THE UNIVERSITY OF ARIZONA
University Libraries

UA CAMPUS REPOSITORY

NO THERMAL INVERSION AND A SOLAR WATER ABUNDANCE FOR THE HOT JUPITER HD 209458B FROM HST /WFC3 SPECTROSCOPY

Item Type	Article
Authors	Line, Michael R.; Stevenson, K. B.; Bean, Jacob L.; Desert, Jean-Michel; Fortney, Jonathan J.; Kreidberg, Laura; Madhusudhan, Nikku; Showman, Adam P.; Diamond-Lowe, Hannah
Citation	NO THERMAL INVERSION AND A SOLAR WATER ABUNDANCE FOR THE HOT JUPITER HD 209458B FROM HST /WFC3 SPECTROSCOPY 2016, 152 (6):203 The Astronomical Journal
DOI	10.3847/0004-6256/152/6/203
Publisher	IOP PUBLISHING LTD
Journal	The Astronomical Journal
Rights	© 2016. The American Astronomical Society. All rights reserved.
Download date	25/08/2022 19:06:32
Item License	http://rightsstatements.org/vocab/InC/1.0/
Version	Final published version
Link to Item	http://hdl.handle.net/10150/622434

The Astronomical Journal, 152:203 (14pp), 2016 December
doi:10.3847/0004-6256/152/6/203

© 2016. The American Astronomical Society. All rights reserved.

NO THERMAL INVERSION AND A SOLAR WATER ABUNDANCE FOR THE HOT JUPITER HD 209458B FROM *HST*/WFC3 SPECTROSCOPY

Michael R. Line[1,2,3,10], Kevin B. Stevenson[4,11], Jacob Bean[4], Jean-Michel Desert[5], Jonathan J. Fortney[6], Laura Kreidberg[4], Nikku Madhusudhan[7], Adam P. Showman[8], and Hannah Diamond-Lowe[9]
[1] NASA Ames Research Center, Moffet Field, CA 94035, USA
[2] Bay Area Environmental Research Institute, 625 2nd Street, Suite 209, Petaluma, CA 94952, USA
[3] School of Earth & Space Exploration, Arizona State University, USA
[4] Department of Astronomy & Astrophysics, University of Chicago, 5640 S Ellis Avenue, Chicago, IL 60637, USA
[5] University of Amsterdam, The Netherlands
[6] Department of Astronomy & Astrophysics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
[7] Institute of Astronomy, University of Cambridge, Cambridge CB3 0HA, UK
[8] Department of Planetary Sciences and Lunar and Planetary Laboratory, University of Arizona, 1629 E. University Blvd., Tucson, AZ 85721, USA
[9] Department of Astronomy, Harvard Smithsonian Center for Astrophysics, 60 Garden Street, MS-10, Cambridge, MA 02138, USA

Received 2016 May 27; revised 2016 September 29; accepted 2016 September 29; published 2016 December 2

ABSTRACT

The nature of the thermal structure of hot Jupiter atmospheres is one of the key questions raised by the characterization of transiting exoplanets over the past decade. There have been claims that many hot Jupiters exhibit atmospheric thermal inversions. However, these claims have been based on broadband photometry rather than the unambiguous identification of emission features with spectroscopy, and the chemical species that could cause the thermal inversions by absorbing stellar irradiation at high altitudes have not been identified despite extensive theoretical and observational effort. Here we present high-precision *Hubble Space Telescope* WFC3 observations of the dayside thermal emission spectrum of the hot Jupiter HD 209458b, which was the first exoplanet suggested to have a thermal inversion. In contrast to previous results for this planet, our observations detect water in absorption at 6.2σ confidence. When combined with *Spitzer* photometry, the data are indicative of a monotonically decreasing temperature with pressure over the range of 1–0.001 bars at 7.7σ confidence. We test the robustness of our results by exploring a variety of model assumptions, including the temperature profile parameterization, presence of a cloud, and choice of *Spitzer* data reduction. We also introduce a new analysis method to determine the elemental abundances from the spectrally retrieved mixing ratios with thermochemical self-consistency and find plausible abundances consistent with solar metallicity (0.06–10 × solar) and carbon-to-oxygen ratios less than unity. This work suggests that high-precision spectrophotometric results are required to robustly infer thermal structures and compositions of extrasolar planet atmospheres and to perform comparative exoplanetology.

Key words: methods: statistical – planets and satellites: atmospheres – planets and satellites: composition – planets and satellites: gaseous planets – planets and satellites: individual (HD 209458b) – techniques: spectroscopic

1. INTRODUCTION

Hubble Space Telescope (*HST*) Wide Field Camera 3 (WFC3) spectrophotometric observations of transiting exoplanets are rapidly transforming our knowledge of planetary atmospheres (Berta et al. 2012; Swain et al. 2013; Deming et al. 2013; Huitson et al. 2013; Line et al. 2013b; Mandell et al. 2013; Wakeford et al. 2013; Ehrenreich et al. 2014; Fraine et al. 2014; Knutson et al. 2014; Kreidberg et al. 2014a, 2014b, 2015; McCullough et al. 2014; Ranjan et al. 2014; Stevenson et al. 2014c; Wilkins et al. 2014; Sing et al. 2015; Evans et al. 2016). As the highest-precision and most robust spectroscopy obtainable with current instruments, WFC3 data allow us to infer the thermal structures, energy transport, molecular abundances, and cloud properties in exoplanet atmospheres. Determining these properties allows us to then begin to understand the fundamental processes occurring in planetary atmospheres over a wide range of conditions.

The thermal structures of hot Jupiter atmospheres, and in particular the question of thermal inversions, have driven much of the characterization work on these planets over the past decade (for reviews, see Burrows & Orton 2010; Madhusudhan et al. 2014b, p. 739; Crossfield et al. 2015). Thermal inversions typically occur in the presence of a high-altitude heating source, usually a strong short-wave absorber. Most solar system bodies with substantial atmospheres possess a thermal inversion (e.g., Robinson & Catling 2014), due to ozone in the case of Earth, or aerosol heating in the case of the giant planets. Hot Jupiters were theorized to have thermal inversions due to the possible presence of strong gas-phase optical absorbers (Hubeny et al. 2003; Burrows et al. 2008), for example, TiO and VO, which can persist at high temperatures ($T > 2000$ K at 1 bar; Fortney et al. 2006, 2008), or polysulfur aerosols (Zahnle et al. 2009).

Observations using five broadband channels on the *Spitzer Space Telescope* claimed the first detection of a thermal inversion in a hot Jupiter atmosphere, in the prototype first transiting planet HD 209458b (Burrows et al. 2007; Knutson et al. 2008), presumably confirming the presence of a high-altitude absorber, the leading candidates being TiO and VO given their presence in the gas phase at HD 209458b's temperatures. Evidence for the inversion was based on the relatively high fluxes of the 4.5 and 5.8 μm channels compared with the 3.6 and 8.0 μm channels. These results were

---

[10] Hubble Postdoctoral Fellow.

[11] Sagan Postdoctoral Fellow.

challenging to interpret theoretically as, first, for plausible compositions, the contribution functions of the four *Spitzer* channels greatly overlap, so it is extremely difficult to get significantly differing brightness temperatures between the channels, and second, the infrared photosphere (optical depth of unity) tends to occur at deeper layers than the layers at which most of the visible layers are absorbed. An inversion is formed because at the altitudes at which the visible radiation is strongly absorbed the infrared emissivity is low, requiring hotter temperatures to maintain radiative equilibrium (e.g., Robinson & Catling 2014). Nevertheless, Madhusudhan & Seager (2009) and Line et al. (2014), using retrieval methods, were able to fit the Knutson et al. (2008) *Spitzer* data by invoking an inversion above the ~1 bar level and large abundances of CO. The large abundance of CO is required to decouple the 4.5 μm (and to a lesser extent 5.8 μm) weighting functions from the rest of the channels, permitting the large temperature differences. It is unclear, however, how well the mean infrared opacity required to produce the retrieved thermal structures agrees with those derived from the abundances.

In addition to the Knutson et al. (2008) results, the secondary eclipse of HD 209458b has been observed and processed repeatedly by many different groups with a variety of different space-based (*Spitzer* IRS: Richardson et al. 2007; Swain et al. 2008; *Spitzer* MIPS: Deming et al. 2005; *Spitzer* IRAC: Diamond-Lowe et al. 2014; Zellem et al. 2014b; Evans et al. 2015; *HST* NICMOS: Swain et al. 2009; Crossfield et al. 2012) and ground-based (Richardson et al. 2003; Zellem et al. 2014a; Schwarz et al. 2015) instruments, many studies of which questioned the possible existence of the previously claimed inversion, though the robustness of some of these earlier data is questionable (e.g., Gibson et al. 2011, 2012).

As an example, Diamond-Lowe et al. (2014) reanalyzed the Knutson et al. (2008) data set combined with new data taken with more optimal observing modes at 3.6, 4.5, and 8.0 μm. This resulted in a significant reduction of the 4.5 and 5.8 μm fluxes (which drove the need for an inversion) and showed that a thermal inversion was no longer required to model those data. Additionally, with an entirely different method altogether, Schwarz et al. (2015), using high-dispersion spectroscopy with the CRIRES instrument combined with cross-correlation data reduction techniques, were able to rule out an inversion between 1 bar and 1 mbar. However, they also did not make a strong detection of molecular absorption indicative of a normal thermal structure.

Furthermore, evidence for inversions in other planets of similar or hotter temperatures is currently lacking. In the years since the HD 209458b observations, retrieval methods of interpreting spectra have shown that often models without temperature inversions could explain wide-band *Spitzer* data as readily as models with inversions (Madhusudhan & Seager 2009; Line et al. 2014). A growing body of work has tried to explain why inversions may not occur. Spiegel et al. (2009) and Showman et al. (2009) investigated how condensation in cold traps could lead to removal of TiO and VO from the gas phase. Madhusudhan et al. (2011) suggested that the possible lack of inversions could be due to high carbon-to-oxygen (C/O) ratios. High C/O would deplete the oxygen required to build the strongly absorbing oxides. However, Line et al. (2014) and Benneke (2015) showed that there is no strong evidence for high C/O ratios over a wide range of planetary conditions–including those in which inversions are expected. An alternative hypothesis for the apparent lack of inversions, brought forth by Knutson et al. (2010), suggests that high incident UV flux could potentially photolyze high-altitude absorbers.

Recently, Haynes et al. (2015) suggested the presence of an inversion along with potential evidence for TiO emission in the hot Jupiter WASP-33b by combining *HST*, *Spitzer*, and ground-based data; while tantalizing evidence, a bulk of the inversion evidence is driven by the ground-based z'-band point and the two bluest channels in WFC3. Furthermore, WASP-33 is a delta-scuti variable, thus making the data analysis and interpretation extremely challenging.

The power of spectroscopy, compared to photometry, is that the planetary thermal emission is probed across a range of wavelengths, and hence a range of depths in the atmosphere, simultaneously. The interpretation of such observations is in principle much more straightforward than in wide-band photometry. Molecular features seen in emission can be interpreted as a telltale sign of an inverted temperature structure. However, a high signal-to-noise ratio (S/N) spectrum is required. As part of *HST* GO program 13467 we have endeavored to produce such high-S/N spectral data to bring clarity to the characterization of exoplanet atmospheres.

In what follows, we describe our WFC3 observations in Section 2, followed by a global dayside atmospheric retrieval analysis Section 3, and finally we discuss implications of our observations and conclude in Section 4.

2. *HST*/WFC3 OBSERVATIONS AND DATA ANALYSIS

2.1. *Observations and Reduction*

As part of a large *HST* Treasury Program (GO 13467), we used the Wide Field Camera 3 (WFC3) to observe HD 209458b during secondary eclipse over five visits from 2014 September to December. WFC3's G141 grism disperses light from 1.12 to 1.65 μm onto its detector, which we operated in 256 × 256 subarray mode.

Each visit comprises five *HST* orbits: the first orbit achieves instrument stability and is excluded from our analysis following standard procedure (Berta et al. 2012; Deming et al. 2013; Kreidberg et al. 2014b), the second and fifth orbits determine the out-of-eclipse baseline, and the third and fourth orbits contain the secondary eclipse. Each *HST* orbit begins with a direct image of HD 209458 for calibration purposes, is followed by 43 spectroscopic frames (each lasting 14.971 s in duration), and concludes with a buffer dump to maximize observing efficiency. During the acquisition of each spectroscopic frame, we utilize the bidirectional spatial scan mode at a rate of $1\farcs15$ s^{-1}, thus spreading the light over ~140 pixel rows perpendicular to the spectral dispersion. The median peak pixel count across all frames and all visits is 40,615e, which is close to the recommended threshold level of ~40,000e for optimal performance (Berta et al. 2012; Swain et al. 2013; Wilkins et al. 2014).

Our data reduction and analysis closely follow the techniques described in our previous WFC3 papers (Stevenson et al. 2014a, 2014b, 2014c). Here we provide a brief description of our methods. For the reduction of spatial scan data, we use the calibrated "_ima" frames provided by the STScI Archive. Using a 2D Gaussian, we compute the centroid of the direct image from the start of each *HST* orbit, calculate the trace of the equivalent stare spectra, and then determine the field-dependent wavelength solution of the scanned spectra (Kuntschner et al. 2009). We apply spectroscopic flat-field corrections using coefficients from the calibration file `WFC3.IR.G141.flat.2.fits`.

Each frame consists of three evenly spaced, nondestructive reads. We compute the difference between pairs of adjacent reads, use optimal extraction (Horne 1986) to produce two 1D spectra, and then combine their values to arrive at a single spectrum per frame. Our extraction window consists of 110 pixel rows centered on the differenced spectra and flanked by an additional 110 pixel rows for background subtraction (220 pixel rows total). This optimized configuration produces the most precise light curves. We divide the spectra into 10 spectrophotometric bins of width ~0.0525 μm and spanning 1.125–1.655 μm.

Our extracted light curves exhibit similar systematics to those noted previously for WFC3 data (e.g., Berta et al. 2012; Deming et al. 2013; Stevenson et al. 2014c). These include a ramp (or hook) at the beginning of each *HST* orbit and a visit-long trend. We follow a two-step process to extract HD 209458b's emission spectrum. First, we model the band-integrated (white) light curves to compute nonanalytic, transit-removed systematic model components and to evaluate the consistency of the measured eclipse depths between visits. Second, we divide the spectroscopic light curves by our nonanalytic model components and then determine the best fit for each channel to derive wavelength-dependent eclipse depths with uncertainties. Below we discuss each of these steps in detail.

2.2. White Light-curve Fits

Our model fitting procedure consists of finding the best solution by simultaneously fitting all free parameters using a Levenberg–Marquardt minimizer and then using a Differential Evolution Markov Chain algorithm (ter Braak & Vrugt 2008) to estimate parameter uncertainties. For our final white light-curve fits, we use 10 chains and terminate after 4 × 10^5 steps when the Gelman & Rubin (1992) diagnostic for all free parameters has dropped to within 1% of unity. For more details on error estimation and convergence, see Stevenson et al. (2012). We find no evidence for correlated noise in the best-fit residuals as determined by plotting the rms versus bin size for each scan direction of each visit. We test various model components and shared parameters and then determine the best combination by comparing their Bayesian Information Criterion (BIC) values (Cornish & Littenberg 2007; Trotta 2008).

Our final white light-curve model takes the form

$$F(\Updownarrow, \phi, \Phi) = F_s(\Updownarrow)E(\phi)R(\phi, \Updownarrow)H(\Phi, \Updownarrow), \tag{1}$$

where $F(\Updownarrow, \phi, \Phi)$ is the modeled flux for scan direction, $\Updownarrow$, at planet orbital phase, ϕ, and *HST* orbital phase, Φ; $F_s(\Updownarrow)$ is the scan-direction-dependent out-of-eclipse system flux; $E(\phi)$ is the eclipse model (Kreidberg 2015); $R(\phi, \Updownarrow) = 1 + r_{1,\Updownarrow}(\phi - 0.5) + r_{2,\Updownarrow}(\phi - 0.5)^2$ is the time-dependent quadratic trend model component; and $H(\Phi, \Updownarrow) = 1 - e^{-h_0\Phi + h_1} + h_{2,\Updownarrow}\Phi$ is the *HST* orbital-phase-dependent rising exponential ramp with linear trend model component. Scan-direction- and visit-dependent free parameters include $F_s(\Updownarrow)$, $r_{1,\Updownarrow}$, $r_{2,\Updownarrow}$, and $h_{2,\Updownarrow}$. In our final fits, the eclipse depth, h_0, and h_1 free parameters have the same values for both scan directions and all visits. We fix the eclipse midpoint ($\phi = 0.5$), total eclipse duration (3.06 hr), and ingress/egress times (0.42 hr; Seager & Mallén-Ornelas 2003) because these parameters are not only poorly constrained due to the long observing gaps during Earth occultation but also precisely known from previous observations of this canonical planet (Zellem et al. 2014b).

Table 1
White Light-curve Eclipse Depths

Visit #	Observation Date	Eclipse Depth (ppm)	χ^2_ν
1	2014 Sep 16	113 ± 36	1.11
2	2014 Oct 18	121 ± 35	0.87
3	2014 Nov 15	50 ± 28	1.43
4	2014 Dec 10	6 ± 35	1.18
5	2014 Dec 31	122 ± 34	1.15
(1, 2, 3, 4, 5)	⋯	82 ± 15	1.14
(1, 2, 3, 5)	⋯	99 ± 17[a]	1.13

Note.
[a] Adopted value.

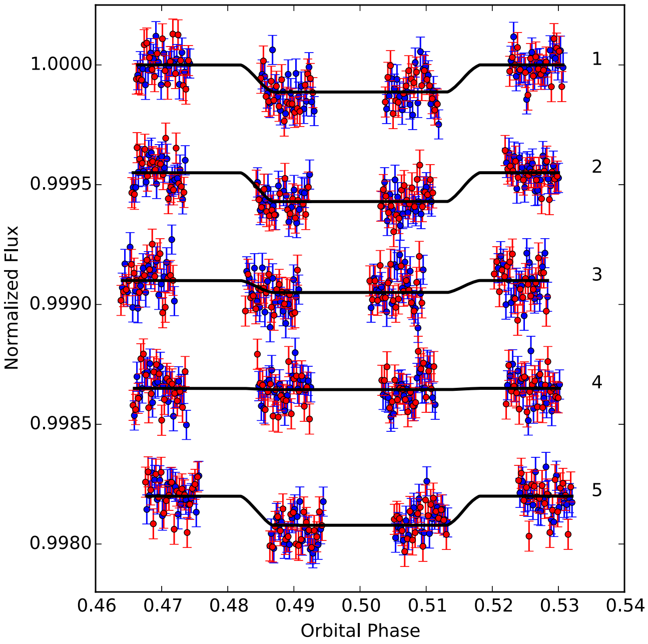

Figure 1. White light curves of HD 209458b. The blue and red data points depict the systematics-removed normalized flux values from the forward and reverse scan directions, respectively. The best-fit models (black lines) highlight the indistinguishable eclipse depth during the fourth visit (labels indicate the visit #).

To test the repeatability of our measurements, we apply our white light-curve model to the data from each visit and report the best-fit eclipse depths and χ^2_ν values in Table 1. Despite having a relatively good fit ($\chi^2_\nu = 1.18$) and contrary to the other visits, the fourth visit contains no discernible secondary eclipse (6 ± 35 ppm; Figure 1). As noted previously (Stevenson et al. 2014c), there is a strong correlation between the eclipse depth and the quadratic term in $R(\phi, \Updownarrow)$. As a test, we include data from the latter half of the first *HST* orbit and derive a deeper eclipse depth (54 ppm) and shallower $r_{2,\Updownarrow}$ values (the other free parameters are unaltered). We conclude that this parameter degeneracy is the most likely source of the fourth visit's eclipse depth inconsistency. As an additional test, we

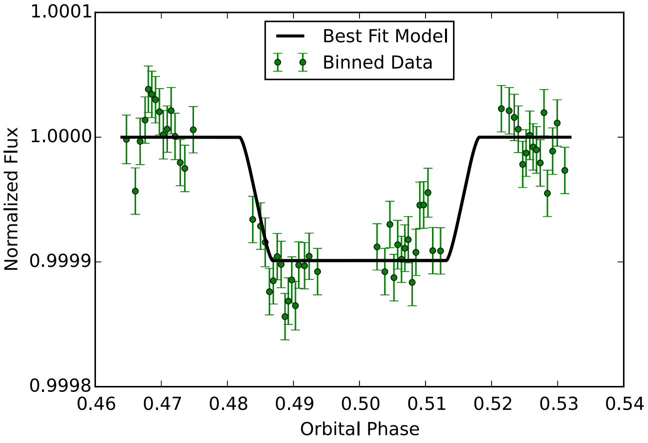

Figure 2. Binned white light curve of HD 209458b. The green points combine data from visits 1, 2, 3, and 5; the best-fit model is from our joint fit.

explored the impact of both linear and quadratic visit-long model components on the Visit 3 eclipse depth. Fitting only visit 3 while excluding the first orbit, we obtain eclipse depths of 62 and 283 ppm for the quadratic and linear models, respectively.[12] Adding the last half of the first orbit increases the eclipse depth from 62 to 119 ppm for the quadratic model, which is consistent with visits 1, 2, and 5, but also increases the reduced chiq-squared value from 1.39 to 1.64. Many WFC3 primary transit analyses emphasize obtaining robust relative transit depths; however, secondary eclipse analyses rely on the absolute depth to determine the planet's temperature. This places additional emphasis on correctly handling the white light-curve systematics.

When we exclude the first *HST* orbit and include all of the visits in a joint fit, the best-fit model favors a shared eclipse depth of 82 ± 15 ppm. Adding data from the latter half of the first *HST* orbit to all visits increases the shared eclipse depth to 97 ± 15 ppm. This value is comparable to a fit that excludes the first *HST* orbit and the fourth visit (99 ± 17 ppm). We choose to adopt this final value as our estimate of the true white light-curve eclipse depth (Table 1). Figure 2 depicts systematics-removed, binned data (excluding visit 4) with a best-fit light-curve model from our joint fit.

2.3. Spectroscopic Light-curve Fits

We apply the `Divide-White` technique (Stevenson et al. 2014a) to remove the wavelength-independent systematics from the spectroscopic light curves. This is accomplished by dividing the spectroscopic light curves from each visit by their corresponding, transit-removed white light curves, which serve as nonanalytic models of the wavelength-independent systematics. Our final spectroscopic light-curve model takes the form

$$F(\lambda, \Updownarrow, \phi) = F_s(\lambda, \Updownarrow)E(\lambda, \phi)R(\lambda, \phi)W(\phi), \tag{2}$$

where $F(\lambda, \Updownarrow, \phi)$ (the wavelength-dependent modeled flux) no longer depends on *HST*'s orbital phase, the definitions for $F_s(\lambda, \Updownarrow)$ and $E(\lambda, \phi)$ are unchanged, $R(\lambda, \phi) = 1 + r_{\lambda,1}(\phi - 0.5)$ is now a time-dependent linear trend with no scan direction dependence on $r_{\lambda,1}$, and $W(\phi)$ is the nonanalytic, transit-removed white light-curve model. The simplicity of this model relative to Equation (1) highlights the effectiveness of using the `Divide-White` technique to analyze *HST*/WFC3 data. We show in Figure 3 the residual rms versus bin size for the spectroscopic light-curve fits, which demonstrates that we have successfully removed time-dependent systematics.

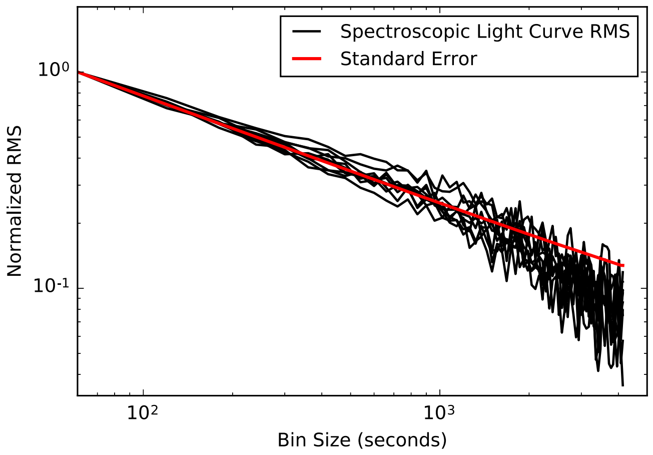

Figure 3. The rms of the residuals to each spectroscopic light-curve fit vs. time-series bin size. The red line is the expectation from Gaussian-distributed residuals. The consistency of the rms trend with the expectation from white noise suggests minimal impact due to time-dependent systematics. If the rms goes below the standard error, then the data are "cleaner" than expected. Typically, this can happen if there is correlated noise on very short timescales or the data have been smoothed, which is not the case here. For these particular observations, we suspect that the ~50 minutes of data followed by ~45-minute gaps per *HST* orbit add additional flexibility in our model fits that enables the rms values to fall slightly below the standard error for bin sizes greater than ~2000 s (30 minutes). In general, the rms values for sets of noncontiguous data points can become unreliable when attempting to bridge the gaps.

Table 2
Spectroscopic Light-curve Eclipse Depths

Wavelength (μm)	Eclipse Depth (ppm)	χ^2_ν
1.125–1.178	43 ± 16	0.98
1.178–1.230	97 ± 15	1.01
1.230–1.282	115 ± 15	1.02
1.282–1.335	109 ± 15	1.13
1.335–1.388	80 ± 15	1.00
1.388–1.440	43 ± 15	0.96
1.440–1.492	86 ± 15	0.84
1.492–1.545	107 ± 17	1.11
1.545–1.598	167 ± 17	1.20
1.598–1.650	170 ± 18	1.12

In Table 2, we provide the spectroscopic eclipse depths from a joint fit encompassing visits 1, 2, 3, and 5. Adding the fourth visit effectively reduces the eclipse depths by an average of ~17 ppm and imparts no additional wavelength modulation. The listed eclipse depth uncertainties are relative values; therefore, a wavelength-independent offset (Table 1) should be applied when fitting atmospheric models that consider additional data. In this case, a Gaussian prior of width 17 ppm (the white light-curve eclipse depth uncertainty) is appropriate. Table 2 also lists mean χ^2_ν values for each channel; the overall mean is 1.04 times the photon plus read-noise limit. In the next section we interpret these measurements within an atmospheric retrieval framework.

---

[12] The first number is slightly different from the value listed in Table 1 (50 ppm) because, for this test, we did not perform a joint fit with all of the visits.

3. ATMOSPHERIC RETRIEVAL ANALYSIS

Our new WFC3 measurements, combined with *Spitzer* broadband photometry[13] (Diamond-Lowe et al. 2014), are interpreted within the context of an atmospheric retrieval framework. In what follows we first describe our nominal model setup (Section 3.1) and results (Section 3.2). We also test the robustness of our nominal model assumptions by exploring a different temperature profile parameterization (Section 3.3), the inclusion of a dayside cloud (Section 3.4), and the impact of different *Spitzer* data reduction methods (Section 3.5). Finally, we explore the chemical plausibility of our retrieved abundances through a new method, *chemical retrieval-on-retrieval* (Section 3.6).

3.1. Atmospheric Parameterization

We use a derivative of the emission portion of the CHIMERA retrieval suite (Line et al. 2013a, 2014; Diamond-Lowe et al. 2014; Kreidberg et al. 2014a; Stevenson et al. 2014c) to interpret the full dayside emission data of HD 209458b. The CHIMERA emission forward model computes the upwelling disk-integrated thermal emission given a temperature structure and molecular abundances and has been validated against other models in two publications (Line et al. 2013a; Morley et al. 2015). This forward model is coupled with the powerful `pymultinest` routine (Buchner et al. 2014), a `python` wrapper to the commonly used `multinest` nested sampling algorithm (Skilling 2006; Feroz et al. 2009), to perform the parameter estimation and Bayesian evidence computation.

The temperature–pressure (TP) profile is parameterized (with five parameters) using the analytic approximation from Parmentier & Guillot (2014; see Line et al. 2013a for implementation details). The parameterization is flexible enough to permit a wide range of thermal structures, including inversions, while also resulting in physically realistic profiles consistent with radiative energy balance as the analytic approximation enforces radiative equilibrium in its derivation (see also Hansen 2008; Guillot 2010; Robinson & Catling 2012; Heng et al. 2014). For this investigation we retrieve seven thermochemically plausible gases that absorb over the wavelengths of interest: H_2O, CH_4, CO, CO_2, NH_3, HCN, and C_2H_2 (e.g., Madhusudhan et al. 2011), with uniform-in-log mixing ratio priors spanning from -12 to 0. Molecular hydrogen and helium (in solar proportions) are assumed to comprise the remaining gas such that all species sum to unity. We use the absorption cross-section database described in Freedman et al. (2014). For this analysis we make no a priori assumption about the realism of our retrieved abundances, but will later show that they are consistent with expectations from thermochemical equilibrium, suggesting that both the data and model reasonably reflect reality. All emission model results presented in this work assume cloud-free atmospheres unless otherwise stated. We also assume a free parameter offset between the WFC3 and *Spitzer* data to account for uncertainties in the absolute calibration in the white light transit depth, though the impact of this parameter is minimal in our retrievals (found to be consistent with no offset, 3 ± 10 ppm).

3.2. Fiducial Results

Figures 4 and 5 summarize our nominal retrieval results under the aforementioned model assumptions, denoted as the *FULL* model scenario. The spread in the models and fits suggests that the model and the data are well matched (median chi-squared per data point of 1.03), much like our previous results on WASP-43b, suggesting that our atmospheric parameterization is adequate; hence, parameter estimates from this data–model combination have meaning. It comes as no surprise that the presence of a deep water vapor absorption feature in the WFC3 bandpass results in a monotonically decreasing TP profile over the planet's infrared photosphere (1 bar–1 mbar). Water is the only robustly constrained molecular absorber, with an abundance constraint of ±1 order of magnitude nearly centered about solar composition (68% confidence interval: 4.0×10^{-5}–5.4×10^{-3}). We can only retrieve upper limits or long unconstrained tails (CO_2) on the other gases. The resulting abundance distributions (Figure 5) appear physically plausible (with the exception of CO_2; more below), or at least they encompass physically plausible molecular combinations and are consistent with or near solar abundance to within 3σ.

In order to further demonstrate the lack of need for a thermal inversion and to identify which molecules are actually "detected" in this spectrum, we perform a standard nested Bayesian model comparison using the full Bayesian evidence calculation from `pymultinest` (Trotta 2008; Benneke & Seager 2013; Swain et al. 2014; Kreidberg et al. 2015; Waldmann et al. 2015a, 2015b). The Bayesian evidence is what the commonly used BIC approximates via a truncated Laplace approximation (Cornish & Littenberg 2007; Trotta 2008). A parameter is deemed "necessary" or "detected" if the improvement in likelihoods from the addition of that given parameter (or parameters) outweighs the increase in prior volume (Chapter 3.5, Gregory 2005). This is a straightforward method for determining whether or not a gas is detected as we simply remove that gas from the model and rerun the retrieval to compute the new evidence. The ratio of the evidences of two models (usually a *FULL* model with all parameters compared with subsets of that model) is the Bayes factor, which can be directly converted into a confidence interval/detection significance (Equation (27), Table 2 in Trotta 2008, reproduced in Benneke & Seager 2013).

Table 3 summarizes the detection significances of the various nested models. We find that water is detected at 6.2σ confidence. This detection, as with our WASP-43b WFC3 observations (Kreidberg et al. 2014b), is strongly driven by the 1.4 μm water feature covered by WFC3. The reduced gases, expected to occur in cooler planets (CH_4, NH_3) or high carbon-to-oxygen ratio atmospheres (HCN, C_2H_2), are not detected, and in fact their inclusion is penalized due to the increase in prior volume without the accompanying improvement in fit quality, consistent with retrieving only upper limits on their abundances. CO+CO_2 are detected at 4.1σ confidence. This detection is largely driven by the 4.5 μm IRAC observation. We cannot decipher the difference between CO and CO_2 as they have nearly overlapping features over the 4.5 μm bandpass; thus, the removeal of one is compensated by the other.

To compute the "detection" (or lack thereof) of a thermal inversion, we perform a separate retrieval in which the prior five-parameter analytic TP profile forces TP profiles that have zero to positive temperature gradients with increasing altitude.

---

[13] While there are numerous sources of data sets available (*HST* NICMOS, *Spitzer* IRS, MPIS, NASA IRTF, etc.), we chose to keep it simple by combining a reasonably new reduction of mid-infrared space-based *Spitzer* data with our new WFC3 observations. A more comprehensive analysis exploring the impact of the many possible data sets and/or combinations thereof is beyond the scope of this paper.

r

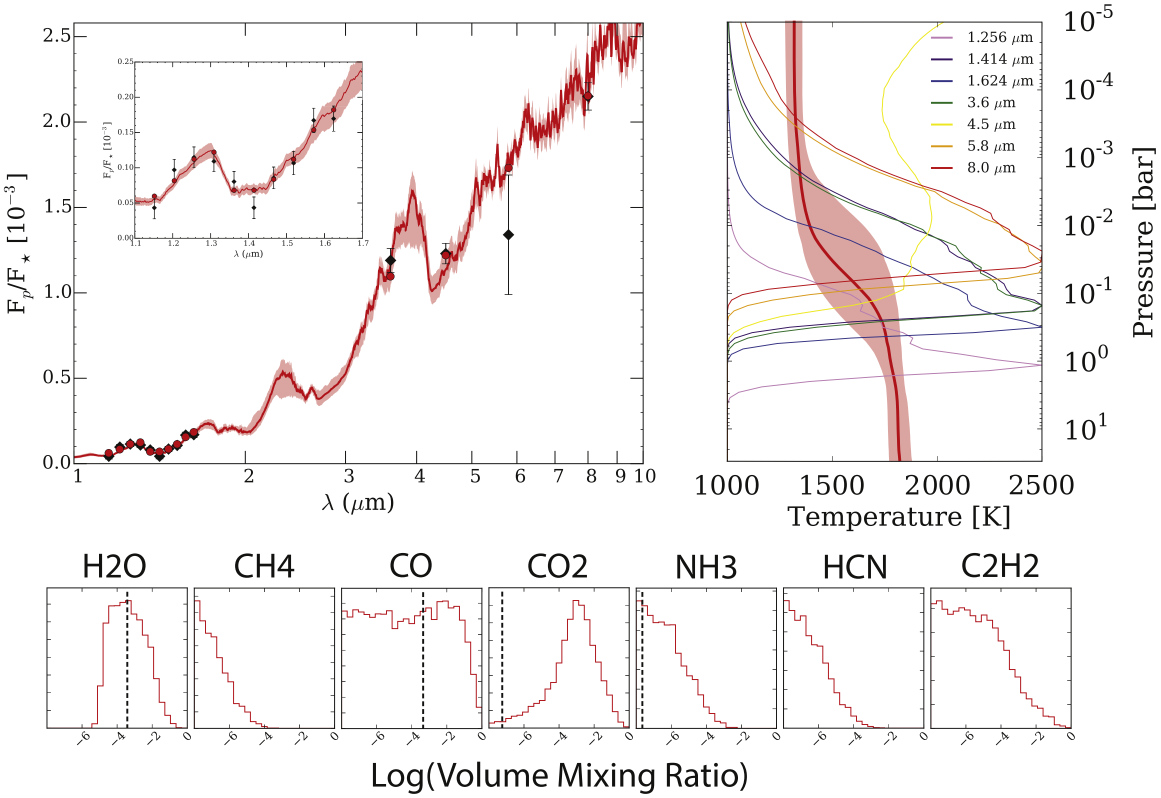

Figure 4. Nominal spectral fits, temperature profile, and molecular abundance retrieval results from the "FULL" model (all molecules and possible TP profiles permitted). Left: model fits to both the WFC3 and *Spitzer* data. The data (WFC3, this work, and *Spitzer* from Diamond-Lowe et al. 2014) are shown as black diamonds with error bars. The light-red shaded regions are the 1σ spread in the spectra resulting from 1000 draws from the posterior, and the solid red curve is the median of those 1000 spectra. The red points are the medians of the model spectra integrated over the photometric bandpasses (median χ^2 per data point of 1.03). A zoom-in over the WFC3 region is shown in the inset. Right: range of allowed TP profiles that best explain the data. The light-red shaded regions are the 1σ spread in the TP profiles resulting from 1000 draws from the posterior, and the solid red curve is the median of those 1000 TP profiles. This is a family of solutions for which all of the TP profiles are monotonically decreasing (no evidence for inversion). Normalized thermal emission contribution functions (Chapman functions) are shown for select wavelengths. Note how the contribution functions are concentrated between ∼1 bar and 1 mbar. The TP profile outside of this range is largely driven by the parameterization and should not be taken as absolute truth. Bottom: marginalized posterior distributions for each of the gases included in the nominal model. The vertical dashed line in each panel is the predicted thermochemical equilibrium mixing ratio at 1700 K, 0.1 bars under solar elemental abundances. The solar composition values for CH_4, C_2H_2, and HCN are less than the plot range lower limit. We find strong water vapor absorption in the WFC3 spectrum, with an abundance consistent with solar composition (within ±1 dex), and a monotonically decreasing temperature profile between 1 bar and 1 mbar.

In other words, we restrict the prior search space to exclude monotonically decreasing profiles. We have not removed or added parameters; we simply reduced the prior volume by restricting the prior ranges. Upon comparing the Bayesian evidence of the two models, we find that a monotonically decreasing TP profile is required at the 7.7σ level.

In summary, we have robustly detected absorption due to water at or near what is expected from solar elemental abundances (±1 dex), a detection of some combination of CO and CO_2, and strong evidence for a monotonically decreasing TP profile over the pressure levels robustly[14] probed by the observations (1 bar–1 mbar). In the following sections we will explore how various model and data assumptions will impact these conclusions.

3.3. *Impact of TP Profile Parameterization*

Given that one of the controversies surrounding the dayside spectrum of HD 209458b is whether or not it possesses a stratospheric thermal inversion, it is thus prudent to explore how our thermal profile parameterization could influence our conclusion. In the nominal model, described above, we used an analytic five-parameter gray radiative equilibrium description for the thermal profile. Here we use a much simpler profile with far fewer assumptions (similar to Schwarz et al. 2015). The atmosphere is divided up into three regions: a deep isothermal region defined by a deep temperature parameter (similar to Madhusudhan & Seager 2009) and two "T linear-in-log(P)" regions defined by a slope and two log-pressure free parameters defining the break between the three regions, for a total of five parameters. Continuity between the regions is enforced. This simple parameterization does not abide by any physical rules (e.g., radiative equilibrium) and is strictly guided by the data and parameter priors. All other aforementioned model assumptions remain the same.

Figure 6 compares the results from this alternate TP profile parameterization with our nominal results from Section 3.2. In general, the conclusions are unchanged: water is observed to be at or near solar composition, upper limits obtained for CH_4, NH_3, HCN, and C_2H_2, and there is similar temperature gradient over the photosphere (1 bar–1 mbar). In fact, upon inspecting the spectra in Figure 6, we find virtually no difference between the fiducial model fits and the simple TP profile fits. There are a

---

[14] By robustly, we mean where most of the integrated area of the weighting functions lies. There is some very weak contribution at lower pressure levels in the broad 4.5 μm IRAC bandpass.

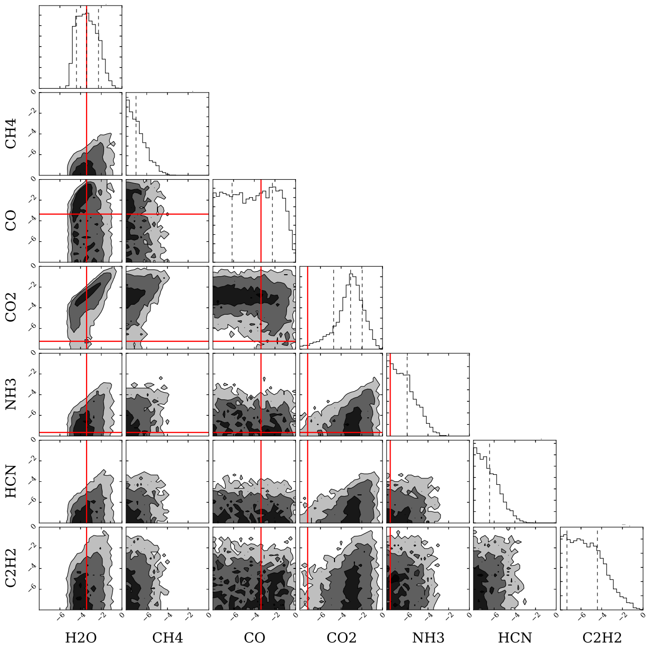

Figure 5. Nominal gas retrieval results from the "FULL" model showing the correlations among each gas. The black, dark-gray, and light-gray regions in the 2D histograms correspond to the 1σ (39.3%), 2σ (86.5%), and 3σ (98.9%) confidence intervals, respectively. The dashed vertical lines in the histograms are the marginalized 16th, 50th, and 84th percentiles. The vertical and horizontal red lines in each panel are the solar composition molecular abundances at 1700 K and 0.1 bars, a representative photospheric temperature (see Figure 4). The water mixing ratio is constrained to ±1 order of magnitude. Note the "elbow"-shaped correlation between CO and CO_2. This is because CO and CO_2 both absorb over the 4.5 μm *Spitzer* bandpass and are thus highly degenerate.

Table 3
Nested Bayesian Model Comparison Results

Scenario (Included Gases)	$\ln B$	Det. Sig.(σ)
FULL (all gases)	⋯	⋯
Molecules other than water (H_2O)	3.96	3.3
CO+CO_2 (H_2O, CH_4, NH_3, HCN, C_2H_2)	6.90	4.1
CH_4+NH_3 (H_2O, CO, CO_2, HCN, C_2H_2)	−0.92	undefined
HCN+C_2H_2 (H_2O, CH_4, CO, CO_2, NH_3)	−0.71	undefined
H_2O (CH_4, CO, CO_2, NH_3, HCN, C_2H_2)	17.10	6.2
Monotonically decreasing T (all gases)	27.10	7.7

Note. The scenarios indicate which molecules/setup we are detecting. Water, CO+CO_2, and a decreasing temperature profile are detected at high confidence. The full compliment of gases includes H_2O, CH_4, CO, CO_2, NH_3, HCN, and C_2H_2.

few subtle differences, however: CO is now favored over CO_2 (though both are still largely unconstrained), the decreasing TP profile extends to slightly lower pressures and is better constrained, and the water abundance constraint is 30% smaller (±0.7 orders of magnitude). The change in the CO/CO_2 abundance distributions is likely due to the slightly steeper temperature gradient over the 4.5 μm bandpass resulting from the simple TP profile parameterization, though solar composition abundances are captured well within their 68% confidence interval. Again, any inference about these molecules is entirely driven by a single *Spitzer* point and is thus highly sensitive to any model assumptions. This is because the spectral slopes over the 4.5 μm bandpass are quite steep. Changing the abundance of both increases the depth of the "V"-shaped absorption feature, but each has a different effect on the spectral slopes. CO more strongly affects the blue edge of the bandpass, whereas CO_2 affects the red edge. The temperature gradient affects the slope at both edges. The integrated flux is dependent on the depth and both the red and blue slopes, making for a complicated degeneracy. Any resulting conclusions about the CO/CO_2 abundances are therefore not robust. There also appears to be an "inversion" above the 1 mbar pressure level. This turnover in temperature is largely outside the photosphere—where the data are probing—and is largely driven by the prior on the temperature gradient in this region of this particular parameterization. Care must be taken when interpreting temperature profile results outside of the photosphere. For instance, our current data would not be able to rule out the presence of a *thermosphere* inversion, which occurs at a few microbars and lower pressures, driven by upper atmospheric energetics (e.g., Yelle 2004).

3.4. Impact of Dayside Clouds

The impact of clouds on dayside emission spectra has largely been ignored, given that there has not yet been any observational evidence to justify otherwise. This is surprising since it is widely accepted (both observationally and theoretically) that clouds strongly impact the spectra of similar-temperature brown dwarfs (e.g., Burrows 2001; Kirkpatrick 2005; Marley & Robinson 2015), along with the mounting evidence for clouds in transmission spectra and in reflected light curves from *Kepler* (e.g., Demory et al. 2013; Esteves et al. 2015; Shporer & Hu 2015). While clouds can dominate transmission spectra due to the enhanced path length through the atmosphere (Fortney 2005), dayside inhomogeneities (or patchiness) driven by circulation (Parmentier et al. 2016) could weaken the effect of clouds if they are present. The appearance of the strong water absorption feature over the WFC3 bandpass suggests that the impact of clouds on the dayside emission spectrum of HD 209458b is minimal. However, to be thorough, we test this hypothesis by introducing a simple, gray, nonscattering cloud in emission with a mass opacity, κ_c, given by

$$\kappa_c(P) = \kappa_{P_0}\left(\frac{P}{P_0}\right)^{\alpha}, \tag{3}$$

where P_0 is the cloud base pressure, κ_{P_0} is the specific absorption coefficient (area/mass) at the base of the cloud, and α is the cloud fill factor index (Burrows et al. 2006). Figure 7 summarizes the results. As anticipated, we find that the cloud has negligible impact on the retrieved abundances and thermal structure, though the retrieved cloudy TP profile is ever so slightly shifted "upward" in pressure and has greater uncertainty in the deep atmosphere. The Bayes factor between the cloudy model and clear model (0.15—extremely insignificant) suggests that a cloud is not required or justified. The histogram inset in Figure 7 shows the integrated cloud optical depth[15] over the pressure levels probed by the WFC3 water band for 1000 randomly drawn samples. Most of the optical

---

[15] The photosphere for each sample occurs at a different pressure level. Therefore, to compute the integrated cloud optical depth, we convolved the cloud optical depth *profiles* with the normalized 1.256 (continuum) and 1.414 (water band) μm contribution functions. This is a rough approximation to the integrated optical depth over the pressure levels probed by the WFC3 observations.

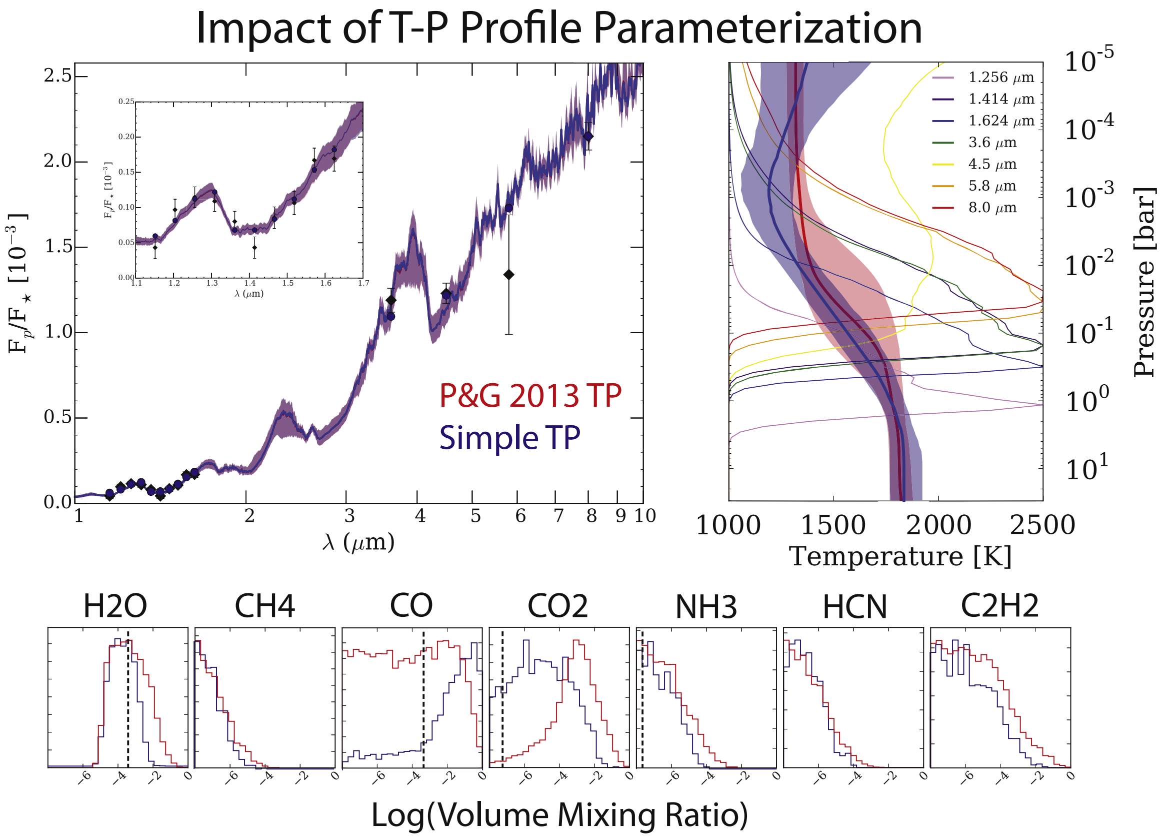

Figure 6. Impact of the temperature profile parameterization on the retrieval results. Plot structure similar to Figure 4. In all panels, red corresponds to the nominal Parmentier & Guillot (2014) analytic radiative equilibrium parameterization described in Section 3.1, and blue to the "simple" TP parameterization described in Section 3.3. The spectra resulting from each TP profile parameterization are nearly indistinguishable, but there are noticeable differences in the retrieved TP profiles and molecular abundances. The temperature gradients over the photosphere (~1–0.001 bars) are largely consistent, though the uncertainties are smaller within the simple TP parameterization. The turnover in the TP profile above the 1 mbar level (outside most of the photosphere) in the "simple TP parameterization" is largely driven by the prior on the slope in that region. Finally, the abundances are largely consistent with the exception of CO and CO_2. See Section 3.3 for details.

depths are below unity, suggesting small to negligible cloud opacities, again consistent with our Bayes factor results. Though no impact/evidence for dayside clouds is found here, this may not always be the case; thus, a more thorough examination of the emission spectra of multiple planets is necessary to determine the impact of dayside clouds.

3.5. Impact of *Spitzer* Data Reduction Technique

The robustness of published *Spitzer* results has been called into question on multiple occasions (Hansen et al. 2014 [and comparison of references therein]; Diamond-Lowe et al. 2014; Lanotte et al. 2014), and photometric flux values for many planets have changed by several sigma. As such, we feel that it is necessary to explore how the differences in published dayside photometry impact our results. As an illustrative example, we compare the retrieved quantities when using the *Spitzer* photometry presented in Diamond-Lowe et al. (2014) (and what we use as our nominal) with those resulting from the photometry published in Evans et al. (2015) (Figure 8). The measured eclipse depths from Diamond-Lowe et al. (2014) and Evans et al. (2015) are consistent (to within 1σ), but the latter report larger uncertainties in all channels. This is typical of Gaussian processes (GP, used in Evans et al. 2015) relative to other methods, as determined in a recent comparison of *Spitzer* data reduction techniques (Ingalls et al. 2016). GP also ranked lower in repeatability and reliability compared to BLISS mapping (Stevenson et al. 2012), which was the systematic correction technique used by Diamond-Lowe et al. (2014).[16] For these reasons, we adopt the eclipse depths reported by Diamond-Lowe et al. (2014) to derive our nominal results.

The subtle differences in the two reductions are reflected in the fits, as the amplitude of the CO/CO_2 absorption feature at 4.5 μm is shallower with the Evans et al. (2015) data. These differences in general have a small effect on the retrieved quantities. The largest noticeable difference, however, is in the shape of the CO_2 histogram, which has "filled" out more at low abundances. The detailed shape of the water histogram changed slightly, with the median shifting by 5% and the 68% confidence interval widening by 20% (in $\log(H_2O)$). This increased water uncertainty is due to the increased temperature uncertainty in the TP profile around ~100 mbar, where the in-water WFC3 contribution functions partially overlap with the 4.5 μm *Spitzer* contribution function. We also experimented with the full-phase-derived 4.5 μm eclipse depth from Zellem et al. (2014b). The eclipse depth falls between the Diamond-Lowe et al. (2014) and Evans et al. (2015) measurements and thus results in retrieved values (not shown) that fall in between those derived from the Diamond-Lowe et al. (2014) and Evans et al. (2015) data. This exercise illustrates the complex sensitivity of the retrieved quantities to the observations, especially when the constraint on a given species (e.g., CO, CO_2) is entirely dependent on a single

---

[16] However, the GP analysis of Evans et al. (2016) was able to produce consistent results for all epochs, whereas the Diamond-Lowe et al. (2014) paper discarded the 2005 and 2010b 4.5 μm results.

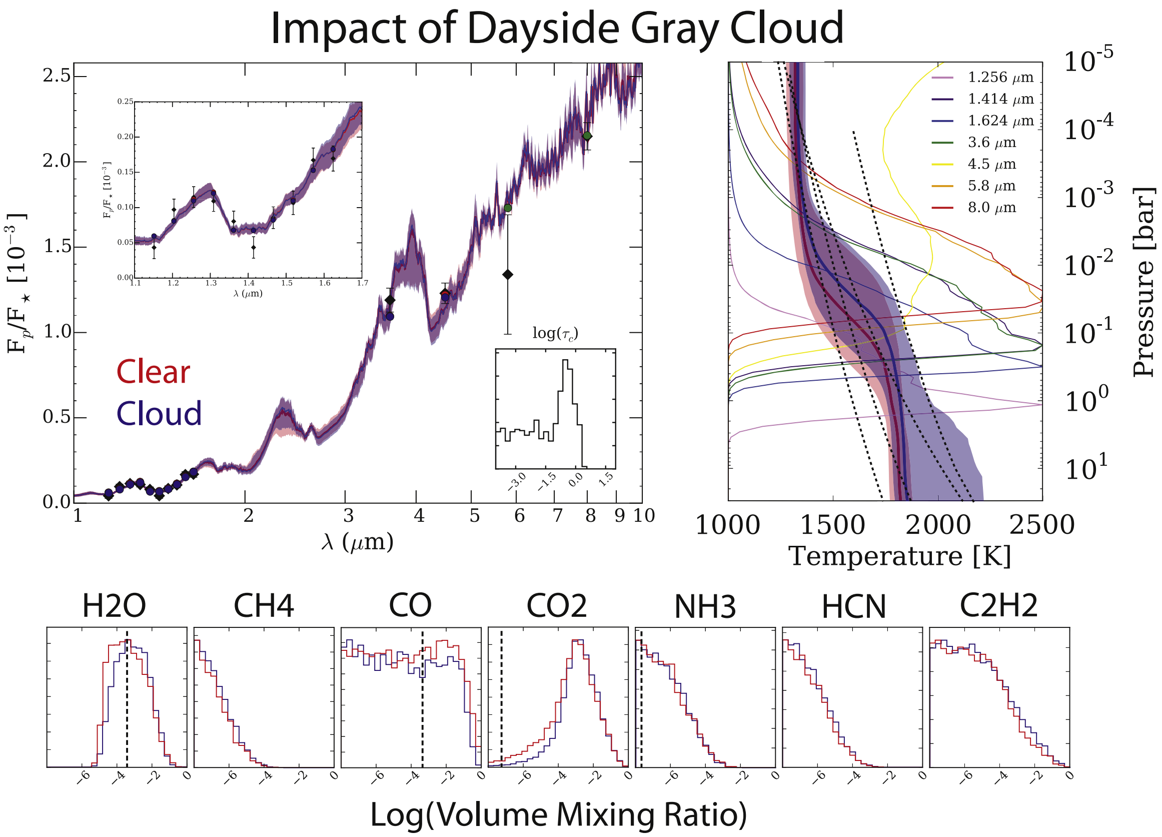

Figure 7. Impact of a dayside cloud. Plot structure similar to Figure 4. In all panels, red corresponds to the nominal cloud free model (clear) and blue to the model that includes a cloud (cloud). The log(τ_c) inset shows the integrated cloud optical depth over the pressure levels probed by the water band in the WFC3 data. The integrated optical depths typically fall below unity, indicative of a fairly transparent atmosphere. Condensate curves for $MgSiO_3$, $MgSiO_4$, Fe, and TiO (from cold to hot) are shown as the dashed lines. While the permitted range of TP profiles crosses many condensation curves, we find, to high confidence, no evidence for a dayside cloud within the photospheric region. See Section 3.4 for details.

photometric datum. Nevertheless, the main conclusions of this paper (water abundance consistent with solar and no thermal inversion) are unchanged no matter which *Spitzer* data are used.

3.6. Elemental Abundance Analysis: Chemical Retrieval-on-retrieval

In order to more rigorously determine the derived *elemental* abundances from the molecular abundances (parameterized through the metallicity, [M/H], and C/O), we introduce a new analysis method, *chemical retrieval-on-retrieval*. The chemical retrieval-on-retrieval approach straddles the line between the classic "fully data driven" (Madhusudhan & Seager 2009; Barstow et al. 2013; Lee et al. 2012; Line et al. 2012, 2014; Benneke & Seager 2013) and "a priori imposition of physical processes" (Fraine et al. 2014; Knutson et al. 2014; Kreidberg et al. 2015; Benneke 2015). This allows us both to obtain the full range of possible molecular combinations that can explain the data and to address a posteriori what regions of that parameter space are plausible. We note that this is the approach taken within the solar system community: molecular abundances (or their profiles) are directly retrieved from the data (whether in situ or remotely) and then compared to chemical models (e.g., Greathouse et al. 2005; Moses et al. 2005; Visscher et al. 2010; Orton et al. 2014).

The *chemical retrieval-on-retrieval* is a two-step process (as its name implies). First, we retrieve the molecular abundances and TP profile directly from the spectroscopic/photometric data, making no assumptions about the relationship between the molecular abundances (Section 3.2). We then treat the retrieved *marginalized* mixing ratio posterior distributions (from Section 3.2) as independent data points to which we fit a thermochemical equilibrium model (again using `pymultinest`). From a statistical perspective, this is very much how one treats wavelength-dependent eclipse depths as independent data points when performing a classic atmospheric retrieval. The thermochemical model (NASA CEA2; Gordon & McBride 1994; Line et al. 2011; Moses et al. 2011; Kreidberg et al. 2015; Mollière et al. 2015) computes the molecular mixing ratios along a given pressure–temperature profile given [M/H] and C/O, which are the retrieved parameters.

We consider all species (~2000) in the CEA2 thermodynamic library that contain H, He, C, N, O, S, P, Fe, Ti, V, Na, and K and assume pure equilibrium (no rainout). Rather than assume Gaussian distributions for the mixing ratio data (e.g., as one does when using chi-squared), we evaluate the probability density of each thermochemically computed mixing ratio with the normalized histograms themselves, through interpolation. When computing the net log-likelihood for multiple gases, we simply sum the log of their probability densities together.[17] We propagate the uncertainty in the retrieved TP profile from the atmospheric retrieval into the retrieved metallicity and C/O by drawing a random TP profile from the TP posterior at each `pymultinest` likelihood evaluation. This ignores the

---

[17] This process is the equivalent of computing chi-square for normal distributions.

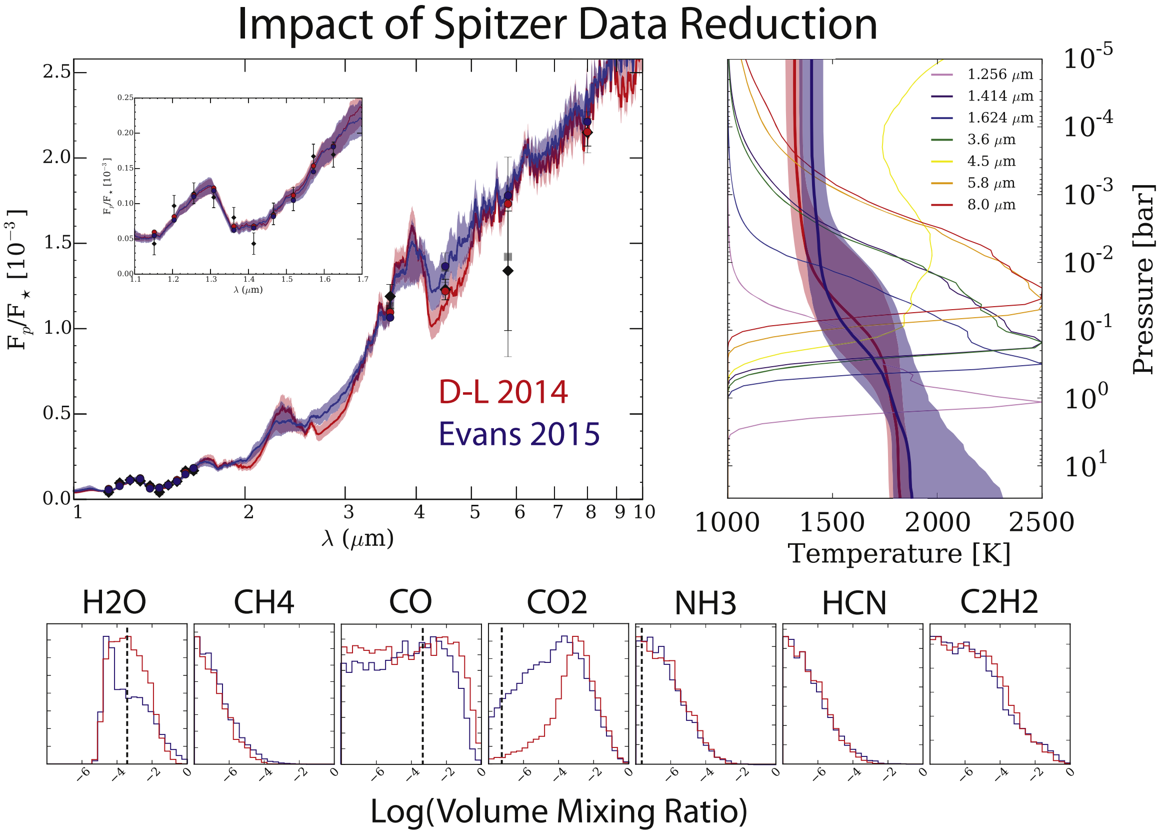

Figure 8. Impact of choice of *Spitzer* photometry on the retrieval results. Plot structure similar to Figure 4. In all panels, red corresponds to the nominal model using the Diamond-Lowe et al. (2014) data (thick black diamonds; Section 3.2) and blue corresponds to the results derived when using the Evans et al. (2015) data (gray squares). Note the shallower "V"-shaped feature near 4.5 μm resulting from the increased $F_p/F_{\rm star}$ in the Evans et al. (2015) data (blue). The Evans et al. (2015) data result in somewhat larger temperature uncertainty in the deeper atmosphere and a slightly warmer upper atmosphere. Overall the molecular abundances are consistent, with the exception of a widening of the CO_2 histogram, owing to the larger 4.5 μm uncertainty on the Evans et al. (2015) datum, and slightly deeper eclipse depth.

correlations between the *elemental* abundances and the TP profile; however, this is acceptable as the marginalized molecular abundances contain that information. We integrate out vertical profile information by column-averaging the thermochemically derived mixing ratio profiles over the infrared photosphere (3 bars–1 mbar), as low-resolution spectra are largely insensitive to vertical profile information (e.g., Lee et al. 2012; Barstow et al. 2013).

Within this framework, we investigate how the constraints on each molecular species contribute to our knowledge of the metallicity and C/O. We first ask how well we can constrain the metallicity and C/O from the retrieved water abundance alone, followed by water combined with methane, and finally water combined with methane and carbon dioxide. Figure 9 summarizes the results. The left panel shows the metallicity and C/O constraints obtained for each of the three scenarios. The right panel shows the mixing ratios derived from the retrieved metallicity and C/O under thermochemical equilibrium (colored histograms), within the photosphere, compared with the mixing ratios retrieved directly from the data (thick black histograms).

When using only water to constrain the elemental abundances, we find a metallicity range of 0.06–9.8 × solar.[18] There still remains, at 3σ, a possibility of higher metallicities. We also find, unsurprisingly, that the water histogram derived from the chemical retrieval is nearly identical to that retrieved from the data, that is, the chemical model is providing a "good fit" to the mixing ratio "data." Because there are no carbon species fit for in this scenario, there is a small trickle of probability beyond C/O = 1. In the predicted mixing ratio distributions this low-probability high C/O results in a low-probability tail of appreciable amounts of HCN and C_2H_2. However, when we include the upper limit derived from the methane mixing ratio, the high C/O tail is eliminated, also eliminating the HCN and C_2H_2 possibilities. Water and methane together put a firm upper limit on the C/O ratio, but offer little to constrain any lower limit since there is no lower limit to the methane mixing ratio retrieved from the spectrum. The retrieved water and methane distributions are consistent with the predicted distributions when fit together, suggesting that the retrieved water and methane mixing ratios are physically plausible.

The predicted CO mixing ratios from the water and water +methane scenarios are well within the spectrally retrieved CO upper limit. However, we do find that that most of the chemically predicted CO_2 probability is significantly below the retrieved CO_2 distribution. This suggests that the chemically retrieved metallicity and C/O, while self-consistently able to explain all the other spectrally retrieved molecular species, are unable to adequately explain the CO_2 abundance. Finally, when including the retrieved CO_2 mixing ratio, with methane and water, we find that a higher metallicity (1.3–123 × solar) is

---

[18] This is a more accurate reflection of the true atmospheric metallicity than simply comparing the retrieved water abundance to what the water abundance should be at solar composition in thermochemical equilibrium, as this method takes into account the degeneracy between the C/O and metallicity on the water abundance. Lower C/O values require a lower metallicity to give the same water abundance, and vice versa.

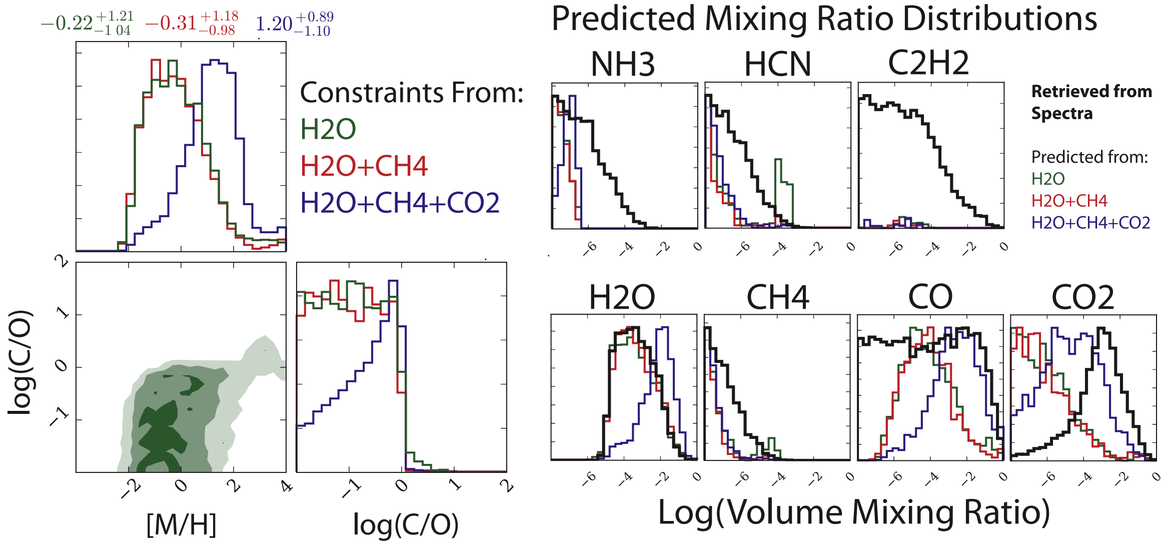

Figure 9. Results of the chemical retrieval-on-retrieval method used to derive [M/H] and C/O. [M/H] and log(C/O) are determined using three different combinations of the spectrally retrieved molecular abundances: H_2O only (green), H_2O+CH_4 (red), and $H_2O+CH_4+CO_2$ (blue). Left: pairs plot illustrating the full [M/H]–log(C/O) posterior for the three scenarios. The green 2D histogram shows the correlation between [M/H] and log(C/O) when using only the spectrally retrieved H_2O histogram. Right: predicted mixing ratio distributions based on the chemically retrieved [M/H] and log(C/O) when using H_2O only (green), H_2O+CH_4 (red), and $H_2O+CH_4+CO_2$ (blue) as the data. The thick black histograms are the spectrally retrieved mixing ratio distributions (Figure 4). These results indicate that all spectrally retrieved species, except CO_2, are consistent with chemical equilibrium near solar elemental abundances. The spectrally retrieved CO_2 abundances require somewhat higher metallicities, but result in somewhat higher water abundances relative to what is spectrally retrieved.

required, which is unsurprising since CO_2 is strongly sensitive to metallicity. The requirement to fit the relatively high CO_2 abundance via high metallicity also pulls the chemically predicted water abundance to higher values, to the point of not being a particularly good fit to the spectrally retrieved water abundance distribution. In short, the spectrally retrieved CO_2 abundance and water abundances are not particularly consistent with each other.

We note that we obtained a similarly high CO_2 abundance for our WASP-43b dayside results (Kreidberg et al. 2014a) when using the 3.6 and 4.5 μm *Spitzer* data. Obviously it is unphysical for CO_2 abundances to be larger than the CO abundance for solar metallicities in thermochemical equilibrium (Prinn & Barshay 1977; Heng & Lyons 2016). Perhaps the unusually high CO_2 abundance is because we are missing some chemical process that enhances CO_2 beyond thermochemical equilibrium. This is unlikely, however, as the most likely processes to do so are chemical quenching via vertical mixing, or photochemistry. Both of these processes are unlikely to significantly enhance the CO_2 abundance to this level in an atmosphere of these temperatures (Cooper & Showman et al. 2006; Moses et al. 2011, Figure 3). Dynamical-chemical effects are also unlikely to significantly impact the abundances (Cooper & Showman et al. 2006; Agundez et al. 2014). We cannot rule out that there could be some additional unknown chemical mechanism that we cannot think of. More likely, however, is that the retrieved CO_2 abundance is not robust. This is supported by the high degree of sensitivity of the CO_2 posterior to our choice of TP profile parameterization or *Spitzer* data, something to consider for future analyses. We note in an independent analysis (Feng et al. 2016, submitted) with synthetic data of a similar observational setup (WFC3+*Spitzer* IRAC 3.6 and 4.5 μm channels) that the retrieved CO_2 abundance is highly sensitive to the particular random noise draw of the 4.5 μm point. Again, this suggests that spurious abundances derived from a *single* broadband measurement should be taken with a grain of salt. The nonlinearities of infrared radiative transfer, combined with ultralow spectral resolution, conspire to produce unphysical solutions. We are thus more inclined to believe the chemically retrieved elemental abundances derived via the other molecular species, which are less impacted by our assumptions.

To summarize, we find that the chemical retrieval-on-retrieval method produces self-consistent chemical abundances, in agreement with most of our spectrally retrieved abundances at 0.06–9.8 × solar and $C/O < 1$. The method is not able to self-consistently reproduce the spectrally retrieved CO_2. However, we question the robustness of the spectrally retrieved CO_2 abundances, which are largely derived from a single broadband measurement and are therefore highly sensitive to our model assumptions.

4. DISCUSSION AND CONCLUSIONS

We present the first high-precision *HST* WFC3 measurements of the dayside emission of the canonical hot Jupiter HD 209458b. These measurements achieve an ultrahigh, photon-noise-limited precision of 15 ppm, providing one of the most robust detections of water (6.2σ) in a dayside emission spectra. The deep water feature, combined with four broadband *Spitzer* observations, suggests a monotonically decreasing temperature profile at 7.7σ confidence, firmly ruling out the presence of a thermal inversion between 1 bar and 1 mbar. We tested the robustness of our atmospheric inferences by exploring multiple model assumptions, such as the impact of the temperature profile parameterization, influence of a gray dayside cloud, and impact of the source of *Spitzer* photometry. We found that the decreasing temperature profile and water abundance were largely invariant to these assumptions, but that the retrieved CO and CO_2 abundances were strongly assumption dependent, owing to their dependence on a single broadband *Spitzer* measurement (4.5 μm).

In Figure 10 we compare the two versions of the retrieved TP profile to two profiles from a self-consistent radiative

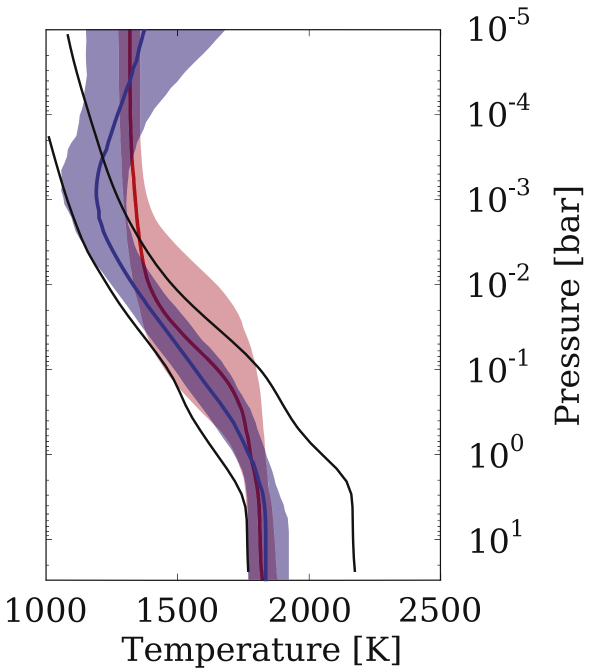

Figure 10. Comparison of retrieved temperature profiles (under both temperature profile parameterizations; red and blue) with self-consistent radiative-convective-thermochemical equilibrium models (black curves; Fortney et al. 2008). The cooler self-consistent TP profile corresponds to full redistribution, whereas the hotter profile corresponds to dayside-only reradiation. The retrieved profiles fall between the two, suggesting a redistribution that falls in between, broadly consistent with previous phase-curve observations (Zellem et al. 2014b). Overall, the agreement is fairly good in terms of the temperature gradient over the regions probed by the observations (1 bar–1 mbar).

equilibrium model (Fortney 2005; Fortney et al. 2008). The hotter black profile assumes only redistribution of energy over the dayside. The cooler profile is a planet-wide average. These self-consistent models include equilibrium chemistry and solar abundances and yield Bond albedos of 0.08. The retrievals yield a profile intermediate between the radiative equilibrium cases, suggesting either a larger Bond albedo for the planet or modest energy redistribution to the nightside. The very low geometric albedo from *MOST* for this planet (Rowe et al. 2008) suggests a low Bond albedo, which favors the explanation that energy is being advected to the nightside, as seen in the 4.5 μm light curve from Zellem et al. (2014b).

This relatively “cool” dayside profile is in contrast to our findings for WASP-43b from Stevenson et al. (2014c, Figure S5). For that planet we found a profile consistent with little energy loss to the nightside, which was in agreement with the low nightside flux seen in the full-orbit phase curve. The two planets have nearly equal equilibrium temperatures, but other significant differences: a 4.3 times longer rotation period for HD 209458b (assuming tidal locking), and at 4.8 times larger surface gravity for WASP-43b. These differences could be important in understanding the energy balance and redistribution of the two planets.

As discussed in Section 1, the claimed detection of a thermal inversion relatively deep in HD 209458b’s infrared photosphere launched an active area of research to look for thermal inversions in other planets, search for the short-wave absorber responsible for causing them, and elucidate the phenomenon theoretically. We had previously suggested that the thermal inversion hypothesis for this prototype was not correct based on a reevaluation of the *Spitzer* photometry that the original inference was based on. Now we show definitively that HD 209458b does not have a thermal inversion in its infrared photosphere using high-precision spectroscopy. The existence of similar thermal inversions in other conventional hot Jupiters (i.e., those with similar temperatures to HD 209458b, say, 800 K $< T_{\rm eq} <$ 2000 K), which all have lower-quality data, should now be questioned. Parmentier et al. (2016) suggest that thermal inversions should not exist at $T <$ 1900 K due to a deep $CaTiO_3$ cold trap, removing titanium from the atmosphere. The recent possible detection of a thermal inversion for WASP-33b (Haynes et al. 2015) is not inconsistent with this statement, as this planet exhibits an average dayside brightness temperature of 3000 K, which means that it is in a very different physical and chemical regime than more typical hot Jupiters like HD 209458b.

In retrospect, the thermal inversions proposed for hot Jupiters were always difficult to understand. Perhaps the most serious issue for the thermal inversion hypothesis is that the *Spitzer* photometric bands typically used to discriminate for inversions are expected to probe similar pressure levels for solar-like composition atmospheres in chemical equilibrium (Burrows et al. 2007, 2008; Showman et al. 2009). Therefore, the altitude and absorption strength of the short-wave observer had to be carefully tuned to give the large temperature difference over a small region of the atmosphere that was needed to match the data. The original leading candidate chemical species (TiO and VO) are now thought to not be present in the dayside atmospheres of conventional hot Jupiters because of the strong vertical mixing to keep these heavy molecules aloft at low pressures (Spiegel et al. 2009) and/or the rainout of larger condensate droplets on the cooler nightside (Parmentier et al. 2013). Furthermore, forced inclusion of these species in general circulation models (GCMs) could not reproduce the data anyway (Showman et al. 2009).

In addition to studying the thermal structure of HD 209458b, we also introduced a new elemental abundance analysis, chemical retrieval-on-retrieval, where we fit a posteriori the retrieved molecular abundances with a thermochemical model to determine the atmospheric metallicity and carbon-to-oxygen ratio. From this analysis we find that the measured water abundance from our spectrum of HD 209458b is consistent with 0.06–9.8 × the solar metallicity, and we firmly rule out C/O values greater than 1. The metallicity derived from the water abundance measurement (from both the chemical retrieval-on-retrieval method and comparison of the retrieved water abundance to solar) is consistent with the trend of increasing atmospheric metallicity for decreasing planet mass that is observed in our solar system (see Figure 11). This result extends the agreement that is seen between the solar system trend and exoplanet atmosphere abundances, which was first investigated by Kreidberg et al. (2014a). The C/O upper limit of 1 buttresses the similarly strict limits that were recently obtained for HD 209458b and other hot Jupiters from the interpretation of transit transmission spectra while accounting for clouds (Benneke 2015; Kreidberg et al. 2015).

Our solar water abundance determination for HD 209458b is inconsistent with the strongly subsolar (0.007–0.05 × solar) abundance interpretation of the planet’s WFC3 transmission spectrum by Madhusudhan et al. (2014a). However, Madhusudhan et al. (2014a) did not account for clouds in their retrieval. We note that the original Deming et al. (2013)

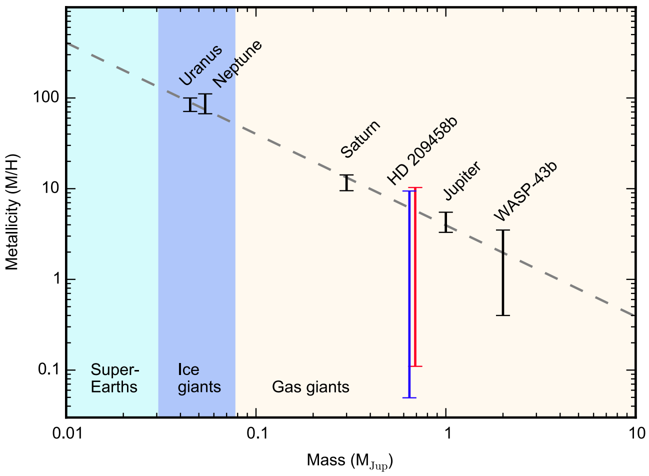

Figure 11. Established mass vs. atmospheric metallicity trend seen in the solar system. We show our previous combined emission+transmisson water-derived metallicity result from WASP-43b, along with our derived abundances in this work. For HD 209458b we show two measurements: the red line is the water-derived metallicity assuming solar carbon-to-oxygen ratios, whereas the blue line is the metallicity derived from the chemical retrieval-on-retrieval method taking into account the metallicity-C/O degeneracy, based on the retrieved water abundance. These measurements are consistent with the trend.

analysis suggested that a cloud is likely what is muting the water feature amplitude. Benneke (2015) and our own unpublished analysis have shown that neglecting clouds in the interpretation of the planet's WFC3 transmission spectrum yields spuriously low and precise abundances because clouds are truncating the water absorption feature (e.g., Iyer et al. 2016) compared to the expectation for a clear atmosphere. The water abundance we derive from the dayside emission spectrum of HD 209458b is consistent with, but more precise than, the water abundance derived from the transmission spectrum by Benneke (2015). The consistency between water abundance values derived from dayside emission and transmission spectra is expected for hot Jupiters on thermochemical grounds (Moses et al. 2011) and was previously seen for the only other planet for which this comparison has been made (WASP-43b; Kreidberg et al. 2014a). We also find that our retrieved molecular abundances are thermochemically self-consistent, and thus "physical," with the exception of CO_2, which is best explained with metallicities $>10\times$ solar. However, again, the retrieved CO_2 abundance is strongly model and data dependent, suggesting that robust inference of CO_2 is not yet possible.

Of the seven planets with published *HST* WFC3 *emission* spectra (WASP-12b, Swain et al. 2013; WASP-4b, TrES-3b, Ranjan et al. 2014; HD 189733b, Crouzet et al. 2014; WASP-43b, Kreidberg et al. 2014a; WASP-33b, Haynes et al. 2015), only two, WASP-43b and HD 209458b (this work), report a robust detection of water over the WFC3 bandpass. Detecting or not detecting water has significant implications for the nature of the thermal structure and/or composition. Over the WFC3 bandpass alone, it is difficult to determine whether or not the lack of a water absorption feature is due to isothermal atmospheres (or nearly isothermal), an intrinsic depletion of water, due to either high C/O or low metallicity, or an optically thick dayside cloud at high altitudes. This ambiguity emphasizes the need to push for higher-precision emission observations to determine the presence, or lack thereof, of water absorption over a wide range of stellar compositions (C/O, metallicity) and planetary equilibrium temperatures. Furthermore, by combining emission and transmission observations (e.g., Kreidberg et al. 2014a), we can begin to break the temperature profile–water–cloud degeneracy.

These observations were made under the GO Treasury Program 13467 with the NASA/ESA *Hubble Space Telescope* at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under the contract NAS 5-26555. M.R.L. acknowledges support provided by NASA through Hubble Fellowship grant 51362 awarded by the Space Telescope Science Institute. K.B.S. acknowledges support from the NASA Exoplanet Science Institute Sagan Postdoctoral Fellowship. J.L.B. acknowledges support from the David and Lucile Packard Foundation. We also thank Dan Foreman-Mackey for making the *corner.py* plotting routine available to the public and Johannes Buchner for making `pymultinest` publicly available. We also thank Vivien Parmentier, Kevin Heng, Tom Evans, Drake Deming, Adam Burrows, and Heather Knutson for useful discussion and comments on the manuscript. We thank Drake Deming, Peter McCullough, Adam Burrows, Sara Seager, David Charbonneau, and Derek Homeier for being co-investigators on the *HST* observing proposal. Finally, we thank Roxana Lupu for support with the molecular absorption cross sections.

REFERENCES

Agúndez, M., Parmentier, V., Venot, O., Hersant, F., & Selsis, F. 2014, A&A, 564, A73
Barstow, J. K., Aigrain, S., Irwin, P. G. J., et al. 2013, MNRAS, 430, 1188
Benneke, B. 2015, arXiv:1504.07655
Benneke, B., & Seager, S. 2013, ApJ, 778, 153
Berta, Z. K., Charbonneau, D., Désert, J.-M., et al. 2012, ApJ, 747, 35
Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125
Burrows, A., Budaj, J., & Hubeny, I. 2008, ApJ, 678, 1436
Burrows, A., Hubbard, W. B., Lunine, J. I., & Liebert, J. 2001, RvMP, 73, 719
Burrows, A., Hubeny, I., Budaj, J., Knutson, H. A., & Charbonneau, D. 2007, ApJL, 668, L171
Burrows, A., & Orton, G. 2010, Exoplanets, 419
Burrows, A., Sudarsky, D., & Hubeny, I. 2006, ApJ, 640, 1063
Cooper, C. S., & Showman, A. P. 2006, ApJ, 649, 1048
Cornish, N. J., & Littenberg, T. B. 2007, PhRvD, 76, 083006
Crossfield, I. J. M. 2015, PASP, 127, 941
Crossfield, I. J. M., Knutson, H., Fortney, J., et al. 2012, ApJ, 752, 81
Crouzet, N., McCullough, P. R., Deming, D., & Madhusudhan, N. 2014, ApJ, 795, 166
Deming, D., Seager, S., Richardson, L. J., & Harrington, J. 2005, Natur, 434, 740
Deming, D., Wilkins, A., McCullough, P., et al. 2013, ApJ, 774, 95
Demory, B.-O., de Wit, J., Lewis, N., et al. 2013, ApJL, 776, L25
Diamond-Lowe, H., Stevenson, K. B., Bean, J. L., Line, M. R., & Fortney, J. J. 2014, ApJ, 796, 66
Ehrenreich, D., Bonfils, X., Lovis, C., et al. 2014, A&A, 570, A89
Esteves, L. J., De Mooij, E. J. W., & Jayawardhana, R. 2015, ApJ, 804, 150
Evans, T. M., Aigrain, S., Gibson, N., et al. 2015, MNRAS, 451, 680
Evans, T. M., Sing, D. K., Wakeford, H. R., et al. 2016, ApJL, 822, L4
Feng, Y. K., Line, M. R., Fortney, J. J., et al. 2016, ApJ, 829, 52
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Fortney, J. J. 2005, MNRAS, 364, 649
Fortney, J. J., Lodders, K., Marley, M. S., & Freedman, R. S. 2008, ApJ, 678, 1419
Fortney, J. J., Saumon, D., Marley, M. S., Lodders, K., & Freedman, R. S. 2006, ApJ, 642, 495
Fraine, J., Deming, D., Benneke, B., et al. 2014, Natur, 513, 526
Freedman, R. S., Lustig-Yaeger, J., Fortney, J. J., et al. 2014, ApJS, 214, 25
Gelman, A., & Rubin, D. B. 1992, StaSc, 7, 457
Gibson, N. P., Pont, F., & Aigrain, S. 2011, MNRAS, 411, 2199
Gibson, N. P., Aigrain, S., Roberts, S., et al. 2012, MNRAS, 419, 2683

Gordon, S., & McBride, B. J. 1994, NASRP, 1311
Greathouse, T. K., Lacy, J. H., Bézard, B., et al. 2005, Icar, 177, 18
Gregory, P. C. (ed.) 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with "Mathematica" Support (Cambridge: Cambridge Univ. Press), 200445930
Guillot, T. 2010, A&A, 520, A27
Hansen, B. M. S. 2008, ApJS, 179, 484
Hansen, C. J., Schwartz, J. C., & Cowan, N. B. 2014, MNRAS, 444, 3632
Haynes, K., Mandell, A. M., Madhusudhan, N., Deming, D., & Knutson, H. 2015, ApJ, 806, 146
Heng, K., & Lyons, J. R. 2016, ApJ, 817, 149
Heng, K., Mendonça, J. M., & Lee, J.-M. 2014, ApJS, 215, 4
Horne, K. 1986, PASP, 98, 609
Hubeny, I., Burrows, A., & Sudarsky, D. 2003, ApJ, 594, 1011
Huitson, C. M., Sing, D. K., Pont, F., et al. 2013, MNRAS, 434, 3252
Ingalls, J. G., Krick, J. E., Carey, S. J., et al. 2016, arXiv:1601.05101
Iyer, A. R., Swain, M. R., Zellem, R. T., et al. 2016, ApJ, 823, 109
Kirkpatrick, J. D. 2005, ARA&A, 43, 195
Knutson, H. A., Benneke, B., Deming, D., & Homeier, D. 2014, Natur, 505, 66
Knutson, H. A., Charbonneau, D., Allen, L. E., Burrows, A., & Megeath, S. T. 2008, ApJ, 673, 526
Knutson, H. A., Howard, A. W., & Isaacson, H. 2010, ApJ, 720, 1569
Kreidberg, L. 2015, PASP, 127, 1161
Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014a, ApJL, 793, L27
Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014b, Natur, 505, 69
Kreidberg, L., Line, M. R., Bean, J. L., et al. 2015, ApJ, 814, 66
Kuntschner, H., Bushouse, H., Kümmel, M., & Walsh, J. R. 2009, ST-ECF Instrument Science Report WFC3-2009-17
Lanotte, A. A., Gillon, M., Demory, B.-O., et al. 2014, A&A, 572, A73
Lee, J.-M., Fletcher, L. N., & Irwin, P. G. J. 2012, MNRAS, 420, 170
Line, M. R., Knutson, H., Deming, D., Wilkins, A., & Desert, J.-M. 2013a, ApJ, 778, 183
Line, M. R., Knutson, H., Wolf, A. S., & Yung, Y. L. 2014, ApJ, 783, 70
Line, M. R., Wolf, A. S., Zhang, X., et al. 2013b, ApJ, 775, 137
Line, M. R., Zhang, X., Vasisht, G., et al. 2012, ApJ, 749, 93
Madhusudhan, N., Crouzet, N., McCullough, P. R., Deming, D., & Hedges, C. 2014a, ApJL, 791, L9
Madhusudhan, N., Knutson, H., Fortney, J. J., & Barman, T. 2014b, in Protostars and Planets VI, ed. H. Beuther et al. (Tuscon, AZ: Univ. Arizona Press)
Madhusudhan, N., Mousis, O., Johnson, T. V., & Lunine, J. I. 2011, ApJ, 743, 191
Madhusudhan, N., & Seager, S. 2009, ApJ, 707, 24
Mandell, A. M., Haynes, K., Sinukoff, E., et al. 2013, ApJ, 779, 128
Marley, M. S., & Robinson, T. D. 2015, ARA&A, 53, 279
McCullough, P. R., Crouzet, N., Deming, D., & Madhusudhan, N. 2014, ApJ, 791, 55
Mollière, P., van Boekel, R., Dullemond, C., Henning, T., & Mordasini, C. 2015, ApJ, 813, 47
Morley, C. V., Fortney, J. J., Marley, M. S., et al. 2015, ApJ, 815, 110
Moses, J. I., Fouchet, T., Bézard, B., et al. 2005, JGRE, 110, E08001
Moses, J. I., Visscher, C., Fortney, J. J., et al. 2011, ApJ, 737, 15
Orton, G. S., Moses, J. I., Fletcher, L. N., et al. 2014, Icar, 243, 471
Parmentier, V., Fortney, J. J., Showman, A. P., Morley, C. V., & Marley, M. S. 2016, arXiv:1602.03088
Parmentier, V., & Guillot, T. 2014, A&A, 562, A133
Prinn, R. G., & Barshay, S. S. 1977, Sci, 198, 1031
Ranjan, S., Charbonneau, D., Désert, J.-M., et al. 2014, ApJ, 785, 148
Richardson, L. J., Deming, D., Horning, K., Seager, S., & Harrington, J. 2007, Natur, 445, 892
Richardson, L. J., Deming, D., & Seager, S. 2003, ApJ, 597, 581
Robinson, T. D., & Catling, D. C. 2012, ApJ, 757, 104
Robinson, T. D., & Catling, D. C. 2014, NatGe, 7, 12
Rowe, J. F., Matthews, J. M., Seager, S., et al. 2008, ApJ, 689, 1345
Schwarz, H., Brogi, M., de Kok, R., Birkby, J., & Snellen, I. 2015, A&AP, 576, A111
Seager, S., & Mallén-Ornelas, G. 2003, ApJ, 585, 1038
Showman, A. P., Fortney, J. J., Lian, Y., et al. 2009, ApJ, 699, 5645
Shporer, A., & Hu, R. 2015, AJ, 150, 112
Sing, D. K., Wakeford, H. R., Showman, A. P., et al. 2015, MNRAS, 446, 2428
Skilling, J. 2006, BayAn, 1, 833
Spiegel, D. S., Silverio, K., & Burrows, A. 2009, ApJ, 699, 1487
Stevenson, K. B., Bean, J. L., Madhusudhan, N., & Harrington, J. 2014a, ApJ, 791, 36
Stevenson, K. B., Bean, J. L., Seifahrt, A., et al. 2014b, AJ, 147, 161
Stevenson, K. B., Désert, J.-M., Line, M. R., et al. 2014c, Sci, 346, 838
Stevenson, K. B., Harrington, J., Fortney, J. J., et al. 2012, ApJ, 754, 136
Swain, M., Deroo, P., Tinetti, G., et al. 2013, Icar, 225, 432
Swain, M. R., Bouwman, J., Akeson, R. L., Lawler, S., & Beichman, C. A. 2008, ApJ, 674, 482
Swain, M. R., Line, M. R., & Deroo, P. 2014, ApJ, 784, 133
Swain, M. R., Tinetti, G., Vasisht, G., et al. 2009, ApJ, 704, 1616
ter Braak, C. J., & Vrugt, J. A. 2008, Stat. Comput., 18, 435
Trotta, R. 2008, ConPh, 49, 71
Visscher, C., Moses, J. I., & Saslow, S. A. 2010, Icar, 209, 602
Wakeford, H. R., Sing, D. K., Deming, D., et al. 2013, MNRAS, 435, 3481
Waldmann, I. P., Rocchetto, M., Tinetti, G., et al. 2015a, ApJ, 813, 13
Waldmann, I. P., Tinetti, G., Rocchetto, M., et al. 2015b, ApJ, 802, 107
Wilkins, A. N., Deming, D., Madhusudhan, N., et al. 2014, ApJ, 783, 113
Yelle, R. V. 2004, Icar, 170, 167
Zahnle, K., Marley, M. S., Freedman, R. S., Lodders, K., & Fortney, J. J. 2009, ApJL, 701, L20
Zellem, R. T., Griffith, C. A., Deroo, P., Swain, M. R., & Waldmann, I. P. 2014a, ApJ, 796, 48
Zellem, R. T., Lewis, N. K., Knutson, H. A., et al. 2014b, ApJ, 790, 53