Preprint typeset using LaTeX style emulateapj v. 04/17/13

HOW TO DISTINGUISH BETWEEN CLOUDY MINI-NEPTUNES AND WATER/VOLATILE-DOMINATED SUPER-EARTHS

Björn Benneke and Sara Seager
Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; USA

Abstract

One of the most profound questions about the newly discovered class of low-density super-Earths is whether these exoplanets are predominately H_2-dominated mini-Neptunes or volatile-rich worlds with gas envelopes dominated by H_2O, CO_2, CO, CH_4, or N_2. Transit observations of the super-Earth GJ 1214b rule out cloud-free H_2-dominated scenarios, but are not able to determine whether the lack of deep spectral features is due to high-altitude clouds or the presence of a high mean molecular mass atmosphere.

Here, we demonstrate that one can unambiguously distinguish between cloudy mini-Neptunes and volatile-dominated worlds based on the differences in the wing steepness and relative depths of water absorption features in moderate-resolution NIR transmission spectra ($R \sim 100$). In a numerical retrieval study, we show for GJ 1214b that an unambiguous distinction between a cloudy H_2-dominated atmosphere and cloud-free H_2O atmosphere will be possible if the uncertainties in the spectral transit depth measurements can be reduced by a factor of ~ 3 compared to the published *HST WFC3* and *VLT* transit observations by Berta et al. (2012) and Bean et al. (2010). We argue that the required precision for the distinction may be achievable with currently available instrumentation by stacking $10-15$ repeated transit observations. We provide a scaling law that scales our quantitative results to other transiting super-Earths and Neptunes such as HD 97658b, 55 Cnc e, and GJ 436b.

The analysis in this work is performed using an improved version of our Bayesian atmospheric retrieval framework. The new framework not only constrains the gas composition and cloud/haze parameters, but also determines our confidence in having detected molecules and cloud/haze species through Bayesian model comparison. Using the Bayesian tool, we demonstrate quantitatively that the subtle transit depth variation in the Berta et al. (2012) data is not sufficient to claim the detection of water absorption.

1. INTRODUCTION

Super-Earth exoplanets, with masses between 1 and 10 Earth masses, lie in the intermediate mass range between terrestrial planets and gas and ice giants in the solar system. Compelling questions arise as to the composition and nature of these objects and whether they are capable of harboring life. According to theoretical studies, (e.g., Seager et al. 2007; Rogers & Seager 2010a; Nettelmann et al. 2011) many super-Earth exoplanets show a bulk density that is high enough to require a larger ice or rock fraction than the solar system ice giants, but far too low to be explained by an entirely Earth-like rocky composition. Rogers & Seager (2010b) showed that their bulk density may, instead, be explained by either planets that have accreted and maintained a thick H_2/He envelope atop an ice and rock core or, alternatively, by a new class of "water worlds" which contain a large fraction of water or ices in their interior and are surrounded by a dense water vapor atmosphere (Kuchner (2003) and Léger et al. (2004)).

One way of answering questions about the nature and habitability of super-Earths is to identify their atmospheric thicknesses and molecular compositions by observing their transmission and/or thermal emission spectra. Miller-Ricci et al. (2009) showed that cloud-free, hydrogen-dominated atmospheres would display absorption features in the transmission spectrum that are several times larger than those expected for atmospheres dominated by water vapor, CO_2, CO, CH_4, or N_2 due to the lower mean molecular mass and resulting larger scale height.

Many observational attempts to detect an atmosphere around the super-Earth GJ 1214b and characterize its composition have been made, but the individual observational data sets were found to be insufficient to identify the presence of atmospheric features to within observational uncertainty (Bean et al. 2010, 2011; Désert et al. 2011; Carter et al. 2011; Crossfield et al. 2011; Berta et al. 2011, 2012; de Mooij et al. 2012; Teske et al. 2013). An initial finding of a difference in the transit depths between the J and Ks band by Croll et al. (2011) could not be confirmed by Bean et al. (2011).

The absence of deep features in the transmission spectrum of GJ 1214b rules out the presence of a cloud-free hydrogen-dominated atmosphere, but the obtained observational data were shown to be compatible with high mean molecular mass atmospheres, such as water vapor-dominated atmospheres, as well as with a hydrogen-dominated atmosphere in the presence of high altitude clouds (Berta et al. 2012). The interpretation of the observational data for GJ 1214b revealed the limitations of the absorption feature depths as a measure of the atmosphere's hydrogen content in the presence of clouds. Theoretical studies of the atmosphere loss were conducted attempting to understand the stability of different atmospheres on highly irradiated super-Earths (e.g., Heng & Kopparla 2012; Fortney et al. 2013; Kurokawa & Kaltenegger 2013). It appears inevitable, however, that

bbenneke@mit.edu

eventually we need observational proof to understand or confirm the nature of low-density super-Earths.

For general atmospheres that may contain clouds, Benneke & Seager (2012) showed that the unambiguous effect of the mean molecular mass $\mu_{\rm ave}$ on the transmission spectrum is not the overall depths of the molecular features, but the gradient $\mathrm{d}R_{P,\lambda}/\mathrm{d}(\ln\sigma_\lambda)$ with which the observed planetary radius $R_{P,\lambda}$ changes as the atmospheric opacity σ_λ changes across the spectrum:

$$\mu_{\rm ave} = \frac{k_B T}{g_{\rm p}} \left(\frac{\mathrm{d}R_{P,\lambda}}{\mathrm{d}(\ln\sigma_\lambda)} \right)^{-1} \times \left(1 \pm \frac{\delta T}{T} \right). \qquad (1)$$

The surface gravity $g_{\rm p}$ can be determined directly from the transit light curve and radial velocity measurements (Winn 2011), k_B is Boltzmann's constant, and the atmospheric temperature T can be approximated by the equilibrium temperature or modeled by a radiative-convective model. The term $(1 \pm \delta T/T)$ accounts for the inherent uncertainty on the mean molecular mass due to the uncertainty, δT, in estimating the temperature T at the planetary radius $R_{P,\lambda}$.

In practice, an estimate of the mean molecular mass can be determined at visible or NIR wavelengths by measuring the steepness of molecular feature wings or comparing the relative transit depths in two or more absorption features of the same molecule. The decrease in opacity σ_λ from the center of the absorption features to wing can be modeled based on molecular line list of the absorber.

Measuring the slope of gaseous Rayleigh scattering signature at UV/visible wavelengths also provide constraints on the mean molecular mass; however, cloud and haze opacities can mask the signature of gaseous Rayleigh scattering and the distinction between a shallow slope due to a high mean molecular mass or clouds may be difficult.

Equation (1) demonstrates that, if sufficient observations of the transmission spectrum are available, the mean molecular mass can determined to the same relative precision as that at which we are able to estimate the atmospheric temperature T. Since the mean molecular mass varies by a factor of ~ 8 or more between hydrogen-rich atmospheres and atmospheres dominated by water vapor or other volatiles, we can distinguish between cloudy hydrogen-rich and water-rich atmospheres even if the temperature T is known only with an uncertainty of several tens of percent.

The ability to measure the mean molecular mass at NIR wavelengths is of particular interest for the near-term characterizations of super-Earths because planets orbiting small M-dwarfs are the ones with the strongest transit signatures and are most amenable to study at infrared wavelengths due to the low stellar flux of M-dwarfs at short wavelengths. In this work, we demonstrate using quantitative simulations that NIR transit observations near the peak of the stellar spectrum of M-stars provide a practical means to distinguish between cloudy hydrogen-dominated atmospheres and atmospheres dominated by water vapor or other volatiles. We determine what precision in the transit depth measurements is required (1) to detect the absorption features in the transmission spectrum of a water-dominated super-Earth and (2) to unambiguously distinguish between water-dominated atmospheres and hydrogen-dominated atmospheres based on the different wing steepnesses of the water absorption bands.

The paper outline is as follows. In Section 2, we describe the Bayesian framework used to identify the presence of molecular species, quantify the statistical significances of molecular detections, and constrain the abundance of the molecular species in the atmosphere. Section 2 presents a quantitative picture of what observations and noise levels are required for the super-Earth GJ 1214b to detect the presence of water vapor with high confidence as well as to distinguish between cloudy, hydrogen-dominated atmospheres and water-dominated atmospheres. In Section 4, we present a retrieval analysis of the spectral observations by *HST WFC3* spectrum of GJ 1214b by Berta et al. (2012). Section 5 provides a scaling law to scale the quantitative results to super-Earth and Neptunes such as HD 97658b, 55 Cnc e, and GJ 436b. We also discuss the near-term feasibility of the proposed study. Finally, we present our summary and conclusions in Section 6.

2. METHODS

The main goal of this work is to determine the level of precision in NIR super-Earth transmission spectra that is required to unambiguously distinguish between atmospheres dominated by hydrogen/helium and atmospheres dominated by water vapor or other volatiles such as CO_2, CO, CH_4, or N_2. We address this question by computing *synthetic* transit observations derived from super-Earth model transmission spectra and analyzing them using a Bayesian atmospheric retrieval framework.

The Bayesian retrieval method employed in this work builds upon ideas introduced in Benneke & Seager (2012), but was extended by a Bayesian model comparison framework that enables one to rationally decide which molecular species and types of aerosols are present in the atmosphere. We employ the nested sampling technique (Skilling 2004; Feroz & Hobson 2008) to efficiently compute and compare the Bayesian evidences of retrieval models with different complexities. Once a retrieval model is identified that is adequate in light of the data, the constraints on atmospheric parameters are inferred from the joint posterior probability distribution obtained as a by-product of the nested sampling calculation.

2.1. *Atmospheric "Forward" Model*

We use the 1D exoplanetary atmosphere "forward" model originally described in Benneke & Seager (2012) to compute model transmission spectra and synthetic observations. Our model uses line-by-line radiative transfer in local thermodynamic equilibrium, hydrostatic equilibrium, and a temperature-pressure profile consistent with the atmospheric composition.

The input to the atmospheric "forward" model is an adaptable set of free model parameters describing the gas composition and aerosol properties in the model atmosphere. A difference between this model and the model used in Benneke & Seager (2012) is that the aerosols considered in this work encompass small particle hazes as well as larger cloud particles. Absorption and scattering of the particles are modeled using Mie theory for

spherical particles (Hansen & Travis 1974). The complex refractive indices of the condensate species Potassium Chloride (KCl) or Zinc Sulfide (ZnS) used in this work for GJ 1214b are taken from Querry (1987).

Synthetic observations are generated by first specifying the chemical composition and aerosol opacities of the individual atmospheric layers. Synthetic observations are then derived from the model transmission spectra by integrating the model spectra over flat instrument response functions with spectral coverage equivalent to the spectral points from the *VLT* observations by Bean et al. (2010) and the *HST WFC3* observations by Berta et al. (2012). Gaussian noise ranging from ~ 180 ppm down to 35 ppm is added to the data to generate synthetic observations of different precisions.

The two main scenarios we aim to unambiguously distinguish in this work are water- or volatile-dominated atmospheres and hydrogen-rich atmospheres with high-altitude clouds. In addition, we would like to contrast these two scenarios from a flat transmission spectra which would be observed if the planet lacks a gaseous atmosphere or clouds are present at extremely high altitude such that all molecular features are muted. As representative example for these three scenarios, we select a cloud-free water-dominated atmosphere composed (95% H_2O + 5% CO_2), a methane-depleted solar metallicity atmosphere with a cloud deck at 10 mbar, and a featureless spectrum. The level of the high altitude clouds in the solar composition scenario is chosen such that the overall depths of the absorption features resemble that of a cloud-free water-rich atmosphere. Methane is removed from the atmospheric scenario to demonstrate that the distinction between water/volatile-dominated scenarios and hydrogen-dominated scenarios is possible based solely on the wing steepnesses and relative sizes of the water features.

The atmospheric scenarios for the synthetic observations are chosen to be plausible scenarios, but are not calculated from a fully self-consistent model. The goal of this work is to demonstrate the retrieval method for exoplanetary atmospheres for which we do not have a full understanding of the physical and chemical behavior prior to the observations.

2.2. *Bayesian Atmospheric Retrieval*

Atmospheric retrieval aims at solving the inverse problem of atmospheric "forward" modeling: "Given an observed planetary spectrum, what are the properties of the planetary atmosphere?" The atmospheric retrieval problem can be solved using *Bayesian parameter estimation* by first choosing a retrieval model that defines the hypothesis space in the form of a set of atmospheric parameters and then deriving constraints on those parameters (Benneke & Seager 2012).

In choosing the atmospheric retrieval model, however, questions arise as to how much complexity and how many free parameters should be included in the retrieval model. In this work, we introduce *Bayesian model comparison* to determine which molecular gases and types of aerosols can be inferred from the data and need to be included in the analysis. The approach enables us to rationally adapt the complexity of the retrieval model to the available data. Exquisite observational data with high signal-to-noise ratio (S/N) and high spectral resolution, as are currently only available for solar system planets, allow the inference of the detailed abundance profiles of the molecular species, the temperature structure, and the presence of particles in the atmospheres. At lower S/N and sparse spectral coverage, however, complex models with large numbers of free parameters would overfit the available data and introduce strong degeneracies. It is therefore necessary to adjust the number of free parameters in the retrieval model according to the amount and precision of observational data available.

2.2.1. *Bayesian Estimation of Atmospheric Parameters*

An atmospheric retrieval model M_i in our Bayesian framework is defined as a set of atmospheric parameters $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_N]$ and their joint prior probability distribution $\pi(\boldsymbol{\theta}|M_i)$. The atmospheric parameters considered in the retrieval models for transmission spectra are the mole fractions (or volume mixing ratios) of the atmospheric gases, the haze or cloud top pressure, and the radius at a reference pressure level. Since we generally have little prior knowledge of the state of the exoplanetary atmosphere, we assign non-informative (uniform) priors to the atmospheric parameters (Table 1). We include the Planetary Bond albedo as a free parameters to account for the uncertainty in the temperature structure introduced by our ignorance about how much of solar radiation is reflected by the planet.

Once a retrieval model is identified that is adequate in light of the data, we can infer the constraints on the model parameters by computing their joint posterior probability distribution $p(\boldsymbol{\theta}|M_i, \boldsymbol{D})$ (Benneke & Seager 2012). Using Bayes' Law, we write the posterior distribution of the atmospheric parameters as

$$p(\boldsymbol{\theta}|M_i, \boldsymbol{D}) = \frac{\pi(\boldsymbol{\theta}|M_i)\mathcal{L}\left(\boldsymbol{D}|M_i, \boldsymbol{\theta}\right)}{\mathcal{Z}\left(\boldsymbol{D}|M_i\right)}, \tag{2}$$

where $\mathcal{L}\left(\boldsymbol{D}|M_i, \boldsymbol{\theta}\right)$ is the likelihood function and $\mathcal{Z}\left(\boldsymbol{D}|M_i\right)$ is the Bayesian evidence of the model M_i. The likelihood function $\mathcal{L}\left(\boldsymbol{D}|M_i, \boldsymbol{\theta}\right)$ is the probability of observing the data $\boldsymbol{D}$, given that the atmospheric parameters are $\boldsymbol{\theta}$. The likelihood function is modeled using the atmospheric "forward" model (Section 2.1). For independent Gaussian errors in the spectral observations, the likelihood function is

$$\mathcal{L}\left(\boldsymbol{D}|M_i, \boldsymbol{\theta}\right) = \prod_{k=1}^{N} \frac{1}{\sigma_k\sqrt{2\pi}} \exp\left\{-\frac{\left[D_{k,obs} - D_{k,model}\left(\boldsymbol{\theta}\right)\right]^2}{2\sigma_k^2}\right\}, \tag{3}$$

where $D_{k,obs}$ is the k-th observational data point in the spectrum, $D_{k,model}\left(\boldsymbol{\theta}\right)$ is the model prediction for the k-th data point given a set of atmospheric parameters $\boldsymbol{\theta}$, and N is the total number of data points in the observed spectrum. The denominator $\mathcal{Z}\left(\boldsymbol{D}|M_i\right)$ is the Bayesian evidence and merely a normalization constant in Equation (2). It is, however, central to Bayesian model comparison as described next.

2.2.2. *Identification of Gases and Aerosols using Bayesian Model Comparison*

One of the questions that arises when interpreting spectra of planetary atmospheres is which molecular species

Parameter		Prior	Lower Bound	Upper Bound
Planet-to-star radius ratio	$R_{\rm P}/R_*$	Uniform on log-scale	$R_{\rm Mars}/R_*$	1
Centered-log-ratio transform of mole fractions	$\boldsymbol{\xi}$	Uniform in $\boldsymbol{\xi}$-space	bound by condition that the mole fractions of all present molecular species are greater than 1 ppb [1]	
Cloud-top pressure	$P_{\rm surf}$	Uniform on log-scale	1 μbar	100 bar
Planetary Bond albedo	$A_{\rm B}$	Uniform	0	1

TABLE 1
Prior probability of atmospheric retrieval parameters. [1]The effect of molecular species with mole fractions below 1 ppb is negligible given currently available observations.

and aerosol types can be inferred from a given data set. Spectral signatures of molecular gases or aerosols are often hidden in the noisy observations of exoplanets. In addition, strong overlap between absorption features of different molecules and the lack of a flat continuum for thick atmospheres may further complicate the identification of individual gases.

Bayesian model comparison offers a rational way of determining which atmospheric species can be inferred from the data. It has been widely used in cosmology where noisy data and little prior knowledge require sophisticated statistical tools (see Trotta 2008; Hobson & Jaffe 2009, for excellent reviews). In the exoplanet context, Bayesian model comparison was employed by Gregory (2007) to detect planets in radial velocity data.

In our atmospheric retrieval application, we quantify our confidence in having detected the presence of a particular atmospheric constituent m by computing the posterior odds ratio (or Bayes factor) between a simpler retrieval model that neglect the presence of the species m and a more complex retrieval model that includes an additional parameter to describe the abundances of the constituent m. If the posterior odds ratio is strongly in favor of the more complex model, we can safely conclude that the atmospheric constituent is present.

In practice, the required comparisons for all relevant atmospheric constituents are achieved by computing the Bayesian evidence for one retrieval model that covers the full prior hypothesis space and comparing it to a list of retrieval models for which individual molecular species and types of aerosols were removed from the prior hypothesis space. The approach ensures that a high confidence in the presence of a particular atmospheric constituent is reported only if no other constituent in the prior hypothesis space can have resulted in the observed data.

Bayesian evidence— The quantity central to all Bayesian model comparisons is the Bayesian evidence defined as

$$\mathcal{Z}\left(\boldsymbol{D}|M_i\right) = \int_{all\,\boldsymbol{\theta}} \pi(\boldsymbol{\theta}|M_i)\,\mathcal{L}\left(\boldsymbol{D}|M_i,\boldsymbol{\theta}\right)\,\mathrm{d}^N\boldsymbol{\theta}. \qquad (4)$$

The Bayesian evidence quantifies the adequacy of a retrieval model M_i, specified by a set of N atmospheric parameters $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_N]$ and their prior probability distribution $\pi(\boldsymbol{\theta}|M_i)$, in the light of the observational data $\boldsymbol{D}$. Comparing the Bayesian evidences of alternative retrieval models with parameterizations of lower and higher complexities enables one to assess which atmospheric constituents and effects are important and which model complexities ought to be removed.

The underlying idea of the Bayesian model comparison is thereby that additional complexity of the parameter space ought to be avoided whenever a simpler model provides an adequate representation of the data. While it is obvious that a best fit of a retrieval model with more free parameters will always be better than (or at least as good as) the best fit of a model with fewer parameters, the Bayesian evidence only favors the more complex model if the improvement in the best fit due to additional parameters is large enough to overcome the so-called Occam penalty for the more complex model parameter space.

Posterior odds ratio / Bayes factor— Using Bayes' law we can express the posterior odds ratio between two alternative models $\mathrm{M}_{\text{“m present”}}$ and as $\mathrm{M}_{\text{“m not present”}}$

$$\frac{p\left(\mathrm{M}_{\text{“}m\text{ is present”}}|\boldsymbol{D}\right)}{p\left(\mathrm{M}_{\text{“}m\text{ not present”}}|\boldsymbol{D}\right)} = B_m \frac{p\left(\mathrm{M}_{\text{“}m\text{ is present”}}\right)}{p\left(M_{\text{“}m\text{ not present”}}\right)} \qquad (5)$$

where the Bayes factor B_m is the ratio of the model's Bayesian evidences

$$B_m = \frac{\mathcal{Z}\left(\boldsymbol{D}|M_{\text{“}m\text{ present”}}\right)}{\mathcal{Z}\left(\boldsymbol{D}|M_{\text{“}m\,not\,present”}\right)}, \qquad (6)$$

and $p\left(\mathrm{M}_{\text{“m present”}}\right)/p\left(M_{\text{“m not present”}}\right)$ is the prior odds ratio, which we assume to be 1 in the absence of prior information.

The posterior probability for the present species m is

$$p\left(\text{“}m\text{ is present”}|\boldsymbol{D}\right) = \frac{B_m}{1+B_m}. \qquad (7)$$

The Bayes factor B_m plays a pivotal role in a Bayesian model comparison. A value of $B_m > 1$ means that the data provide support in favor of the presence of the atmospheric constituent m. Bayes factors are generally interpreted against the Jeffrey scale (ref). Values of the odds of 3:1, 12:1, and 150:1 represent weak, moderate, and strong support in favor of the presence of additional molecules (Table 2).

Calibration between Bayesian and Frequentist Detections— The increase in availability of computational resources over the last decades has led to a widespread of Bayesian techniques in the analysis of astrophysical observations. Parameter estimation problems are frequently solved in a Bayesian framework using Markov Chain Monte Carlo (MCMC) techniques. Yet, the significance of detections of atmospheric absorption features is generally still reported in terms of the frequentist measures of confidence such as the p-value or the “sigma”-significance. As a result, many scientists are accustomed to the interpretation of sigma significance, but not Bayes factors.

A useful calibration between the Bayes factor and the frequentist measures of confidence is provided by the expression

p-value	B_m	$\ln B_m$	"sigma"	interpretation
0.05	2.5	0.9	2.0σ	
0.04	2.9	1.0	2.1σ	"weak" detection
0.01	8.0	2.1	2.6σ	
0.006	12	2.5	2.7σ	"moderate" detection
0.003	21	3.0	3.0σ	
0.001	53	4.0	3.3σ	
0.0003	150	5.0	3.6σ	"strong" detection
$6 \cdot 10^{-7}$	43000	11	5.0σ	

TABLE 2
Translation table between frequentist significance values (p-values) and the Bayes factor (B_m) in favor of the more complex model. Adopted from Trotta (2008). A Bayes factor of 150 can be considered a "strong" detection, corresponding to approximately 3.6σ significance in the frequentist's framework.

$$B_m \leq -\frac{1}{\mathrm{e} \cdot \rho \cdot \ln \rho}, \tag{8}$$

where e is the exponential of one and ρ is the p-value(Sellke et al. 2001). The p-value, in turn, can be related to the sigma significance n_σ using the expression

$$\rho = 1 - \mathrm{erf}\left(\frac{n_\sigma}{\sqrt{2}}\right), \tag{9}$$

where erf is the error function. Representative values for a range of confidence values are listed in Table 2. Equation (8) presents an upper bound on the Bayes factor, e.g., a Bayes factor $B_m = 21$ corresponds, at least, to a 3.0σ detection. Equation (8) is valid for $\rho < \mathrm{e}^{-1}$ and under the assumption of a mild principle of indifference as to the value of the added parameter in the more complex model (Sellke et al. 2001).

2.3. *Nested Sampling for Atmospheric Retrieval*

We employ the multimodal nested sampling algorithm (MultiNest) to efficiently compute the Bayesian evidences of alternative retrieval models (Skilling 2004; Feroz & Hobson 2008; Feroz et al. 2009). The joint posterior probability distribution of the atmospheric parameters for a given retrieval model is obtained as a by-product. The MultiNest algorithm was developed as a Bayesian inference tool for cosmology and particle physics, and we find that it provides substantial advantages for atmospheric retrieval over techniques based on Markov Chain Monte Carlo (MCMC).

2.3.1. *Nested Sampling versus MCMC*

Parameter estimations in many astrophysical contexts are typically performed using the MCMC technique with the Metropolis-Hastings algorithm or its variants such as the Gibbs sampler. The two main disadvantages of the MCMC techniques are, however, that MCMC does not directly provide the Bayesian evidence for comparing retrieval models of different complexities, and that MCMC becomes inefficient when sampling from a posterior distribution with multiple separate modes or elongated curving degeneracies. In Benneke & Seager (2012), we were able to search for separated modes in the posterior distribution using the parallel tempering MCMC technique. Calculating the Bayesian evidences for Bayesian model comparisons would, however, require a second step, such as a Restricted Monte Carlo (RMC) integration. Restricted Monte Carlo (RMC) becomes extremely inefficient, however, when the joint posterior distribution shows strongly curved correlations, such as those encountered in atmospheric retrieval problems.

2.3.2. *The Nested Sampling Algorithm*

Nested sampling is an alternative Monte Carlo technique for Bayesian inference. It enables one to efficiently compute the Bayesian evidence and provides the posterior distribution for parameter estimations as a by-product. A full discussion is provided in (Skilling 2004; Feroz & Hobson 2008). Here, we provide a brief description of the main ideas. For clarity, we do not explicitly state the model M_i and the data $\boldsymbol{D}$ because the calculation of the Bayesian evidence is conducted for each model and data set individually.

The basic idea behind nested sampling is to transform the multi-dimensional integral for the computation of the Bayesian evidence (Equation (4)) into the one-dimensional integral

$$\mathcal{Z} = \int_0^1 \mathcal{L}^*(V)\, dV. \tag{10}$$

The integration variable V in Equation (10) is the "prior volume" defined as

$$V(\mathcal{L}^*) = \int_{\mathcal{L}(\boldsymbol{\theta}) > \mathcal{L}^*} \pi(\boldsymbol{\theta}) \mathrm{d}^N \boldsymbol{\theta}. \tag{11}$$

The prior volume $V(\mathcal{L}^*)$ is the prior probability density integrated over all regions in the parameter space for which the likelihood function $\mathcal{L}(\boldsymbol{D}|\boldsymbol{\theta})$ exceeds the value $\mathcal{L}^*$. It is a monotonically decreasing function of the likelihood limit $\mathcal{L}^*$ because the volume in the parameter space that meets the criteria $\mathcal{L}(\boldsymbol{\theta}) > \mathcal{L}^*$ decreases as the likelihood limit $\mathcal{L}^*$ is increases. The extreme values are $V = 1$ for $\mathcal{L}^* = 0$, in which case the integration is performed over the entire prior parameter space, and $V = 0$ for $\mathcal{L}^* > \max(\mathcal{L}(\boldsymbol{D}|\boldsymbol{\theta}))$. The function $\mathcal{L}^*(V)$ in Equation 10 is the inverse function of the prior volume $V(\mathcal{L}^*)$ and is thus also monotonically decreasing.

Figure 1 provides a graphical interpretation of the prior volume V and equation (10) for a two-dimensional parameter space. For two dimensional parameter spaces with the uniform prior probability $\pi(\theta_1, \theta_2) = \pi_{\mathrm{uni}}$, the Bayesian evidence can be regarded as the geometric volume between the θ_1,θ_2-plane and the surface of the likelihood function $\mathcal{L}(\boldsymbol{D}|\theta_1, \theta_2)$, multiplied by π_{uni} (see equation (4)). The prior volume $V(\mathcal{L}^*)$ can be interpreted as the area within iso-likelihood contour $\mathcal{L}(\boldsymbol{D}|\theta_1, \theta_2) = \mathcal{L}^*$ projected onto the θ_1, θ_2 -plane, again multiplied by uniform prior π_{uni}.

The Bayesian evidence in Equation 10 can be approximated using standard quadrature methods if the inverse function $\mathcal{L}^*(V)$ of Equation (11) can be evaluated at a sequence of values $0 \leqslant V_M < \cdots < V_i < \cdots < V_1 \leqslant 1$. Using the trapezium rule, we obtain

$$\mathcal{Z} = \sum_{i=1}^{M} w_i \mathcal{L}_i. \tag{12}$$

where $\mathcal{L}_i$ is the likelihood limit $\mathcal{L}^*(V_i)$ corresponding

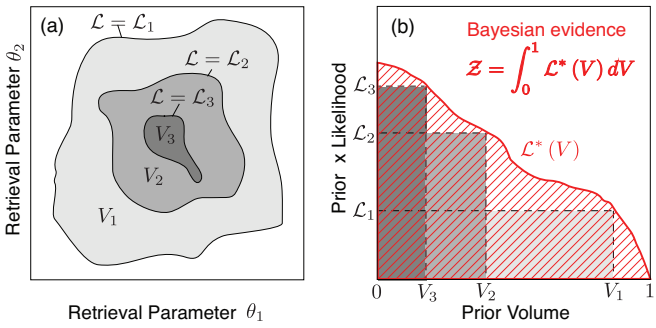

FIG. 1.— Graphical interpretation of the prior volume and Bayesian evidence for a two dimensional parameter space. The prior volume $V_i = V(\mathcal{L}_i)$ is the prior probability density integrated over all regions in the parameter space for which the likelihood function $\mathcal{L}(\theta_1, \theta_2)$ exceeds the value $\mathcal{L}_i$. For a two-dimensional parameter space with a uniform prior, the prior volume is the projected area of the region in the parameter space for which $\mathcal{L}(\theta_1, \theta_2)$ exceeds the value $\mathcal{L}_i$ (panel (a)). The Bayesian evidence $\mathcal{Z}$ can be regarded as the geometric volume between the θ_1,θ_2-plane and the likelihood surface $\mathcal{L}(\theta_1, \theta_2)$. Once a sequence of $\mathcal{L}_i, V_i$ pairs is known, the Bayesian evidence can be approximated using standard quadrature methods (panel (b)).

to V_i, the weights w_i are given by $w_i = \frac{1}{2}(V_{i-1} - V_{i+1})$, and M is the number of values in the sequence.

In nested sampling, a sequence of $\mathcal{L}^*(V_i)$ values is generated as follows. The algorithm is initialized by randomly drawing a user-specified number of "active samples" ($N \approx 50 \ldots 10000$) from the full prior probability distribution $\pi(\boldsymbol{\theta})$ and computing their likelihood $\mathcal{L}(\boldsymbol{\theta})$ according to Section 3. These samples are distributed randomly across the full prior parameter space and the prior volume $V_0 = V(\mathcal{L}^* = 0)$ is 1. The first iteration is started by sorting the samples in order of their likelihood. The sample with the lowest likelihood $\mathcal{L}_1$, i.e. the sample with the worst fit to the data, is then removed and replaced by a new sample. The new sample is again drawn from the prior distribution, but subject to the constraint that the likelihood of the new sample is higher than $\mathcal{L}_1$. If necessary, the drawing of the new sample is repeated until a sample is found with $\mathcal{L}(\boldsymbol{\theta}) > \mathcal{L}_1$. After the replacement, all active samples meet the condition $\mathcal{L}(\boldsymbol{\theta}) > \mathcal{L}_1$ and they are uniformly distributed within the iso-likelihood contour $\mathcal{L}^* = \mathcal{L}_1$. The new prior volume V_1 contained within the iso-likelihood contour $\mathcal{L}^* = \mathcal{L}_1$ can be written as $V_1 = t_1 V_0$, where the shrinkage ratio t_1 is a random number smaller than 1. The probability distribution of the shrinkage, $p(t_1)$, is the distribution for the largest of n samples drawn from a uniform distribution between 0 and 1. The distribution $p(t_1)$ describes the decrease in prior volume between subsequent iso-likelihood contours $\mathcal{L}^* = V_0$ and $\mathcal{L}^* = V_1$ in a probabilistic way because the active samples are uniformly distributed within the iso-likelihood contours.

In subsequent iterations, the process of replacing the sample with the lowest likelihood $\mathcal{L}_i$ is repeated, and the corresponding prior volume $V(\mathcal{L}_i)$ is repeatedly tightened. The nested sampling algorithm thus progresses through nested shells of iso-likelihood contours with continually decreasing prior volumes, until the regions of the highest likelihoods are localized. The prior volume V_i after iteration i can be approximated as

$$V_i = \exp(-i/N). \tag{13}$$

Equation (13) approximate the prior volume at iteration i because each value of t_i is independent and the mean and standard deviations of $\log(t)$ are $\mathrm{E}[\log(t)] = -1/N$ and $\sigma[\log(t)] = 1/N$, resulting in $\log(V_i) \approx (i \pm \sqrt{i})/N$.

2.3.3. *Simultaneous Ellipsoidal Nested Sampling*

The main challenge in efficiently computing the Bayesian evidence is to efficiently generate random samples within the iso-likelihood contour $\mathcal{L}^* = \mathcal{L}_i$. If one continued to draw random samples from the full prior probability distribution $\pi(\boldsymbol{\theta})$ as in the first iteration, the chance of randomly finding one that is within the iso-likelihood contour $\mathcal{L}^* = \mathcal{L}_i$ would decrease exponentially as the likelihood limit $\mathcal{L}_i$ is raised.

The algorithm employed in this work uses the simultaneous ellipsoidal nested sampling method developed in Feroz et al. (2009). At each iteration, the full set of active samples is partitioned according to local clustering and an optimum number of ellipsoids are constructed such that the union of the ellipsoidal volumes tightly encompasses all samples. At early iterations, a small number of large ellipsoids cover almost the entire prior parameter space. As the active sample become confined to the regions within tighter iso-likelihood contours, the ellipsoids encompass only the region(s) of high likelihood $\mathcal{L}(\boldsymbol{\theta})$. New random samples are drawn from within the ellipsoids, thus the random samples have a high likelihood $\mathcal{L}(\boldsymbol{\theta})$ with higher probability, resulting in a higher sample acceptance rate and higher efficiency of the nested sampling algorithms. The algorithm is robust to multimodal posterior distributions and elongated curving degeneracies, while remaining efficient for simpler distributions. Refer to Feroz et al. (2009) for a detailed description of the algorithm employed for partitioning and construction of the ellipsoids.

2.3.4. *Convergence and Numerical Uncertainty*

The iterations are continued until the Bayesian evidence is determined to a specified precision $\Delta\mathcal{Z}$. In this work, we terminated the iterations once the logarithm of the evidence did not change by more than $\Delta(\log(\mathcal{Z})) = 0.001$. The final uncertainty can be estimated following Sivia & Skilling (2006) as

$$\log(\mathcal{Z}) \approx \log\left(\sum_{i=1}^{M} w_i \mathcal{L}_i\right) \pm \sqrt{\frac{\mathcal{H}}{N}}, \tag{14}$$

where $\mathcal{H}$ is the *information,* or *negative entropy,*

$$\mathcal{H} \approx \sum_{i=1}^{M} \frac{w_i \mathcal{L}_i}{\mathcal{Z}} \log\left(\frac{\mathcal{L}_i}{\mathcal{Z}}\right). \tag{15}$$

2.3.5. *Parameter Estimation*

Once the Bayesian evidence $\mathcal{Z}$ is determined, the joint posterior distribution of the atmospheric parameters can be constructed using all active and discarded samples that were generated during the nested sampling iterations. Each sample is assigned a weight according to

$$W_i = \frac{w_i \mathcal{L}_i}{\mathcal{Z}}. \tag{16}$$

The weighted samples can be used to plot the posterior distributions and calculate the statistical measures such as the mean and covariances matrix.

3. RESULTS: DISTINGUISHING BETWEEN H_2-DOMINATED AND H_2O-DOMINATED ATMOSPHERES

In this Section, we demonstrate that a promising strategy to distinguish between cloudy hydrogen-dominated atmospheres and clear water- or volatile-dominated atmospheres is to observe individual water absorption features in the planet's transmission spectrum at moderate spectral resolution ($R \approx 100$). The unambiguous distinction is possible based on the effect of the mean molecular mass on the wing slopes of the water features and the relative depths of different water absorption features. As a case study, we investigate, for the super-Earth GJ 1214b, how much the observational uncertainty needs to be improved compared to published transmission spectra to reliably detect NIR water absorption features and to unambiguously distinguish between the hydrogen-dominated and water-dominated scenarios. All results derived in this Section can be generalized to other super-Earths and Neptunes such as HD 97658b, 55 Cnc e, and GJ 436b using the scalings provided in Section 5.1 and 5.2.

3.1. *Distinction Based on NIR Transmission Spectroscopy*

Our main finding is that an unambiguous distinction will be possible for a GJ 1214b-like planet if the observational uncertainty in the transit depth measurements can be reduced by a factor of ~ 3 compared with the published *VLT* observations by Bean et al. (2010) and *HST WFC3* observations by Berta et al. (2012). Alternatively, the distinction can be achieved using only *HST WFC3* observations if the uncertainty can be reduced by a factor of ~ 5 compared to the previous result by Berta et al. (2012) (Section 3.2).

For the following discussions, we consider synthetic observations of three alternative scenarios for the atmosphere of GJ 1214b. The scenarios are a cloud-free, water-dominated atmosphere composed of 95% H_2O and CO_2 (Figure 2), a hydrogen-dominated atmosphere with solar metallicity ($\sim$ 400 ppm H_2O) and high-altitude clouds (Figure 3), and flat transmission spectra due to the lack of atmospheres or the presence of an extremely high cloud deck (Figure 4). The hydrogen-dominated scenario is depleted in methane and clouds are present at 10 mbar such that depths of the water absorption resemble the ones of the cloud-free water-dominated scenario. These choices ensure that the distinction between the water-dominated and hydrogen-dominated scenarios is based the effect of the mean molecular mass on relative sizes and shapes of the water absorption features and independent of the abundance or the presence of other molecular species.

We find that spectral observations covering the spectral ranges $0.78-1\,\mu$m (Bean et al. 2010) and $1.1-1.8\,\mu$m (Berta et al. 2012), can unambiguously distinguish between water-dominated atmospheres and hydrogen-dominated atmospheres with water absorption if the observational uncertainty can be reduced to 60 ppm at a spectral resolving power of $R = 70$ ((Figure 2-4). This corresponds to an improvement of ~ 3 compared with the published uncertainties of $180 - 200$ ppm by Bean et al. (2010) and Berta et al. (2012).

(a) Water-dominated scenario.— If GJ 1214b is surrounded by a cloud-free water-dominated atmosphere (Figure 2), transit depth measurements with 60 ppm uncertainty will constrain the mean molecular mass to $\mu_{\rm atm} > 13$ at 99.7% probability (3σ). Such observations would conclusively rule out a hydrogen-rich nature of GJ 1214b. Taking a simple two-component model atmosphere composed of water vapor and hydrogen gas, we could infer at 3σ that $> 70\%$ of the atmosphere would need to be water vapor, leaving only a maximum of 30% for hydrogen-helium gas (Figure 2(b)). From 60 ppm observations, we would also infer at 3σ that the atmosphere is cloud-free down to at least the 20 mbar level. The posterior probability density would be maximum for water fractions above 90% and cloud top pressures above 100 mbar. Changes of the water fractions above 90% have little effect on the water absorption features because the mean molecular mass remains largely unchanged. An increase in cloud top pressure above 100 mbar, similarly, has negligible effects on the observable spectrum because the high opacity of water vapor across the full NIR spectrum prevents probing deeper atmospheric layers through transmission spectroscopy.

(b) Hydrogen-rich scenario with high-altitude clouds.— If, alternatively, GJ 1214b is surrounded by a hydrogen gas envelope (Figure 3), transit depth measurements with 60 ppm uncertainty will be able to confirm the hydrogen-dominated nature of the atmosphere, even if the depth of water absorption features resembles the one of a water-dominated scenario. 60 ppm transit observations would provide sufficient information on the relative strengths of the absorption features and the steepnesses of the feature wings to infer a mean molecular mass below 10 at 3σ (Figure 3(b)). A mean molecular mass below 10 would mean that hydrogen/helium gas makes up at least 50% of the atmosphere. The observations would also confirm the presence of an upper cloud deck with a cloud-top pressure between 0.2 mbar and 200 mbar at 99.7% confidence.

Stronger independent constraints on the water mole fraction in the hydrogen-dominated atmosphere would not be available despite a 7σ detection of water absorption because a strong correlation exists between the water mole fraction and the cloud-top pressure for the transmission spectra of a hydrogen-dominated atmosphere. The strong correlation is present because atmospheric scenarios with different combinations of mean molecular masses and cloud-top pressures can lead to identical feature depths.

It is interesting to note that even though hydrogen/helium gas shows no direct absorption features, the presence of hydrogen/helium gas in the atmosphere can be inferred from the NIR observations if hydrogen/helium gas is present in sufficient amounts to affect the mean molecular mass (Figure 3(c)). The reason is that no other plausible atmospheric component can explain the mean molecular mass below that of water ($\mu = 18$) and methane ($\mu = 16$) in the temperature range expected for GJ1214b. Clouds and hazes can be inferred from transmission spectra if present at high altitude because they lower the transit depth variations below that

expected for a clear atmosphere of a given atmospheric composition.

(c) Featureless spectrum.— A planet's transmission spectrum may be featureless if the planet lacks a gaseous atmosphere or clouds are present at high altitude. If a featureless spectrum with an observational uncertainty of 60 ppm was observed for GJ 1214b (Figure 4), the observations would be sufficient to rule out the presence of cloud-free, water-rich atmospheres at high significance (Figure 2(e-f)). For the full range of water mole fractions between solar composition ($X_{\mathrm{H_2O}} \sim 0.04\%$) and water-dominated ($X_{\mathrm{H_2O}} \rightarrow 100\%$), we could conclude that cloud top pressures must be below 1 mbar. The observations would additionally rule out scenarios with cloud-top pressures down to 0.01 mbar for water mole fractions between 1% and 10%.

As a consistency check, Figure 4(c) shows that no molecular species can be inferred from the flat spectrum. It is interesting to note that the Bayes factor for the presence of clouds is not conclusive despite the flat spectrum. The reason for this is that high altitude clouds are only one explanation for measuring a flat spectrum. For example, the absence of strong absorber such as water and methane in combination with a high mean molecular mass would similarly explain the lack of any features at the 60 ppm level. A comparison between Figures 4(c) and 3(c), in fact, shows that the presence of weak molecular features may lead to a higher confidence in the presence of clouds than a flat spectrum. Absorption features enable us to infer the composition and the depths and shapes of the features then enable us to determine whether clouds must be present or not. This information is not available if a flat spectrum is observed.

3.2. *Distinction Based Solely on HST WFC3 Observations*

The distinction between hydrogen-dominated and water-dominated scenarios is possible from *HST WFC3* observations if the observational uncertainty can be reduced to below $\sim$ 35 ppm (Figure 5). The main advantage of augmenting the *HST WFC3* observation with observations at $0.8 - 1\,\mu$m (such as the ones provided by *VLT*) is that $0.8 - 1\,\mu$m contains weaker absorption features that, together with the strong absorption feature at 1.38 μm, enable a good comparison between the relative depths of the strong and weak water absorption features. These relative depths, in turn, constrain the mean molecular mass and thus enable the distinction between hydrogen-dominated atmospheres and those dominated by water or other volatiles. If only *HST WFC3* observations are available, the observational uncertainty needs to be sufficient to capture the feature wings between $1.1 - 1.7\,\mu$m at higher precision because the information from the relative depths of the features will not be available (Figure 5(a)).

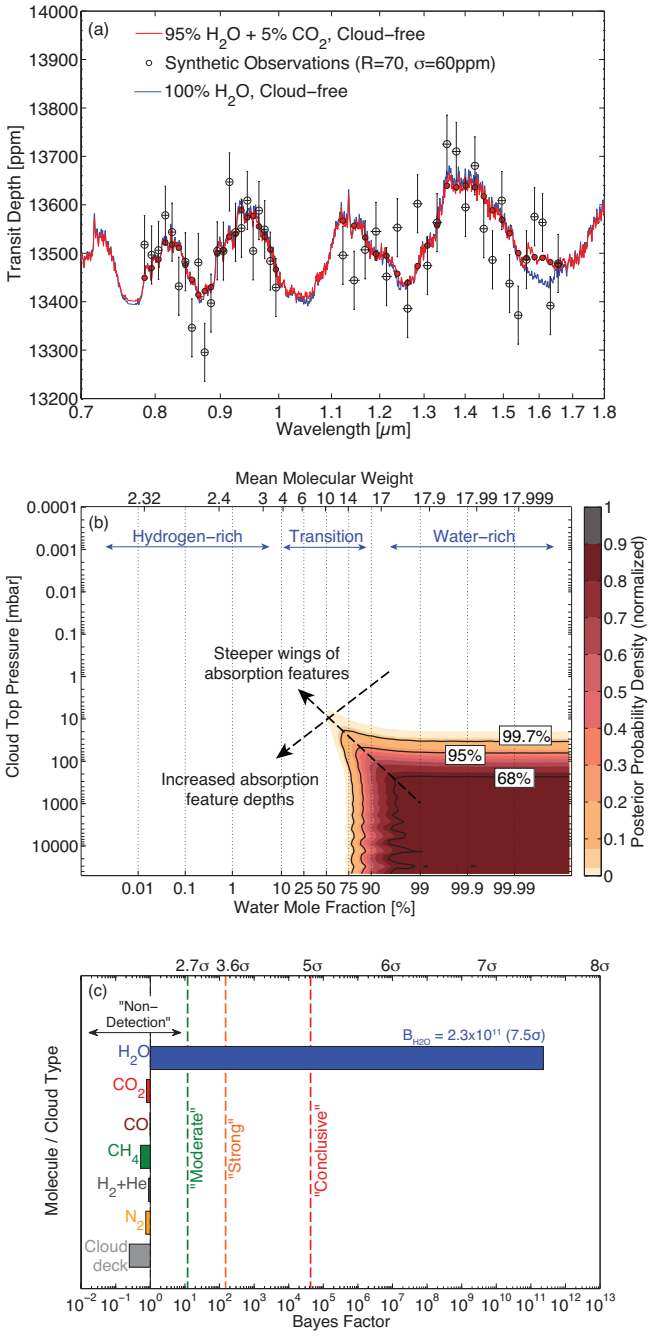

FIG. 2.— Synthetic 60 ppm observations and retrieval results for a cloud-free, water-dominated atmosphere on GJ 1214b. Panel (a) shows synthetic observations of a 95% H_2O + 5% CO_2 atmosphere covering the spectral range of the *VLT* and *HST WFC3* observations by Bean et al. (2010) and Berta et al. (2012). Panel (b) illustrates the posterior probability distribution derived from the synthetic observations as a function of cloud top pressure and mean molecular weight for a simple two-component atmosphere composed of hydrogen and water vapor. The black contour lines indicate the 68%, 95%, and 99.7% Bayesian credible regions. The colored shading illustrates the regions of high posterior probability density. The horizontal axis is scaled as the centered-log-ratio parameter $\xi_1 = \log\left(X_{H2O}/\sqrt{X_{H2O} \cdot X_{H2}}\right)$. 60 ppm observations would constrain the mean molecular mass to $\mu_{\mathrm{atm}} > 13$ at 99.7% probability (3σ). Such observations would conclusively rule out a hydrogen-rich nature of the atmosphere. Panel (c) illustrates the Bayes factors describing the detection confidences of molecular gases and clouds based on the synthetic observations. 60 ppm observations would enable a robust $B_{\mathrm{H_2O}} = 2 \cdot 10^{12}$ (7.8σ) detection of water absorption. The 5% CO_2 remain undetectable because strong water absorption blocks the CO_2 signature.

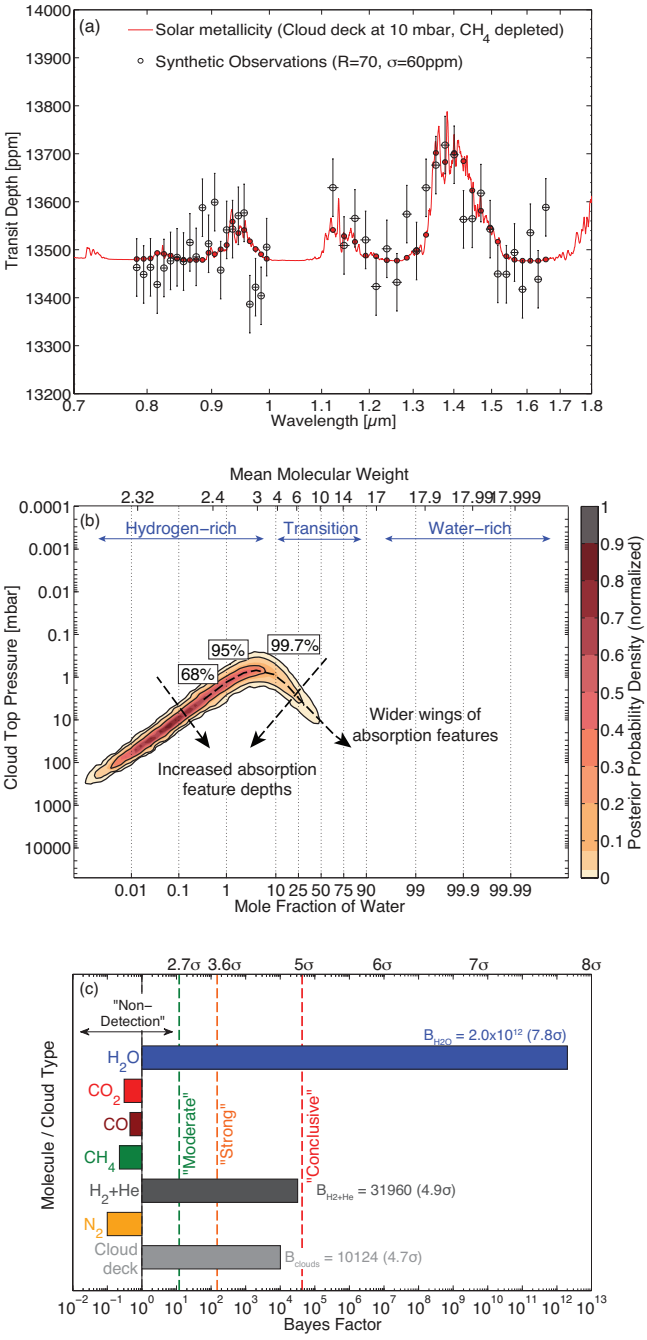

FIG. 3.— Synthetic 60 ppm observations and retrieval results for a H_2-dominated atmosphere with high-altitude clouds on GJ 1214b. Panel (a) shows synthetic observations of a methane-depleted, solar metallicity atmosphere with thick gray clouds extending up to the 10 mbar level. The spectral range is equivalent to the *VLT* and *HST WFC3* observations by Bean et al. (2010) and Berta et al. (2012). The methane depletion and vertical extent of the clouds were chosen to ensure that the distinction between the H_2O-dominated and H_2-dominated scenarios is independent of other molecular absorbers. Panels (b) and (c) display the retrieval results as explained in Figure 2. 60 ppm observations would be able to confirm the hydrogen-dominated nature by setting an 3σ upper limit on mean molecular mass at $\mu_{\rm ave} = 10$. The observations would be sufficient to robustly infer the presence of water and hydrogen, and clouds at high confidence.

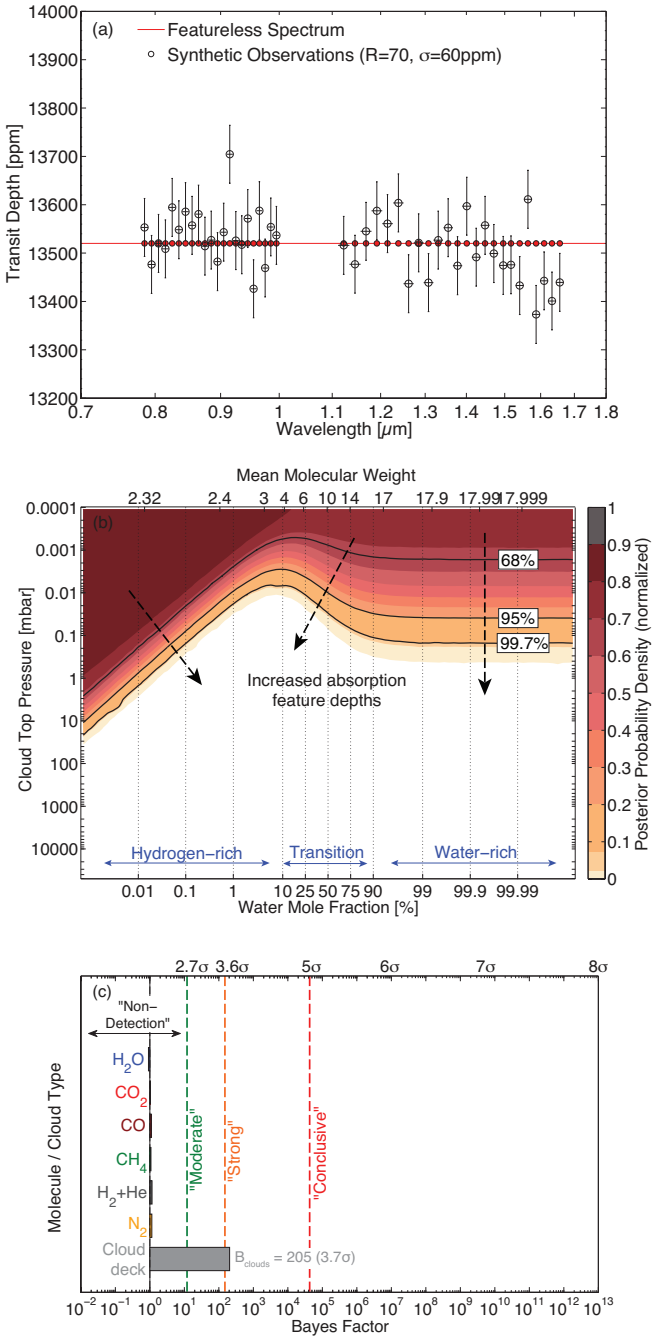

FIG. 4.— Synthetic 60 ppm observations and retrieval results for a featureless spectrum of GJ 1214b. Panel (a) displays synthetic observations GJ 1214b as it will appear if a thick cloud deck surrounds the planet at above the 0.1 mbar pressure level or no atmosphere is present. Panels (b) and (c) display the retrieval results derived from synthetic observations as explained in the caption of Figure 2. At 60 ppm, a flat spectrum would rule out thick, cloud-free scenarios for all water mole fractions between the one expected for solar composition ($X_{\rm H_2O} \sim 0.04\%$) and water-dominated ($X_{\rm H_2O} \rightarrow 100\%$). At 3σ, water-dominated scenarios are possible only if a cloud deck is present above the 0.1 mbar pressure level. Hydrogen-dominated scenarios are compatible with the flat spectrum at 60 ppm only if the partial pressure of H_2O is below 2 μbar at the uppermost cloud deck surface. Panel (c) shows that no molecular detections can be inferred from the featureless spectrum. Clouds are likely but cannot conclusively be inferred because atmospheric scenarios without NIR absorbers could theoretically match the featureless spectrum.

3.3. *Detectability of NIR Water Absorption*

It is important to recognize that detections of water absorption are a necessary, but not sufficient condition to determine the water fraction and nature of low-density super-Earths. Section 3.1 showed that transit observations with $\lesssim$ 50 ppm observational uncertainty are required to distinguish hydrogen-dominated and water-dominated scenarios. The presence of water absorption in a water-dominated atmosphere can already be detected at high confidence for transit observations with up to 80 ppm uncertainty. 80 ppm observations at $R = 70$ would provide a $B_{\rm H_2O} = 40000$ (5σ) detection of water absorption and would only require an improvement by a factor of 2.5 compared to Bean et al. (2010) and Berta et al. (2012).

3.3.1. *Characteristic Degeneracy between the Mean Molecular Mass and Cloud Top Pressure*

Transit observations with larger transit depth uncertainties, e.g. 80 ppm, can leave the water mole fraction unconstrained in the entire range between 100% and fractions of 1% despite a robust $B_{\rm H_2O} = 2040$ (4.3σ) detection of water absorption (Figure 6). High posterior probabilities would exist along a line of constant feature depth, resulting in a degeneracy between the mean molecular mass and the cloud top pressure. This degeneracy is characteristic for transmission spectra that enable the detection of the absorber, but do not provide sufficient constraints on the steepness of the feature wings or depths of other features to constrain the mean molecular mass independently.

In the hydrogen-rich regime, the feature size remains constant along lines of constant water column density. The cloud top pressure compatible with the synthetic observations decreases as the water mole fraction in the atmosphere is increased (negative correlation). In the transition region between 10% and 90% water mole fraction, the change in the mean molecular mass dominates the change in the feature size. Maintaining the same feature sizes requires higher cloud top pressures as the water fraction is increased (positive correlation). Changes in the feature sizes are small if the water mole fraction is increased because the mean molecular mass remains largely unchanged and deeper layers in the atmosphere are optically thick regardless of the exact water mole fraction.

3.3.2. *Overlapping Molecular Bands in WFC3 Bandpass: H_2O or CH_4 Absorption?*

A challenge in the spectral range of the *HST WFC3* bandpass is that methane and water have two strongly overlapping absorption bands at 1.15 μm and 1.4 μm. An unambiguous detection of water vapor requires not only the detection of the absorption bands, but also requires one to determine whether the absorption was caused by water, methane, or both. Distinction between the water and methane cases is possible based on subtle differences in the absorption feature shapes, the slope red of 1.6 μm, and the different opacities in the bandpass of the VLT observations. It is obvious, however, that better observations are required to distinguish between water features and similarly shaped methane features than is necessary to distinguish between water features and a flat spectrum.

It is worth noting that the results from the Bayesian analysis in this work inherently accounts for the overlap between water and methane absorption features. The framework assigns a high probability for the presence of water vapor only if the observed feature resembles the expected shape for a water feature considerably closer than it resembles the expected shape of a methane feature. Taking again the example of our 80 ppm synthetic observations (Figure 6**),** we find that our confidence in having detected water absorption is lowered by the possibility that methane is responsible for the absorption features. The Bayes factor as a measure of our confidence in the presence of water is $B_{\rm H_2O} = 2060$ (4.3σ) while our confidence that either water or methane is present is $B_{\rm H_2O\,or\,CH_4} = 41756$ (5.0σ). In other words, it is easier to detect absorption features at 1.15 μm and 1.4 μm than it is to conclusively state that the features were caused by water absorption, and this needs to be accounted for when claiming the detection of water absorption. Another way to look at it is that we could detect at $B = 43000$ (5σ) that either water or methane is present with 80 ppm observations, but we need $\sim$70 ppm precision to conclude at the same confidence that the feature was caused by water absorption.

3.4. *Effects of Non-Gray Aerosols*

In Sections 3.1-3.3, we assumed gray clouds when demonstrating the distinction between cloud-free water-rich atmospheres and cloudy hydrogen-rich atmospheres. Here, we demonstrate that the fundamental approach of determining the mean molecular mass based on wing slopes of molecular absorption bands or by comparing features of the same absorber remains viable in the general case of non-gray aerosols. We model cloud scattering using Mie theory and find that the cores of molecular absorption bands in transmission spectra are largely unaffected by the type and spectrum of the particles. The molecular absorption bands therefore provide unambiguous constraints on the mean molecular mass as long as there are significant detectable molecular absorption features penetrating the "continuum" spectrum of the haze or cloud particles.

We demonstrate that the wing slopes and relative sizes of water absorption features are good measures of the mean molecular mass by presenting two spectra of hydrogen-dominated atmospheres with high altitude hazes and comparing them to cloud-free water atmospheres (Figure 7). We consider high-altitude haze particles composed of either condensed Potassium Chloride (KCl) or Zinc Sulfide (ZnS) as two possible scenarios that would sufficiently mute the spectral features of a hydrogen-dominated atmosphere to match the published observations of GJ 1214b. KCl and ZnS are considered for GJ 1214b because the he temperature pressure profile in the atmosphere of GJ 1214b is likely to cross their condensation curves, suggesting that KCl and ZnS may be able condense in the atmosphere of GJ 1214b (Morley et al. 2012, 2013).

The particle size distributions and vertical extent of KCl and ZnS condensates in the upper atmosphere of a super-Earths depend sensitively on the atmospheric dynamics. Atmospheric dynamics on exoplanets are, however, widely uncharacterized. The goal of this section is, therefore, not to self-consistently model the size distri-

Retrieval model	Retrieval model parameters	Evidence $\ln(\mathcal{Z}_i)$	Best-fit $\chi^2_{\rm best-fit}$	Bayes factor $B_i = \mathcal{Z}_0/\mathcal{Z}_i$
Full hypothesis space	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CO_2}, X_{\rm CH_4}, X_{\rm CO}, X_{\rm H_2/He}, X_{\rm N_2})$	-284.57	55.17	Ref.
H_2O removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm CO_2}, X_{\rm CH_4}, X_{\rm CO}, X_{\rm H_2+He}, X_{\rm N_2})$	-292.21	69.90	$B_{\rm H_2O} = 2060$ (4.3σ)
CO_2 removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CH_4}, X_{\rm CO}, X_{\rm H_2/He}, X_{\rm N_2})$	-284.63	58.23	$B_{\rm CO_2}$ =1.06
CH_4 removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CO_2}, X_{\rm CO}, X_{\rm H_2/He}, X_{\rm N_2})$	-284.31	55.04	$B_{\rm CH_4}$ =0.77
CO removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CO_2}, X_{\rm CH_4}, X_{\rm H_2/He}, X_{\rm N_2})$	-284.43	55.31	$B_{\rm CO}$ =0.86
N_2 removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CO_2}, X_{\rm CH_4}, X_{\rm CO}, X_{\rm N_2})$	-284.31	56.61	$B_{\rm N_2}$ =0.77
H_2/He mix removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CO_2}, X_{\rm CH_4}, X_{\rm CO}, X_{\rm H_2/He})$	-284.50	55.12	$B_{\rm H_2/He}$ =0.93
H_2O & CH_4 removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm CO_2}X_{\rm CH_4}, X_{\rm CO}, X_{\rm H_2/He}, X_{\rm N_2})$	-295.21	83.49	$B_{\rm H_2O\,or\,CH_4}$ =41756 (5.0σ)
Clouds removed	$\frac{R_p}{R_*}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CO_2}, X_{\rm CH_4}, X_{\rm CO}, X_{\rm H_2/He}, X_{\rm N_2})$	-283.63	55.34	$B_{\rm Clouds}$ =0.39

TABLE 3
RESULTS OF BAYESIAN MODEL COMPARISON FOR SYNTHETIC 80 PPM OBSERVATIONS OF THE WATER-DOMINATED SCENARIO ON GJ 1214B. THE SYNTHETIC OBSERVATIONS ARE DEPICTED IN FIGURE 6(A). VALUES THAT ARE DIRECTLY REFERRED TO IN SECTION 3 ARE MARKED IN BOLD. THE VECTOR $\boldsymbol{\xi}$ IS THE CENTER-LOG-RATIO (CLR) TRANSFORMATION OF THE MOLE FRACTIONS OF THE ATMOSPHERIC GASES.

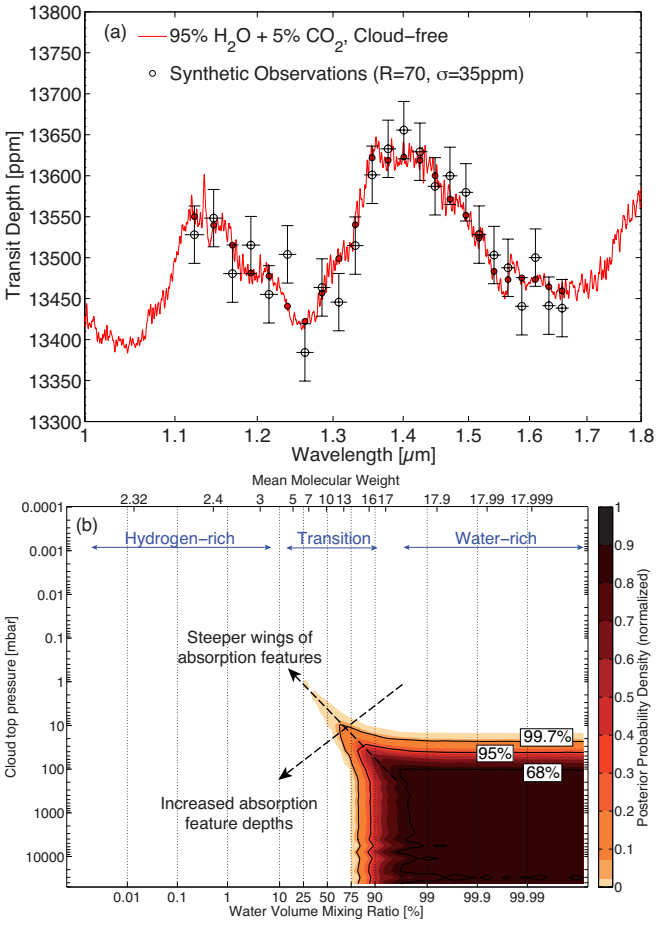

FIG. 5.— Synthetic 35 ppm observations and retrieval results for a cloud-free, H_2O-dominated atmosphere on GJ 1214b. The atmospheric scenario (95% H_2O + 5% CO_2) is identical to the one in Figure 2, however, the synthetic observations cover only the spectral range of *HST WFC3*. Panel (b) illustrates the posterior probability distribution as a function of cloud top pressure and mean molecular weight as explained in Figure 2. 35 ppm observations in the spectral range of *HST WFC3* ($1.1-1.7\,\mu$m) would constrain the mean molecular mass to $\mu_{\rm atm} > 11$ at 99.7% probability (3σ). Such observations would conclusively rule out a hydrogen-rich nature of the atmosphere. Smaller observational uncertainties are required compared to Figure 2 because the weaker water absorption bands at $0.8-1\,\mu$m are not captured by *HST WFC3* observations alone.
.

butions and vertical extent, but to demonstrate that the proposed approach to estimate the mean molecular mass remains viable for non-gray clouds.

For demonstrative purposes, we model thin, high-altitude KCl and ZnS hazes using a standard analytical size distribution, a variation of the gamma distribution introduced by (Hansen 1971),

$$n(r) = \text{constant} \times r^{(1-3v_{\rm eff})/v_{\rm eff}} e^{-r/r_{\rm eff}v_{\rm eff}}. \quad (17)$$

The mean particle size $r_{\rm eff}$ is set to $0.3\,\mu$m and $1\,\mu$m for the KCl and ZnS haze scenarios, respectively. The variance of the size distributions, $v_{\rm eff}$, is 0.1. The hazes extend up to 5 mbar pressure level. The ratio of the condensed mass and the gas mass in the upper atmosphere is 1 ppm. The vertical extent of the hazes was chosen to obtain hydrogen-rich scenarios that show water absorption features with transit depth variations similar to those of water worlds, and, therefore, are most difficult to distinguish from water-dominated scenarios. The particle sizes and condensed mass fraction were selected to obtain haze spectra that considerably deviate from the assumption of gray clouds while simultaneously providing a reasonable fit to previously obtained transit depth measurements of GJ 1214b. Particles smaller than $0.3\,\mu$m would increasingly lead to steep slopes at near-infrared wavelengths. Larger particles would lead to an increasingly gray appearance of the clouds in transmission, and thus brings us back to the assumption made in Section 7.

Figure 7 shows that the cores of the strong molecular absorption features are largely unaffected by the wavelength dependence of the cloud/haze opacities. Estimates of the mean molecular mass made based on the feature cores in transmission spectra are, therefore, largely independent of the spectral properties of the cloud properties. The cores of molecular absorption features in exoplanet transmission spectra are largely independent of the cloud properties for two reasons. First, for transmission spectra, the observed transit depth at a given wavelength is almost exclusively determined by the strongest opacity source at that wavelength. This is as a result of the grazing geometry in which the transmission spectrum is formed (Brown 2001), combined with the exponential decrease in gas density with altitude. Second, aerosol opacities generally change more gradually with wavelength than molecular opacities. Molecular opacities at low pressures are dominated by sharp absorption lines and bands that result from quantum mechanical transitions between discrete vibrational and rotational states in the molecules. Light extinction by aerosols, on the other hand, is a result of interference of light that

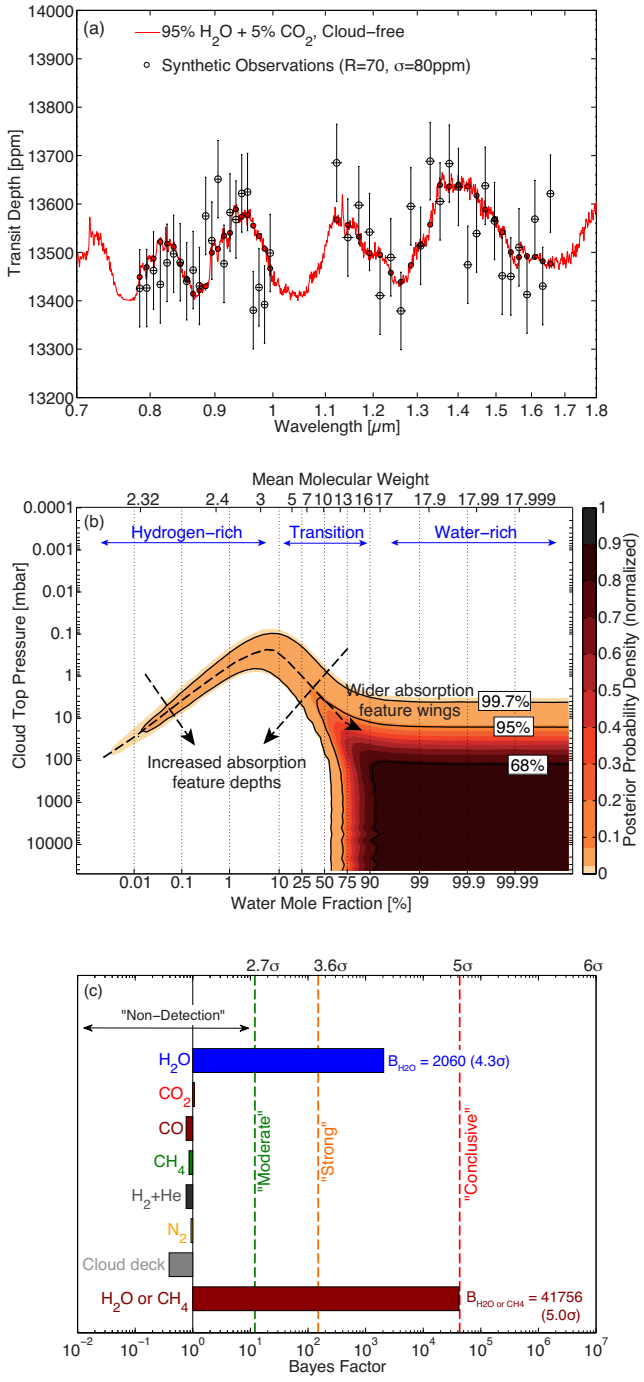

FIG. 6.— Synthetic 80 ppm observations and retrieval results for a cloud-free, water-dominated atmosphere on GJ 1214b. The atmospheric scenario and panels (a)-(c) are identical to the ones described Figure 2. 80 ppm observations of a water-dominated atmosphere would lead to a robust detection of water vapor ($B_{\rm H_2O}$ = 2060), but would be insufficient to robustly determine whether the atmosphere is water-dominated or hydrogen-dominated. Atmospheric scenarios are in agreement with the data along a contour of constant feature depth, resulting in a degeneracy between the mean molecular mass and cloud top pressure. The water mole fraction remains unconstrained between 0.02% and 100% at 3σ. The degeneracy is characteristic for transmission spectra that enable the detection of a single absorption, but do not provide sufficient constraints on the steepness of the feature wings or depth of other features to constrain the mean molecular mass.

was scattered, refracted, or diffracted by a generally continuous distribution of different particles sizes. Spectral features arise as result of the interference or the wavelength dependence of the complex refractive index, and the effects of the on the transmission spectrum are more gradual than molecular state transitions.

Second, while spectral features due to condensed phase absorption can be prominent in reflective an thermal emission spectra, modeling of exoplanet transmission spectra reveals that the features due to condensed phase absorption are generally far less pronounced. The reason for the difference is that multi-scattering of light can play a dominant role in the approximately nadir-viewing geometry associated with reflective and thermal emission spectra. The long pathways associated with light that is scattered multiple times within in a cloud can result in strong absorption features in reflective spectra of clouds. Water clouds, for example, will appear highly reflective at visible wavelengths for which the imaginary part of the refractive index is low (low absorption), while it will appear almost black at some near-infrared wavelength for which the imaginary part of the refractive index is high (high absorption). In the exoplanet transit geometry, however, any single scattering events will prevent the grazing light beams from the host star to arrive at the observer due to large distance between the target exoplanet and Earth.

3.5. *Probing the Composition of Volatile-Rich Atmospheres*

The distinction between H_2-dominated sub-Neptunes and water or ice-rich worlds described in Section 3 is solely based on the sharp contrast in molecular masses between hydrogen gas (H_2) and the ices (H_2O, CO_2, CO, CH_4, N_2, etc). The basic argumentation is that primordial H_2-dominated scenarios can be excluded if the mean molecular mass deviates significantly from $\mu_{\rm ave} = 2.3$. Measuring a high mean molecular mass, however, does not unambiguously determine the abundances of the individual volatile species in the atmosphere.

Figure 8 illustrates the difficulty of determining the mole fraction of the individual ice species for two vastly different atmospheres with identical molecular masses ($\mu_{ave} = 23.2$). Following the argumentation in Benneke & Seager (2012), the relative abundances of the ices that have IR absorption features, e.g., H_2O, CO_2, CO, CH_4, and NH_3 can be determined by comparing the transit depths in the strongest absorption bands of the different ices. An atmosphere with 80% H_2O + 20% CO_2 ($H_2O/CO_2 = 4$), for example, can be distinguished from an atmosphere with 90% H_2O and 10% CO_2 ($H_2O/CO_2 = 9$) because the transit depth within the CO_2 band at 4.3 μm would be higher relative to the transit depths in the H_2O bands.

Scenarios with similar relative abundances of absorbing gases, however, are practically indistinguishable through moderate-resolution ($R \sim 100$) infrared observations. Figure 8 shows that the NIR spectrum of a volatile-dominated atmosphere can remain virtually unchanged when the mole fraction of water is reduced from 80% to 8%. Spectrally inactive gases may be present in the correct ratios for the mean molecular mass to remain unchanged. A distinction between 80% water and 8% water therefore requires observations at short wavelengths.

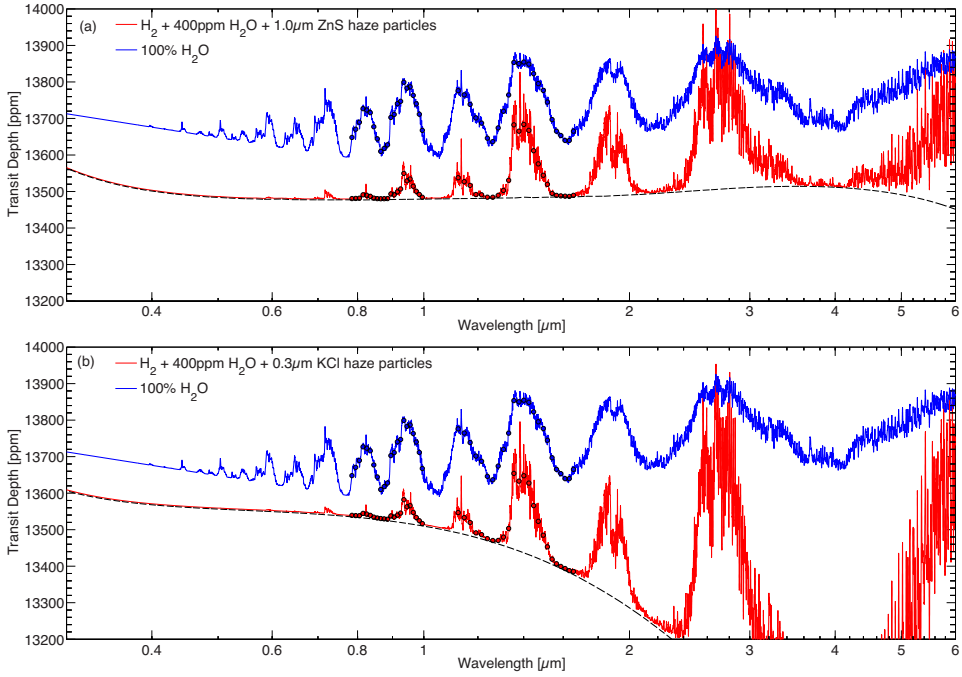

FIG. 7.— Comparison of model transmission spectra of water worlds and cloudy sub-Neptunes with non-gray high-altitude clouds. The red model spectra in panels (a) and (b) are model spectra for two alternative scenarios for hydrogen-rich atmospheres with thin, high-altitude hazes on GJ 1214b. Panel (a) assumes ZnS particles with a mean radius of 1 μm, while panel (b) assumes KCl particles with a mean radius of 0.3 μm. The particles sizes and cloud top pressure (5 mbar) are chosen such that the main water features in the HST WFC3 are of similar size as the ones predicted for water vapor atmospheres (blue spectra). The non-gray effects of the hazes only are shown by the black dashed spectra for which the gaseous absorption was set to zero. The cores of the strong molecular absorption features remain largely unaffected by the non-gray effects of the hazes. The steepness of the feature wings and the relative depths between different absorption features remain measures of mean molecular mass, even in the presence high-altitude non-gray hazes, as long as significant absorption features can be observed. The spectra are offset for clarity by slightly modifying the planet's radius at the reference pressure level.

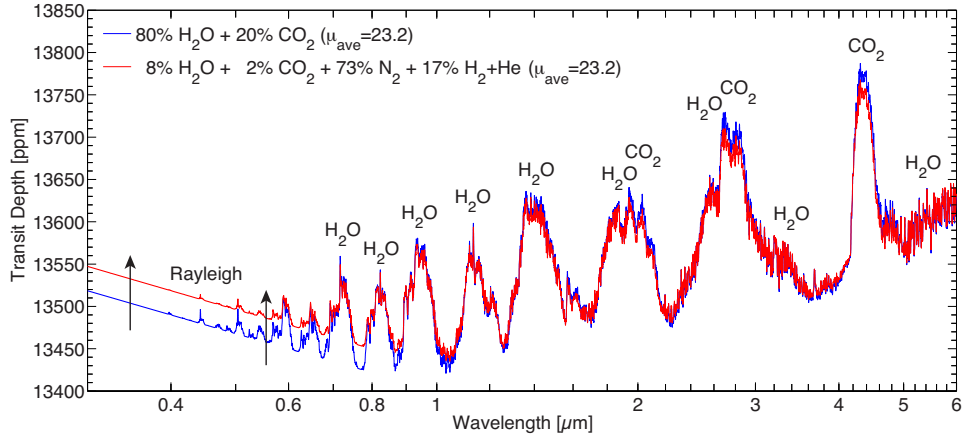

FIG. 8.— Challenge in distinguishing water-dominated atmospheres (80% H_2O) and water-rich atmospheres (80% H_2O) with similar mean molecular masses. Despite the vastly different compositions, the H_2O/CO_2-dominated atmosphere (80% H_2O and 20% CO_2) and H_2O/CO_2-rich, but N_2-dominated atmosphere (8% H_2O and 2% CO_2) show virtually identical infrared transmission spectra. Atmospheric sceanrios with the same mean molecular masses and the same relative abundances of the absorbers cannot be distinguished using moderate-resolution NIR transmission spectra alone (Benneke & Seager 2012). Distinction is possible only at short wavelengths ($\lambda < 1\,\mu$m) based on the increased Rayleigh scattering due to the presence of N_2 and $H_2 + He$. This is an example that even a strong detection of a water feature is not sufficient to determine whether the atmosphere is dominated by water vapor or only a small fraction of the gas is water vapor.

4. RESULTS: ANALYSIS OF PUBLISHED HST WFC3 SPECTRA OF GJ 1214B

A careful look at the observed *HST WFC3* spectrum by Berta et al. (2012) reveals that the transit depth variation resembles the trends expected for a clear, water-dominated atmosphere (Figure 9). The resemblance was also pointed out by Howe & Burrows (2012), but no assessment of the significance of the water absorption was provided. In this Section, we assess the significance of the trend in the transit depth variation in a statistically robust Bayesian way. We also assess which constraints on cloud-top pressure and water/hydrogen abundances can be derived from the *HST* observation by Berta et al. (2012). We demonstrate that the standard frequentist hypothesis testing based on attempting to reject the "flat spectrum" null-hypothesis or comparing data points "within" and "outside" the suggested features generally leads to ambiguous or misleading results.

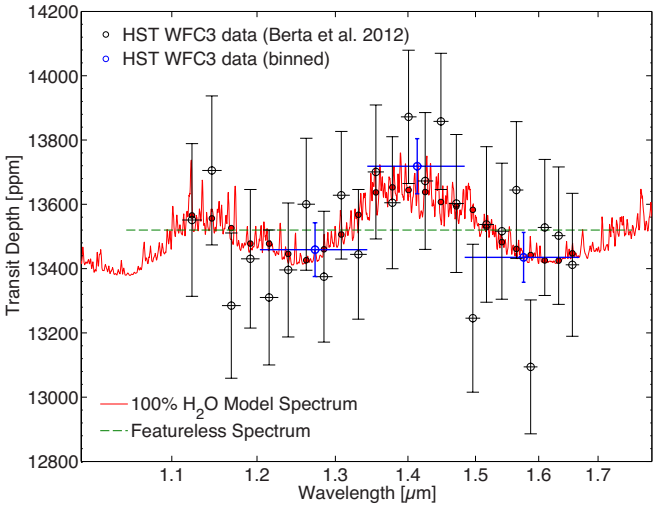

FIG. 9.— Observed *HST WFC3* observation of GJ 1214b by Berta et al. (2012) compared to model spectra. A featureless spectrum (green) and a model spectrum of a 100% water vapor atmosphere (red) are shown. Binned data points (blue) are illustrated to indicate the proposed transit depth variation near the 1.38 μm water feature. The binned data points would suggest a 2.44σ. The Bayesian analysis, however, indicates that there is there is little statistical evidence of water absorption in the *HST WFC3* observations. We emphasize that model-independent approaches to detect molecular absorption features by binning data points "inside" the suspected feature and comparing the transit depth to the surrounding "continuum" generally lead to ambiguous results because spectra of thick atmospheres lack clear separations between features and surrounding "continuum" (Table 5).

4.1. *Testing for the Presence of Water Absorption*

Our Bayesian analysis shows that, despite the suggested trend in the spectral data points, there is little statistical evidence of water absorption in the *HST WFC3* observations by Berta et al. (2012) (Table 4). The Bayesian factor $B_{\rm H_2O}$ describing our confidence in the presence of water is only 3.34, which is at best a weak suggestion that water absorption may be present. The Bayes factor for all other molecule and clouds are below ~ 2, indicating that no molecular absorption can be inferred from the data set.

It is important to note that standard model-independent approaches based on binning data points "within" and "outside" can lead to ambiguous or even misleading results for a spectral data set such as that provided by Berta et al. (2012). The reason is that molecular absorption features of thick atmospheres are generally not box-shaped. The gradual shape of the absorption feature wings make it impossible to unambiguously assign which data points are "inside" and "outside" of the water absorption features. Figure 9 and Table 5 show that the contrast between binned data points "inside" and "outside" can vary between $\sim 1.4\sigma$ and $\sim 2.65\sigma$, depending on which data points are considered within the feature. A finding of an absorption feature at 2.65σ could then be misinterpreted as strong suggestion of water absorption.

The Bayesian framework presented in this work does not lead to ambiguity because the data are taken as given. High confidence in the detection of water absorption is only assigned if the original data points follow the spectral shapes expected for atmospheres with water absorption, and no other molecular absorber can explain the data.

4.2. *Atmospheric Constraints*

The observed transit spectrum by Berta et al. (2012) leads to correlated constraints on the mean molecular mass (or water mole fraction) and the cloud top pressure that are characteristic for transit observations that show a weak trend of an absorption feature (Figure 10). The highest posterior probabilities exist along a curve of constant feature size corresponding to the apparent variation in the Berta et al. (2012) data. The 99.7% (3σ) extends, however, all the way to virtually featureless scenarios with low cloud top pressures below 0.0001 mbar.

Cloud-free hydrogen-rich scenarios (bottom left region of Figure 10) are disfavored by the observational data because the observed transmission spectra do not show the deep absorption features expected for clear hydrogen-rich atmospheres. At 99.7% probability (3σ), we conclude that cloud free atmospheres are only in agreement with the data if the mean molecular mass is above 7, corresponding to more than 30% H_2O in an H_2O-H_2 atmosphere.

Hydrogen-dominated scenarios with less than 2% water are possible only if clouds are present at low pressure such that the partial pressure of water vapor at the cloud surface does not exceed $p_{\rm H_2O} \sim 0.025$ mbar. Above 2% water vapor, the water vapor significantly increases the mean molecular mass, thereby reducing the size of the molecular absorption features. As a result, the 3σ contour of the Bayesian credible region peaks at $X_{\rm H_2O} \approx 2\%$ and $p_{\rm surf} = 10$ mbar and then continues to fall to higher cloud top pressures as the water mole fraction is increased further.

Our findings are in qualitative agreement with the list of implications for the atmospheric composition provided by Berta et al. (2012). One major difference is that we provide Bayesian credible regions describing our knowledge about the atmospheric composition in a statistically robust probabilistic way. Berta et al. (2012) derived implications on the atmospheric composition using χ^2 hypothesis testing and sequentially rejecting (or failing to reject) predefined atmospheric scenarios.

In additional difference is that we used the new mass, radius, and surface gravity estimates by Carter et al. (2013, in preparation). The new estimate for the planet's surface gravity is 53% higher than the one published by Charbonneau et al. (2009). As a result, the scale height and modeled feature depth for a given atmospheric scenario decreases by 34.5%.

5. DISCUSSION

5.1. *Scaling Laws for Required Observation Precision*

Planet and Star Scenario— The quantitative results obtained for the super-Earth GJ 1214b in Section 3 can be generalized for transiting super-Earth exoplanets with different bulk properties and different host stars using the scaling law

$$\frac{\sigma_{\rm req}}{\sigma_{\rm ref}} \approx \frac{R_P}{R_{P,\rm ref}} \frac{T_{eq}}{T_{eq,\rm ref}} \frac{g_{\rm p,ref}}{g_{\rm p}} \frac{R^2_{*,\rm ref}}{R^2_*}. \qquad (18)$$

The required precision $\sigma_{\rm req}$ is the maximum allowed observational uncertainty in the spectral transit depth measurements, R_p is the planetary radius, T_{eq} is the equilibrium temperature, $g_{\rm p}$ is the surface gravity of the

Retrieval model	Retrieval model parameters	Evidence $\ln(\mathcal{Z}_i)$	Best-fit $\chi^2_{\rm best-fit}$	Bayes factor $B_i = \mathcal{Z}_0/\mathcal{Z}_i$
Full hypothesis space	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CO_2}, X_{\rm CH_4}, X_{\rm CO}, X_{\rm H_2/He}, X_{\rm N_2})$	-162.96	12.24	Ref.
H_2O removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm CO_2}, X_{\rm CH_4}, X_{\rm CO}, X_{\rm H_2+He}, X_{\rm N_2})$	-163.16	14.31	$B_{\rm H_2O} = 3.34$
CO_2 removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CH_4}, X_{\rm CO}, X_{\rm H_2/He}, X_{\rm N_2})$	-164.16	12.20	$B_{\rm CO_2}$ =1.26
CH_4 removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CO_2}, X_{\rm CO}, X_{\rm H_2/He}, X_{\rm N_2})$	-163.19	12.18	$B_{\rm CH_4}$ =1.63
CO removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CO_2}, X_{\rm CH_4}, X_{\rm H_2/He}, X_{\rm N_2})$	-163.44	12.28	$B_{\rm CO}$ =1.70
N_2 removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CO_2}, X_{\rm CH_4}, X_{\rm CO}, X_{\rm N_2})$	-163.49	12.72	$B_{\rm N_2}$ =1.82
H_2/He mix removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CO_2}, X_{\rm CH_4}, X_{\rm CO}, X_{\rm H_2/He})$	-163.56	12.45	$B_{\rm H_2/He}$ =2.01
H_2O & CH_4 removed	$\frac{R_p}{R_*}$, $P_{\rm clouds}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm CO_2} X_{\rm CH_4}, X_{\rm CO}, X_{\rm H_2/He}, X_{\rm N_2})$	-163.65	17.56	$B_{\rm H_2O\,or\,CH_4} = 3.65$
Clouds removed	$\frac{R_p}{R_*}$, $\boldsymbol{\xi} = {\rm clr}(X_{\rm H_2O}, X_{\rm CO_2}, X_{\rm CH_4}, X_{\rm CO}, X_{\rm H_2/He}, X_{\rm N_2})$	-164.25	12.51	$B_{\rm Clouds}$ =1.23

TABLE 4
Results of Bayesian model comparison for observed *HST WFC3* transmission spectrum of GJ 1214b by Berta et al. (2012). The observations are depicted in Figure 9(a).

Binning scheme	Indices of data points "inside" the feature	Indices of data points in surrounding "continuum"		1 - p-Value		Detection significance	
#1 (Figure 9)	11-16	5-10 (left)	17-24 (right)	96.9%	98.5%	$2.16\,\sigma$ (left)	$2.44\,\sigma$ (right)
#2	10-17	5-9 (left)	18-24 (right)	82.9%	85.6%	$1.37\,\sigma$ (left)	$1.46\,\sigma$ (right)
#3	11-16	5-10 and 17-24		99.2%		$2.65\,\sigma$	
#4	10-17	5-9 and 18-24		90.5%		$1.67\,\sigma$	

TABLE 5
Ambiguity in the model-independent detection significances for non-box-shaped spectral features. The table presents the detection significances of an increased transit depth in the $1.38\,\mu$m water band derived from the published transmission spectrum of GJ 1214b by Berta et al. (2012). The indices of the data points are counted from the shortest wavelength data point #1 at $1.123\,\mu$m to the longest data point #24 at $1.656\,\mu$m (Figure 9). For thick atmospheres, there is no clear cutoff to determine which data points belong to the feature and which do not. The model-independent detection significance ranges between 1.46σ and 2.65σ depending on which data points are considered to be inside the water band and which data points are considered to be in the surrounding continuum.

planet, and R_* is the radius of the host star. Equation (18) is derived by relating the area $2R_pH$ of an annulus around the planet and a width of one scale height $H \propto T_{eq}/g$ to the cross sectional area of the stellar disk πR_*^2.

Equation (18) is valid for clear, hazy, and cloudy atmospheres to within a few percent error as long as identical scenarios for the atmospheric compositions and clouds/hazes properties are compared. This simple scaling law is valid for all types of atmospheres because differences between the transit depths observed at two different wavelengths are directly related to the difference in the altitudes Δz at which the atmosphere becomes opaque to a grazing light beam at the two wavelengths. The difference in the altitudes Δz, in turn, is related to the pressure scale on which the atmospheric scenarios, e.g., cloud top pressure and compositional profiles, are defined by $\Delta z = -H \cdot \Delta P/P$. Equation (18) does not account for the increase in atmospheric path length of grazing light beams due to an increase in the planet radius. The longer path length, however, generally leads to an approximately uniform increase in transit depths across the entire spectrum and has little effect on the transit depth variations.

Equation (18) shows that the required observational precision depends most strongly on the radius of the host star due to large differences in the stellar radii and the square dependence on that parameter. The stellar radius of the M-dwarf GJ 1214 is only 18.8% of the radius of a Sun-like star. As a result, the required photometric precision to characterize a planet orbiting a Sun-like star is ∼28 times larger than the required precision for the characterization of the same planet orbiting GJ 1214. Atmospheric characterizations of a planet orbiting a Sun-like star with the same apparent brightness would therefore require more than 500 times more observational time than a similar atmospheric characterization requires for a planet orbiting GJ 1214. The effect of the atmospheric temperature and planetary radius on the required precision scales only linearly. The required precision for the atmospheric characterization of an temperate super-Earth orbiting a nearby M-dwarf, for example, is only higher by a factor of $2-3$ higher compared to GJ 1214b (see Section 5.4 for a discussion). Similarly, the required precision would only increase by a factor of 2.4 if an Earth-sized planet was studied rather than GJ 1214b.

Spectral Resolution of Observations— The quantitative results of the photometric precision in Section 3 are presented for a resolving power $R = \lambda/\Delta\lambda = 70$. Higher resolution spectra, whose uncertainties are dominated by white noise, can provide the same information, even if the uncertainties of the individual data points are higher. Assuming that the observational uncertainties are dominated by white noise, the requirements on the precision of each single data point scale as

$$\sigma_{\rm HR,\,req} \approx \sigma_{\rm ref}\sqrt{\frac{\Delta\lambda_{\rm ref}}{\Delta\lambda_{\rm HR}}} = \sigma_{\rm ref}\sqrt{\frac{R_{\rm HR}}{R_{\rm ref}}}, \qquad (19)$$

for $R_{\rm HR} > R_{\rm ref}$. Equation 19 describes in a formal way that the required precision, $\sigma_{\rm HR,\,req}$, of spectra with a higher resolving power, $R_{\rm HR}$, can be binned to reduce the observational uncertainty of individual data points to $\sigma_{\rm ref}$ at a reference resolving power of $R_{\rm ref}$. The higher resolution spectra, therefore, contain at least as much information about the atmosphere as the spectra with ob-

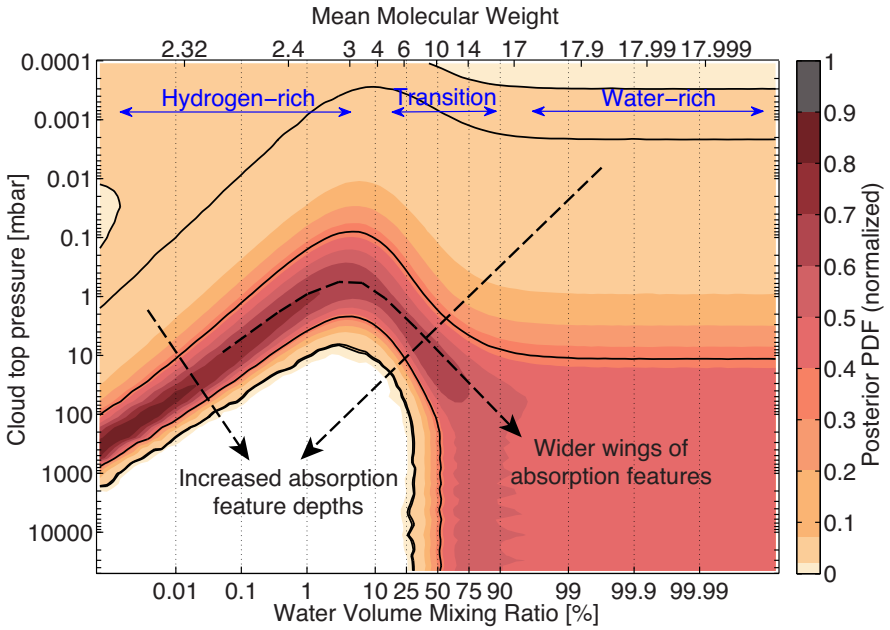

FIG. 10.— Atmospheric constraints derived from the observed transmission spectrum by Berta et al. (2012). The shading illustrates the marginalized posterior probability as a function of mean molecular mass and cloud top pressure for a two-component H_2O+H_2 atmosphere. The black contour lines indicate the 68%, 95%, and 99.7% Bayesian credible regions. The 95% Bayesian credible region extends over large parts of the parameter space, preventing an unambiguous characterization of the atmosphere of GJ 1214b with the currently available data. Starting from the most robust statement, we can rule out clear hydrogen-dominated atmospheres with hydrogen mole fractions above 40% at $> 3\sigma$. Water-dominated and cloudy, hydrogen-dominated atmospheres are, however, in agreement with the data at 1σ due to degeneracy between the water mole fraction and the cloud top pressure (compare Figure 6). The prior probability is uniform in the parameter space spanned by $\log(p_{surf})$, $\xi_1 = \log(X_{H2O}/\sqrt{X_{H2O} \cdot X_{H2}})$.

servational uncertainty $\sigma_{\rm ref}$ and reference resolving power $R_{\rm ref}$.

Equation (19) can be regarded as a conservative scaling law because observations with significantly higher resolving power than $R = \lambda/\Delta\lambda = 70$ may capture individual peaks within the water absorption features that are not captured at low resolution. High resolution observations may therefore provide additional information to better constrain the gradient $dR_{P,\lambda}/d(\ln\sigma_\lambda)$ and thus the mean molecular mass (equation1). We confirm in numerical studies that equation (19) is valid for resolving power between $\sim$ 50 and several hundreds. A detailed numerical exploration of the effect of the resolving power is beyond the scope of this study.

5.2. *HD 97658b, 55 Cnc e, and GJ 436b*

Besides GJ 1214b, the transiting exoplanets in the mass regime of super-Earths and Neptunes that are most amenable to study are HD 97658b, 55 Cnc e, and GJ 436b. Their bulk densities are high enough to require a larger ice or rock fraction than the solar system ice giants, but far too low to be explained by an entirely Earth-like rocky composition. Similar to GJ 1214b, the question arises whether these planets are surrounded by a thick hydrogen-dominated envelope or a ice-dominated gas envelope.

We find from Equation (18) that the precision required to distinguish between water and hydrogen-rich scenarios for HD 97658b, 55 Cnc e, and GJ 436b is significantly higher (Table). The main driver in increasing the required precision is the larger host star diameters. It should be noted, however, the photon flux received from HD 97658 and 55 Cnc is orders of magnitude higher due to the brightness of the stars (Table 6). Significantly higher transit depth precision may therefore be achievable per transit if photon limited observations and a high integration efficiency can be achieved. The implications of the host star brightness on the instrumental effects and integration are specific to each instrument and beyond the discussion in this work.

Transmission spectra of the exoplanet HD 97658b are displayed in Figure 11 to further guide future observations. The planet was recently found to transit are displayed (Henry et al. 2011; Dragomir et al. 2013). Cloud-free hydrogen-dominated atmospheres would results in transit variation of $100 - 200$ ppm that would readily be detectable with currently available instrumentation. A confident detection of water vapor or other volatile-rich atmospheric scenarios would require a precision below 10 ppm at moderate spectral resolution (Table 6).

5.3. *Uncertainty Reduction through Stacking Transit Observations*

In Section 3, we made the assumption that the uncertainty in the spectral data points scales inversely with the number of transits observed and photons collected. Based on this assumption, we calculated how many transit observations need to be stacked with current observational techniques to detect water absorption in the atmospheres of super-Earths and to distinguish between hydrogen-rich sub-Neptunes and water worlds. The assumption that the uncertainty scales inversely with the square root of the number of observed transits is true for observational uncertainties that are dominated by white noise, such as photon-noise or stellar granulation noise. Instrumental effects or long-period stellar variability may ultimately set a lower limit on achievable precision for transit observations. To date, however, it is not clear

Planet	Scaling factor	Required precision σ_{req} at $R = 70$: Detection of H_2O atmosphere at $B \approx 2000$ ($> 4\,\sigma$)	Required precision σ_{req} at $R = 70$: Unambiguous distinction between clear H_2O and H_2 atmospheres	Host star brightness: K	Host star brightness: V
GJ 1214b	Reference	~ 80 ppm	~ 60 ppm	8.78	14.67
HD 97658b	0.092	7.4 ppm	5.5 ppm	5.73	7.71
55 Cnc e	0.101	8.1 ppm	6.1 ppm	4.02	5.95
GJ 436b	0.367	29.4 ppm	22.0 ppm	6.07	10.59

TABLE 6
Scaling factors and estimates of required precision in transit depth measurements to study the atmospheres of HD 97658b, 55 Cnc e, and GJ 436b. Stellar and planet parameters are taken from Carter et al., 2013 for GJ 1214b, Dragomir et al. (2013) for HD 97658b, Demory et al. (2011) and Gillon et al. (2012) for 55 Cnc e, and Torres et al. (2008) and Knutson et al. (2011) for GJ 436b.

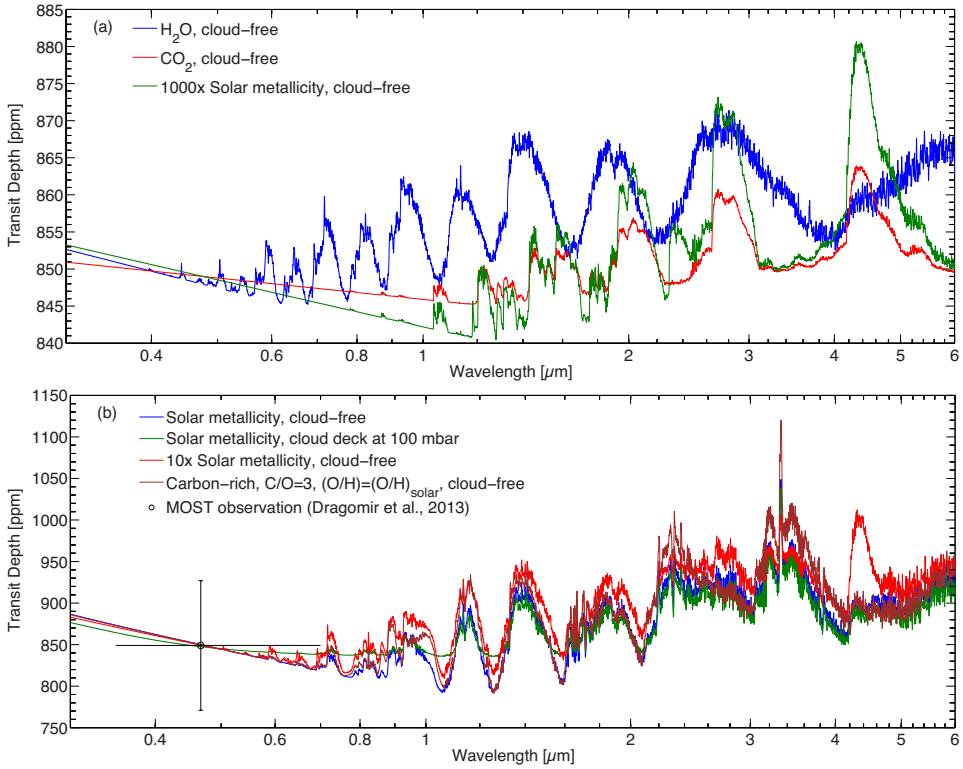

Fig. 11.— Model transmission spectra for the super-Earth HD 97658b. Panel (a) shows spectra for atmospheres dominated by volatiles with high mean molecular masses. Panel (b) shows spectra for hydrogen-dominated scenarios. All spectra are modeled to match the transit depth measurement in the *MOST* bandpass by Dragomir et al. (2013) (black).

whether and after how many transits the transit depth uncertainty will reach a lower limit.

On the contrary, there is impressive empirical evidence from space-based observatories such as *Kepler* and *HST* that the transit depth uncertainty continues to decrease as the number of stacked transits increases to a few, tens, and even several hundred transits. A prominent example is the continuous *Kepler* observations of the hot-Jupiter TrES-2b from which the visible broadband transit depth was determined to a precision of 1.7 ppm by combining hundreds of transits (Barclay et al. 2012). Spectroscopic observations of three transits of GJ 1214b using *HST WFC3* by Berta et al. (2012), reaching a precision within 10% of the photon limit, suggest that a similar trend is possible for spectroscopic transit observations. More observations are required to fully understand whether the uncertainty of exoplanet transmission spectra is ultimately limited by instrumental and/or astrophysical noise.

5.4. *Atmospheric Characterizations of Potentially Habitable Worlds around M-dwarfs*

One compelling feature of the scaling law, Equation (18), is that the photometric precision required to perform atmospheric characterizations scales only linearly with the atmospheric temperature and planet size. The weak scaling illustrates the great potential of transmission spectroscopy to characterize temperate and small exoplanets (Deming et al. 2009; Seager & Deming 2010). In principle, transit observations with currently available instrumentation have the potential to characterize the atmospheres of potentially habitable super-Earths orbiting nearby M-dwarfs if observational errors continue to decrease as the number of transits is increased to tens of transits (see Section 5.3).

For example, for a super-Earth planet that orbits

a GJ 1214-like star in the habitable zone at a moderate temperature of $T_{eq} = 300\,K$, the requirements on observational time will rise by only a factor of $(T_{eq,P}/T_{eq,\mathrm{GJ1214b}})^2 \approx 3.3$ compared to the ones presented in Section 3. Given a sufficiently large observational program that covers $30-50$ transits, we would, in principle, have the capability to find water or methane features with currently available instrumentation, assuming that observational uncertainty remains to be dominated by the photon noise. Considering that the orbital period of a temperate planet orbiting a GJ 1214-like star is only ∼9 days, up to 40 transit observations could theoretically be performed over the course of a year.

5.5. *Distinction Based on Rayleigh Scattering Slope*

Observations of the slope of the Rayleigh scattering at short wavelengths provide an alternative way of constraining the mean molecular mass. In practice, however, the Rayleigh slope provides only an upper limit on the mean molecular mass, thus only a lower limit on the hydrogen fraction in the atmosphere. The reason is the presence of clouds or large particle hazes can reduce the observed slope at visible and UV wavelengths.

A hydrogen-dominated atmosphere could be identified if a steep negative slope is observed in the UV-visible part of the transmission spectrum. If there is no steep slope, however, the distinction between hydrogen-dominated and volatile-dominated is not possible because the lack of a steep slope at short wavelength may be due to a high mean molecular mass or the presence of large particle clouds. The upper limit on the mean molecular mass can therefore be determined as

$$\mu_{\mathrm{mix}} \lesssim \frac{4k_B T}{gR_*} \frac{\ln\left(\frac{\lambda_1}{\lambda_2}\right)}{\left(\frac{R_p}{R_*}\right)_{\lambda_2} - \left(\frac{R_p}{R_*}\right)_{\lambda_1}}, \qquad (20)$$

where we derived equation (20) from Equation (1) using that the Rayleigh cross section σ is proportional to λ^{-4} . We assumed that, at least, two measurements of the transit depth R_p/R_* are available at wavelengths λ_1 and λ_2 at which Rayleigh scattering dominates, and we incorporated the uncertainty factor $(1 \pm \delta T/T)$ from Equation (20) into the approximately smaller sign.

As an example, we consider a haze-free water-rich scenario on GJ 1214b in comparison to a range of hydrogen-dominated scenarios with high-altitude ZnS hazes with different size distributions (Figure 7(a)). In the absence of hazes, the slope of the Rayleigh scattering signature at short wavelengths can determined from equation (20) using an "approximately equal" sign instead of the "approximately smaller" sign. In the presence of the ZnS hazes, however, the slope at short wavelengths depends not only on the mean molecular mass, but also on the size distribution of the haze particles, the vertical extent of the particles, and the amount of condensed mass. The slope is therefore not an unambiguous measure of the mean molecular mass. For haze particles that are small compared to the wavelength, the slope of the Rayleigh signature remains unchanged. As the haze particles become bigger, the slope decreases if the particles are present at high altitude in sufficient abundance. In the limit of gray, large particle clouds, more and more of the spectrum becomes flat.

An upper limit on the mean molecular mass can be determined based on the detection of a straight Rayleigh signature in the UV-visible spectrum because, at least in the limits of Mie scattering theory of spherical particles, the opacities of particles do not change with a slope greater than than $\sigma \propto \lambda^{-4}$ for any realistic particle size distributions. Besides, we are not aware of any condensate substances for which changes in the real or imaginary refractive index at UV-visible wavelengths would steepen the slope across a significant portion of the UV-visible spectrum. As a result, haze particles would flatten not steepen the Rayleigh slope at UV-visible transmission spectrum and we can provide a lower limit, but not an upper limit on the mean molecular mass.

6. SUMMARY AND CONCLUSIONS

We demonstrated that one can unambiguously distinguish between cloudy Neptunes with hydrogen atmospheres and low-density super-Earths with water/volatile-dominated atmospheres by observing the relative sizes and wing steepnesses of the absorption features in the planet's NIR transmission spectrum. The proposed observational distinction offers a promising approach to break the compositional degeneracy in the interior modeling of low-density super-Earths and Neptunes. We argue that the distinction can be achieved efficiently for super-Earths orbiting M-dwarfs by observing water features at moderate spectral resolution ($R \sim 100$) near the brightness peak of the M-dwarf's spectrum.

In this work, we use the super-Earth GJ 1214b as a case study and provide a scaling law that scales our quantitative results to other transiting super-Earths or Neptunes such as HD 97658b, 55 Cnc e, and GJ 436b. For GJ 1214b, we show quantitatively that an unambiguous distinction between cloud-free water-dominated atmospheres and cloudy hydrogen-dominated atmospheres is possible if the observational uncertainties can be reduced by a factor of ~ 3 compared to the published *HST WFC3* and *VLT* transit observations by Berta et al. (2012) and Bean et al. (2010). Similar results can be achieved using *HST WFC3* alone if the observational uncertainty can be reduced to 35 ppm at $R = 70$.

All results and spectra presented for GJ 1214b are derived using the new mass, radius, and surface gravity estimates for GJ 1214b by Carter et al. (2013, in preparation). The 53% increase in surface gravity leads to a 35% decrease in scale height, thereby significantly altering the strengths of spectral signatures compared to previously published models.

The required decrease in observational uncertainty may be achievable for GJ 1214b by observing $10-20$ transits in large programs with *Hubble* and/or ground-based telescopes. For example, observing 15 transits with *HST WFC3* and increasing the integration efficiency from 10% to ∼50% using the new spatial scan mode on *WFC3* (Deming et al. 2013) would increase the total number of detected photons by a factor of 25 compared to Berta et al. (2012). Assuming that the observations remain dominated by white-noise (see Section 5.3 for a discussion), such an observational program would be able to unambiguously distinguish between a cloud-free water-dominated atmosphere and a cloudy hydrogen-dominated atmosphere.

The results in this work were obtained using an advanced Bayesian retrieval framework. The framework not only constrains the atmospheric parameters, but also determines in a statistically robust way which molecular species and types of clouds can be inferred from the observational data. One main advantage over model-independent approaches to infer molecular absorbers is that the Bayesian approach inherently accounts for overlapping spectral features. It assigns a high probability for the presence of a particular gas or cloud type only if that gas or cloud type represents the only explanation for the observed data. The Bayesian approach also enables us to quantify our confidence in the presence of molecular species if the spectral features are not box-shaped and are not surrounded by a flat continuum. Spectra of thick atmospheres generally lack a flat continuum and the signatures of water vapor and methane resemble characteristic changes in the transit depth across the whole spectrum rather than distinct features. Finally, the observational data in the Bayesian framework are analyzed at full resolution, thereby avoiding any information loss and ambiguity introduced by binning the data.

The super-Earth GJ 1214b is currently the super-Earth most amenable to spectroscopic characterizations. The diminutive stellar radius (0.189 $R_{\odot}$) and the high planetary temperature (>500 K) result in relatively large transit depth variations if an atmospheric envelope is present. The atmospheric characterization of other currently known transiting super-Earths, such as 55 Cnc e and HD 97658b, is significantly more challenging because the larger host star radius decreases the transit signal by a factor of ten or more.

Large efforts for the discovery of GJ 1214b analogues are, however, ongoing. Ground-based transit surveys such as *MEarth* (Nutzman & Charbonneau 2008) currently present the most promising pathway to detect super-Earths around nearby M-dwarfs that are most amenable to study. The *TESS* mission expected to launch in 2017 will survey the full sky for planets transiting nearby M-dwarfs. If an exoplanet like GJ 1214b is found in the habitable zone of a close-by M-dwarf, this work indicates that we may have the capability to spectroscopically probe the atmosphere of a potentially habitable planet - not only in the next decades with *JWST* or *TPF*-like missions, but today with *HST* and ground-based telescopes.

We thank Jacob Bean, Zachory Berta, Jean-Michel Désert, and Andras Zsom for valuable discussions. We thank the anonymous referee for thorough reading and thoughtful comments that improved the manuscript.

REFERENCES

Barclay, T., et al. 2012, The Astrophysical Journal, 761, 53
Bean, J. L., Kempton, E. M.-R., & Homeier, D. 2010, Nature, 468, 669
Bean, J. L., et al. 2011, The Astrophysical Journal, 743, 92
Benneke, B., & Seager, S. 2012, The Astrophysical Journal, 753, 100
Berta, Z. K., Charbonneau, D., Bean, J., Irwin, J., Burke, C. J., Désert, J.-M., Nutzman, P., & Falco, E. E. 2011, The Astrophysical Journal, 736, 12
Berta, Z. K., et al. 2012, The Astrophysical Journal, 747, 35
Brown, T. M. 2001, The Astrophysical Journal, 553, 1006
Carter, J. A., Winn, J. N., Holman, M. J., Fabrycky, D., Berta, Z. K., Burke, C. J., & Nutzman, P. 2011, The Astrophysical Journal, 730, 82
Charbonneau, D., et al. 2009, Nature, 462, 891
Croll, B., Albert, L., Jayawardhana, R., Kempton, E. M.-R., Fortney, J. J., Murray, N., & Neilson, H. 2011, The Astrophysical Journal, 736, 78
Crossfield, I. J. M., Barman, T., & Hansen, B. M. S. 2011, The Astrophysical Journal, 736, 132
de Mooij, E. J. W., et al. 2012, Astronomy & Astrophysics, 538, A46
Deming, D., et al. 2009, Publications of the Astronomical Society of the Pacific, 121, 952
—. 2013, arXiv:1302.1141
Demory, B.-O., et al. 2011, Astronomy & Astrophysics, 533, A114
Désert, J.-M., Kempton, E. M.-R., Berta, Z. K., Charbonneau, D., Irwin, J., Fortney, J., Burke, C. J., & Nutzman, P. 2011, The Astrophysical Journal, 731, L40
Dragomir, D., Matthews, J. M., Winn, J. N., Rowe, J. F., & Team, M. S. 2013, arXiv:1302.3321
Feroz, F., & Hobson, M. P. 2008, Monthly Notices of the Royal Astronomical Society, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, Monthly Notices of the Royal Astronomical Society, 398, 1601–1614
Fortney, J. J., Mordasini, C., Nettelmann, N., Kempton, E. M.-R., Greene, T. P., & Zahnle, K. 2013, A Framework for Characterizing the Atmospheres of Low-Mass Low-Density Transiting Planets, arXiv e-print 1306.4329
Gillon, M., et al. 2012, Astronomy & Astrophysics, 539, A28
Gregory, P. C. 2007, Monthly Notices of the Royal Astronomical Society, 374, 1321
Hansen, J. E. 1971, Journal of the Atmospheric Sciences, 28, 1400
Hansen, J. E., & Travis, L. D. 1974, Space Science Reviews, 16, 527
Heng, K., & Kopparla, P. 2012, The Astrophysical Journal, 754, 60
Henry, G. W., Howard, A. W., Marcy, G. W., Fischer, D. A., & Johnson, J. A. 2011, arXiv:1109.2549
Hobson, M. P., & Jaffe, A. H. 2009, Bayesian Methods in Cosmology (Cambridge University Press)
Howe, A. R., & Burrows, A. S. 2012, The Astrophysical Journal, 756, 176
Knutson, H. A., et al. 2011, The Astrophysical Journal, 735, 27
Kuchner, M. J. 2003, The Astrophysical Journal, 596, L105
Kurokawa, H., & Kaltenegger, L. 2013, Atmospheric mass loss and evolution of short-period exoplanets: the examples of CoRoT-7b and Kepler-10b, arXiv e-print 1306.0973
Léger, A., et al. 2004, Icarus, 169, 499
Miller-Ricci, E., Seager, S., & Sasselov, D. 2009, The Astrophysical Journal, 690, 1056
Morley, C. V., Fortney, J. J., Kempton, E. M.-R., Marley, M. S., Visscher, C., & Zahnle, K. 2013, Quantitatively Assessing the Role of Clouds in the Transmission Spectrum of GJ 1214b, arXiv e-print 1305.4124
Morley, C. V., Fortney, J. J., Marley, M. S., Visscher, C., Saumon, D., & Leggett, S. K. 2012, The Astrophysical Journal, 756, 172
Nettelmann, N., Fortney, J. J., Kramm, U., & Redmer, R. 2011, The Astrophysical Journal, 733, 2
Nutzman, P., & Charbonneau, D. 2008, Publications of the Astronomical Society of the Pacific, 120, 317, ArticleType: research-article / Full publication date: March 2008 / Copyright © 2008 The University of Chicago Press
Querry, M. 1987, 333
Rogers, L. A., & Seager, S. 2010a, The Astrophysical Journal, 712, 974
—. 2010b, The Astrophysical Journal, 716, 1208
Seager, S., & Deming, D. 2010, Annual Review of Astronomy and Astrophysics, 48, 631
Seager, S., Kuchner, M., Hier-Majumder, C. A., & Militzer, B. 2007, The Astrophysical Journal, 669, 1279
Sellke, T., Bayarri, M. J., & Berger, J. O. 2001, The American Statistician, 55, 62
Sivia, D. S., & Skilling, J. 2006, Data analysis: a Bayesian tutorial (Oxford University Press)
Skilling, J. 2004, AIP Conference Proceedings, 735, 395

Teske, J. K., Turner, J. D., Mueller, M., & Griffith, C. A. 2013, Monthly Notices of the Royal Astronomical Society, 431, 1669

Torres, G., Winn, J. N., & Holman, M. J. 2008, The Astrophysical Journal, 677, 1324

Trotta, R. 2008, Contemporary Physics, 49, 71

Winn, J. N. 2011, in Exoplanets (University of Arizona Press)