



# A *Hubble* PanCET Study of HAT-P-11b: A Cloudy Neptune with a Low Atmospheric Metallicity

Yayaati Chachan<sup>1</sup> , Heather A. Knutson<sup>1</sup>, Peter Gao<sup>2,15</sup> , Tiffany Kataria<sup>3</sup> , Ian Wong<sup>4,15</sup> , Gregory W. Henry<sup>5</sup> , Bjorn Benneke<sup>6</sup> , Michael Zhang<sup>1</sup> , Joanna Barstow<sup>7</sup> , Jacob L. Bean<sup>8</sup> , Thomas Mikal-Evans<sup>9</sup>, Nikole K. Lewis<sup>10</sup> , Megan Mansfield<sup>8</sup> , Mercedes López-Morales<sup>11</sup> , Nikolay Nikolov<sup>12</sup> , David K. Sing<sup>13</sup> , and Hannah Wakeford<sup>14</sup>

<sup>1</sup> California Institute of Technology, 1200 E. California Boulevard, Pasadena, CA 91125, USA; [ychachan@caltech.edu](mailto:yachachan@caltech.edu)

<sup>2</sup> Department of Astronomy, University of California, Berkeley, CA 94720, USA

<sup>3</sup> Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, USA

<sup>4</sup> Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>5</sup> Center of Excellence in Information Systems, Tennessee State University, Nashville, TN 37209, USA

<sup>6</sup> Department of Physics, Université de Montréal, Montréal, QC, Canada

<sup>7</sup> Physics and Astronomy, University College London, London, UK

<sup>8</sup> Department of Astronomy, University of Chicago, Chicago, USA

<sup>9</sup> MIT Kavli Institute for Astrophysics and Space Research, Vassar Street, Cambridge, MA, USA

<sup>10</sup> Department of Astronomy and Carl Sagan Institute, Cornell University, 122 Sciences Drive, Ithaca, NY 14853, USA

<sup>11</sup> Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

<sup>12</sup> Astrophysics Group, School of Physics, University of Exeter, Exeter, EX4 4QL, UK

<sup>13</sup> Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD, USA

<sup>14</sup> Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

Received 2019 May 29; revised 2019 September 30; accepted 2019 October 15; published 2019 November 22

## Abstract

We present the first comprehensive look at the 0.35–5  $\mu\text{m}$  transmission spectrum of the warm ( $\sim 800$  K) Neptune HAT-P-11b derived from 13 individual transits observed using the *Hubble* and *Spitzer Space Telescopes*. Along with the previously published molecular absorption feature in the 1.1–1.7  $\mu\text{m}$  bandpass, we detect a distinct absorption feature at 1.15  $\mu\text{m}$  and a weak feature at 0.95  $\mu\text{m}$ , indicating the presence of water and/or methane with a combined significance of  $4.4\sigma$ . We find that this planet’s nearly flat optical transmission spectrum and attenuated near-infrared molecular absorption features are best matched by models incorporating a high-altitude cloud layer. Atmospheric retrievals using the combined 0.35–1.7  $\mu\text{m}$  *Hubble Space Telescope* (*HST*) transmission spectrum yield strong constraints on atmospheric cloud-top pressure and metallicity, but we are unable to match the relatively shallow *Spitzer* transit depths without underpredicting the strength of the near-infrared molecular absorption bands. HAT-P-11b’s *HST* transmission spectrum is well matched by predictions from our microphysical cloud models. Both forward models and retrievals indicate that HAT-P-11b most likely has a relatively low atmospheric metallicity ( $<4.6 Z_{\odot}$  and  $<86 Z_{\odot}$  at the  $2\sigma$  and  $3\sigma$  levels respectively), in contrast to the expected trend based on the solar system planets. Our work also demonstrates that the wide wavelength coverage provided by the addition of the *HST* STIS data is critical for making these inferences.

*Unified Astronomy Thesaurus concepts:* [Exoplanet atmospheres \(487\)](#); [Exoplanet atmospheric composition \(2021\)](#)

*Supporting material:* machine-readable table

## 1. Introduction

The atmospheric compositions of extrasolar gas giant planets are expected to vary depending on their formation locations and accretion histories. Variation in composition of disk gas and solids as well as the availability of polluting solids at different locations leaves an imprint on a planet’s atmosphere (e.g., Oberg et al. 2011; Lambrechts et al. 2014; Venturini et al. 2016; Pudritz et al. 2018). By measuring the wavelength-dependent transit depth when one of these planets passes in front of its host star (the planet’s “transmission spectrum”), we can detect atmospheric absorption features that directly constrain the mean molecular weight and relative abundances of molecules including water, methane, carbon monoxide, and carbon dioxide. Although some planets with strong and clear absorption features have been thus characterized (e.g., WASP 96b, Nikolov et al. 2018a; WASP 39b, Wakeford et al. 2018; WASP 107b, Kreidberg et al. 2018), large observing campaigns using the *Hubble Space Telescope* (*HST*) have

revealed the presence of high-altitude clouds that attenuate the expected absorption signal in a majority of the close-in gas giant planets observed to date (e.g., Sing et al. 2016; Fu et al. 2017; Tsiaras et al. 2018; Wakeford et al. 2019). This problem is even more acute for the current sample of Neptune-sized planets, whose relatively small radii, high surface gravities, and low temperatures all combine to reduce the expected amplitude of atmospheric absorption as compared to their better-studied Jovian counterparts (e.g., Crossfield & Kreidberg 2017). This limits our ability to search for trends in atmospheric properties with other parameters of the system, e.g., planet mass, radius, and temperature—all of which are crucial for improving our understanding of planet formation and evolution.

Although the current body of observed transmission spectra clearly require the presence of high-altitude scattering particles, there is considerable debate about the nature and origin of these particles. At high temperatures, we expect refractory species such as metal oxides, silicates, and sulphides to condense in exoplanetary atmospheres (e.g., Morley et al. 2012; Helling 2019; Powell et al. 2018). However, cloud formation is a

<sup>15</sup> 51 Pegasi b Fellow.

complex process that depends on both microphysical processes, such as sedimentation, nucleation, and growth, and the material properties of the condensing species, many of which are highly uncertain or unknown (Helling 2019). Consequently, the use of different underlying assumptions can lead to significantly different cloud properties, severely limiting the predictive power of these models.

While some of these questions may be resolved by ongoing laboratory experiments (Johnson et al. 2017, 2018; He et al. 2018a, 2018b; Hörst et al. 2018), observational constraints on the properties of clouds in exoplanetary atmospheres provide complementary leverage to further refine and develop microphysical cloud models. The nature of these constraints varies depending on the wavelength of the observations: optical and near-infrared transmission spectroscopy can be used to investigate the sizes, number density, and vertical distribution of cloud particles, while vibrational modes in the mid-infrared can be used to directly determine the compositions of cloud particles (e.g., Wakeford & Sing 2015; Pinhas & Madhusudhan 2017; Kitzmann & Heng 2018).

Although clouds represent a substantial challenge for compositional inferences from transmission spectroscopy, previous *HST* studies have demonstrated that we can nonetheless obtain reasonable constraints on atmospheric composition for planets with detectable near-infrared water features by utilizing information at optical wavelengths to break degeneracies between cloud-top pressure and atmospheric metallicity (e.g., HAT-P-26b, Wakeford et al. 2017a; WASP 39b, Wakeford et al. 2018). Spectroscopic observations in the near-infrared have been instrumental in the detection of molecular absorption in exoplanetary atmospheres but they are usually unable to put tight constraints on the composition, i.e., the absolute mixing ratios, of these molecules. This is because the transmission spectra of an atmosphere with a deep cloud and low mixing ratios is statistically indistinguishable (with currently available precision) from an atmosphere with a high cloud and high mixing ratios. These distinct scenarios can be distinguished by their differing spectral behavior in the optical. In this spirit, the Panchromatic Comparative Exoplanet Treasury (PanCET) survey is a multi-cycle *HST* treasury program whose primary goal is to characterize the atmospheres of a sample of 20 transiting gas giant planets at wavelengths ranging from the ultraviolet to the near-infrared (e.g., Evans et al. 2017, 2018; Wakeford et al. 2017b; Alam et al. 2018; Bourrier et al. 2018; Nikolov et al. 2018b). In this study we present new optical *HST* STIS PanCET observations of HAT-P-11b, a warm Neptune-sized planet with a radius of 4.4 Earth radii and mass of 23 Earth masses on a 4.88 days orbit around a  $0.81 M_{\odot}$ ,  $0.68 R_{\odot}$  K4 star ( $T_{\text{eff}} = 4780 \pm 50$  K). This planet has a significantly eccentric orbit ( $e = 0.218$ ) and as a result its predicted equilibrium temperature varies between  $\sim 700$  and  $900$  K (Bakos et al. 2010; Deming et al. 2011; Yee et al. 2018). The planet therefore crosses multiple condensation lines, which enhances its potential for cloud formation.

HAT-P-11b has been previously observed with both ground- (e.g., Bakos et al. 2010; Sanchis-Ojeda & Winn 2011) and space-based telescopes (e.g., Deming et al. 2011; Fraine et al. 2014; Huber et al. 2017) and is one of the most favorable Neptune-sized planets for atmospheric characterization due to its large atmospheric scale height and host star brightness ( $V \sim 9$ ). It is one of the smallest planets with a published detection of water absorption in its  $1.1\text{--}1.7 \mu\text{m}$  *HST* WFC3

transmission spectrum (Fraine et al. 2014, hereafter F14). Although there is an optical detection of the planet’s secondary eclipse using *Kepler* photometry (Huber et al. 2017), no corresponding infrared detection has been reported to date. Measurements of absorption in the He metastable  $10830 \text{ \AA}$  line during transit provide complementary constraints on the size of the planet’s exosphere and corresponding mass loss rate (Allart et al. 2018; Mansfield et al. 2018). Although the relatively high activity level of HAT-P-11b’s K dwarf primary can bias the shape of the planet’s measured transmission spectrum (e.g., Sing et al. 2011; Rackham et al. 2018, 2019), the planet’s nearly polar orbit (Winn et al. 2010; Hirano et al. 2011) has enabled exquisitely detailed studies of the starspot distribution and active latitudes (e.g., Deming et al. 2011; Sanchis-Ojeda & Winn 2011; Morris et al. 2017a, 2017b) that can be used to effectively correct for these effects.

Here, we combine previously published transit observations from *HST* WFC3 ( $0.8\text{--}1.7 \mu\text{m}$ ) and *Spitzer* ( $3.6, 4.5 \mu\text{m}$ ) (Fraine et al. 2014; Mansfield et al. 2018) with new optical *HST* STIS observations to obtain the first comprehensive look at HAT-P-11b’s transmission spectrum between  $0.35$  and  $5 \mu\text{m}$ . We compare the resulting transmission spectrum to predictions from forward models for cloud condensation and use retrievals to independently constrain the planet’s atmospheric composition and cloud properties. Sections 2 and 3 describe our spectral and photometric extraction methods, while Section 4 discusses instrumental and astrophysical noise sources in our data. Section 5 details our fits to these data, and Section 6 discusses predictions from forward models for HAT-P-11b’s atmosphere. Adopting some material properties and tools from this section, we then use simple models to directly fit the observed transmission spectrum in order to derive statistical constraints on atmospheric parameters in Section 7, which we compare to the forward models in Section 8.

## 2. Observations

A summary of the observations used in our analysis is given in Table 1. We analyze 13 transits in total and describe each of them below.

We observed three transits of HAT-P-11b with the Space Telescope Imaging Spectrograph (STIS) on the *HST* (PI Sing & López-Morales, GO 14767). Two observations were conducted using the G430L grism ( $0.29\text{--}0.57 \mu\text{m}$ ) on UT 2017 February 22 and UT 2017 May 26, while a third visit on UT 2017 April 12 used the G750L grism ( $0.524\text{--}1.027 \mu\text{m}$ ). All of our observations were obtained using the  $52'' \times 2''$  slit. This was done to minimize slit losses and the effect of telescope breathing. Each visit consists of five *HST* orbits. Short (1 s) exposures were taken before each orbit to mitigate the severity of the exponential ramp at the beginning of each orbital light curve, but this step did not appear to be effective for these observations. The wavelength calibration and flat field exposures were taken during the occultation of *HST* by Earth during the last orbit. Along with the *HST* STIS data, we independently re-reduce and fit all of the prior data collected with *HST* and *Spitzer* as part of our updated global analysis, which we discuss below in chronological order.

HAT-P-11b was observed with *HST*’s Wide Field Camera 3 (WFC3) instrument in 2012 (PI Deming, GO 12449) using the G141 grism in the  $256 \times 256$  subarray mode, which provides a low resolution spectrum in the  $1.1\text{--}1.7 \mu\text{m}$  wavelength range. The data were collected over four *HST* orbits using only

**Table 1**  
Observations

Date (UT)	Start Time	Duration	Observatory	Band Pass ( $\mu\text{m}$ )	Integration Time (s)	Exposures	Reference
2011 Jul 11	23:11:41	7.43 hr	<i>Spitzer</i>	3.05–3.95	0.4	62,592	Fraine et al. (2014)
2011 Aug 5	07:02:48	7.43 hr	<i>Spitzer</i>	4.05–4.95	0.4	58,112	Fraine et al. (2014)
2011 Aug 15	01:49:20	7.43 hr	<i>Spitzer</i>	3.05–3.95	0.4	52,633	Fraine et al. (2014)
2011 Aug 29	17:37:18	7.43 hr	<i>Spitzer</i>	4.05–4.95	0.4	62,592	Fraine et al. (2014)
2012 Oct 18	04:58:38	6.87 hr	<i>Hubble</i>	1.1–1.70	44.4	113	Fraine et al. (2014)
2016 Sep 14	10:36:07	5.65 hr	<i>Hubble</i>	0.8–1.15	81.1	99	Mansfield et al. (2018)
2016 Oct 13	18:44:21	5.83 hr	<i>Hubble</i>	0.8–1.15	81.1	99	Mansfield et al. (2018)
2016 Nov 7	05:12:38	5.88 hr	<i>Hubble</i>	0.8–1.15	81.1	99	Mansfield et al. (2018)
2016 Nov 26	18:22:55	5.83 hr	<i>Hubble</i>	0.8–1.15	81.1	99	Mansfield et al. (2018)
2016 Dec 26	02:17:30	5.77 hr	<i>Hubble</i>	0.8–1.15	81.1	99	Mansfield et al. (2018)
2017 Feb 22	17:04:39	7.17 hr	<i>Hubble</i>	0.29–0.57	140	82	This work
2017 Apr 12	14:15:13	7.67 hr	<i>Hubble</i>	0.524–1.027	140	81	This work
2017 May 26	13:50:02	7.28 hr	<i>Hubble</i>	0.29–0.57	140	81	This work

forward scans (McCullough & Mackenty 2012) with a scan rate of  $0''.3891 \text{ s}^{-1}$ . The second orbit covers part of ingress. A buffer dump occurred during the third orbit, which partially resets the ramp that is used to model the instrumental behavior (Deming et al. 2013; Knutson et al. 2014; Kreidberg et al. 2014); see Section 4 for more details. These data were originally published in F14.

Four transits of HAT-P-11b were observed in 2011 using the Infrared Array Camera (IRAC) mounted on *Spitzer Space Telescope*, with two transits in each of the two warm-*Spitzer* channels (3.6 and 4.5  $\mu\text{m}$ ). The observations were taken in the subarray mode, which yielded  $32 \times 32$  pixel images with an integration time of 0.4 s. The *Spitzer* data were published along with the WFC3 G141 data in F14.

Finally, five transits of HAT-P-11b were also observed using the WFC3 G102 grism (0.8–1.15  $\mu\text{m}$ ) in the  $256 \times 256$  subarray mode (PI Bean, GO 14793) on UT 2016 September 14, 2016 October 13, 2016 November 7, 2016 November 26, and 2016 December 26. This grism is complementary to the G141 observations, as both grisms together span a series of adjacent and overlapping water and methane bands. During each visit, the planet was observed in scan mode over four orbits. The use of forward and backward scans and longer exposure times for G102 observations yielded a higher observational efficiency ( $\sim 75\%$ ) than the 2012 G141 observations ( $\sim 50\%$ ). These data were published in Mansfield et al. (2018); (hereafter M18), who reported a strong helium absorption from escaping gas in the planet’s outer atmosphere but did not see the expected molecular (water) absorption features in this bandpass.

### 3. Spectral and Photometric Extraction

We use the ExoTEP framework (Benneke et al. 2019) for the extraction and fitting of all data sets. The extraction process for each of the instruments is described below.

#### 3.1. HST STIS Spectroscopy

We correct for cosmic ray hits and other transient phenomena by stacking all of the images from a given visit and examining flux as a function of time at each pixel position. Because these data have relatively sparse time sampling ( $<100$  images per visit) and time-correlated instrumental effects, we find that we obtain optimal results when we flag  $4\sigma$  outliers in

each pixel’s time series and replace them with the median pixel value. We then estimate the background in each image by taking the median pixel value in two rectangular regions located far enough from the spectral trace to avoid contamination. We optimize the aperture width (in the cross-dispersion direction) for extraction of one-dimensional (1D) spectra and decide whether or not to remove the background by minimizing the scatter in the white light residuals after subtracting the best-fit transit and instrumental noise model for each visit (e.g., Deming et al. 2013). For each visit, we consider aperture sizes of 7, 9, 11, and 13 pixels. In the G750L visit we obtain optimal results when we use a 9 pixel wide aperture centered on the peak of the point-spread function and do not subtract the background. For the G430L observations, we prefer to subtract the background and utilize 13 and 11 pixel wide apertures for the first and second visits, respectively. We find that in all visits the white light transit depths and transmission spectral shapes are relatively insensitive to our choice of aperture width.

Data taken with the G750L grism exhibit a fringing effect due to internal reflection within individual pixels. We correct for this effect using a fringe flat field obtained contemporaneously with our data following the methods outlined in Nikolov et al. (2014, 2015) and Sing et al. (2016). Using the first frame as a template, we then fit for the shift in position in the dispersion direction and relative amplitude of all subsequent frames in order to align the frames in wavelength. These best-fit relative amplitudes give us the normalized white light curve for each visit. For the wavelength-dependent light curves, we sum the flux within a 200 pixel wide bin for the G750L grism and a 100 pixel wide bin for the G430L grism. We also check for the presence of sodium and potassium absorption in the G750L bandpass by extracting the flux in two narrow bandpasses centered on the corresponding absorption lines (588.7–591.2 and 770.3–772.3 nm, respectively).

#### 3.2. HST WFC3 Spectroscopy

We reduce data from both the G102 and G141 grisms following the method outlined in Tsiaras et al. (2016). Unlike that study, we begin with the bias- and dark-corrected `ima` images produced by the standard `calwfc3` pipeline rather than calibrating the `raw` images ourselves. Each of the exposures consists of five nondestructive reads. We create difference subexposures by subtracting consecutive reads (e.g., Deming et al. 2013; Kreidberg et al. 2014; Evans et al. 2016). We



determine the extent of the subexposure in the scan direction by finding the rows where the median flux profile in the spatial scan direction falls to 20% of the peak flux and add an additional buffer of 15 pixels above and below these rows. The extraction is not very sensitive to the number of pixels used for this buffer and any value between 10 and 20 suffices. We then mask out the rows exterior to this  $y$  pixel range and estimate the background using a 20 column wide rectangular region within the subexposure spanning columns between the end of the spectral trace and the edge of the array, taking care to avoid any secondary sources in the image. We remove any bad pixels by discarding  $3\sigma$  outliers from this background region and then subtract the median of the remaining pixels from the unmasked part of the image. We then create a combined full frame image by coadding all of the background subtracted subexposures.

Although the pointing of WFC3 is generally very stable, our scanned observations nonetheless exhibit small image-to-image variations in the position of the spectral trace in the  $x$  (dispersion) direction. By default, we estimate the magnitude of these shifts relative to the first frame by summing each image in the  $y$  direction and using this rough 1D-extracted spectrum to calculate the corresponding  $x$  offset. We find that the magnitude of this shift is less than 0.1 pixel over the entire duration of the WFC3 G102 visits. The WFC3 G141 data were taken shortly after the spatial scanning mode was first implemented on *HST* and exhibit a larger shift of approximately one pixel over the visit, most likely due to the suboptimal scanning strategy utilized in these older observations. We find that using the centroid of each exposure and determining the horizontal offset relative to the centroid of the first exposure significantly decreases the scatter in the best-fit residuals for the G141 visit. We then use the wavelength and trace calibration functions provided by STScI (Kuntschner & Bushouse 2009a; Kuntschner et al. 2009b) for each grism to calculate the full 2D wavelength solution for each image.

We flat-field all frames using the calibration files provided by STScI (Kuntschner et al. 2011) following the method outlined in Wilkins et al. (2014) and identify bad pixels in each individual image using a  $6\sigma$  moving median filter in both the  $x$  and  $y$  directions. Although we also consider lower filter thresholds, we find that these result in overly aggressive outlier correction. We replace these outliers with the mean value within the moving filter and repeat the same filtering a second time to ensure that we have identified and removed all outliers.

The width of the spectral trace in the dispersion direction varies with the  $y$ -position of the star on the detector. As a result, lines of constant wavelength are slanted relative to the columns of the detector. For the wavelength-dependent light curve extraction, we follow the method outlined in Tsiaras et al. (2016) and use the wavelength solution to determine the boundaries of each slanted wavelength bin and sum the flux within each bin. When the bins intersect with pixels, we use a second-order 2D polynomial to interpolate and integrate the flux over each partial-pixel region. This procedure ensures flux conservation and leads to a small reduction in the photometric scatter relative to other commonly employed methods, which usually smooth the data in the dispersion direction before light curve extraction (e.g., Deming et al. 2013; Fraine et al. 2014; Knutson et al. 2014).

For the wavelength-dependent light curves obtained with the G141 grism, we use 30 nm wide bins spanning the wavelength range 1.1–1.7  $\mu\text{m}$ . F14 utilized narrower wavelength bins, but

also convolved their 1D spectra with a 4 pixel wide Gaussian filter prior to binning. Since we do not smooth our data, we adopt a lower wavelength resolution. For the G102 data, we utilize bins with a width of  $\sim 23.3$  nm spanning the wavelength range 0.8–1.15  $\mu\text{m}$ , identical to those adopted by M18. The white light curve is simply obtained by summing the flux from all the spectroscopic light curves.

### 3.3. Spitzer 3.6 and 4.5 $\mu\text{m}$ Photometry

We extract the photometric light curve for each *Spitzer* visit following the method described in Knutson et al. (2012), Wong et al. (2016), and Zhang et al. (2018a). We determine the star’s position in each  $32 \times 32$  pixel *Spitzer* subarray image by iteratively calculating the flux-weighted centroid within a circular aperture with a radius of 3 pixels. To estimate the sky background, we first mask pixels located within a 12 pixel radius of the star’s position and then iteratively trim  $3\sigma$  outliers (e.g., Knutson et al. 2012). We calculate the mean value of the remaining background pixels using the biweight location method (Astropy Collaboration et al. 2013, 2018) and subtract it from each image. We then use the `photutils` package (Bradley et al. 2018) to extract the photometry using circular apertures with radii ranging from 1.5 to 3 pixels in 0.1 pixel increments and 3–5 pixels in 0.5 pixel increments. We select the optimal aperture for each visit by minimizing the scatter in the best-fit residuals, which are binned in 60 s intervals (see Section 4.3 for more information). This procedure gives extraction apertures of 2.8 and 2.3 pixels for first and second transit in the 3.6  $\mu\text{m}$  channel, respectively, and 2.3 and 2.6 pixels for the first and second transit in the 4.5  $\mu\text{m}$  channel, respectively.

We iteratively trim outliers in the resulting time series using a 50 point moving median filter and discarding photometric points that lie more than  $3\sigma$  away. We also fit third-order polynomials to the star’s  $x$  and  $y$  positions and discard any photometric points more than  $3\sigma$  away from the polynomial model position during the observation, as these points are not well-corrected by our instrumental noise model. The number of points removed in each of these steps ranges between 0.09% and 0.97% for each individual visit and is commensurate with expectations for normally distributed data.

## 4. Systematics and Astrophysical Models

### 4.1. HST/STIS Instrumental Model

We remove the first orbit in each of the STIS data sets as the instrumental systematics are notably worse than they are in subsequent orbits. This difference is attributed to the thermal relaxation of *HST* following target acquisition due to the change in incidence angle of solar radiation. In addition, we remove the first exposure within each orbit as it has a much lower flux that is not well matched by our parametric model. Both of these steps are standard practice for STIS data sets (e.g., Sing et al. 2011; Nikolov et al. 2015; Wakeford et al. 2017a). For the instrumental systematics model, we use a fourth-order polynomial in orbital phase and a linear trend in time (Sing et al. 2008). We also fit for a linear trend in the  $x$  (dispersion) position of the star on the array for the G750L visit and the first G430L visit as it significantly reduces the Bayesian Information Criterion (BIC: change of 35 and 8, respectively) and lowers the residual scatter in our light curve fits from 1.51



and 1.6 times the photon noise limit to 1.26 and 1.53 times, respectively.

As discussed in Sing et al. (2019), we find that the scatter in our white light residuals is further reduced if we decorrelate against additional parameters related to variations in telescope pointing. We find that the white light residuals from our initial fit exhibit a strong correlation with the recorded R.A. and decl., V2 and V3 roll, and latitude and longitude from the image file headers. However, these parameters are highly correlated with each other and we therefore use Principal Component Analysis (PCA) to reduce the number of independent fit parameters. We start with six principal components and retain those that capture  $\geq 95\%$  of the systematic variation in the light curves. Using this criterion, we retain four and three parameters for the first and second visit in the G430L bandpass, respectively and three parameters in the G750L bandpass. We include linear contributions from these PCA parameters as part of our final systematics model. The addition of these linear jitter parameters has a negligible effect on the BIC ( $|\text{BIC}| < 2$ ) for all three visits but it reduces the scatter in our residuals by 5%–8%. The full systematics model  $S(t)$  is given as

$$S(t) = c + vt_v + mx + \sum_i j_i p_{\text{jitter}} + \sum_{k=1}^4 p_k t_{\text{orb}}^k, \quad (1)$$

where  $t_v$  is the time from the beginning of the visit,  $t_{\text{orb}}$  is the time from the beginning of an orbit,  $p_{\text{jitter}}$  are the PCA vectors that describe the telescope pointing jitter, and  $c$ ,  $v$ ,  $m$ ,  $j_i$  and  $p_i$  are free parameters in the fit.

## 4.2. HST/WFC3 Instrumental Model

### 4.2.1. G141 grism

F14 used the spectral template fitting method to derive wavelength-dependent transit depths for the WFC3 data. Here, we fit the time series for each individual spectroscopic light curve independently following the method described in Tsiaras et al. (2016). As with the STIS data, we trim the first orbit and the first exposure of each orbit, as they are not well matched by our instrumental noise model.

Although there is an alternative physically motivated model that would in theory allow us to fit these data (Zhou et al. 2017), we do not expect that this would improve the precision of our transit depth measurement as we already have an out-of-transit baseline that is comparable in duration to our in-transit data. We fit the remaining orbits using a linear function of time and an exponential function of orbital phase, which is needed in order to correct for charge-trapping in the array (e.g., Deming et al. 2013; Zhou et al. 2017).

Our WFC3 systematics model  $S(t)$  is

$$S(t) = (c + vt_v) + (1 - e^{-at_{\text{orb}} - b - D_t}), \quad (2)$$

where  $c$ ,  $v$ ,  $a$ , and  $b$  are free parameters in the fit,  $t_v$  is the time from the beginning of the visit,  $t_{\text{orb}}$  is the time from the beginning of an orbit, and  $D_t$  is a vector (same length as  $t_v$ ) that is used to add duration-specific non-zero phase offsets. We use it to model the partial reset of the exponential ramp after a midorbit buffer dump in the third orbit (free parameter  $e$ ) and to account for the slightly different ramp amplitude of the first fitted orbit (free parameter  $d$ , see Kreidberg et al. 2015). The parameters  $c$  and  $v$  characterize the linear dependence of systematic noise on time. For the exponential dependence,  $a$

controls the dependence on  $t_{\text{orb}}$ , and  $b$  sets the overall time-independent amplitude of the exponential term.

### 4.2.2. G102 grism

Unlike the G141 data, which only scanned in a single direction, the G102 observations were taken with an alternating scan direction. The behavior of the ramp is slightly different for each scan direction, likely due to small offsets in the relative position of the scanned spectrum on the array. We carry out an initial fit in which we allow the full exponential ramp model to vary independently for each of the scan directions and find that all parameters except the constant  $c$  in Equation (2) are consistent with a single common value. We therefore carry out our final fits assuming the same slope  $v$  and exponential ramp coefficients  $a$ ,  $b$ , and  $D_t$  for both scan directions, but assign the forward and backward directions separate constant terms  $c_f$  and  $c_b$  that are allowed to vary independently.

## 4.3. Spitzer Instrumental Model

*Spitzer* 3.6 and 4.5  $\mu\text{m}$  photometry exhibits a ramp-like behavior (e.g., Lewis et al. 2013; Wong et al. 2016; Zhang et al. 2018a) at the start of each new observation. Rather than fitting this ramp with a model, we simply trim the first 0.5–2 hr of data and find that the optimum trim duration for each visit that minimizes the scatter in our binned best-fit residuals is 1 hr. Even after truncation, we find that the second visit in the 3.6  $\mu\text{m}$  bandpass possesses a significant ramp. Fitting this visit with the standard systematics model we adopt (see Equation (3) below) yields a much shallower transit depth and larger BIC ( $\Delta \text{BIC} \sim 20$ ) compared to the values we obtain when we fit for this ramp. We do not use the ramp model for the other *Spitzer* visits because it changes the transit depths by  $\lesssim 1\sigma$  and increases the BIC. Prior to fitting we bin the data in 60 s intervals. This binning results in a lower level of time-correlated noise in our best-fit residuals while still resolving the transit ingress and egress (for a discussion of binning practices with *Spitzer* data, see Deming et al. 2015 and Kammer et al. 2015).

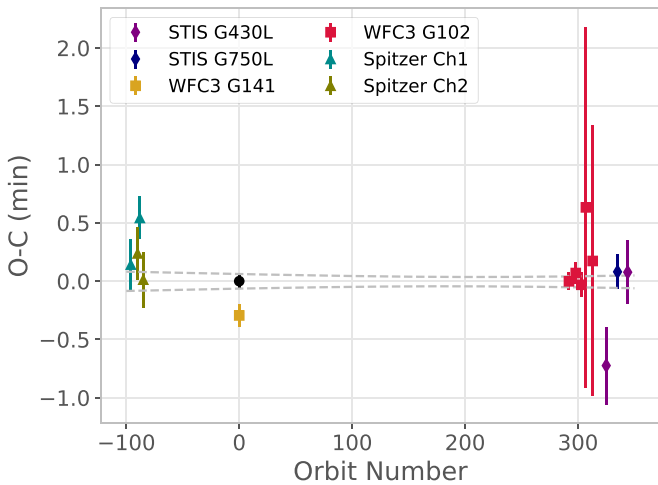
The primary instrumental noise source in the 3.6 and 4.5  $\mu\text{m}$  *Spitzer* arrays is intrapixel sensitivity variations combined with telescope pointing jitter. We model this behavior using Pixel-Level Decorrelation (PLD) following Deming et al. (2015):

$$S(t) = 1 + vt_v + \sum_{i=1}^9 w_i P_i(t), \quad (3)$$

where  $t_v$  is the time from the beginning of the visit,  $P_i$  is the normalized pixel count in the  $3 \times 3$  array around the source, and  $w_i$  are the weights assigned to each of these arrays, which are determined using linear regression after dividing out the transit light curve at each step in the fit. The slope parameter  $v$  is left to vary as a free parameter. For the second visit in the 3.6  $\mu\text{m}$  bandpass, we have an additional ramp term in the model with an amplitude  $A$  and decay timescale  $\tau$ :  $Ae^{-t_v/\tau}$ .

## 4.4. Transit Model

We use the BATMAN package (Kreidberg 2015) to model the transit light curve. The astrophysical model depends on the planet-star radius ratio  $R_p/R_*$ , planet semimajor axis to stellar radius ratio  $a/R_*$ , impact parameter  $b$ , period  $P$ , and transit center time  $T_c$ . We fit for all of these parameters in our global



**Figure 1.** Observed minus calculated midtransit times from fits to individual visits, where the color indicates the instrument. Predicted transit times are calculated using the best-fit ephemeris from the global fit, with  $1\sigma$  uncertainties indicated by the dashed gray lines. Visits with minimal data during ingress or egress have significantly larger uncertainties.

fit, but use fixed values for  $P$ ,  $a/R_*$ , and  $b$  when fitting individual transits. We fix the orbital eccentricity  $e = 0.218$  and longitude of periastron  $w = 199^\circ$  to the best-fit values from Yee et al. (2018). We validate our assumption of a linear ephemeris by comparing the best-fit midtransit times from individual visits with the best-fit ephemeris from our global fit in Figure 1. The best-fit midtransit times for all visits are consistent with a linear ephemeris at the  $2\sigma$  level or better.

Our updated ephemeris is consistent with the values reported in Deming et al. (2011) and Southworth (2011) to within  $0.2\sigma$ . However, there is only moderate agreement with the values reported in Sanchis-Ojeda & Winn (2011) and Huber et al. (2017). Curiously enough, the values reported by Huber et al. (2017), Sanchis-Ojeda & Winn (2011), and Southworth (2011) are for the same epoch and they disagree at the  $10\sigma$  level. We suspect this is due to errors in reporting of the midtransit time in the stated time convention. For example, Southworth (2011) and Sanchis-Ojeda & Winn (2011) report almost identical values for the midtransit time, but the former report it in BJD UTC while the latter do so in BJD TDB. These two time conventions differ by 66.184 s (an additional leap second was added in the first month of *Kepler*’s quarter 14). Similarly, the value reported by Huber et al. (2017), supposedly in BJD UTC, matches that of Southworth (2011) converted to BJD TDB. Careful accounting of these errors might resolve the paradoxes posed by these differing midtransit times.

As part of ExoTEP, we employ the Python package LDTk (Parviainen & Aigrain 2015) to calculate limb darkening coefficients for all of our observations except the *Spitzer* transits. LDTk queries spectral intensity profiles from the PHOENIX library (Husser et al. 2013) and computes a mean limb darkening profile for a star given its effective temperature, surface gravity, and metallicity (and associated uncertainties). We then fit this profile with a four-parameter nonlinear limb darkening model, and we fix the limb darkening coefficients to the model values in our light-curve fits. PHOENIX profiles extend from 50 to 2600 nm in wavelength space and can therefore only supply limb darkening coefficients for the *HST* bandpasses. For the *Spitzer* bandpasses, we use the (four-parameter nonlinear) limb darkening coefficients tabulated by

Sing (2010, assuming  $T_{\text{eff}} = 4750$  K,  $\log g = 4.5$ ,  $[\text{Fe}/\text{H}] = 0.3$ ), which are calculated from ATLAS models. We investigate the importance of our choice of limb darkening models in the *Spitzer* bands by refitting the *Spitzer* light curves with quadratic limb darkening coefficients as free parameters. We obtain transit depths that agree to within  $0.5\sigma$  with those obtained with ATLAS limb darkening coefficients. We therefore conclude that our use of ATLAS models instead of PHOENIX models at 3.6 and 4.5  $\mu\text{m}$  has a negligible effect on our results.

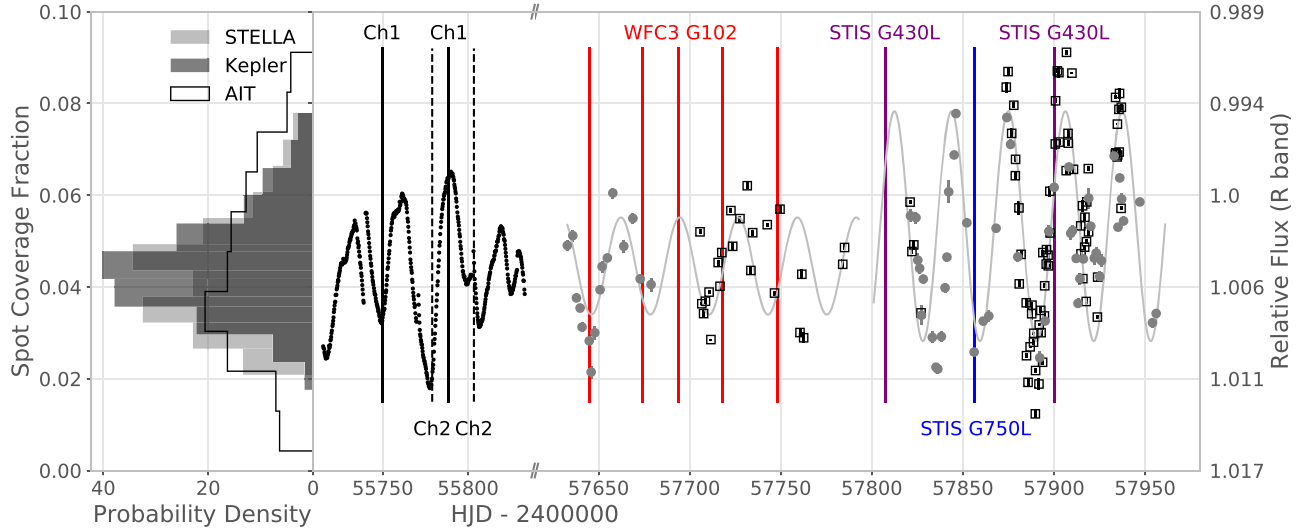
#### 4.5. Stellar Activity

HAT-P-11 is a relatively active K dwarf with a Ca II H & K emission line strength of  $\log(R'_{\text{HK}}) = -4.57$  (Knutson et al. 2010), and it is therefore important to address the impact of its activity on the transmission spectrum (McCullough et al. 2014; Morris et al. 2017a, 2017b). Both occulted and unocculted spots introduce wavelength-dependent biases in the transmission spectrum (e.g., Pont et al. 2008; Sing et al. 2011; Rackham et al. 2018). These biases must be corrected to combine transit depth measurements from different epochs and different wavelength bandpasses.

We find no evidence for any spot crossings during the *HST* observations included in this analysis, with the exception of two G102 visits. Following M18, we simply trim the data associated with the spot occultation rather than including this effect in our models. While two of the four *Spitzer* transits with contemporaneous *Kepler* transit photometry included a spot occultation, this occultation was evident only in the *Kepler* light curve. Given the relatively small chromatic effect of spot crossing at infrared wavelengths, F14 concluded that these spots would have had a negligible effect on the measured *Spitzer* transit depths.

Unocculted spots are usually much harder to correct for as accounting for their effect requires knowledge of the fractional surface area of the star that is covered by the spots as well as the average spot temperature. Fortunately, HAT-P-11 has some of the best constraints on spot properties among all stars that host transiting planets. This is because HAT-P-11b orbits its star from pole to pole (its orbit is misaligned with the stellar spin axis by  $106^\circ$ ; Deming et al. 2011; Sanchis-Ojeda & Winn 2011) and the star was monitored by *Kepler* in a broad optical bandpass from 2009 to 2012, allowing us to observe more than 200 transits of the planet. This essentially provides us with a latitude–longitude map of the entire stellar surface and constrains the spot covering fraction of the stellar surface to be  $3^{+6}_{-1}\%$  (Morris et al. 2017a, 2017b).

The *Kepler* data span the epoch of the *Spitzer* transit observations and although the G141 observations were taken in 2012, they unfortunately coincided with a gap in the *Kepler* coverage (F14). We also obtained photometric monitoring data in the Cousins *R* bandpass with the Celestron 14 inch (C14) Automated Imaging Telescope (AIT) at Fairborn Observatory (Sing et al. 2015) and in the Johnson *B* and *V* filters from the 1.2 m robotic STELLA telescope at Izaña Observatory (Strassmeier et al. 2004; data taken from M18). These data were obtained between 2015 and 2017, covering the epochs of the WFC3 G102 and STIS observations but not the 2012 WFC3 G141 observations. This introduces a source of uncertainty, as there is no uniform source of monitoring data spanning the epochs of all of the data sets included in our global analysis.



**Figure 2.** Change in HAT-P-11’s *R* band flux and corresponding spot coverage fraction in 2011 and from late 2015 to early 2017. Points are calculated using photometric monitoring data obtained in the Cousins *R* bandpass using the AIT telescope at Fairborn Observatory, in the Johnson *B* and *V* bands using the STELLA telescope at Izaña Observatory, and from the *Kepler* telescope. We assume that the relative flux baseline for all three telescopes corresponds to a spot coverage fraction of 4.4%, and use a photospheric temperature of 4780 K and spot temperature of 4500 K to convert these observations to the equivalent *R* band fluxes. Visit times for *Spitzer* 3.6  $\mu\text{m}$  (black) and 4.5  $\mu\text{m}$  channels (black dashed), *HST* WFC3 G102 (red), STIS G430L (purple), and STIS G750L (blue) observations are indicated by vertical lines. The gray curves are sinusoidal functions that best match the observed variability at different epochs and are used to infer spot coverage fractions for *HST* visits that do not have contemporaneous ground-based monitoring.

We use the *Kepler* and ground-based photometric monitoring data to estimate the spot coverage fraction during the *Spitzer*, *HST* WFC3 G102, and *HST* STIS observation epochs. We assume that the baseline of the relative flux from each telescope corresponds to a median spot coverage fraction  $\bar{\epsilon}$  and calculate the absolute values of  $\epsilon$  for all the other relative flux values. We account for the difference in the telescope bandpasses while calculating the spot coverage fraction. In Figure 2, we show the photometric data, relative flux in *R* band, and the corresponding spot coverage fraction from *Kepler*, STELLA, and the AIT for a median spot coverage fraction  $\bar{\epsilon}$  of 4.4% and average spot temperature of 4500 K. Histograms for the inferred spot coverage fraction from the photometric data are consistent with each other and with the  $3^{+6}_{-1}\%$  estimate obtained by Morris et al. (2017b). We find that during the STIS observations, the stellar variability is best matched by a sine curve with a period of 30 days and peak-to-peak relative flux of about 1.5%. The star appears to have been somewhat less active and variable during the epoch of the WFC3 G102 observations with peak-to-peak relative flux of 0.7% and a period of 33 days. These observations imply that there is almost a 1%–2% difference in the relative transit depth between epochs due to changes in stellar brightness. These periods and variability are also in good agreement with inferences from *Kepler*.

The spot coverage fraction  $\epsilon$ , stellar photospheric temperature, and spot temperature determine the ratio of the observed ( $D_{\lambda,\text{obs}}$ ) to true ( $D_{\lambda}$ ) transit depths (Rackham et al. 2018):

$$D_{\lambda,\text{obs}} = \frac{D_{\lambda}}{1 - \epsilon(1 - F_{\lambda,\text{spots}}/F_{\lambda,\text{star}})}, \quad (4)$$

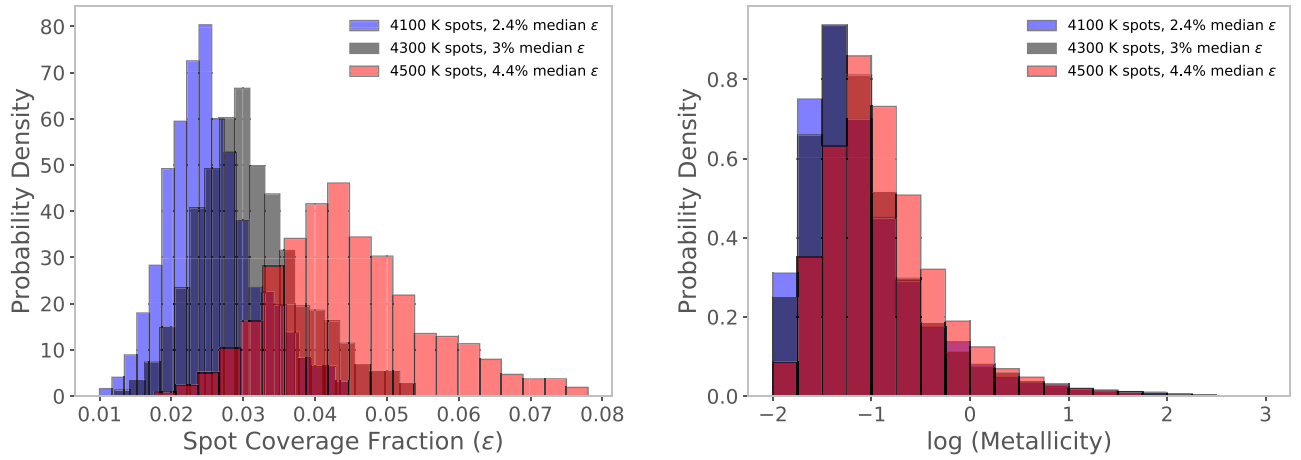
where  $F_{\lambda,\text{spots}}$  and  $F_{\lambda,\text{star}}$  are the stellar intensity profiles corresponding to the temperature of the spots and the unspotted stellar photosphere, respectively. We apply this correction by rescaling the model transit light curves at each step in our fits by the denominator in Equation (4). We do not include faculae

in our model because they produce a distinct spectral signature in the optical region of the transmission spectrum (e.g., Zhang et al. 2018b), and we observe no such effect in our three *HST* STIS visits (see Section 5).

To model the starspots and the surface fluxes, we use BT-NextGen (AGSS2009) stellar models (Allard et al. 2012) and fix the photospheric temperature to 4780 K. The brightness contrasts estimated from spot crossings in the *Kepler* light curves give a range for spot temperatures. We explore the effect of changing median spot coverage fraction  $\bar{\epsilon}$  and spot temperature on the retrieved atmospheric metallicity of the planet. We choose combinations of spot temperatures and  $\bar{\epsilon}$  such that the absolute corrections to the transit depths in the *Kepler* bandpass are identical. Spot temperatures of 4100, 4300, and 4500 K are thus combined with  $\bar{\epsilon}$  of 2.4%, 3%, and 4.4%, respectively. Figure 3 shows histograms for spot coverage fractions for the range of variability observed in the *Kepler* light curves and the corresponding atmospheric metallicity constraints for HAT-P-11b obtained from retrievals. We find that the metallicity posterior is relatively insensitive to our choice of spot temperature. Following the more detailed stellar activity study of HAT-P-11 conducted by Morris et al. (2017b) and spot temperature characterization by M18, we choose to adopt a spot temperature of 4500 K in the rest of this study.

For the *HST* WFC3 G141 data, we assume a fixed spot coverage fraction of 4.4% as this visit is not covered by any photometric observation. For the *Spitzer*, WFC3 G102, and STIS visits, we apply a visit-specific correction. We fit periodic curves to the spot coverage fraction to determine its value for the third G102 visit and the first G430L visit as ground based data at these epochs are scarce. For the other WFC3 G102 and STIS visits, we use the closest observation to obtain an estimate of the spot coverage fraction, if the next closest observation is more than 0.5 days away (i.e., on a different night). Otherwise, we use the average of the two nearest observations.





**Figure 3.** We vary starspot temperatures and spot coverage fractions such that they produce the same absolute correction in the *Kepler* bandpass. The spot coverage fractions in the left panel are deduced from *Kepler* long cadence photometry. We fit light curves for these different stellar spot properties and quantify their effect on the retrieved atmospheric metallicity. In the right panel, we show that the posterior for metallicity is relatively insensitive to our choice of spot temperature. We adopt a value of 4500 K in the rest of this study following Morris et al. (2017b).

### 5. Analysis

The log-likelihood  $\mathcal{L}$  (logarithm of the posterior probability) of our astrophysical transit model  $M$  and systematics model  $S$  given data  $D$  with uncertainty  $\sigma$  is

$$\mathcal{L} = \sum_{i=1}^n \left[ \left( \frac{D_i - (M_i \times S_i)}{\sqrt{2} \sigma_i} \right)^2 + \ln(\sqrt{2\pi} \sigma_i) \right]. \quad (5)$$

We use the Markov Chain Monte Carlo (MCMC) method to fit the white light time series for each visit individually and then carry out a joint fit where the same transit shape and ephemeris parameters are used for all data sets, while the planet-star radius ratio is allowed to vary across different bandpasses. In all cases we fit an independent instrumental systematics model for each individual transit. We carry out our fits using the `emcee` package, which is an affine-invariant ensemble sampler (Foreman-Mackey et al. 2012).

We first fit each data set individually to obtain an initial set of best-fit parameters. For these individual fits, in addition to fitting for astrophysical and systematics model parameters, we allow the measurement error  $\sigma$  to vary as a free parameter to ensure we obtain a reduced  $\chi^2$  of unity and to accurately model uncertainties in the parameters due to the intrinsic scatter in the light curves. We then use the results of these individual fits as initial guesses in the joint fit and fix the measurement error  $\sigma$  for each visit to the best-fit value obtained from its corresponding individual fit. We run an initial burn-in phase with 2000 steps for individual data sets and 40,000 steps for the global fit. We identify and discard walkers that become trapped in local minima by removing any chain whose maximum likelihood value is lower than median likelihood value of any of the other chains. We set the initial number of walkers to four times the number of free parameters and typically reject  $\lesssim 10\%$  of these walkers. Whenever an odd number of walkers remains, we randomly remove a walker from the remaining set. After burn-in, the fit is continued with the remaining walkers for another 3000 steps for individual fits and 60,000 steps for the global fit. We assume flat priors within a suitable range for each parameter. We check for convergence by inspecting the chain plots and running these fits with long chains three times. We find that the parameter estimates are consistent at the  $0.5\sigma$  level

or better and the transmission spectrum is consistent to within  $0.5\sigma$ .

We fit a total of 13 individual transits in our global analysis, each with their own instrumental systematics model. This corresponds to a total of 93 free parameters, which is too large to reliably explore with MCMC. We therefore utilize linear optimization to reduce the number of free parameters in our MCMC fit. At each step in the fit, we calculate new best-fit values for all linear parameters in the global systematics model using linear regression while keeping all other model parameters fixed to their values at that step in the MCMC. This reduces the number of free parameters in the MCMC fit to 48. We additionally fix the  $\sigma$  parameters for all visits in our global fit to the values obtained in our individual fits, which reduces the number of free parameters to 35. This is small enough to ensure reliable convergence within a reasonable number of steps. We acknowledge that in principle this approach might cause us to underestimate the uncertainties in our astrophysical model parameters, as we are optimizing rather than marginalizing over the linear instrumental model parameters (see, e.g., Benneke et al. 2019). However, we find that in practice these linear instrumental model parameters contribute negligibly to the uncertainties in our astrophysical model parameters. Optimizing the linear instrumental parameters in a global fit to the data excluding G102 light curves reduces the uncertainties in  $R_p/R_*$  by less than 5%.

#### 5.1. White Light Curve Fits

We confirm that the individual transit depths in bandpasses with multiple visits agree to within  $2\sigma$  after correcting for the effects of unocculted starspots, as discussed in Section 4.5. We therefore report the global best-fit transit depths for each band in Table 2. The best-fit transit light curves and their residuals are shown in Figures 4–6. The white light curve depths for the WFC3 G141 visit and G102 visits agree with the values reported by F14 and M18 at the  $1.5\sigma$  and  $0.6\sigma$  level respectively. Our visit-averaged  $3.6$  and  $4.5 \mu\text{m}$  *Spitzer* transit depths are in good agreement ( $1.5\sigma$  lower and  $0.5\sigma$  higher respectively) with the values obtained by F14. The residuals from our *Spitzer* fits display the predicted root- $n$  scaling expected for Gaussian noise.

**Table 2**  
Global Broadband Light Curve Fit Results<sup>a</sup>

Parameter	Instrument	Band Pass ( $\mu\text{m}$ )	Value
Planet radius, $R_p/R_*$	STIS G430L	0.29–0.57	$0.05806^{+0.00036}_{-0.00028}$
Planet radius, $R_p/R_*$	STIS G750L	0.524–1.027	$0.05783^{+0.00034}_{-0.00035}$
Planet radius, $R_p/R_*$	WFC3 G102	0.8–1.15	$0.05788^{+0.00016}_{-0.00011}$
Planet radius, $R_p/R_*$	WFC3 G141	1.1–1.7	$0.05847^{+0.00016}_{-0.00015}$
Planet radius, $R_p/R_*$	IRAC Channel 1	3.16–3.93	$0.05778^{+0.00024}_{-0.00026}$
Planet radius, $R_p/R_*$	IRAC Channel 2	3.97–5.02	$0.05811^{+0.00028}_{-0.00027}$
Transit center time $T_c$ (BJD <sub>TDB</sub> )	...	...	$2456218.866182^{+0.000042}_{-0.000044}$
Period $P$ (days)	...	...	$4.88780228^{+0.00000016}_{-0.00000018}$
Impact parameter $b$	...	...	$0.135^{+0.064}_{-0.078}$
Relative semimajor axis $a/R_*$	...	...	$17.46^{+0.14}_{-0.20}$
Inclination <sup>b</sup> $i$	...	...	$89.56^{+0.26}_{-0.22}$

**Notes.**

<sup>a</sup> The  $R_p/R_*$  values reported here have been corrected for unocculted spots assuming a photosphere temperature of 4780 K, spot temperature of 4500 K, and spot covering fraction of 4.4%.

<sup>b</sup> Calculated from posteriors for  $b$  and  $a/R_*$ .

We find that both our 3.6 and 4.5  $\mu\text{m}$  *Spitzer* transit depths are somewhat lower than our WFC3 G141 white light transit depth. The difference in white light curve depths between the WFC3 G141 observations and the *Spitzer* observations is consistent with the results reported by F14. F14 attributed this difference to stellar activity and used an offset of 93 ppm for the WFC3 spectrum to obtain their best-fit model. However, this difference cannot be explained by stellar activity for plausible starspot properties. For the *Spitzer* transit depths to be  $\gtrsim 100$  ppm higher than the *HST* measurements, HAT-P-11 would need to be 3% brighter during the *Spitzer* epochs than the *HST* ones, which is larger than the observed peak-to-peak variability of the star. For representative spot temperatures of 4500 and 4300 K, the spot coverage fraction would need to be different by  $>10\%$  and  $\sim 5\%$ , respectively, to obtain such a large relative correction to the transit depths. Finally, such a large correction to the *HST* measurements would strongly distort the transmission spectrum from 0.3 to 1.7  $\mu\text{m}$  and impart an almost unphysical upward slope (with increasing wavelength) to it. We discuss this difference in the *HST* and *Spitzer* transit depths and our efforts to interpret it in Section 7.

## 5.2. Wavelength-dependent Light Curves

When fitting for the wavelength-dependent radius ratio  $R_p/R_*$  within each *HST* STIS and WFC3 bandpass, we fix the orbital parameters  $P$ ,  $T_c$ ,  $a/R_*$ , and  $b$  to the best-fit values from the global fit. We refit the full systematics model in each individual bandpass without recourse to values obtained from the white light fit. We found that fitting the individual spectroscopic light curves with the full systematics model significantly improved the quality of the fit as compared to using the (scaled) systematics models from the global fit. For the *HST* STIS data, all parameters in the full systematics model are obtained by linear optimization and we simply use this model for the spectroscopic light curves as well. We find that fitting the individual spectroscopic light curves with the full systematics model as compared to using the (scaled) systematics models from the global fit significantly improves the quality of the fit for the WFC3 G102 data ( $\Delta\text{BIC} > 10$  for 8 out of 12 wavelength bins) but not for the WFC3 G141 data. Applying a common-mode correction to the spectroscopic light

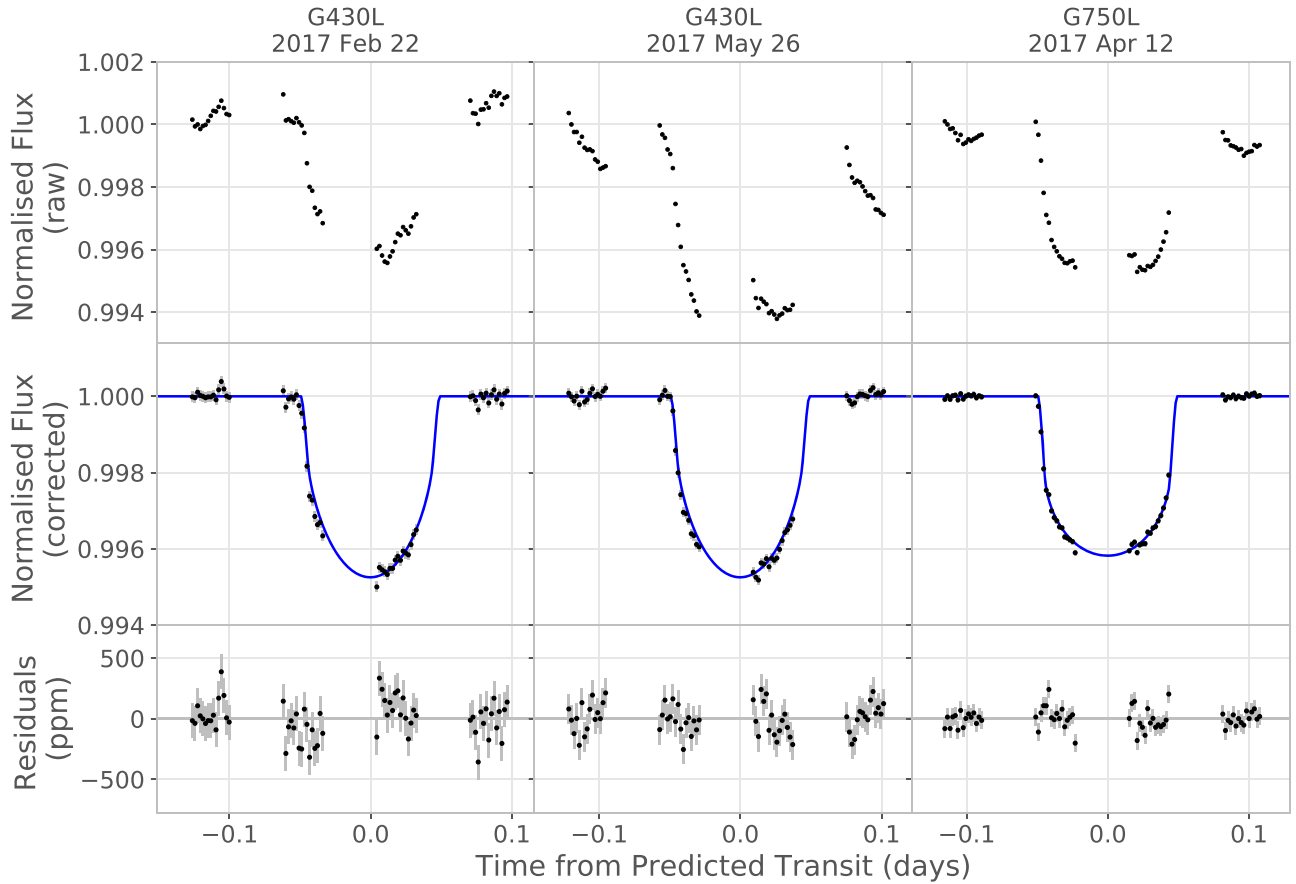
curves obtained by dividing the white light curve flux with the best-fit transit model (e.g., Deming et al. 2013) and employing a simpler model for the residual systematics in the spectroscopic light curves is strongly favored ( $\Delta\text{BIC} > 10$  for 16 out of 19 wavelength bins). Our simple model for the WFC3 G141 spectroscopic light curves is a linear function of the measured shift  $(x - x_o)$  in the dispersion direction relative to the first exposure with offset  $c$  and slope  $v$ :

$$S(t) = c + v(x - x_o). \quad (6)$$

We present  $R_p/R_*$  and associated errors for each bandpass in Table 3, the transmission spectrum in Figure 7, and show the corresponding wavelength-dependent light curves in the Appendix.

In Figure 7, we show both stellar-activity corrected and uncorrected transit depths. We obtain the uncorrected depths by fixing the orbital parameters  $b$  and  $a/R_*$  to values obtained from the global white light curve fit (shown in Table 2) and fitting the light curves without any wavelength- or epoch-dependent correction. This allows us to isolate the effect of activity correction on the transit depths. We note that activity correction is crucial for obtaining correct inferences from the optical data. The uncorrected upward slope in the STIS G430L bandpass would dramatically affect our interpretation of the planet's atmospheric properties. In addition, the magnitude of the correction is commensurate with values necessary to produce a consistent and connected spectrum across multiple bandpasses. For example, the uncorrected STIS G750L depths are fairly low compared to the STIS G430L measurements, but STIS G750L observations are taken at a time when spot coverage of the star is at a minimum and the STIS G430L measurements are obtained when the star is fairly spotted (see Figure 2). This produces a small correction for the STIS G750L measurements and a large one for the STIS G430L depths, as one would expect.

We see evidence for molecular absorption in the WFC3 G141 bandpass, in good agreement with the results from F14. Our spectrum is not as smooth as that of F14, but this is likely due to their use of a 4 pixel wide smoothing kernel (Figure 7). Our spectrum agrees within  $\sim 1\sigma$  with the previously published spectrum in almost all the wavelength bins. Stellar-activity correction introduces a slightly different slope than that of F14,



**Figure 4.** STIS white light transit light curves before (top) and after (middle) dividing out the best-fit instrumental systematics model. The best-fit transit light curve is shown in blue for comparison, and the fit residuals are shown at the bottom.

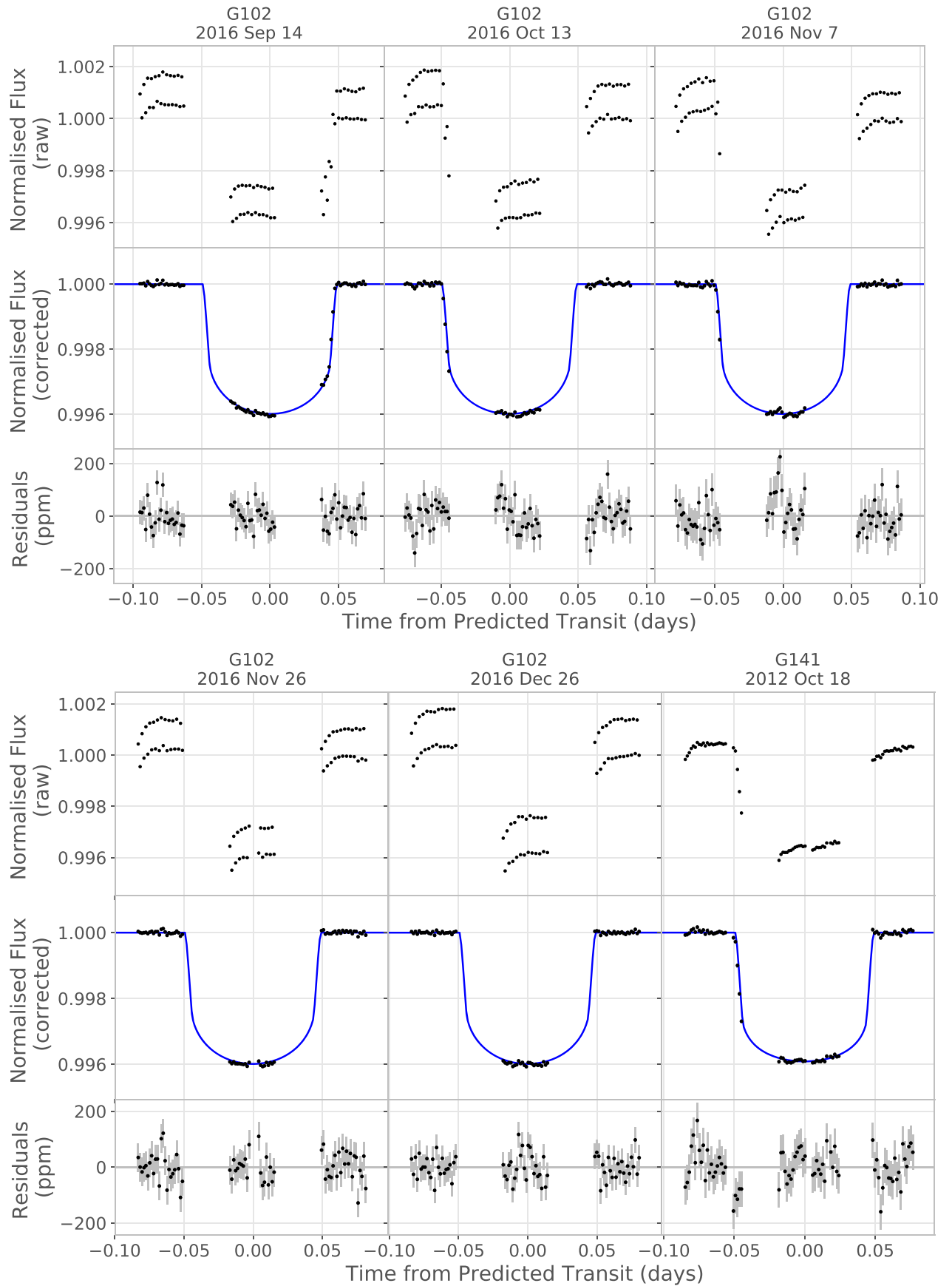
with shallower transit depths at short wavelengths and larger transit depths at longer wavelengths. Notably, our updated spectrum (both with and without correction) possesses a steeper rise longward of  $1.5\ \mu\text{m}$  compared with F14’s, suggesting the presence of methane in the planet’s atmosphere (see Section 7.2).

Our WFC3 G102 spectrum differs from the version published by M18 in subtle but significant ways (see inset, Figure 7). We diagnose the reason for this discrepancy by carrying out an additional set of fits using our models applied to the light curves from M18. We find that a majority of the observed vertical offset between the spectrum published in M18 and our fit to M18’s light curves is due to differences in the assumed values for the orbital parameters. We fit for period, while fixing impact parameter and  $a/R_*$  to the best-fit values from our global fit, and eccentricity and argument of pericenter values to the values obtained from Yee et al. (2018). In contrast, M18 fix the period and eccentricity to values from Huber et al. (2017) and use impact parameter and  $a/R_*$  values from F14 with Gaussian priors. Small differences in the stellar-activity correction were found to be insignificant. Our spectrum is not a perfect match for the one we derive using M18’s light curves. The spectral shape of our fit to M18’s light curves is intermediate to that of our spectrum and the published spectrum. This implies that although our choice of systematics model (especially the use of an additional ramp delay parameter  $d$  for the first fitted orbit) and global fitting of orbital parameters improves the agreement between our spectra, some differences must partly arise due to choices

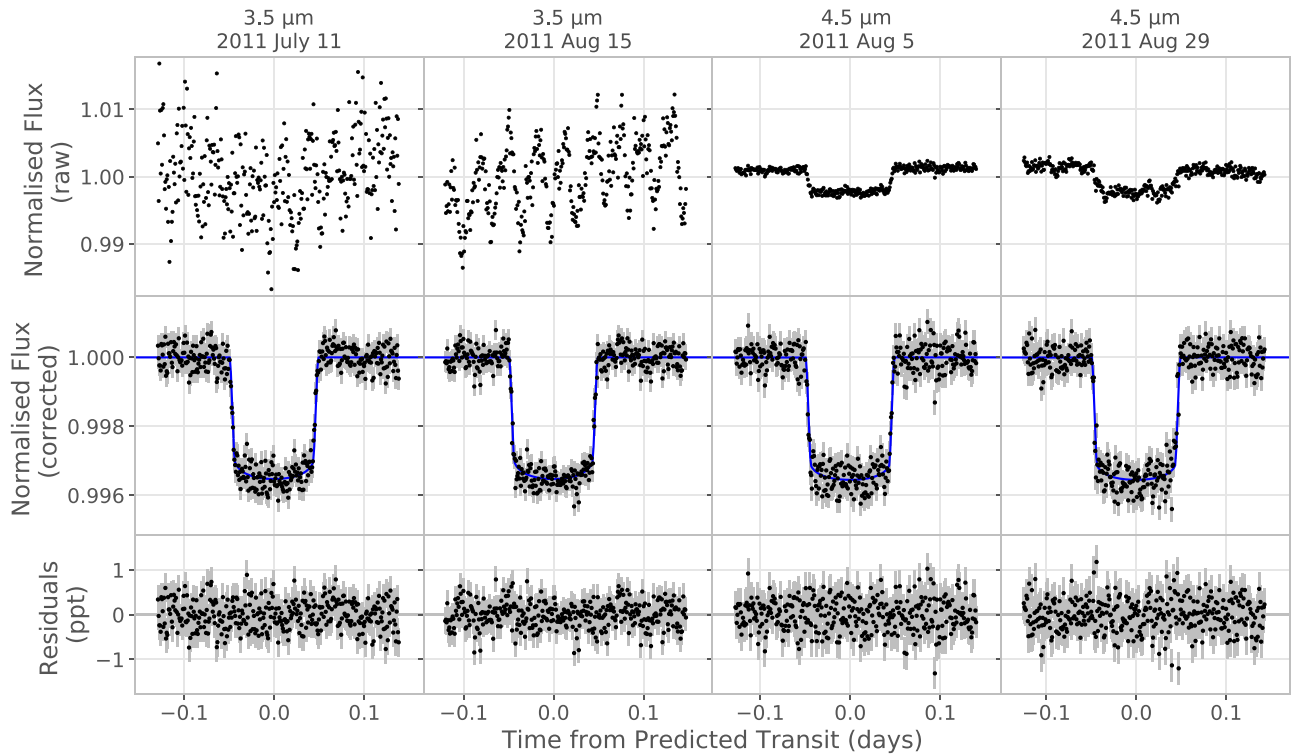
made in the light-curve extraction. In particular, there are significant differences in our light curves for the first visit, which arise due to M18’s decision to exclude the last nondestructive read (for forward scan, first read for backward) of the scan. These differences are important, as the absorption features at  $1.15$  and  $0.95\ \mu\text{m}$  are barely discernible in the spectrum published by M18. In our updated spectrum, the combination of WFC3 G102 and G141 data reveals three molecular absorption features: two strong features centered at  $1.15$  and  $1.4\ \mu\text{m}$  and a weak feature at  $0.95\ \mu\text{m}$  (Figure 7). This allows us to infer the presence of water and/or methane with a combined significance of  $4.4\sigma$  (see Section 7.2).

Our new STIS observations indicate that HAT-P-11b has a relatively featureless transmission spectrum at optical wavelengths with a hint of increasing transit depth with decreasing wavelength (scattering slope). This is in agreement with recently reported measurements obtained from ground-based observations (Murgas et al. 2019). As mentioned above, a careful accounting for the effects of unocculted spots produces a much flatter optical transmission spectrum than the uncorrected version. This plays an important role in constraining atmospheric metallicity and places constraints on the effective size and number density of the particles responsible for scattering in the atmosphere. We see no evidence for narrow-band sodium or potassium absorption, although these features are expected to form at relatively low pressures where cloud opacity should be less important. This is not surprising, as HAT-P-11b’s atmosphere is predicted to be too cold for these elements to remain in vapor form (e.g., Lodders 1999).





**Figure 5.** WFC3 G102 and G141 white light transit light curves before (top) and after (middle) dividing out the best-fit instrumental systematics model. The best-fit transit light curve is shown in blue for comparison, and the fit residuals are shown at the bottom.



**Figure 6.** *Spitzer* transit light curves before (top) and after (middle) dividing out the best-fit instrumental systematics model. The best-fit transit light curve is shown in blue for comparison, and the fit residuals are shown at the bottom.

Additionally, we do not see the jump in transit depth at  $0.8\ \mu\text{m}$  that Lothringer et al. (2018) report for GJ 436b and note for HAT-P-26b.

## 6. Comparison to Forward Models

We next compare HAT-P-11b’s observed transmission spectrum to predictions from a 1D microphysical cloud model originally developed for use with solar system planets (e.g., Toon et al. 1979, 1992; James et al. 1997; Colaprete et al. 1999; Gao et al. 2017). These cloud models require a temperature-pressure profile and a prescription for the vertical mixing in the atmosphere as inputs. We draw both of these profiles from results of a 3D general circulation model (GCM) for HAT-P-11b. We discuss the details of both models in the following two subsections.

### 6.1. General Circulation Model

We use a GCM to put constraints on the extent of (1D) mixing in the atmosphere. This allows us to take into account the effect of three-dimensional (3D) dynamics on the 1D atmospheric profiles used in transmission spectroscopy studies. This is particularly important for eccentric short-period planets like HAT-P-11b, which are presumed to be tidally locked and therefore may have a pressure and temperature structure that varies significantly with longitude. The appreciable eccentricity of HAT-P-11b also leads to the convolution of latitudinal structure and orbital phase of the planet. We take the planet’s eccentricity into account in our GCM and use atmospheric profiles (for temperature, pressure, eddy diffusion coefficient) from the planet’s transit. In this case, we utilize the Substellar and Planetary Radiation and Circulation (SPARC) model (Showman et al. 2009; Kataria et al. 2016), which couples the MITgcm dynamical core (Adcroft et al. 2004)

with a plane-parallel, two-stream version of the multi-stream radiation code developed by Marley & McKay (1999). As we will discuss in Section 7, our retrievals using the *HST* data prefer relatively low-metallicity values, so we choose models with atmospheric metallicities of  $1\times$  and  $50\times$  solar (we multiply relative abundances of elements heavier than hydrogen and helium by this metallicity value and renormalize the sum of relative abundances to 1); this range is therefore a good match for the posterior probability distribution for this parameter.

We model vertical mixing as a diffusive process with an effective eddy diffusion coefficient  $K_{zz}$ . Deviations from this diffusive approximation are almost guaranteed for tidally locked planets, which are expected to also have vigorous horizontal transport between the day and night sides (e.g., Zhang & Showman 2018a, 2018b). However, it is nontrivial to accurately capture this horizontal transport, and we therefore neglect it for the moment in order to explore the effects of vertical mixing, which is key for cloud formation. This mixing is typically parameterized as a constant value with or without an inverse dependence on square root of pressure (e.g., Parmentier 2013). We depart from this formalism and instead use the temperature, pressure, and  $K_{zz}$  profiles from the GCM, which should be more representative of the relevant conditions in HAT-P-11b’s atmosphere. We use the GCM results to calculate 1D pressure–temperature profiles that are spatially averaged over the east and west limbs of the planet. We estimate the corresponding pressure/height-dependent  $K_{zz}$  values for these locations using mixing length theory:

$$K_{zz} = w(z)L(z) = \frac{\omega H^2}{P} \quad (7)$$

where  $w(z)$  is the vertical velocity in  $\text{m s}^{-1}$  and  $L(z)$  is a characteristic length scale, in this case the atmospheric pressure

**Table 3**  
Spectroscopic Light Curve Fit Results

Wavelength ( $\mu\text{m}$ )	$R_p/R_*$	$\pm 1\sigma$
STIS G430L		
0.346–0.401	0.05788	0.00117
0.401–0.456	0.05821	0.00045
0.456–0.511	0.05828	0.00031
0.511–0.566	0.05812	0.00029
STIS G750L		
0.528–0.577	0.05903	0.00086
0.577–0.626	0.05719	0.00068
0.626–0.674	0.05787	0.00070
0.674–0.723	0.05766	0.00073
0.723–0.772	0.05587	0.00109
0.772–0.821	0.05763	0.00084
0.821–0.870	0.05789	0.00116
0.870–0.919	0.05732	0.00129
0.919–0.967	0.05597	0.00159
0.967–1.016	0.05687	0.00210
0.589–0.591*	0.06244	0.00361
0.766–0.773*	0.05689	0.00192
WFC3 G102		
0.850–0.873	0.05812	0.00019
0.873–0.897	0.05778	0.00016
0.897–0.920	0.05782	0.00015
0.920–0.943	0.05795	0.00014
0.943–0.967	0.05807	0.00013
0.967–0.990	0.05810	0.00013
0.990–1.013	0.05805	0.00013
1.013–1.037	0.05784	0.00011
1.037–1.060	0.05811	0.00013
1.060–1.083	0.05787	0.00012
1.083–1.107	0.05811	0.00012
1.107–1.130	0.05831	0.00012
WFC3 G141		
1.120–1.150	0.05899	0.00044
1.150–1.180	0.05896	0.00041
1.180–1.210	0.05825	0.00028
1.210–1.240	0.05740	0.00033
1.240–1.270	0.05726	0.00031
1.270–1.300	0.05842	0.00023
1.300–1.330	0.05803	0.00023
1.330–1.360	0.05914	0.00030
1.360–1.390	0.05867	0.00031
1.390–1.420	0.05909	0.00030
1.420–1.450	0.05941	0.00031
1.450–1.480	0.05933	0.00030
1.480–1.510	0.05751	0.00029
1.510–1.540	0.05878	0.00027
1.540–1.570	0.05846	0.00030
1.570–1.600	0.05827	0.00036
1.600–1.630	0.05889	0.00030
1.630–1.660	0.05950	0.00037
1.660–1.690	0.05823	0.00102

(This table is available in machine-readable form.)

scale height. This commonly adopted method (e.g., Moses et al. 2011) gives us a height-dependent  $K_{zz}$  value which we then use in our microphysical cloud models. We show the resulting  $K_{zz}$  and temperature profiles as a function of pressure for the limb average, eastern limb average, and western limb average in Figure 8. As shown in previous GCM studies exploring the effect of atmospheric metallicity (e.g., Lewis et al. 2010; Kataria et al. 2014), the higher metallicity profile of

HAT-P-11b has a higher photosphere due to the higher opacity, which produces a  $K_{zz}$  profile that rises more rapidly with height than the lower metallicity model.

## 6.2. Microphysical Cloud Model

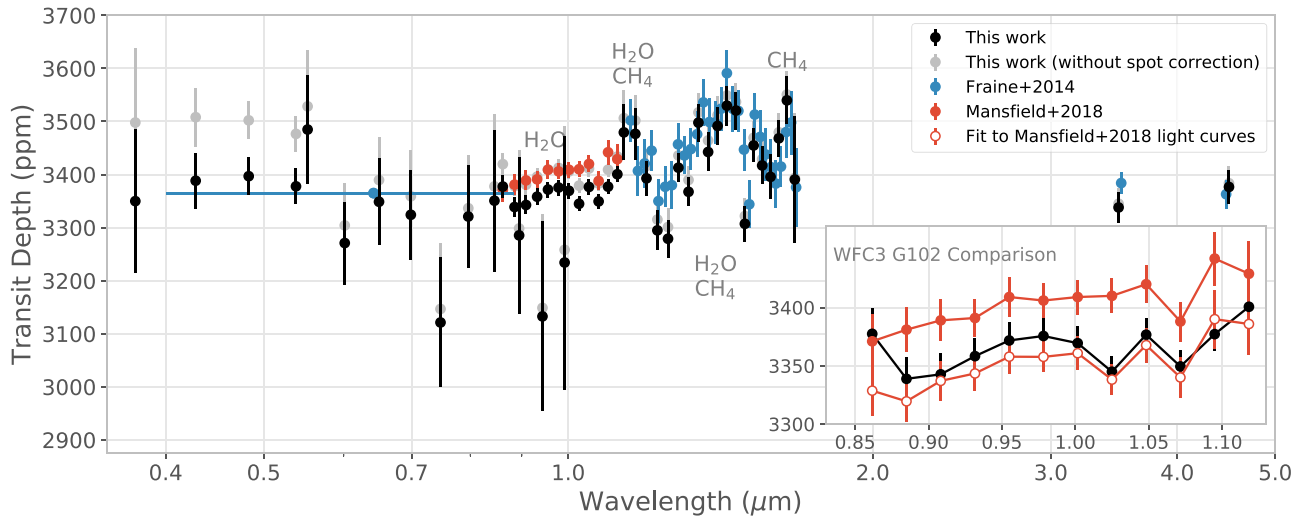
We use the Community Aerosol and Radiation Model for Atmospheres (CARMA) to determine which species are expected to condense in HAT-P-11b’s atmosphere and the corresponding particle-size distribution and abundance. CARMA is a bin-scheme cloud microphysics model that considers microphysical processes such as nucleation, evaporation, condensation, sedimentation, and diffusion. The strength of bin-scheme microphysics is that it uses discrete bins for particle sizes and makes no prior assumption regarding the size distribution, instead allowing the different bins to “interact” (i.e., exchange mass) via the aforementioned microphysical processes. For a thorough exposition of the model, we direct the reader to Gao et al. (2018) and Powell et al. (2018).

We include the following condensable species in our model: Cr, KCl,  $\text{Al}_2\text{O}_3$ ,  $\text{Mg}_2\text{SiO}_4$ , Fe, and  $\text{TiO}_2$ . We also consider condensation of metal sulphides but find it to be unimportant.  $\text{Na}_2\text{S}$ , MnS, and ZnS have high nucleation energy barriers that inhibit the formation of these cloud species (P. Gao et al. 2019, private communication). Another reason ZnS clouds can be neglected is the low abundance of Zn. We assume that KCl, Cr,  $\text{TiO}_2$ , and  $\text{Al}_2\text{O}_3$  can nucleate homogeneously, meaning that they can condense into stable clusters directly from the gas phase and subsequently grow to larger sizes. In contrast, heterogeneous nucleation requires a foreign surface or “seed” onto which vapor can condense. Though the majority of  $\text{Al}_2\text{O}_3$  condensates likely form via heterogeneous surface reactions (e.g., Helling 2019), assuming homogeneous nucleation is unlikely to greatly affect our results, as  $\text{Al}_2\text{O}_3$  condenses at much higher temperatures ( $\sim 2000$  K) than considered here.  $\text{Al}_2\text{O}_3$  is present in small concentrations at the high altitudes that we probe (Figure 9), but its distribution in this region is primarily controlled by transport processes rather than condensation and nucleation (Gao & Benneke 2018). We assume that Fe and  $\text{Mg}_2\text{SiO}_4$  nucleate heterogeneously on  $\text{TiO}_2$  particles, similar to the treatment of Helling (2019) and related works. Although Fe can nucleate homogeneously as well, we do not consider it as this process may not be efficient (Lee et al. 2018).

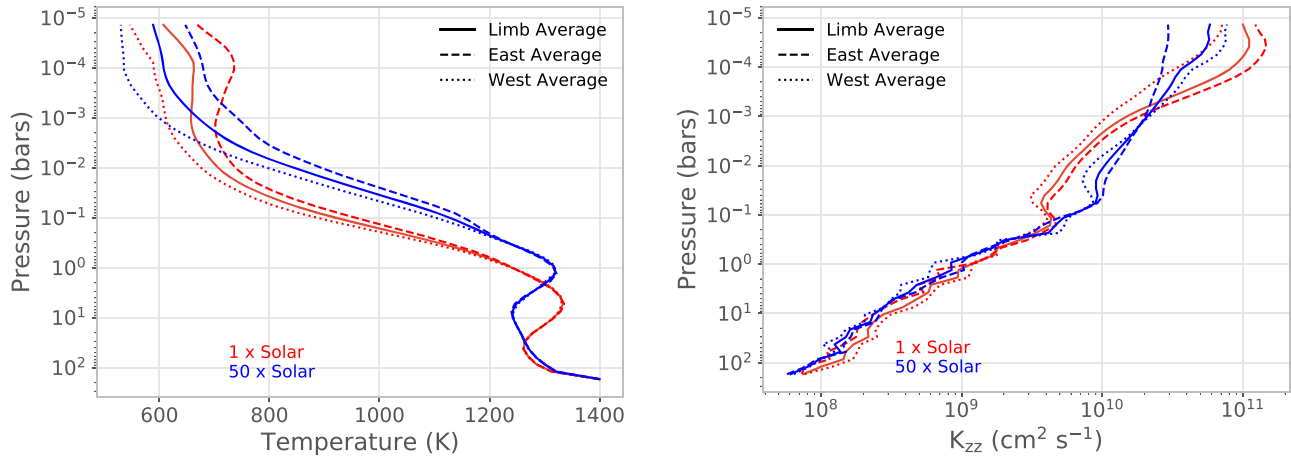
We model the east and west limbs separately, as well as a limb-averaged profile, ( $T$  and  $K_{zz}$ ) for both solar and  $50\times$  solar metallicity atmospheres. We neglect the effect of radiative feedback from condensation and cloud formation on the atmosphere’s  $T$ – $P$  profile. The resulting particle sizes and number densities of the dominant condensate species are shown in Figure 9 as a 2D visualization of a slice of the atmosphere at a pressure of  $\sim 2$  mbar ( $\tau \sim 1$  for transmission spectroscopy) with a path length of 100 cm through the atmosphere. In addition, the area covered by the different condensate species is proportional to the geometric cross section due to each species, thereby visually indicating which species dominate the cloud opacity.

It is immediately evident that for both metallicity cases, the east and west limb-averaged profiles display distinct cloud properties and are dominated by different condensate species. This is primarily due to the temperature difference between the two limbs, which can be as large as 100–200 K (see Figure 8). Most notably, the west limb is cool enough for KCl to condense and contribute dominantly to the opacity whereas the





**Figure 7.** The transmission spectrum of HAT-P-11b both with and without stellar-activity correction. Our transmission spectrum is in good agreement with F14’s published spectrum. In the inset figure, we compare our WFC3 G102 spectrum with a fit to M18’s light curves, as well as M18’s published spectra. Our G102 spectrum deviates most significantly from the published spectrum at 0.86, 1.025, and 1.095  $\mu\text{m}$ , which has the effect of washing out the small absorption feature at 0.95  $\mu\text{m}$  in the published version.



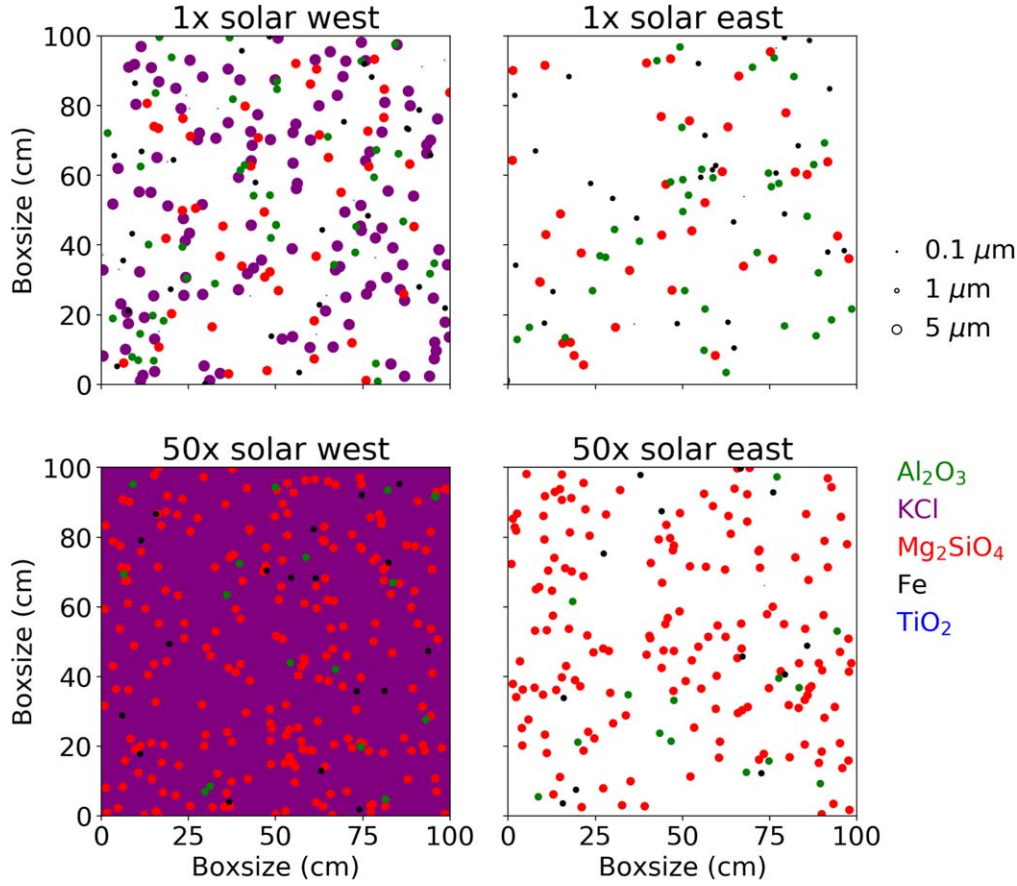
**Figure 8.** Temperature (left) and vertical mixing parameter  $K_{zz}$  (right) profiles as a function of pressure at the orbital phase of the transit (since HAT-P-11b has a significant eccentricity). These profiles are obtained from a SPARC GCM model for HAT-P-11b and are used as inputs in our microphysical cloud models. Transmission spectroscopy probes the atmosphere at pressures roughly between  $10^{-1}$  and  $10^{-4}$  bars.

east limb is completely devoid of condensed KCl. The lower temperature of the west limb also causes more nucleation sites to form, additionally increasing the cloud opacity in this region. The east limb has a significantly lower condensate number density ( $<100 \text{ m}^{-3}$ ) and consists of species that have cloud bases deep in the atmosphere but are carried to pressures probed by transmission spectroscopy by strong vertical mixing (Figure 8). These differences result in distinct predictions for the mid-IR spectra of the two limbs, and suggest that cloud models utilizing the limb-averaged pressure–temperature profile may not produce accurate predictions (e.g., Kempton et al. 2017). Using the average of the transmission spectra rather than the average of the pressure–temperature profile for the two limbs should allow a better comparison of the models with the data. We therefore compare our retrieval results with model transmission spectra generated by averaging the spectra from the east and west limbs.

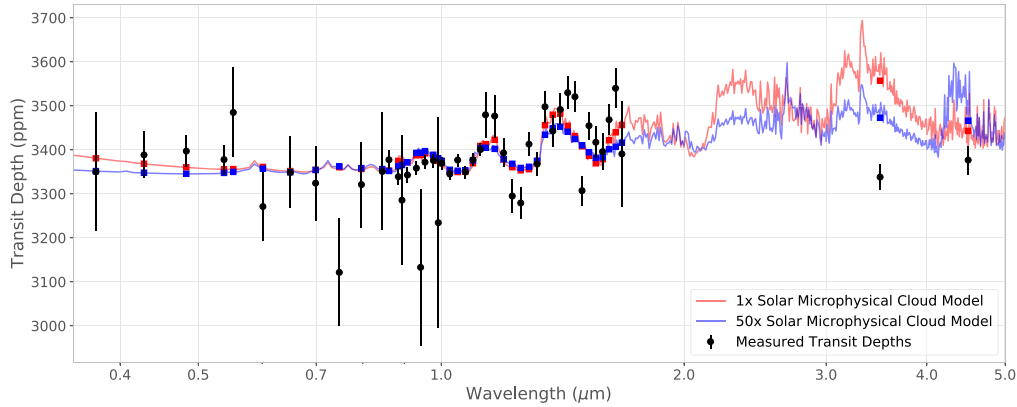
Increasing the metallicity from  $1\times$  solar to  $50\times$  solar increases the abundance of condensates by 1–2 orders of

magnitude. Although the rate of homogeneous nucleation increases when the metallicity increases, the particle sizes tend to be somewhat smaller because there is less gas (per nucleated site) to provide additional condensable material for the growing particle. KCl overwhelms the absorption cross section on the west limb while the east limb is much clearer.

Figure 10 shows transmission spectra generated using CARMA models. The models provide a good match to the observed absorption features at 0.95, 1.15, and 1.4  $\mu\text{m}$  while maintaining a relatively flat optical spectrum without any fine-tuning or fitting. We find that the  $1\times$  solar metallicity atmosphere is a slightly better match for the observed amplitude of the molecular absorption bands and optical scattering between 0.3 and 1.7  $\mu\text{m}$  than the  $50\times$  solar metallicity model (reduced  $\chi^2$  of 1.8 and 2.1, respectively). However, both of these models predict strong methane absorption in the 3.6  $\mu\text{m}$  *Spitzer* band, making them a relatively poor match to the observed transit depth in this band.



**Figure 9.** Plot windows showing 2D slices of the atmospheric condensate compositions for a  $1\times$  solar and  $50\times$  solar metallicity atmosphere. The slices sample the atmosphere on the east and west limbs at  $\tau \sim 1$  and show the number of condensate particles contained in a  $100\text{ cm} \times 100\text{ cm} \times 100\text{ cm}$  volume. Condensates on the two limbs have distinct compositions and increasing the metallicity has a significant effect on condensate number density, especially on the west limb. These plots serve as a visual guide and indicate that the scattering cross section at the wavelengths of interest is mostly dominated by KCl particles.  $\text{Mg}_2\text{SiO}_4$  and  $\text{Al}_2\text{O}_3$  particles also make significant contributions to cloud opacity, especially in the  $1\times$  solar metallicity case.



**Figure 10.** Measured transmission spectrum of HAT-P-11b vs. transmission spectra generated by averaging CARMA models for the east and west limbs. These model spectra fit the measured spectrum quite well without any fine-tuning or parameter fitting. The *HST* data display a slight preference for the  $1\times$  solar metallicity model. However, both the  $1\times$  and  $50\times$  solar metallicity models are unable to reproduce the *Spitzer* transit depths.

## 7. Atmospheric Retrieval: PLATON

We use a simple and highly customizable atmospheric retrieval model, PLATON<sup>16</sup> (Zhang et al. 2019) to constrain HAT-P-11b’s atmospheric properties using its transmission spectrum. PLATON is based on ExoTransmit (Kempton et al. 2017) and uses a fast

Python-based algorithm to compute forward models for planetary atmospheres, which are then compared with the data in a retrieval framework. PLATON includes opacities for 30 different molecular and atomic species (M. Zhang et al. 2019, in preparation), the majority of which are calculated using line lists from ExoMol (Tennyson et al. 2018) and HITRAN (Gordon et al. 2017). We use nested sampling for our retrievals to accurately capture the posteriors of atmospheric model parameters that may display multimodality. More importantly, using nested sampling allows us

<sup>16</sup> Planetary Transmission Atmosphere Tool for Observer Noobs: <https://github.com/ideasrule/platon>.

**Table 4**  
Median Parameters and 68% Confidence Intervals (CI) from PLATON Retrieval

Parameter	Prior	<i>HST</i> WFC3		<i>HST</i> WFC3 + STIS <sup>a</sup>		<i>HST</i> + <i>Spitzer</i>	
		Median	68% CI	Median	68% CI	Median	68% CI
Isothermal Temperature (K)	[500, 1200]	941	[726, 1114]	740	[635, 876]	736	[540, 1026]
log (Metallicity/ $Z_{\odot}$ )	[-2, 3]	-1.39	[-1.79, -0.16]	-0.98	[-1.40, -0.36]	2.04	[0.12, 2.75]
C/O	[0.2, 2]	1.03	[0.62, 1.56]	0.97	[0.51, 1.56]	0.63	[0.30, 1.49]
log (Cloudtop Pressure/Pa)	[1, 6]	4.71	[3.96, 5.46]	4.25	[3.67, 4.88]	2.94	[2.02, 4.77]
log (Particle Size/m)	[-8, -5]	-6.69	[-7.61, -5.62]	-6.67	[-7.59, -5.65]	-6.60	[-7.58, -5.60]
log (Number Density/ $\text{m}^{-3}$ )	[-10, 15]	-1.70	[-7.11, 3.92]	-1.70	[-7.09, 3.67]	-0.59	[-6.65, 5.42]
Error Multiple ( $\sigma_{\text{mult}}$ )	[0, 4]	1.46	[1.29, 1.69]	1.32	[1.19, 1.48]	1.67	[1.51, 1.87]

**Note.**

<sup>a</sup> We regard this to be the most reliable retrieval. See Sections 7.2 and 8.

to compare the Bayesian evidence for different retrievals and rigorously quantify the significance of molecular absorption detection.

We fit for HAT-P-11b’s atmospheric properties assuming an isothermal atmosphere in chemical equilibrium. We allow the planet radius  $R_p$ , temperature  $T$ , atmospheric metallicity  $\log(Z)$ , and the carbon-to-oxygen ratio C/O to vary as free parameters in our fit. We also include scattering from high-altitude clouds, which we discuss in the following section. All of these parameters have flat priors. For  $R_p$  and  $T$  we choose physically motivated lower and upper bounds, while our prior range for metallicity and C/O ratio is dictated by limitations in our model’s precomputed equilibrium chemistry grid (see Table 4). Our grid limits us to  $\log(Z) \geq -1$ , but we linearly extrapolate (in  $Z$ ) abundances of atoms and molecules containing elements heavier than hydrogen and helium to lower metallicities (down to  $\log(Z) = -2$ ) to resolve the posterior distribution on the lower metallicity end. We verify that linear extrapolation in  $Z$  captures the atmospheric composition reasonably well by comparing transmission spectra obtained for atmospheric metallicities between  $0.1\times$  and  $1\times$  solar from extrapolation and from the precomputed abundance grid. We include the stellar radius ( $0.683 \pm 0.009 R_{\odot}$ ; Deming et al. 2011) and planetary mass ( $23.4 \pm 1.5 M_{\oplus}$ ; Yee et al. 2018) as free parameters in our model with Gaussian priors set to the published values. This ensures that we correctly account for the effects of these uncertainties in our model fits. We also include an additional parameter (“Error Multiple”  $\sigma_{\text{mult}}$ , same for all instruments) that multiplies the errors on the data with a constant factor to account for the errors’ under- or overestimation.

### 7.1. Scattering from Clouds

We model scattering particles with five parameters: a cloud-top pressure ( $P_{\text{cloud}}$ ) below which the atmosphere is opaque at all wavelengths (top of a gray cloud), particle number density  $n_0$  at  $P_{\text{cloud}}$ , a lognormal distribution of particle sizes centered on an effective particle size  $a$  with distribution width  $\sigma$ , and the scale height for particle number density as a fraction  $f$  of the gas scale height  $H_{\text{gas}}$ . This allows for a deep gray cloud that begins to thin as the pressure decreases. Alternatively, it can be interpreted as a haze layer lying on top of a gray cloud. The particle-size distribution  $p(r)$  and number density as a function

of height  $n(z)$  are given by:

$$p(r) = \frac{1}{\sqrt{2\pi}\sigma r} \exp\left[-\frac{(\ln r - \ln a)^2}{2\sigma^2}\right], \quad (8)$$

$$n(z) = n_0 \exp[-z/(fH_{\text{gas}})]. \quad (9)$$

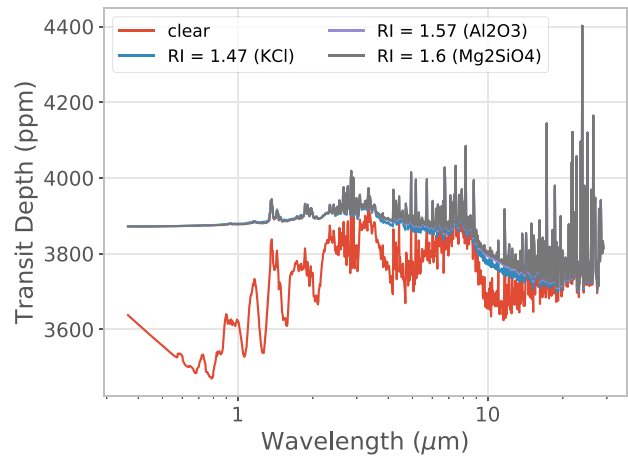
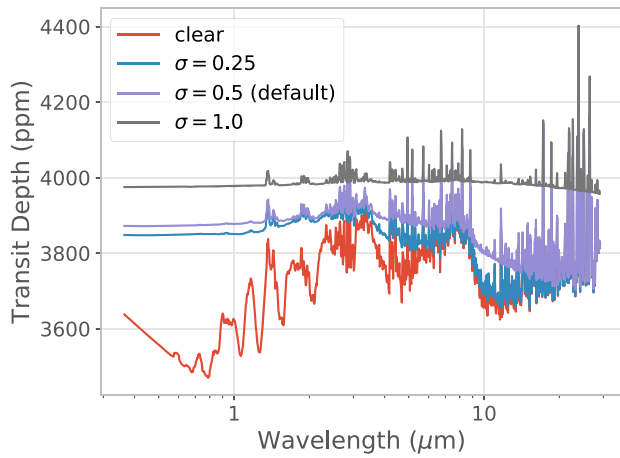
The extinction cross section,  $\sigma_{\text{ext}}$ , from condensate particles is then given as:

$$\sigma_{\text{ext}}(\lambda, z) = n_0 e^{-z/fH_{\text{gas}}} \int p(r) Q_{\text{ext}}(\lambda, r) \pi r^2 dr. \quad (10)$$

We calculate  $Q_{\text{ext}}$ , which depends on the refractive index, using the Mie scattering formalism. The effective particle size  $a$ , number density  $n_0$ , and relative scale height  $f$  play a decisive role in shaping the planetary transmission spectrum. The effective particle size  $a$  determines the wavelength where the Rayleigh slope begins ( $\lambda \sim 2\pi a$ ). The number density  $n_0$  and fractional scale height  $f$  set the overall scale of the opacity contribution from scattering (relative to molecular absorption opacity) and are partially degenerate with each other. We find that  $f$  is almost entirely unconstrained by our data and allowing it to vary in our retrievals does not have any significant effect on the posteriors for the other parameters in our model. We therefore turn to our microphysical cloud models for HAT-P-11b, which indicate the effective particle size is roughly constant in the pressure range 0.1–100 mbar and that the effective number density falls off with the pressure scale height  $H_{\text{gas}}$ . We fix  $f = 1$  unless otherwise specified in order to reduce the number of free parameters and to allow for a more direct comparison with predictions from our microphysical models.

We also keep the value of the refractive index fixed to a single, wavelength-independent value in our fits. Our microphysical cloud models predict that condensate clouds in HAT-P-11b’s atmosphere will include multiple distinct species. However, the refractive indices for all these species apart from Fe have a very weak dependence on wavelength and negligible imaginary parts in the 0.1–5  $\mu\text{m}$  region spanned by our data (e.g., see Kitzmann & Heng 2018). Adopting a wavelength-independent real value for the refractive index also speeds up our model computations enormously, which is a necessary requirement for retrieval codes. Figure 11 shows that the shape of the predicted transmission spectrum is relatively insensitive





**Figure 11.** Effect of varying  $\sigma$  and refractive index in our Mie scattering model. We assume a particle size  $a = 1 \mu\text{m}$ , refractive index of 1.5, fractional scale height  $f = 1$ , particle-size distribution width  $\sigma = 0.5$ , and number density at the base of the atmosphere  $n_0 = 10^4 \text{ cm}^{-3}$  unless specified otherwise.

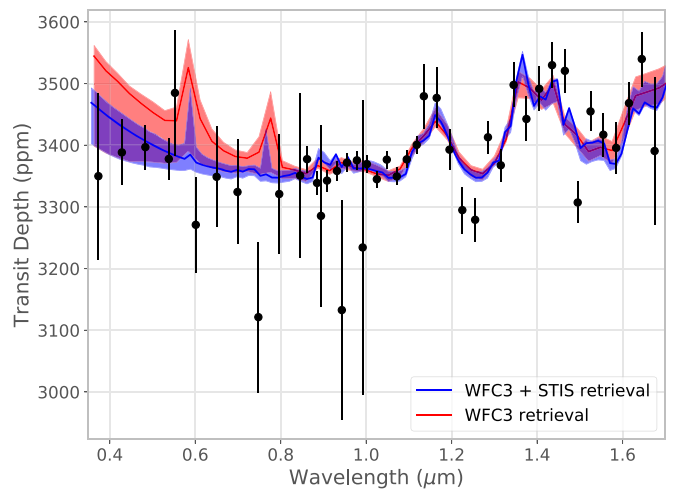
to the exact value we assume for the refractive index in our wavelength range of interest. We set this parameter equal to 1.5, as this is fairly representative of the dominant cloud species (KCl) predicted by our forward models.

Although the particle-size distribution can take an arbitrary functional form, the distribution of large particles that are abundant enough to contribute most significantly to scattering may be captured by a lognormal distribution. We keep the width of the lognormal distribution fixed in our fits. Varying this parameter mimics the effect of increasing particle size as a broader distribution shifts the effective size of the particles to larger values and large particles tend to dominate the cloud opacity (e.g., Wakeford & Sing 2015). Therefore, variations in the distribution width are degenerate with changes in particle-size distributions. Increasing the distribution width makes the spectrum flatter in a given wavelength range, as does increasing the effective particle size (see Figure 11). We fix  $\sigma = 0.5$ , which agrees well with typical values for aerosols in the Earth’s atmosphere (e.g., Pinnick et al. 1978; Ackerman & Marley 2001; Elias et al. 2009; Shen et al. 2015) and produces a scattering behavior that is roughly compatible with that produced by the CARMA model with its nonparameterized particle-size distribution.

## 7.2. Retrieval Results

### 7.2.1. HST WFC3

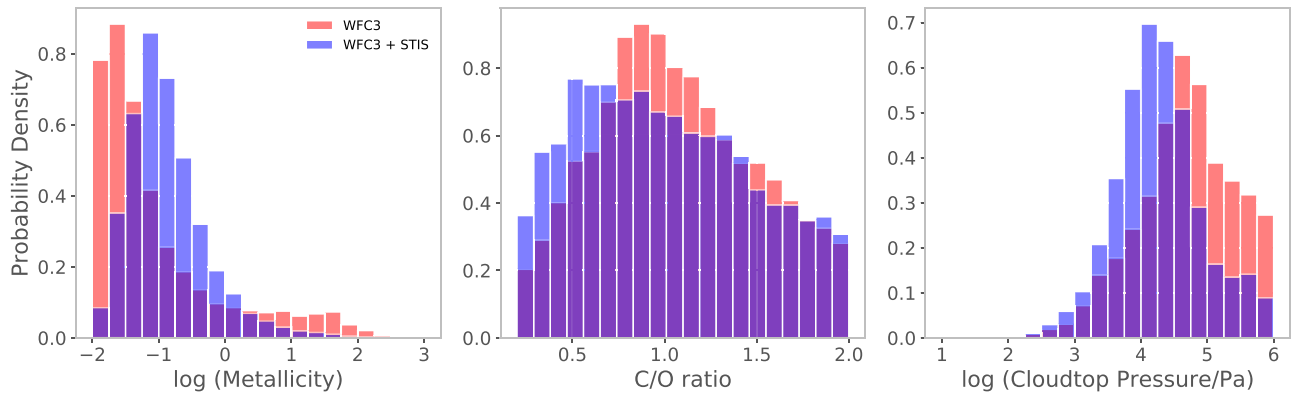
We begin by fitting the molecular absorption features in the WFC3 G102 and G141 bandpasses, as these features provide the strongest constraints on the planet’s atmospheric composition. Because these data span a relatively limited wavelength range, a simplified cloud model with a single opaque cloud deck is adequate. Nonetheless, we “fit” for Mie scattering parameters for later comparison of best-fit models with models that match the entire *HST* transmission spectrum. We fit for temperature, atmospheric metallicity, and C/O ratio as well, assuming chemical equilibrium. The resulting best-fit model is shown in Figure 12 and the corresponding constraints on the model parameters are given in Table 4. The steep rise in transit depth longward of  $1.5 \mu\text{m}$  hints at the presence of methane in the atmosphere. We verify this by confirming that this upward rise disappears if methane is removed from our atmospheric models.



**Figure 12.** Transmission spectrum in the *HST* WFC3 and STIS bandpasses (black filled circles) with best-fit Mie scattering model spectra from PLATON overlaid along with the  $1\sigma$  contours.

We find that HAT-P-11b’s atmospheric parameters, in particular its metallicity, are poorly constrained in these fits (see Figure 13 and Table 4). The limited wavelength range of the WFC3 data limits our ability to uniquely infer the metallicity and cloud-top pressure. As for the C/O ratio, the presence of absorption features due to water does not automatically imply a C/O ratio  $< 0.9$  for planets with equilibrium temperatures  $\lesssim 800\text{--}1000 \text{ K}$  as it does for hot Jupiters<sup>17</sup> (Madhusudhan 2012; Kreidberg et al. 2015; Heng 2018). Below  $\sim 800 \text{ K}$ , methane is the thermodynamically favored carbon-bearing species in hot Neptunes, except at very high atmospheric metallicities (Moses et al. 2013). Adding more carbon relative to oxygen does not therefore increase the abundance of CO at the expense of water. Our models indicate that increasing the C/O ratio (even to values greater than one) at temperatures below  $800 \text{ K}$  has a negligible effect on the water abundance and the methane abundance simply increases linearly with C/O.

<sup>17</sup> The exact transition temperature depends on other properties such as atmospheric metallicity and surface gravity.



**Figure 13.** Marginalized posterior probability distributions for the metallicity, C/O ratio, and cloud-top pressure from a fit to the WFC3 data alone and a fit to the WFC3 + STIS data set.

The results from this retrieval differ significantly from those presented in F14 primarily for three reasons. First, we include WFC3 G102 data here that have small uncertainties and consequently a strong influence on the retrieved posteriors. The addition of WFC3 G102 data shifts the peak of the metallicity posteriors to lower values. When we utilize only the WFC3 G141 data (or WFC3 G141 + *Spitzer* data with an offset for the *Spitzer* data), our retrieved results agree with F14’s. Second, we apply a wavelength-dependent stellar-activity correction that changes the spectrum in such a way that a low-metallicity–deep-cloud solution fits the data. To test whether this shift to low metallicity is due to our stellar-activity correction, we combined the WFC3 G141 spectrum from F14 and our WFC3 G102 spectrum and performed retrieval analysis on the corrected and uncorrected version of the combined spectrum. We found that applying the stellar-activity correction shifts the posteriors to low metallicity. Third, we choose a different prior for atmospheric metallicity and extend it to  $0.01 \times$  solar so as to resolve the posterior for the retrieved metallicity. F14 only explored atmospheric metallicities  $\geq 1 \times$  solar in their retrievals and we find that restricting our prior space to match their results in significantly better agreement. Additionally, our models do not favor atmospheric metallicities  $\gtrsim 100 \times$  solar primarily because our spectrum, unlike the one published in F14, favors the presence of methane in the atmosphere (see Section 7.2.4 for more details).

### 7.2.2. HST WFC3 + STIS

Next, we see how the inclusion of STIS data alters the posteriors for these parameters. Because our data now span a much larger wavelength range, we must include wavelength-dependent scattering in our model (Section 7.1). The best-fit model is shown in Figure 12, parameter constraints are tabulated in Table 4, and the full posteriors for key atmospheric parameters are shown in Figure 14. The data place relatively tight constraints on the cloud-top pressure, indicating that we are probing down to  $\sim 100$  mbar. This is in rough agreement with the inferred (gray) cloud-top pressures of 10–50 mbar for CARMA models. The constraints on atmospheric metallicity are significantly tighter than those provided by WFC3 data alone, with a  $2\sigma$  confidence interval of  $0.02\text{--}4.6 \times$  solar. The posterior for atmospheric metallicity has a skewed shape with a long tail toward high metallicities. We find that the  $3\sigma$  upper

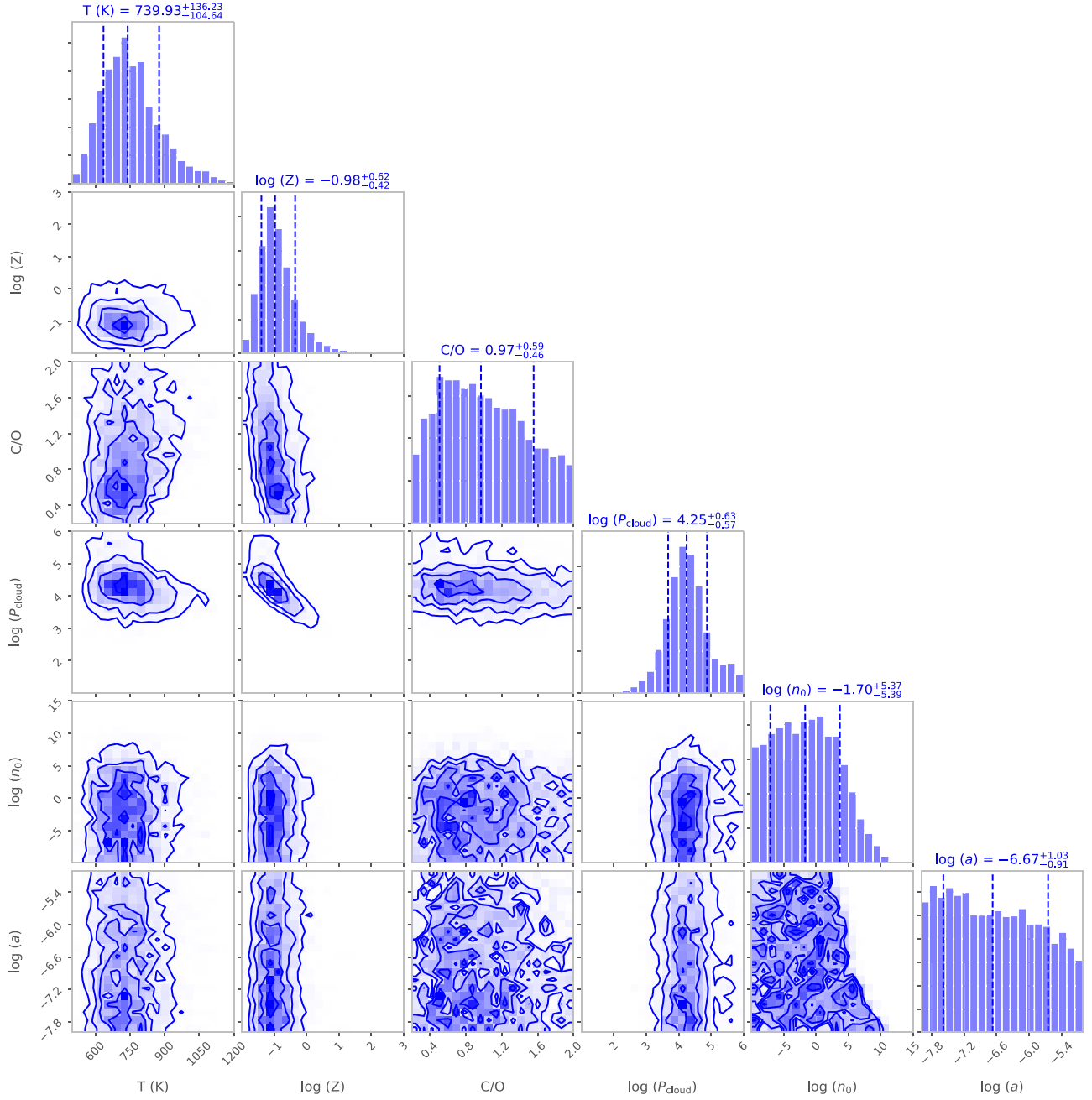
limit for metallicity is  $86 \times$  solar, indicating that enhanced metallicities are still consistent with our data. Unlike F14, our fits prefer lower atmospheric metallicities. Nonetheless, for metallicities greater than the lower prior bound in F14 ( $1 \times$  solar), our metallicity posteriors are in qualitative agreement with the ones published in F14. The addition of the STIS data to WFC3 data limits the degeneracy between cloud-top pressure and atmospheric metallicity (see Figure 14) encountered by F14, resulting in correspondingly narrower constraints on these properties (Benneke & Seager 2012).

We show the marginalized posterior probability distributions for metallicity, C/O, and cloud-top pressure in Figure 13. The *HST* STIS data provide additional constraints on atmospheric properties by disfavoring models with very low metallicity ( $\log(Z) \lesssim -1.5$ ), and correspondingly high cloud-top pressure  $P_{\text{cloud}}$  and high metallicity ( $\log(Z) \gtrsim 1$ ). This is apparent in Figure 12 where we see that the STIS data narrow the range of model transmission spectra that agree within  $\pm 1\sigma$ .

### 7.2.3. HST + Spitzer Data

We carry out a final set of fits including both the *HST* STIS + WFC3 and *Spitzer* transit depths. The full transmission spectrum with the best-fit model from PLATON is shown in Figure 15 and the median and confidence intervals for retrieved parameters are given in Table 4. Our  $3.6 \mu\text{m}$  *Spitzer* transit depth is low relative to the *HST* data and discrepant with the depth predicted by the best-fit model to the *HST* data. We are unable to find a single model that can simultaneously match the observed strength of the WFC3 absorption features while fitting the noticeably shallower *Spitzer* transit depths.

The inclusion of *Spitzer* data worsens the constraints on most atmospheric parameters (Table 4). The acceptable temperature and cloud-top pressure ranges now span the entire prior range. The constraints on metallicity from this fit are inconsistent with results from the *HST*-only fits. The preferred metallicity rises to a few  $100 \times$  solar, which allows the models to fit the flat baseline of the data by reducing the scale height while still maintaining some molecular absorption and reducing the relative abundance of methane in the atmosphere. We find that the particle size and number density are relatively unconstrained in both the *HST*-only and *HST* + *Spitzer* fits. The upper limit on the number density varies as a function of particle size (as expected) and is marginally higher for the

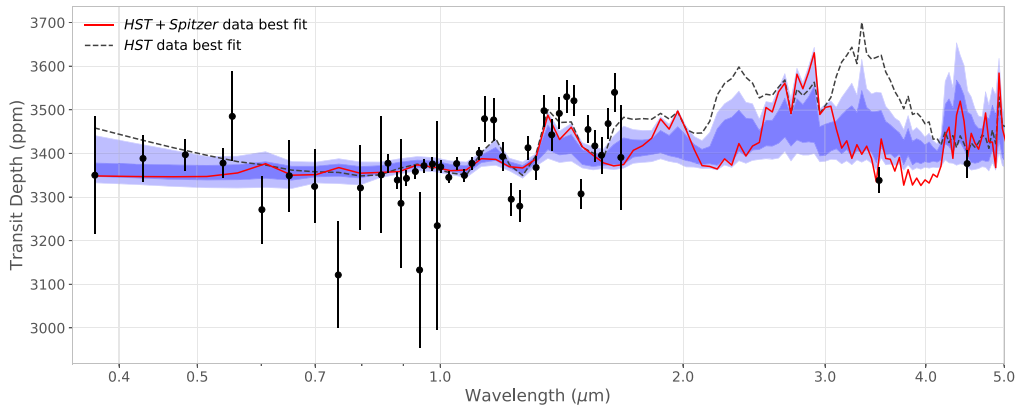


**Figure 14.** Posterior probability distributions for fits of the *HST* data set. Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with vertical dashed lines.

*HST* + *Spitzer* fit. The error multiple ( $\sigma_{\text{mult}}$ ) parameter, which is a measure of how underestimated the errors in the data are, jumps to  $\sim 1.7$ , i.e.,  $>20\%$  larger than the value obtained with *HST* data alone. In addition, the reduced  $\chi^2$  value (calculated using the errors on the transit depth measurements) increases from 1.9 for the *HST*-only fit to 2.8 for the full data set fit. We therefore conclude that our models are unable to provide a satisfactory fit to the full data set. Including an offset of  $\sim 100$ – $150$  ppm could reconcile the *Spitzer* depths with the models that fit the *HST* data. Fitting for this offset in a retrieval framework also yields similar estimates for its magnitude. However, as emphasized in Sections 4.5 and 5.1, such a large stellar-activity correction is incommensurate with the observed stellar variability.

#### 7.2.4. Retrievals Without Methane and/or Water Opacity

We quantify the significance of observed molecular absorption features by using the evidence obtained from nested sampling to compute Bayes factors for model comparisons. To test for the presence of a certain molecule (and the associated confidence/significance), we remove opacity contributions from the molecule and refit the transmission spectrum while keeping the priors unchanged. The ratio of the Bayesian evidence for fits with and without the molecular opacity yields the Bayes factor and allows us to quantify the data's preference for one model over the other (e.g., Benneke & Seager 2013). There is significant overlap between methane and water features in the near-infrared region ( $0.8$ – $1.7 \mu\text{m}$ ), and we therefore perform three additional retrievals for the *HST* data



**Figure 15.** Transmission spectrum for HAT-P-11b including both *HST* and *Spitzer* data (black filled circles) along with the best-fit model from PLATON and corresponding  $1\sigma$  and  $2\sigma$  contours (dark blue and light blue, respectively). The best-fit model for *HST* data is also shown for comparison, which predicts a much larger transit depth at  $3.6\ \mu\text{m}$ . The inclusion of the *Spitzer* transit depths shifts the models toward solutions with high atmospheric metallicity, which suppresses the depth of the absorption features in the WFC3 bands and decreases the overall quality of the fit in this region.

**Table 5**  
*HST* Retrievals Evidence

Model	log (Evidence)	Bayes Factor	$\sigma$
Nominal	$368.9 \pm 0.1$	...	...
Without $\text{CH}_4$ and $\text{H}_2\text{O}$	$361.0 \pm 0.1$	1:2812	4.4
Without $\text{CH}_4$	$364.8 \pm 0.1$	1:64	3.4
Without $\text{H}_2\text{O}$	$366.1 \pm 0.1$	1:17	2.9

along with the nominal case described above. In these three retrievals, we remove both water and methane opacity, just water opacity, and just methane opacity.

The evidence, Bayes factor (relative to the nominal model that includes both methane and water opacity), and equivalent  $\sigma$  significance for each of the three cases are shown in Table 5. The combined significance for the presence of water and methane is  $4.4\sigma$ . The Bayes factor for the two molecules individually is lower than the reported combined significance. The detection significance for each molecule is sensitive to relatively subtle features of the spectrum and may change due to small differences in the shape of the absorption features. Notably, the inclusion of *HST* STIS data makes the case for the presence of water and/or methane stronger. With WFC3 data alone, a similar comparison gives lower values for the Bayes factor for all three retrievals. This is primarily because the relatively flat optical spectrum excludes very low atmospheric metallicity models ( $\log Z \lesssim -1.5$ ), which possess somewhat higher evidence values (in *HST* WFC3-only retrievals) and therefore weaken the case for the presence of these molecules.

This exercise also allows us to investigate whether the disagreement between inferences made from *HST* and *Spitzer* data arises simply due to the absence of methane from the atmosphere. Vertical mixing and quenching could lower the methane abundance by orders of magnitude relative to the equilibrium values (Moses et al. 2011, 2013). However, quantifying this effect for HAT-P-11b requires a more careful analysis as its temperature-pressure profile overlaps with the equal abundance curve of  $\text{CH}_4$ –CO. This picture is further complicated by the planet’s orbital eccentricity (see Visscher 2012). We test whether our fit to the *HST* data without  $\text{CH}_4$  opacity fits the *Spitzer* data any better. We find that removing

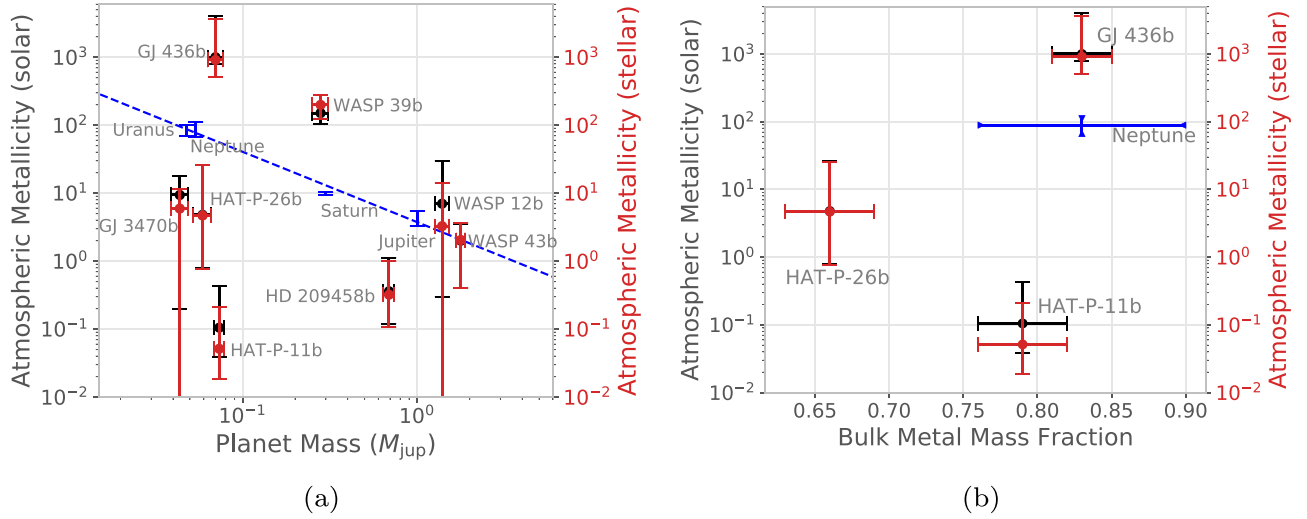
methane’s opacity requires a larger abundance of water to match the strength of the spectral features in the WFC3 bandpass. This pushes the best-fit models to higher metallicities (lower abundances/metallicities are ruled out by the STIS data). The best-fit models thus obtained match the  $3.6\ \mu\text{m}$  depth quite well but the higher atmospheric metallicities imply the presence of a substantial amount of CO and  $\text{CO}_2$  as well, which increases the  $4.5\ \mu\text{m}$  model depth and make it as discrepant with the data as the  $3.6\ \mu\text{m}$  depth is in our nominal model, which includes methane opacity.

## 8. Discussion and Conclusions

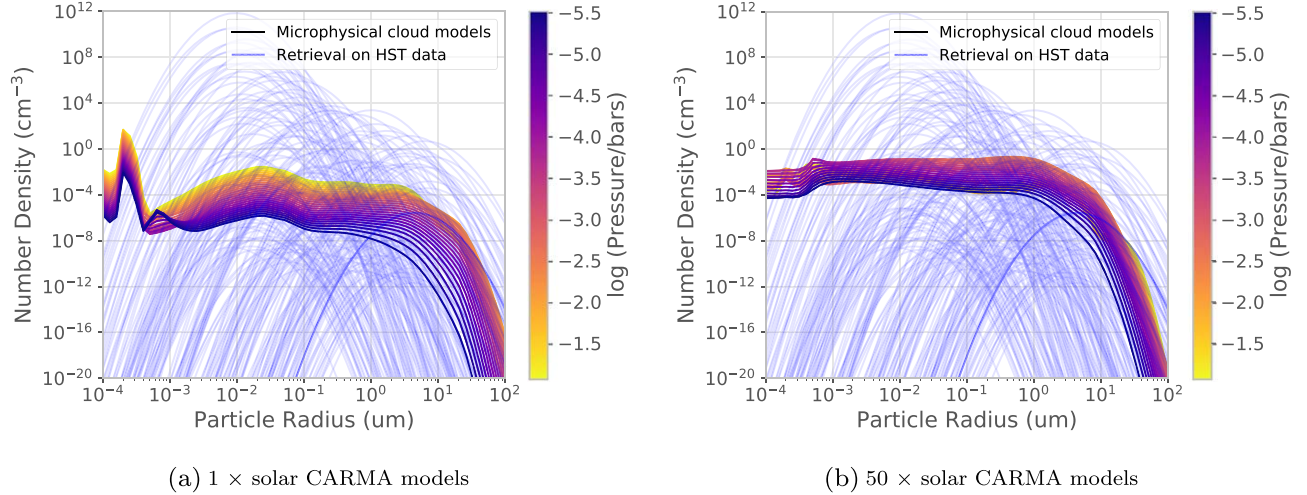
Our picture of HAT-P-11b’s atmosphere is primarily driven by the *HST* observations, which provide a self-consistent, spectrally resolved picture of the planet’s atmosphere over nine separate transit observations. The fact that we see clear evidence for molecular absorption across multiple visits and multiple bands leads us to conclude that any plausible model for this planet’s atmosphere must be able to reproduce the observed shape of these absorption (water + methane) bands. These models all overestimate the observed transit depth in the  $3.6\ \mu\text{m}$  *Spitzer* band; this may indicate that methane is underabundant in HAT-P-11b’s atmosphere as compared to the predictions of our equilibrium chemistry models. However, comparison of Bayesian evidence for *HST* retrievals suggests that methane is indeed present. We are unable to resolve these apparent contradictions with the current data set, but future spectroscopic observations of this planet with the *James Webb Space Telescope* (JWST) should provide a much clearer picture of its transmission spectrum in the mid-infrared wavelengths probed by the *Spitzer* photometry.

If we focus our attention for now on the *HST*-only fits, our updated results point to a significantly lower value for the planet’s atmospheric metallicity than that reported by F14. This runs counter to the trend observed in the solar system (Figure 16(a)): Uranus and Neptune have atmospheric C/H ratios between  $70\times$  and  $100\times$  that of the Sun, while Jupiter’s C/H ratio is just a few times solar (Wong et al. 2004; Fletcher et al. 2009; Karkoschka & Tomasko 2011; Sromovsky et al. 2011, see also, e.g., Kreidberg et al. 2014). Although there are relatively few published constraints on the atmospheric metallicities of Neptune-mass planets around other stars, GJ 436b appears to have an atmospheric metallicity of at least





**Figure 16.** (a) Atmospheric metallicity vs. planet mass for planets observed by *HST* and *Spitzer* (Kreidberg et al. 2014, 2015; Brogi et al. 2017; Morley et al. 2017; Wakeford et al. 2017a, 2018; Benneke et al. 2019). (b) Atmospheric metallicity vs. bulk metallicity (obtained from Thorngren & Fortney 2019) for Neptune-class planets. For Neptune, we plot lower and upper limits rather than  $1\sigma$  error bars (Helled & Guillot 2018). GJ 3470b is not included on this plot because the assumptions used to derive bulk metallicity constraints in the Thorngren et al. (2016) models may not be appropriate for planets with such low masses.



**Figure 17.** Particle number density as a function of radius from our microphysical cloud models at different pressures/heights in the atmosphere. We overplot a sample of lognormal particle-size distributions at 10 mbar from our retrievals for comparison. The best-fit size distribution is highlighted with a dark blue line. All profiles correspond to models with high likelihoods.

200 $\times$  solar (Madhusudhan & Seager 2011; Moses et al. 2013; Morley et al. 2017). However, HAT-P-26b (Wakeford et al. 2017a) provides a counterexample of an extrasolar Neptune with a relatively low atmospheric metallicity ( $4.8^{+21.5}_{-4.0} \times$  solar). Our new observations suggest that HAT-P-11b is more similar to HAT-P-26b than it is to either Neptune or GJ 436b. The low atmospheric metallicity of HAT-P-11b is all the more striking because it orbits a metal-rich star ( $[\text{Fe}/\text{H}] = +0.3$ ). The composition of the planet’s atmosphere therefore verges on being almost identical to that of the primordial gas that formed the star. This diversity in atmospheric composition of Neptune-mass planets suggests that they may not be a homogeneous planet population.

Comparison of atmospheric metallicity with bulk metallicities (mass fraction) calculated by Thorngren & Fortney (2019) indicates that Neptune-class planets may possess low-metallicity envelopes despite having a high bulk metal

fraction (Figure 16(b)). This implies that most of the solids, which have the potential to enrich the envelope, ought to have finished accreting before the initiation of substantial gas accretion from the disk. It also requires mixing in the interior to not be strong enough to significantly enrich the envelope. We expect that the sample of Neptune-mass planets with well-measured atmospheric metallicities will be significantly expanded by *JWST*, providing a much clearer view of the statistical properties of this population of planets.

In addition to providing improved constraints on HAT-P-11b’s atmospheric metallicity, our updated transmission spectrum provides us with an opportunity to explore the properties of the scattering particles in this planet’s atmosphere. We find that transmission spectra for our microphysical cloud models agree quite well with the observed *HST* spectrum (Figure 10). In Figure 17, we compare our retrieved cloud properties to those predicted by the models. The data do not put

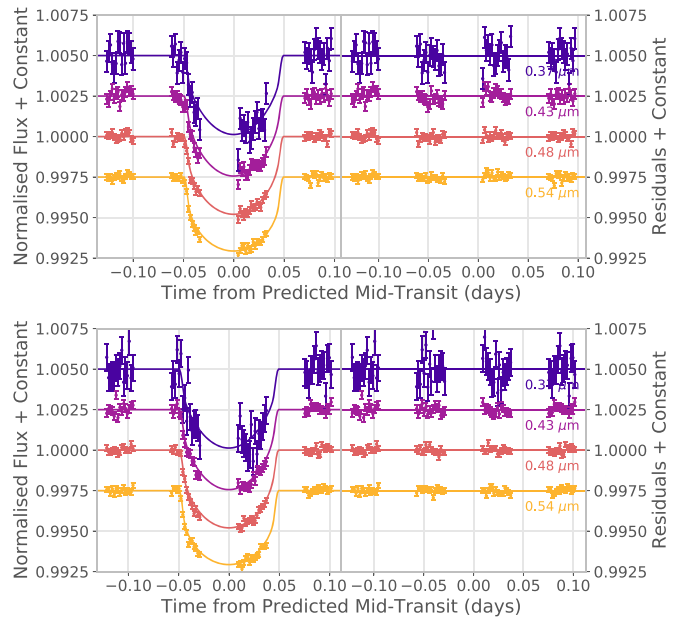
narrow constraints on these retrieved cloud properties and there is a degeneracy between mean particle size and number density (as evident in Figure 14). Regardless, the upper limit on mean particle size and its corresponding number density is roughly commensurate with predictions from the microphysical cloud models. Improved constraints provided by new data in the future should enable us to compare the predictions of the forward model and the retrieved parameters more rigorously. Moreover, the good agreement between the CARMA models and the retrieved models from PLATON (which uses local condensation from GG-chem<sup>18</sup> to deplete the gas phase) is reassuring because it is usually unclear if the amount of retrieved cloud opacity is realistic or not compared to the gas phase chemistry and condensation.

In the future, more accurate microphysical cloud models will be crucial for improving our understanding of the properties of these atmospheres. Better a priori predictions for cloud formation could allow future *JWST* observers to identify and prioritize observations of planets with relatively cloud-free terminators, while model-based constraints on cloud properties would help to limit degeneracies between cloud properties and atmospheric metallicity for planets with cloudy atmospheres. Our observations of HAT-P-11b serve as a useful illustration of both the limitations of our current understanding of cloud formation in these atmospheres, and also the power of spectrally resolved data with broad wavelength coverage to provide useful constraints on atmospheric composition despite our limited understanding of relevant cloud-formation processes.

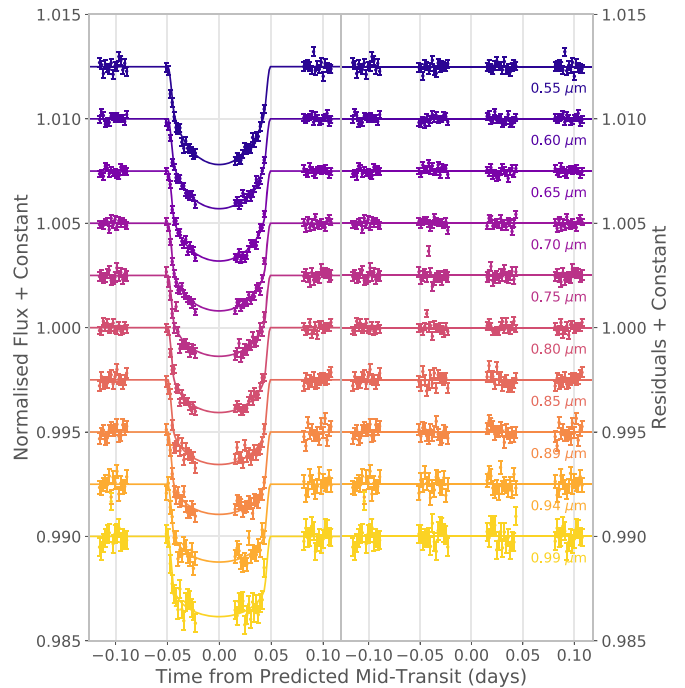
This work is based on observations from the *Hubble Space Telescope*, operated by AURA, Inc. on behalf of NASA/ESA. This work also includes observations made with the *Spitzer Space Telescope*, operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. Support for this work was provided by NASA through Space Telescope Science Institute grants GO-14260 and GO-14767. P. Gao and I. Wong acknowledge the generous support of the Heising-Simons Foundation via the 51 Pegasi b fellowships in Planetary Astronomy. J. K. Barstow acknowledges funding support from the Royal Astronomical Society Research Fellowship. G. W. Henry acknowledges support from NASA, NSF, Tennessee State University, and the State of Tennessee through its Centers of Excellence program. We would like to thank D. Deming and J. Fraine for exchanges regarding the WFC3 G141 data set. Y. C. also thanks C. V. Morley, S. Hörst, Y. Kawashima, Y. Wu, E. Lee, and Y. Yung for inspiring conversations.

## Appendix

Spectroscopic light curves for all the *HST* visits are shown here in Figures 18–23.

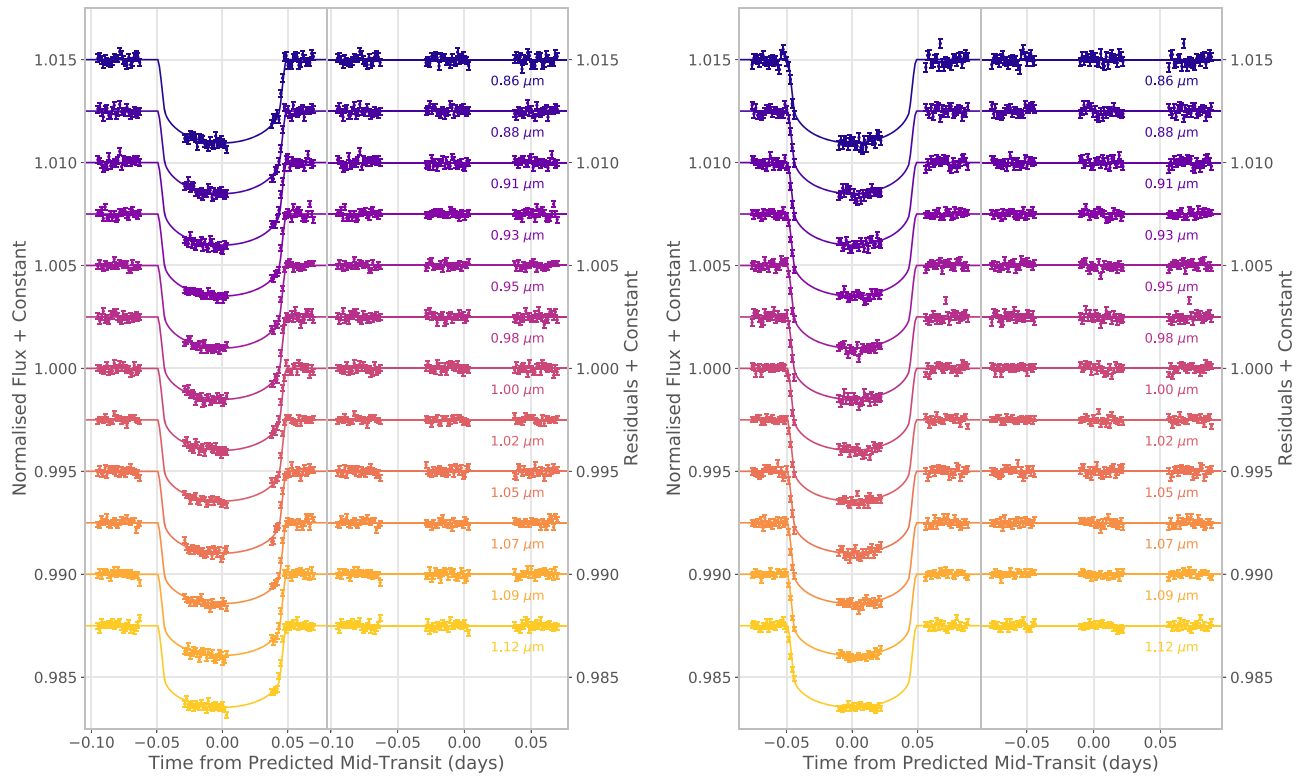


**Figure 18.** *HST* STIS G430L wavelength-dependent light curves for visit 1 and 2.

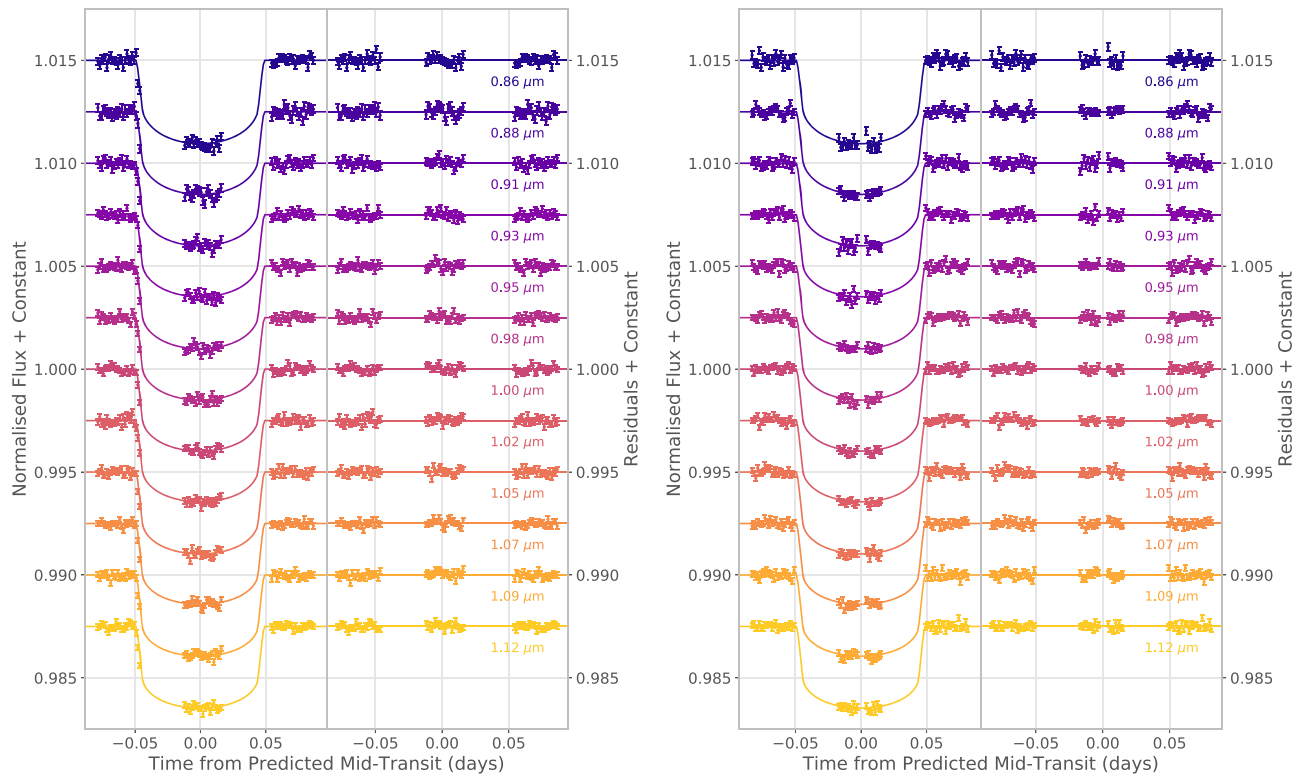


**Figure 19.** *HST* STIS G750L wavelength-dependent light curves.

<sup>18</sup> GG-chem is an open source thermo-chemical equilibrium code that calculates abundances of different molecular and atomic species given gas elemental composition, temperature, and pressure (Woitke et al. 2018, <https://github.com/pw31/GGchem>).



**Figure 20.** *HST* WFC3 G102 wavelength-dependent light curves for visit 1 and 2.



**Figure 21.** *HST* WFC3 G102 wavelength-dependent light curves for visit 3 and 4.

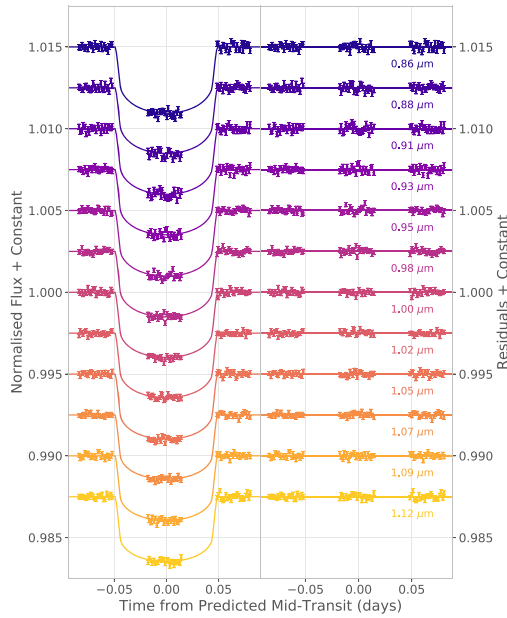


Figure 22. *HST* WFC3 G102 wavelength-dependent light curves for visit 5.

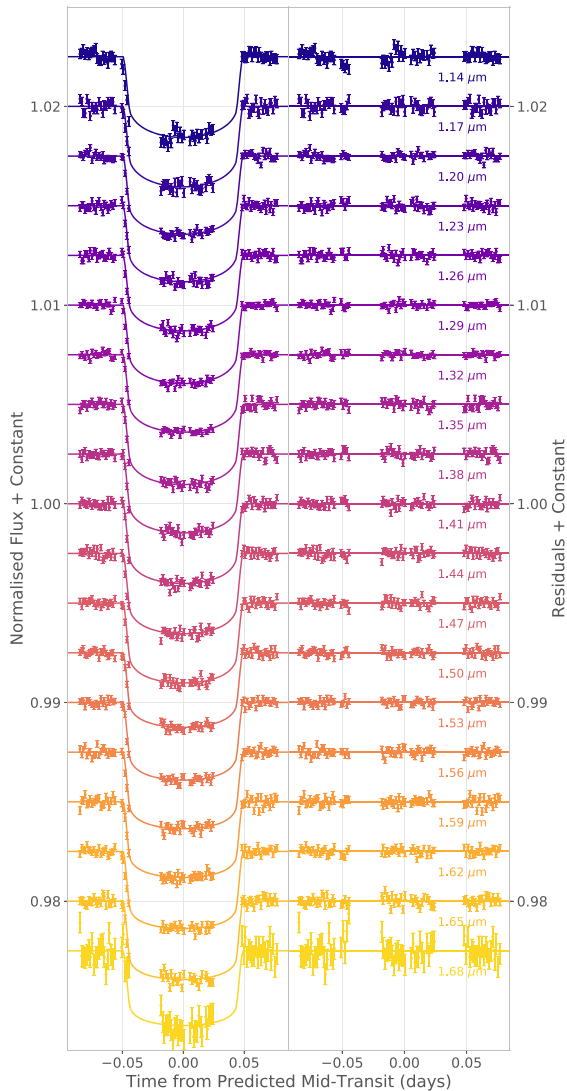


Figure 23. *HST* WFC3 G141 wavelength-dependent light curves.

## ORCID iDs

Yayaati Chachan <https://orcid.org/0000-0003-1728-8269>  
 Peter Gao <https://orcid.org/0000-0002-8518-9601>  
 Tiffany Kataria <https://orcid.org/0000-0003-3759-9080>  
 Ian Wong <https://orcid.org/0000-0001-9665-8429>  
 Gregory W. Henry <https://orcid.org/0000-0003-4155-8513>  
 Bjorn Benneke <https://orcid.org/0000-0001-5578-1498>  
 Michael Zhang <https://orcid.org/0000-0002-0659-1783>  
 Joanna Barstow <https://orcid.org/0000-0003-3726-5419>  
 Jacob L. Bean <https://orcid.org/0000-0003-4733-6532>  
 Nikole K. Lewis <https://orcid.org/0000-0002-8507-1304>  
 Megan Mansfield <https://orcid.org/0000-0003-4241-7413>  
 Mercedes López-Morales <https://orcid.org/0000-0003-3204-8183>  
 Nikolay Nikolov <https://orcid.org/0000-0002-6500-3574>  
 David K. Sing <https://orcid.org/0000-0001-6050-7645>  
 Hannah Wakeford <https://orcid.org/0000-0003-4328-3867>

## References

- Ackerman, A. S., & Marley, M. S. 2001, *ApJ*, **556**, 872  
 Adcroft, A., Campin, J.-M., Hill, C., et al. 2004, *MNRAS*, **352**, 2845  
 Alam, M. K., Nikolov, N., López-Morales, M., et al. 2018, *AJ*, **156**, 298  
 Allard, F., Homeier, D., & Freytag, B. 2012, *RSTP*, **370**, 2765  
 Allart, R., Bourrier, V., Lovis, C., et al. 2018, *Sci*, **362**, 1384  
 Astropy Collaboration, Price-Whelan, A. M., Sipcz, B. M., et al. 2018, *AJ*, **156**, 123  
 Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33  
 Bakos, G. Á., Torres, G., Pál, A., et al. 2010, *ApJ*, **710**, 1724  
 Benneke, B., Knutson, H. A., Lothringer, J. D., et al. 2019, *NatAs*, **3**, 813  
 Benneke, B., & Seager, S. 2012, *ApJ*, **753**, 100  
 Benneke, B., & Seager, S. 2013, *ApJ*, **778**, 153  
 Bourrier, V., Lecavelier des Etangs, A., Ehrenreich, D., et al. 2018, *A&A*, **620**, A147  
 Bradley, L., Sipocz, B., Robitaille, T., et al. 2018, *astropy/photutils*, V0.5, Zenodo, doi:10.5281/zenodo.1340699  
 Brogi, M., Line, M., Bean, J., Désert, J.-M., & Schwarz, H. 2017, *ApJL*, **839**, L2  
 Colaprete, A., Toon, O. B., & Magalhães, J. A. 1999, *JGRE*, **104**, 9043  
 Crossfield, I. J. M., & Kreidberg, L. 2017, *AJ*, **154**, 261  
 Deming, D., Knutson, H., Kammer, J., et al. 2015, *ApJ*, **805**, 132  
 Deming, D., Sada, P. V., Jackson, B., et al. 2011, *ApJ*, **740**, 33  
 Deming, D., Wilkins, A., McCullough, P., et al. 2013, *ApJ*, **774**, 95  
 Elias, T., Haefelin, M., Drobinski, P., et al. 2009, *AtmRe*, **92**, 443  
 Evans, T. M., Sing, D. K., Goyal, J., et al. 2018, *AJ*, **156**, 283  
 Evans, T. M., Sing, D. K., Kataria, T., et al. 2017, *Natur*, **548**, 58  
 Evans, T. M., Sing, D. K., Wakeford, H. R., et al. 2016, *ApJL*, **822**, L4  
 Fletcher, L., Orton, G., Teanby, N., Irwin, P., & Bjoraker, G. 2009, *Icar*, **199**, 351  
 Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2012, *PASP*, **125**, 306  
 Fraine, J., Deming, D., Benneke, B., et al. 2014, *Natur*, **513**, 556  
 Fu, G., Deming, D., Knutson, H., et al. 2017, *ApJL*, **847**, L22  
 Gao, P., & Benneke, B. 2018, *ApJ*, **863**, 165  
 Gao, P., Fan, S., Wong, M. L., et al. 2017, *Icar*, **287**, 116  
 Gao, P., Marley, M. S., & Ackerman, A. S. 2018, *ApJ*, **855**, 86  
 Gordon, I., Rothman, L., Hill, C., et al. 2017, *JQST*, **203**, 3  
 He, C., Hörst, S. M., Lewis, N. K., et al. 2018a, *ApJL*, **856**, L3  
 He, C., Hörst, S. M., Lewis, N. K., et al. 2018b, *AJ*, **156**, 38  
 Helled, R., & Guillot, T. 2018, in *Handbook of Exoplanets*, ed. H. Deeg & J. A. Belmonte (Cham: Springer), 167  
 Helling, C. 2019, *AREPS*, **47**, 583  
 Heng, K. 2018, *RNAAS*, **2**, 128  
 Hirano, T., Narita, N., Shporer, A., et al. 2011, *PASJ*, **63**, 531  
 Hörst, S. M., He, C., Lewis, N. K., et al. 2018, *NatAs*, **2**, 303  
 Huber, K. F., Czesla, S., & Schmitt, J. H. M. M. 2017, *A&A*, **597**, 113  
 Husser, T.-O., Wende-Von Berg, S., Dreizler, S., et al. 2013, *A&A*, **553**, A6  
 James, E., Toon, O., & Schubert, G. 1997, *Icar*, **129**, 147  
 Johnson, A., Zawadowicz, M., Lance, S., & Cizco, D. 2018, AAS, DPS Meeting, **50**, 220.03



- Johnson, A. V., Schneiderman, T. M., Bauer, A. J. R., & Cizco, D. J. 2017, *AGUFM*, **2017**, P43E-2938
- Kammer, J. A., Knutson, H. A., Line, M. R., et al. 2015, *ApJ*, **810**, 118
- Karkoschka, E., & Tomasko, M. G. 2011, *Icar*, **211**, 780
- Kataria, T., Showman, A. P., Fortney, J. J., Marley, M. S., & Freedman, R. S. 2014, *ApJ*, **785**, 92
- Kataria, T., Sing, D. K., Lewis, N. K., et al. 2016, *ApJ*, **821**, 9
- Kempton, E. M. R., Bean, J. L., & Parmentier, V. 2017, *ApJL*, **845**, L20
- Kempton, E. M.-R., Lupu, R., Owusu-Asare, A., Slough, P., & Cale, B. 2017, *PASP*, **129**, 044402
- Kitzmann, D., & Heng, K. 2018, *MNRAS*, **475**, 94
- Knutson, H. A., Dragomir, D., Kreidberg, L., et al. 2014, *ApJ*, **794**, 155
- Knutson, H. A., Howard, A. W., & Isaacson, H. 2010, *ApJ*, **720**, 1569
- Knutson, H. A., Lewis, N., Fortney, J. J., et al. 2012, *ApJ*, **754**, 22
- Kreidberg, L. 2015, *PASP*, **127**, 1161
- Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014, *ApJL*, **793**, L27
- Kreidberg, L., Line, M. R., Bean, J. L., et al. 2015, *ApJ*, **814**, 66
- Kreidberg, L., Line, M. R., Thorngren, D., Morley, C. V., & Stevenson, K. B. 2018, *ApJL*, **858**, L6
- Kuntschner, H., Bushouse, H., Kümmel, M., & Walsh, J. R. 2009a, ST-ECF Instrument Science Report, *WFC3-2009-18*
- Kuntschner, H., Bushouse, H., Kümmel, M., & Walsh, J. R. 2009b, ST-ECF Instrument Science Report, *WFC3-2009-18*
- Kuntschner, H., Kümmel, M., Walsh, J. R., & Bushouse, H. 2011, ST-ECF Instrument Science Report, *WFC3-2011-05*
- Lambrechts, M., Johansen, A., & Morbidelli, A. 2014, *A&A*, **572**, A35
- Lee, G. K. H., Blečić, J., & Helling, C. 2018, *A&A*, **614**, A126
- Lewis, N. K., Knutson, H. A., Showman, A. P., et al. 2013, *ApJ*, **766**, 95
- Lewis, N. K., Showman, A. P., Fortney, J. J., et al. 2010, *ApJ*, **720**, 344
- Lodders, K. 1999, *ApJ*, **519**, 793
- Lothringer, J. D., Benneke, B., Crossfield, I. J. M., et al. 2018, *AJ*, **155**, 66
- Madhusudhan, N. 2012, *ApJ*, **758**, 36
- Madhusudhan, N., & Seager, S. 2011, *ApJ*, **729**, 41
- Mansfield, M., Bean, J. L., Oklopčić, A., et al. 2018, *ApJL*, **868**, L34
- Marley, M. S., & McKay, C. P. 1999, *Icar*, **138**, 268
- McCullough, P., & Mackenty, J. 2012, STScI Instrument Science Report, *WFC3 2012-08*
- McCullough, P. R., Crouzet, N., Deming, D., & Madhusudhan, N. 2014, *ApJ*, **791**, 55
- Morley, C. V., Fortney, J. J., Marley, M. S., et al. 2012, *ApJ*, **756**, 172
- Morley, C. V., Knutson, H., Line, M., et al. 2017, *AJ*, **153**, 86
- Morris, B. M., Hawley, S. L., Hebb, L., et al. 2017a, *ApJ*, **848**, 58
- Morris, B. M., Hebb, L., Davenport, J. R. A., Rohn, G., & Hawley, S. L. 2017b, *ApJ*, **846**, 99
- Moses, J. I., Line, M. R., Visscher, C., et al. 2013, *ApJ*, **777**, 34
- Moses, J. I., Visscher, C., Fortney, J. J., et al. 2011, *ApJ*, **737**, 15
- Murgas, F., Chen, G., Pallé, E., Nortmann, L., & Nowak, G. 2019, *A&A*, **622**, A172
- Nikolov, N., Sing, D. K., Burrows, A. S., et al. 2015, *MNRAS*, **447**, 463
- Nikolov, N., Sing, D. K., Fortney, J. J., et al. 2018a, *Natur*, **557**, 526
- Nikolov, N., Sing, D. K., Goyal, J., et al. 2018b, *MNRAS*, **474**, 1705
- Nikolov, N., Sing, D. K., Pont, F., et al. 2014, *MNRAS*, **437**, 46
- Oberg, K. I., Murray-Clay, R., Bergin, E. A., et al. 2011, *ApJL*, **743**, L16
- Parmentier, V., Showman, A. P., & Lian, Y. 2013, *A&A*, **558**, 91
- Parviainen, H., & Aigrain, S. 2015, *MNRAS*, **453**, 3821
- Pinhas, A., & Madhusudhan, N. 2017, *MNRAS*, **471**, 4355
- Pinnick, R. G., Hoihjelle, D. L., Fernandez, G., et al. 1978, *JatS*, **35**, 2020
- Pont, F., Knutson, H., Gilliland, R. L., Moutou, C., & Charbonneau, D. 2008, *MNRAS*, **385**, 109
- Powell, D., Zhang, X., Gao, P., & Parmentier, V. 2018, *ApJ*, **860**, 18
- Pudritz, R. E., Cridland, A. J., & Alessi, M. 2018, in *Handbook of Exoplanets*, ed. H. Deeg & J. A. Belmonte (Cham: Springer), 1
- Rackham, B. V., Apai, D., & Giampapa, M. S. 2018, *ApJ*, **853**, 122
- Rackham, B. V., Apai, D., & Giampapa, M. S. 2019, *AJ*, **157**, 96
- Sanchis-Ojeda, R., & Winn, J. N. 2011, *ApJ*, **743**, 61
- Shen, X., Sun, J., Zhang, X., et al. 2015, *AtmEn*, **120**, 307
- Showman, A. P., Cho, J. Y.-K., & Menou, K. 2009, *ApJ*, **699**, 564
- Sing, D. K. 2010, *A&A*, **510**, A21
- Sing, D. K., Fortney, J. J., Nikolov, N., et al. 2016, *Natur*, **529**, 59
- Sing, D. K., Lavvas, P., Ballester, G. E., et al. 2019, *AJ*, **158**, 91
- Sing, D. K., Pont, F., Aigrain, S., et al. 2011, *MNRAS*, **416**, 1443
- Sing, D. K., Vidal-Madjar, A., Desert, J. M., des Etangs, A. L., & Ballester, G. 2008, *ApJ*, **686**, 658
- Sing, D. K., Wakeford, H. R., Showman, A. P., et al. 2015, *MNRAS*, **446**, 2428
- Southworth, J. 2011, *MNRAS*, **417**, 2166
- Sromovsky, L., Fry, P., & Kim, J. 2011, *Icar*, **215**, 292
- Strassmeier, K. G., Granzer, T., Weber, M., et al. 2004, *AN*, **325**, 527
- Tennyson, J., Yurchenko, S., Tennyson, J., & Yurchenko, S. N. 2018, *Atoms*, **6**, 26
- Thorngren, D. P., & Fortney, J. J. 2019, *ApJ*, **874**, L31
- Thorngren, D. P., Fortney, J. J., Murray-Clay, R. A., & Lopez, E. D. 2016, *ApJ*, **831**, 64
- Toon, O., McKay, C., Griffith, C., & Turco, R. 1992, *Icar*, **95**, 24
- Toon, O. B., Turco, R. P., Hamill, P., Kiang, C. S., & Whitten, R. C. 1979, *JatS*, **36**, 718
- Tsiaras, A., Waldmann, I. P., Rocchetto, M., et al. 2016, *ApJ*, **832**, 202
- Tsiaras, A., Waldmann, I. P., Zingales, T., et al. 2018, *AJ*, **155**, 156
- Venturini, J., Alibert, Y., & Benz, W. 2016, *A&A*, **596**, 90
- Visscher, C. 2012, *ApJ*, **757**, 5
- Wakeford, H. R., & Sing, D. K. 2015, *A&A*, **573**, A122
- Wakeford, H. R., Sing, D. K., Deming, D., et al. 2018, *AJ*, **155**, 29
- Wakeford, H. R., Sing, D. K., Kataria, T., et al. 2017a, *Sci*, **356**, 628
- Wakeford, H. R., Stevenson, K. B., Lewis, N. K., et al. 2017b, *ApJL*, **835**, L12
- Wakeford, H. R., Wilson, T. J., Stevenson, K. B., & Lewis, N. K. 2019, *RNAAS*, **3**, 7
- Wilkins, A. N., Deming, D., Madhusudhan, N., et al. 2014, *ApJ*, **783**, 113
- Winn, J. N., Johnson, J. A., Howard, A. W., et al. 2010, *ApJL*, **723**, 223
- Woitke, P., Helling, C., Hunter, G. H., et al. 2018, *A&A*, **614**, A1
- Wong, I., Knutson, H. A., Kataria, T., et al. 2016, *ApJ*, **823**, 122
- Wong, M. H., Mahaffy, P. R., Atreya, S. K., Niemann, H. B., & Owen, T. C. 2004, *Icar*, **171**, 153
- Yee, S. W., Petigura, E. A., Fulton, B. J., et al. 2018, *AJ*, **155**, 255
- Zhang, M., Chachan, Y., Kempton, E. M.-R., & Knutson, H. A. 2019, *PASP*, **131**, 034501
- Zhang, M., Knutson, H. A., Kataria, T., et al. 2018a, *AJ*, **155**, 83
- Zhang, X., & Showman, A. 2018a, *ApJ*, **866**, 1
- Zhang, X., & Showman, A. 2018b, *ApJ*, **866**, 2
- Zhang, Z., Zhou, Y., Rackham, B. V., & Apai, D. 2018b, *AJ*, **156**, 178
- Zhou, Y., Apai, D., Lew, B. W. P., & Schneider, G. 2017, *AJ*, **153**, 243