular Astronomy, 22, 275

Sanchis-Ojeda, R., Fabrycky, D. C., Winn, J. N., Barclay, T., Clarke, B. D., Ford, E. B., Fortney, J. J., Geary, J. C., Holman, M. J., Howard, A. W., Jenkins, J. M., Koch, D., Lissauer, J. J., Marcy, G. W., Mullally, F. a nd Ragozzine, D., Seader, S. E., Still, M., & Thompson, S. E. . 2012, Nature, 487, 449

Santerne, A., Hébrard, G., Deleuil, M., Havel, M., Correia, A. C. M., Almenara, J.-M., Alonso, R., Arnold, L., Barros, S. C. C., Behrend, R., Bernasconi, L., Boisse, I., Bonomo, A. S., Bouchy, F., Bruno, G., Damiani, C., Díaz, R. F., Gravallon, D., Guillot, T., Labrevoir, O., Montagnier, G., Moutou, C., Rinner, C., Santos, N. C., Abe, L., Audejean, M., Bendjoya, P., Gillier, C., Gregorio, J., Martinez, P. a nd Michelet, J., Montaigut, R., Poncy, R., Rivet, J.-P. an d Rousseau, G., Roy, R., Suarez, O., Vanhuysse, M., & Verilhac, D. 2014, A&A, 571, A37

Saumon, D. & Marley, M. S. 2008, ApJ, 689, 1327

Scharf, C. & Cronin, L. 2016, Proceedings of the National Academy of Science, 113, 8127

Schmitt, J. R., Agol, E., Deck, K. M., Rogers, L. A., Gazak, J. Z., Fischer, D. A., Wang, J., Holman, M. J. a nd Jek, K. J., Margossian, C., Omohundro, M. R., Winarski, T., Brewer, J. M., Giguere, M. J., Lintott, C., Lynn, S., Parrish, M., Schawinski, K., Schwamb, M. E., Simpson, R., & Smith, A. M. 2014, ApJ, 795, 167

Schopf, J. W., Kudryavtsev, A. B., Czaja, A. D., & Tripathi, A. B. 2007, Precambrian Research, 158, 141

Schwarz, G. 1978, The Annals of Statistics, 6, 461

Seager, S. 2014, Proceedings of the National Academy of Sciences, 111, 12634

Seager, S., Deming, D., & Valenti, J. A. 2009, Astrophysics and Space Science Proceedings, 10, 123

Seager, S. & Hui, L. 2002, The Astrophysical Journal, 574, 1004

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

Seager, S. & Sasselov, D. D. 2000, ApJ, 537, 916

Segura, A., Kasting, J. F., Meadows, V., Cohen, M., Scalo, J., Crisp, D., Butler, R. A., & Tinetti, G. 2005, Astrobiology, 5, 706

Sheppard, S. S. 2016, The Giant Planet & Moon Page

Sheppard, S. S., Jewitt, D., & Kleyna, J. 2005, AJ, 129, 518

—. 2006, AJ, 132, 171

Sheppard, S. S. & Jewitt, D. C. 2003, Nature, 423, 261

Simpson, F. 2016, MNRAS, 456, L59

Sinukoff, E., Howard, A. W., Petigura, E. A., Schlieder, J. E., Crossfield, I. J. M., Ciardi, D. R., Fulton, B. J., Isaacson, H., Aller, K. M., Baranec, C., Beichman, C. A., Hansen, B. M. S., Knutson, H. A., Law, N. M., Liu, M. C., & Riddle, R. 2015, ArXiv e-prints

Smalley, B., Anderson, D. R., Collier-Cameron, A., Doyle, A. P., Fumel, A., Gillon, M., Hellier, C., Jehin, E., Lendl, M., Maxted, P. F. L., Pepe, F., Pollacco, D., Queloz, D., Ségransan, D., Smith, A. M. S., Southworth, J., Triaud, A. H. M. J., Udry, S., & West, R. G. 2012, A&A, 547, A61

Smirnov, N. 1948, The annals of mathematical statistics, 19, 279

Smith, A. M. S. 2015, Acta Astron., 65

Smith, A. M. S., Anderson, D. R., Armstrong, D. J., Barros, S. C. C., Bonomo, A. S., Bouchy, F., Brown, D. J. A., Collier Cameron, A., Delrez, L., Faedi, F., Gillon, M., Gómez Maqueo Chew, Y., Hébrard, G., Jehin, E., Lendl, M., Louden, T. M. an d Maxted, P. F. L., Montagnier, G., Neveu-VanMalle, M., Osborn, H. P., Pepe, F., Pollacco, D., Queloz, D., Rostron, J. W., Segransan, D., Smalley, B., Triaud, A. H. M. J., Turner, O. D., Udry, S., Walker, S. R., West, R. G., & Wheatley, P. J. 2014, A&A, 570, A64

Smith, A. M. S., Anderson, D. R., Bouchy, F., Collier Cameron, A., Doyle, A. P., Fumel, A., Gillon, M., Hébrard, G., Hellier, C., Jehin, E., Lendl, M., Maxted, P. F. L., Moutou, C., Pepe, F., Pollacco, D., Queloz, D., Santerne, A., Segransan, D. a nd Smalley, B., Southworth, J., Triaud, A. H. M. J., Udry, S., & West, R. G. 2013, A&A, 552, A120

Smith, A. M. S., Anderson, D. R., Collier Cameron, A. a nd Gillon, M., Hellier, C., Lendl, M., Maxted, P. F. L. an d Queloz, D., Smalley, B., Triaud, A. H. M. J., West, R. G., Barros, S. C. C., Jehin, E., Pepe, F., Pollacco, D., Segransan, D., Southworth, J., Street, R. A., & Udry, S. 2012, AJ, 143, 81

Sober, E. 2008, Evidence and evolution: The logic behind the science (Cambridge University Press)

Southworth, J. 2008, MNRAS, 386, 1644

—. 2009, MNRAS, 394, 272

—. 2010, MNRAS, 408, 1689

—. 2011, MNRAS, 417, 2166

—. 2012, MNRAS, 426, 1291

Southworth, J., Bruni, I., Mancini, L., & Gregorio, J. 2012a, MNRAS, 420, 2580

Southworth, J., Hinse, T. C., Burgdorf, M., Calchi Novati, S., Dominik, M., Galianni, P., Gerner, T., Giannini, E., Gu, S.-H., Hundertmark, M., Jørgensen, U. G., Juncher, D., Kerins, E., Mancini, L., Rabus, M., Ricci, D., Schäfer, S., Skottfelt, J., Tregloan-Reed , J., Wang, X.-B., Wertz, O., Alsubai, K. A., Andersen, J. M., Bozza, V., Bramich, D. M., Browne, P., Ciceri, S., D'Ago, G., Damerdji, Y., Diehl, C., Dodds, P., Elyiv, A., Fang, X.-S., Finet, F., Figuera Jaimes, R. a nd Hardis, S., Harpsøe, K., Jessen-Hansen, J., Kains, N., Kjeldsen, H., Korhonen, H., Liebig, C., Lund, M. N., Lundkvist, M., Mathiasen, M., Penny, M. T., Popovas, A., Prof., S., Rahvar, S., Sahu, K., Scarpetta, G., Schmidt, R. W., Schönebeck, F., Snodgrass, C., Street, R. A., Surdej, J., & Tsapras, Y. a nd Vilela, C. 2014, MNRAS, 444, 776

Southworth, J., Hinse, T. C., Dominik, M., Fang, X. .-S., Harpsøe, K., Jørgensen, U. G., Kerins, E., Liebig, C., Mancini, L., Skottfelt, J., Anderson, D. R., Smalley, B., Tregloan-Reed, J., Wertz, O., Alsubai, K. A., Bozza, V., Calchi Novati, S., Dreizler, S., Gu, S.-H. a nd Hundertmark, M., Jessen-Hansen, J., Kains, N., Kjeldsen, H., Lund, M. N., Lundkvist, M., Mathiasen, M., Penny, M. T., Rahvar, S., Ricci, D., Scarpetta, G., Snodgrass, C., & Surdej, J. 2012b, MNRAS, 426, 1338

Southworth, J., Mancini, L., Browne, P., Burgdorf, M., Calchi Novati, S., Dominik, M., Gerner, T., Hinse, T. C. ., Jørgensen, U. G., Kains, N., Ricci, D., Schäfer, S., Schönebeck, F., Tregloan-Reed, J., Alsubai, K. A., Bozza, V., Chen, G., Dodds, P., Dreizler, S., Fang, X.-S., Finet, F., Gu, S.-H., Hardis, S., Harpsøe, K., Henning, T., Hundertmark, M., Jessen-Hansen, J., Kerins, E., Kjeldsen, H., Liebig, C., Lund, M. N., Lundkvist, M., Mathiasen, M. a nd Nikolov, N., Penny, M. T., Proft, S., Rahvar, S., Sahu, K., Scarpetta, G., Skottfelt, J., Snodgrass, C. a nd Surdej, J., & Wertz, O. 2013, MNRAS, 434, 1300

Southworth, J., Mancini, L., Ciceri, S., Budaj, J., Dominik, M., Figuera Jaimes, R., Haugbølle, T., Jørgensen, U. G., Popovas, A., Rabus, M., Rahvar, S., von Essen, C., Schmidt, R. W., Wertz, O. an d Alsubai, K. A., Bozza, V., Bramich, D. M., Calchi Novati , S., D'Ago, G., Hinse, T. C., Henning, T., Hundertmark, M. a nd Juncher, D., Korhonen, H., Skottfelt, J., Snodgrass, C., Starkey, D., & Surdej, J. 2015a, MNRAS, 447, 711

Southworth, J., Mancini, L., Maxted, P. F. L., Bruni, I., Tregloan-Reed, J., Barbieri, M., Ruocco, N., & Wheatley, P. J. 2012c, MNRAS, 422, 3099

Southworth, J., Mancini, L., Tregloan-Reed, J., Calchi Novati, S., Ciceri, S., D'Ago, G., Delrez, L. an d Dominik, M., Evans, D. F., Gillon, M., Jehin, E., Jørgensen, U. G., Haugbølle, T., Lendl, M., Arena, C., Barbieri, L., Barbieri, M., Corfini, G., Lopresti, C., Marchini, A., Marino, G., Alsubai, K. A., Bozza, V., Bramich, D. M., Jaimes, R. F., Hinse, T. C., Henning, T., Hundertmark, M., Juncher, D., Korhonen, H., Popovas, A., Rabus, M., Rahvar, S., Schmidt, R. W., Skottfelt, J., Snodgrass, C., Starkey, D., & Surdej, J. a nd Wertz, O. 2015b, MNRAS, 454, 3094

Southworth, J., Tregloan-Reed, J., Andersen, M. I., Calchi Novati, S., Ciceri, S., Colque, J. P., D'Ago, G., Dominik, M., Evans, D., Gu, S.-H., Herrera-Cruces, A., Hinse, T. C., Jorgensen, U. G., Juncher, D., Kuffmeier, M., Mancini, L., Peixinho, N. a nd Popovas, A., Rabus, M., Skottfelt, J., Tronsgaard, R. a nd Unda-Sanzana, E., Wang, X.-B., Wertz, O., Alsubai, K. A. ., Andersen, J. M., Bozza, V., Bramich, D. M., Burgdorf, M. ., Damerdji, Y., Diehl, C., Elyiv, A., Figuera Jaimes, R., Haugbolle, T., Hundertmark, M., Kains, N., Kerins, E. a nd Korhonen, H., Liebig, C., Mathiasen, M., Penny, M. T. a nd Rahvar, S., Scarpetta, G., Schmidt, R. W., Snodgrass, C. ., Starkey, D., Surdej, J., Vilela, C., von Essen, C., & Wang, Y. 2015c, ArXiv e-prints

Spake, J. J., Brown, D. J. A., Doyle, A. P., Hébrard, G., McCormac, J., Armstrong, D. J., Pollacco, D., Gómez Maqueo Chew, Y., Anderson, D. R., Barros, S. C. C., Bouchy, F., Boumis, P., Bruno, G., Collier Cameron, A., Courcol, B., Davies, G. R., Faedi, F., Hellier, C., Kirk, J., Lam, K. W. F., Liakos, A., Louden, T., Maxted, P. F. L., Osborn, H. P., Palle, E., Prieto Arranz, J., Udry, S., Walker, S. R. a nd West, R. G., & Wheatley, P. J. 2015, ArXiv e-prints

Spiegel, D. S., Burrows, A., & Milsom, J. A. 2011, ApJ, 727, 57

Spiegel, D. S., Fortney, J. J., & Sotin, C. 2014, Proceedings of the National Academy of Sciences, 111, 12622

Spiegel, D. S. & Turner, E. L. 2012, Proceedings of the National Academy of Sciences, 109, 395

Spiegelhalter, D. J., Best, N. G., & Carlin, B. P. 2002, Journal of the Royal Statistical Society, Series B, 64, 583

Steffen, J. H. 2016, MNRAS, 457, 4384

Swift, J. J., Bottom, M., Johnson, J. A., Wright, J. T., McCrady, N., Wittenmyer, R. A., Plavchan, P., Riddle, R., Muirhead, P. S., Herzig, E., Myles, J., Blake, C. H., Eastman, J., Beatty, T. G., Barnes, S. I., Gibson, S. R., Lin, B., Zhao, M., Gardner, P., Falco, E., Criswell, S., Nava, C., Robinson, C., Sliski, D. H., Hedrick, R., Ivarsen, K., Hjelstrom, A., de Vera, J., & Szentgyorgyi, A. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 027002

Tasker, E., Tan, J., Heng, K., Kane, S., Spiegel, D., Brasser, R., Casey, A., Desch, S., Dorn, C., Hernlund, J., Houser, C., Laneuville, M., Lasbleis, M., Libert, A.-S., Noack, L., Unterborn, C., & Wicks, J. 2017, Nature Astronomy, 1, 0042

Tegmark, M., Strauss, M. A., Blanton, M. R., Abazajian, K., Dodelson, S., Sandvik, H., Wang, X., Weinberg, D. H., Zehavi, I., Bahcall, N. A., Hoyle, F., Schlegel, D., Scoccimarro, R., Vogeley, M. S., Berlind, A., Budavari, T., Connolly, A., Eisenstein, D. J., Finkbeiner, D., Frieman, J. A., Gunn, J. E., Hui, L., Jain, B., Johnston, D., Kent, S., Lin, H., Nakajima, R., Nichol, R. C., Ostriker, J. P., Pope, A., Scranton, R., Seljak, U., Sheth, R. K., Stebbins, A., Szalay, A. S., Szapudi, I., Xu, Y., Annis, J., Brinkmann, J., Burles, S., Castander, F. J., Csabai, I., Loveday, J., Doi, M., Fukugita, M., Gillespie, B., Hennessy, G., Hogg, D. W., Ivezić, Ž., Knapp, G. R., Lamb, D. Q., Lee, B. C., Lupton, R. H., McKay, T. A., Kunszt, P., Munn, J. A., O'Connell, L., Peoples, J., Pier, J. R., Richmond, M., Rockosi, C., Schneider, D. P., Stoughton, C., Tucker, D. L., vanden Berk, D. E., Yanny, B., & York, D. G. 2004, Phys. Rev. D, 69, 103501

Torres, G., Andersen, J., & Giménez, A. 2010, A&A Rev., 18, 67

Torres, G., Kipping, D. M., Fressin, F., Caldwell, D. A., Twicken, J. D., Ballard, S., Batalha, N. M., Bryson, S. T., Ciardi, D. R., Henze, C. E., Howell, S. B., Isaacson, H. T., Jenkins, J. M., Muirhead, P. S., Newton, E. R., Petigura, E. A., Barclay, T., Borucki, W. J., Crepp, J. R., Everett, M. E., Horch, E. P., Howard, A. W., Kolbl, R., Marcy, G. W. an d McCauliff, S., & Quintana, E. V. 2015, ApJ, 800, 99

Tregloan-Reed, J. & Southworth, J. 2013, MNRAS, 431, 966

Tregloan-Reed, J., Southworth, J., Burgdorf, M., Novati, S. C., Dominik, M., Finet, F., Jørgensen, U. G., Maier, G., Mancini, L., Prof, S., Ricci, D., Snodgrass, C., Bozza, V., Browne, P., Dodds, P., Gerner, T., Harpsøe, K., Hinse, T. C., Hundertmark, M., Kains, N., Kerins, E., Liebig, C., Penny, M. T., Rahvar, S., Sahu, K., Scarpetta, G., Schäfer, S. an d Schönebeck, F., Skottfelt, J., & Surdej, J. 2015, MNRAS, 450, 1760

Tuomi, M. 2012, A&A, 543, A52

Turner, O. D., Anderson, D. R., Collier Cameron, A., Delrez, L., Gillon, M., Hellier, C., Jehin, E., Lendl, M., Maxted, P. F. L., Pepe, F., Pollacco, D., Queloz, D., Sègransan, D., Smalley, B., Smith, A. M. . S., Triaud, A. H. M. J., Udry, S., & West, R. G. 2015, ArXiv e-prints

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

Valley, J. W., Cavosie, A. J., Ushikubo, T., Reinhard, D. A., Lawrence, D. F., Larson, D. J., Clifton, P. H., Kelly, T. F., Wilde, S. A., Moser, D. E., et al. 2014, Nature Geoscience, 7, 219

Van Eylen, V., Agentoft, C., Lundkvist, M. S., Kjeldsen, H., Owen, J. E., Fulton, B. J., Petigura, E., & Snellen, I. 2018, Monthly Notices of the Royal Astronomical Society, 479, 4786

Van Grootel, V., Gillon, M., Valencia, D., Madhusud han, N., Dragomir, D., Howe, A. R., Burrows, A. S., Demory, B.-O. ., Deming, D., Ehrenreich, D., Lovis, C., Mayor, M., Pepe, F., Queloz, D., Scuflaire, R., Seager, S., Segransan, D., & Udry, S. 2014, ApJ, 786, 2

Vanderburg, A., Montet, B. T., Johnson, J. A., Buchhave, L. A., Zeng, L., Pepe, F., Collier Cameron, A. ., Latham, D. W., Molinari, E., Udry, S., Lovis, C., Matthews, J. M., Cameron, C., Law, N., Bowler, B. P. an d Angus, R., Baranec, C., Bieryla, A., Boschin, W., Charbonneau, D., Cosentino, R., Dumusque, X., Figueira, P., Guenther, D. B., Harutyunyan, A., Hellier, C., Kuschnig, R., Lopez-Morales, M., Mayor, M., Micela, G., Moffat, A. F. J., Pedani, M., Phillips, D. F., Piotto, G., Pollacco, D., Queloz, D. a nd Rice, K., Riddle, R., Rowe, J. F., Rucinski, S. M., Sasselov, D., Ségransan, D., Sozzetti, A., Szentgyorgyi, A., Watson, C., & Weiss, W. W. 2015, ApJ, 800, 59

Weiss, L. M. & Marcy, G. W. 2014, ApJL, 783, L6

Weiss, L. M., Marcy, G. W., Rowe, J. F., Howard, A. . W., Isaacson, H., Fortney, J. J., Miller, N., Demory, B.-O., Fischer, D. A., Adams, E. R., Dupree, A. K., Howell, S. B., Kolbl, R., Johnson, J. A., Horch, E. P., Everett, M. E., Fabrycky, D. C., & Seager, S. 2013, ApJ, 768, 14

Welsh, W. F., Orosz, J. A., Carter, J. A., Fabrycky , D. C., Ford, E. B., Lissauer, J. J., Prša, A., Quinn, S. N., Ragozzine, D., Short, D. R., Torres, G. a nd Winn, J. N., Doyle, L. R., Barclay, T., Batalha, N., Bloemen, S., Brugamyer, E., Buchhave, L. A., Caldwell, C., Caldwell, D. A., Christiansen, J. L., Ciardi, D. R., Cochran, W. D., Endl, M., Fortney, J. J., Gautier, III, T. N., Gilliland, R. L., Haas, M. R., Hall, J. R., Holman, M. J., Howard, A. W., Howell, S. B. ., Isaacson, H., Jenkins, J. M., Klaus, T. C., Latham, D. W., Li, J., Marcy, G. W., Mazeh, T., Quintana, E. V., Robertson, P., Shporer, A., Steffen, J. H., Windmiller, G., Koch, D. G., & Borucki, W. J. 2012, Nature, 481, 475

Wesson, P. S. 2010, SSR, 156, 239

West, R. G., Hellier, C., Almenara, J.-M., Anderson , D. R., Barros, S. C. C., Bouchy, F., Brown, D. J. A., Collier Cameron, A., Deleuil, M., Delrez, L., Doyle, A. P., Faedi, F., Fumel, A., Gillon, M., Gómez Maqueo Chew, Y., Hébrard, G., Jehin, E., Lendl, M., Maxted, P. F. L., Pepe, F., Pollacco, D., Queloz, D., Ségransan, D., Smalley, B., Smith, A. M. . S., Southworth, J., Triaud, A. H. M. J., & Udry, S. 2016, A&A, 585, A126

Wilde, S. A., Valley, J. W., Peck, W. H., & Graham, C. M. 2001, Nature, 409, 175

Williams, D. R. 2016, NASA Planetary Fact Sheet

Wolfgang, A., Rogers, L. A., & Ford, E. B. 2016, The Astrophysical Journal, 825, 19

Wolfgang, A., Rogers, L. A., & Ford, E. B. 2016, ApJ, 825, 19

Wolszczan, A. & Frail, D. A. 1992, Nature, 355, 145

Zapolsky, H. S. & Salpeter, E. E. 1969, ApJ, 158, 809

Zeng, L. & Sasselov, D. 2013, PASP, 125, 227

Zeng, L., Sasselov, D. D., & Jacobsen, S. B. 2016, The Astrophysical Journal, 819, 127

Zeng, L., Sasselov, D. D., & Jacobsen, S. B. 2016, ApJ, 819, 127

Zhou, G., Bayliss, D., Hartman, J. D., Bakos, G. Á., Penev, K., Csubry, Z., Tan, T. G., Jordán, A., Mancini, L., Rabus, M., Brahm, R., Espinoza, N., Mohler-Fischer, M., Ciceri, S., Suc, V., Csák, B., Henning, T., & Schmidt, B. 2014a, MNRAS, 437, 2831

Zhou, G., Bayliss, D., Penev, K., Bakos, G. Á., Hartman, J. D., Jordán, A., Mancini, L., Mohler, M., Csubry, Z., Ciceri, S., Brahm, R., Rabus, M., Buchhave, L., Henning, T., Suc, V., Espinoza, N., Béky, B., Noyes, R. W., Schmidt, B. a nd Butler, R. P., Shectman, S., Thompson, I., Crane, J. an d Sato, B., Csák, B., Lázár, J., Papp, I., Sári, P., & Nikolov, N. 2014b, AJ, 147, 144

Zhu, W., Huang, C. X., Zhou, G., & Lin, D. 2014, The Astrophysical Journal, 796, 67

Zuluaga, J. I., Kipping, D. M., Sucerquia, M., & Alvarado, J. A. 2015, ApJL, 803, L14