

## ACCESS: Ground-based Optical Transmission Spectroscopy of the Hot Jupiter WASP-4b

ALEX BIXEL,<sup>1,2</sup> BENJAMIN V. RACKHAM,<sup>1,2</sup> DÁNIEL APAI,<sup>1,2,3</sup> NÉSTOR ESPINOZA,<sup>4,5,6</sup> MERCEDES LÓPEZ-MORALES,<sup>7,2</sup>  
 DAVID OSIP,<sup>8</sup> ANDRÉS JORDÁN,<sup>5,6,4,2</sup> CHIMA MCGRUDER,<sup>7</sup> AND IAN WEAVER<sup>7</sup>

<sup>1</sup>*Department of Astronomy/Steward Observatory, The University of Arizona, 933 N. Cherry Avenue, Tucson, AZ 85721, USA*

<sup>2</sup>*Earths in Other Solar Systems Team, NASA Nexus for Exoplanet System Science*

<sup>3</sup>*Lunar and Planetary Laboratory, The University of Arizona, 1640 E. University Blvd, AZ 85721, USA*

<sup>4</sup>*Max-Planck-Institut für Astronomie, Königstuhl 17, 69117, Heidelberg, Germany*

<sup>5</sup>*Instituto de Astrofísica, Facultad de Física, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 782-0436 Macul, Santiago, Chile*

<sup>6</sup>*Millennium Institute of Astrophysics, Av. Vicuña Mackenna 4860, 782-0436 Macul, Santiago, Chile*

<sup>7</sup>*Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 01238, USA*

<sup>8</sup>*Las Campanas Observatory, Carnegie Institution of Washington, Colina el Pino, Casilla 601, La Serena, Chile*

## ABSTRACT

We present an optical transmission spectrum of the atmosphere of WASP-4b obtained through observations of four transits with *Magellan*/IMACS. Using a Bayesian approach to atmospheric retrieval, we find no evidence for scattering or absorption features in our transit spectrum. Our models include a component to model the transit light source effect (spectral contamination from unocculted spots on the stellar photosphere), which we show can have a marked impact on the observed transmission spectrum for reasonable spot covering fractions (< 5%); this is the first such analysis for WASP-4b. We are also able to fit for the size and temperature contrast of spots observed during the second and third transits, finding evidence for both small, cool and large, warm spot-like features on the photosphere. Finally, we compare our results to those published by Huitson et al. (2017) using *Gemini*/GMOS and May et al. (2018) using IMACS, and find that our data are in agreement.

**Keywords:** planets and satellites: atmospheres — planets and satellites: individual (WASP-4b) — stars: activity — stars: starspots — techniques: spectroscopic

## 1. INTRODUCTION

WASP-4b is a  $1.4 R_J$  hot Jupiter orbiting a G7V star with a period of 1.34 d, equilibrium temperature of  $\sim 1700$  K, and transit depth  $(R_p/R_s)^2 = 2.4\%$  (Wilson et al. 2008). It has been observed in transit over three dozen times, offering strong constraints on its orbit (Hoyer et al. 2013), the spot activity and relative rotation of its host star (Sanchis-Ojeda et al. 2011), and also placing upper limits on its transit timing variation amplitude (Nikolov et al. 2012).

Beerer et al. (2011) used *Spitzer* to observe the planet’s secondary eclipse and place constraints on the temperature profile of its atmosphere, and they conclude that the evidence is consistent with either a weak temperature inversion or none at all. However, even stronger detections of thermal inversions using *Spitzer* have later been called into question (e.g., HD 209458b, Diamond-Lowe et al. 2014). The evidence for an inversion therefore remains marginal, leaving us with

little insight about the composition of the planet’s upper atmosphere.

Given the tight constraints on its transit depth and orbital properties, WASP-4b is a natural target for spectroscopic transit observations. Transit spectroscopy can be used to constrain the composition and structure of the planet’s upper atmosphere and test for the presence of clouds, scattering hazes, and atomic or molecular absorbers (e.g., Seager & Sasselov 2000; Brown 2001; Hubbard et al. 2001). These observations also offer insight into the stellar photosphere, i.e., through the measurement of star spot temperatures within the transit chord (e.g., Pont et al. 2008; Sing et al. 2011; Béky et al. 2014). However, signals from the photosphere can be degenerate with those from the planet’s atmosphere, leading to contrasting interpretations of planetary and stellar origins for features in optical transmission spectra (e.g., Pont et al. 2013; McCullough et al. 2014). It is therefore critical to demonstrate methods for accounting for this degeneracy as the field moves toward smaller targets and increasingly precise observations (e.g., Rackham et al. 2018).

Transit spectroscopy of WASP-4b has been attempted with *HST*/WFC3 (Ranjan et al. 2014), but was unsuccessful due to detector saturation. A *Gemini*/GMOS optical transmission spectrum of WASP-4b has been published by Huitson et al. (2017), who measure a nearly uniform opacity from 440–940 nm, suggesting the presence of high-altitude clouds and a possible sodium absorption feature. More recently, May et al. (2018) have published a transmission spectrum using *Magellan*/IMACS, and have similarly found no evidence for spectral features.

As part of the Arizona-CfA-Católica-Carnegie Exoplanet Spectroscopy Survey (ACCESS, Rackham et al. 2017), we have observed four transits of WASP-4b with *Magellan*/IMACS. We have previously demonstrated the use of this instrument for transit spectroscopy using our custom data reduction pipeline (Jordán et al. 2013; Rackham et al. 2017; Espinoza et al. 2019). In this paper, we present an optical transmission spectrum from 450–900 nm. We interpret our results using a Bayesian retrieval code introduced in Espinoza et al. (2019) and find no evidence for scattering or absorption features. We also fit for the size and temperature of photosphere features occulted during the second and third transits, and derive corrections for their effects on the transit spectrum. Finally, we compare our findings with those of Huitson et al. (2017) and May et al. (2018).

## 2. OBSERVATIONS

We conducted spectroscopic observations of WASP-4 on the nights of 24 September 2013, 17 October 2013, 14 August 2015, and 26 September 2015 (hereafter Transits 1–4) using the Inamori-Magellan Areal Camera & Spectrograph (IMACS, Dressler et al. 2011) on the 6.5m Magellan-Baade telescope at Las Campanas Observatory in Chile. We observed using multi-slit masks in the f/2 mode with 2x2 binning (0.4"/px). Observations of HeNeAr and quartz calibration lamps before and after the observations allowed for wavelength calibration and flat-field correction. The key parameters of our observations are listed in Table 1.

### 2.1. 24 September 2013, 17 October 2013, and 14 August 2015

On the first three nights we used a setup consisting of a 400–1000 nm spectroscopic filter, a 300 lines/mm grism with a blaze angle of 17.5 degrees, and 10" wide by 20" long spectral slits for the target and reference stars. Twelve reference stars were observed, although only one was used in the final data reduction for reasons discussed in Section 3.3. Most of the stellar spectra were dispersed across two chips. While the lunar sky background was minimal on the first and third nights, it was substantial during the second, and we take this into account in our data reduction pipeline. Finally, the observations on 24 September 2013 (Transit 1) did not com-

mence until shortly after ingress, so we were only able to observe a partial transit.

### 2.2. 26 September 2015

Our instrument setup for the fourth observation was modified from the previous three. We used a 570–980 nm order-blocking filter for the purpose of eliminating higher-order interference towards red wavelengths, 10" wide by 10" long slits, and a 150 lines/mm grism with a blaze angle of 18.8 degrees. This setup allowed for more tightly dispersed spectra which fall on a single chip, thereby reducing detector-to-detector variations in the spectra and avoiding chip gaps. The moon was in full phase and separated from the target by 40 degrees, contributing significantly to the ambient sky background, which we again account for in our data reduction pipeline.

## 3. DATA REDUCTION

We reduce the raw data using our custom Python-based pipeline which has been used for similar observations of WASP-6b (Jordán et al. 2013), GJ 1214b (Rackham et al. 2017), and WASP-19b (Espinoza et al. 2019). In the following paragraphs, we give a brief overview of the pipeline functions. A more detailed review can be found in Jordán et al. (2013) and Rackham et al. (2017); the only more recent addition to the pipeline is the correction for the non-uniform sky background described in Section 3.1.

We use quartz lamp images taken with the same configuration as the science images to apply a flat-field correction, and we calculate the bias offset from the overscan region of each chip. We use full-frame flats taken without a mask or grism to identify bad pixels, and we identify cosmic rays on the stellar spectra using a  $3\sigma$  threshold at each row along the direction of spectral dispersion. To trace the stellar spectra, we calculate the centroid in each row along the dispersion direction and fit a second-order polynomial to the centroid values. For Transits 1 and 2, we assume the background is uniform in the spatial direction, measure it using the outermost 14 px in each row along the dispersion direction, and subtract. For Transits 3 and 4, the background is not spatially uniform, so we subtract it using the method outlined in Section 3.1. Finally, to extract the spectra we use the optimal extraction algorithm outlined by Marsh (1989), which involves fitting a third-order polynomial to the spectral profile at each row along the dispersion direction, then using that profile to weight each pixel when summing the flux.

We use HeNeAr arc lamp exposures taken before and after the transit observations to calculate the wavelength solutions for each star, which convert the pixel coordinates along the dispersion direction to wavelength values. We manually identify prominent emission lines and use a sixth-order polynomial to fit the wavelength solution; given the large number of lines used, the danger of over-fitting is minimal. The

**Table 1.** Instrument setup and transit model characteristics for each of our four observations. All observations made use of the IMACS f/2 camera with multi-slit masks. The ratio in column 7 factors in the brightness of the target and single reference star, but not the sky background.

Transit	Date (start of night)	Filter	Grism (l/mm)	Airmass	$R_p/R_s$ (white light)	Scatter / Photon noise
1	24 September 2013	Spectroscopic f/2	300	< 1.3	$0.1528^{+0.0012}_{-0.0011}$	4.9
2	17 October 2013	...	...	< 1.1	$0.1537^{+0.0008}_{-0.0007}$	7.8
3	14 August 2015	...	...	< 1.7	$0.1544^{+0.0002}_{-0.0002}$	2.9
4	26 September 2015	WBP 5694-9819	150	< 1.2	$0.1565^{+0.0004}_{-0.0005}$	6.2

marked lines are iteratively rejected based on their residuals to the fit until the residuals are below  $\sim 0.05$  Å.

Our spectra drift in the dispersion direction over the course of the night, resulting in frame-to-frame offsets in the wavelength solution. To solve this, we cross-correlate the first spectrum of a star with each of the subsequent spectra to determine the shift in wavelength, then fit a third-order polynomial to this shift as a function of time and use it to correct the wavelength solution in each frame.

The output of the pipeline is a set of reduced wavelength-binned and integrated (“white”) light curves for the target and reference stars.

### 3.1. Non-uniform sky background

In some of our IMACS data sets for this and other ACCESS targets, we have noticed that spatially uniform sources of light inherit a non-uniform profile in the spatial direction once the mask and dispersive element are in place. This effect also widens the spectra of point sources. We have attributed this to internal scattering within the instrument, and have found that it is more common in newer images: of the four data sets for WASP-4, only Transits 3 and 4 are affected. We measure the extent of the effect using our flat field images, finding that the profile of spatially uniform light peaks at the slit center and appears  $\sim 10\%$  fainter at the slit edges. Unless we account for this, we will underestimate the sky background when we extract the stellar spectra.

For the two affected data sets, we use quartz lamp exposures taken at the beginning of the night to model the scattered light profile. The peaked profiles are not consistent with common symmetrical functions (e.g., Gaussian), nor are they well-described by a classical high-order polynomial, because high-order polynomials commonly fail at the edges of their fitted intervals (an effect known as Runge’s phenomenon). Instead, we use Chebyshev polynomials which are not as susceptible to this effect, and find that sixth-order polynomials are sufficient to match the data.

The polynomials are fitted independently for every row along the dispersion direction of every slit in the flat field images. Then, in each science image we fit the amplitudes of the polynomials using the background level in the outermost

14 px in each row along the dispersion direction (8 px for Transit 4 due to shorter slits). Finally, we subtract the fitted polynomial from each row.

### 3.2. Flat field correction

Applying the flat-field correction for Transits 1, 3, or 4 does not significantly change our binned light curves, or the transit spectra which we derive in Section 4. However, applying the correction for Transit 2 introduces  $\sim 2\%$ -level, non-linear time-dependent trends into out-of-transit baseline flux of the binned light curves blueward of 530 nm. These trends remain even when normalizing by the reference star as discussed in the following. Therefore, for consistency, we choose not to apply the flat-field correction to any of our data sets.

### 3.3. Reference star selection

The shape of the target’s light curve is complicated by instrumental and atmospheric effects, such as changes in airmass or transparency. To calibrate out these effects, we simultaneously observed 12 reference stars of comparable optical apparent magnitudes and color ratios using multi-slit masks. Of these, two spectra reached the saturation limit of the detector and were not usable.

We use our highest quality data set (Transit 3) to determine which of the ten remaining reference stars to use in our light curve analysis. Our primary consideration is the shape of the out-of-transit baseline for the target light curve when normalized by each star. Eight of the stars leave residual long-term trends at the  $\sim 1\%$  level. Two stars leave trends at the  $\sim 0.1\%$  level, and they happen to be the only two stars occupying the same pair of detector chips as the target. The point-to-point scatters of the baseline points for the two resulting light curves are 0.4 and 0.6 mmag, which are both smaller than the scatters for the eight light curves with larger long-term trends; we therefore further restrict our set of potential reference stars to these two. We perform a similar analysis of the two remaining stars using our three other data sets, and discover that one of the two stars reliably produces a flat baseline, while the other star introduces trends at the  $\sim 0.5\text{--}1\%$  level. Therefore, we discard this star as well.

We are left with a single reference star, 2MASS J23341836-4204509, which reliably produces a flat out-of-transit baseline with a low point-to-point scatter in all of our data sets. We note that this reference star was also used in a similar analysis by Huitson et al. (2017). The spectral type of the reference star has not been previously published, but using parallax measurements from Gaia DR2 (Gaia Collaboration 2018) we calculate that it is intrinsically brighter than the target by  $\sim 80\%$  from 400 - 1000 nm. While this wavelength range does not capture all of the emitted light, it does cover the emission peak for late F and all G stars. Using a simple  $L_* \propto M_*^{3.5}$  scaling relation, and assuming that our wavelength range covers most of the emitted light, this suggests that the reference star is  $\sim 20\%$  more massive than the G7V,  $0.85 M_\odot$  target (Gillon et al. 2009), consistent with an early G or late F dwarf.

### 3.4. Effects of atmospheric dispersion

Magellan-Baade is equipped with an atmospheric dispersion compensator (ADC), which corrects for the effects of differential atmospheric refraction as the airmass of our target changes over the course of the night. If left uncorrected, differential refraction would affect our stellar wavelength solutions in a non-linear manner, “stretching” or “shrinking” the spectra over the course of the night. We test the accuracy of the ADC correction by measuring the distance in calibrated wavelength space between Na and H $\alpha$  in the target and reference spectra as a function of time, which we calculate to vary by no more than 1 Å over the course of the night. Furthermore, the difference between the measured distance for the target and reference stellar spectra changes by no more than 0.5 Å, and less than 0.1 Å on nights with higher quality data. Since the long-term change in this measurement is much lower than the frame-to-frame scatter (up to 5 Å), we reason that any effect on the measured transit depth will be negligible.

### 3.5. Observing efficiency

During Transits 1 and 2, we opted for longer, low-noise readout modes and short exposure times to avoid detector non-linearity. As a result, we spent only  $\sim 20\%$  of our time gathering light. By comparison, we achieved an efficiency of  $\sim 70\%$  and  $\sim 50\%$  during Transits 3 and 4, respectively, as we chose the fastest readout mode and took longer exposures. The difference in read noise between the fastest (5.6 e-) and slowest (2.8 e-) modes is negligible compared to the Poisson noise ( $>100$  e-), and as we show in Section 3.8, the detector is sufficiently linear for our purposes provided the pixel values remain near or below half-well.

Based on this experience, we advise that observers prioritize longer exposure times with fast readouts for transit spectroscopy with IMACS.

### 3.6. Sky background

Transits 2 and 4 suffer from high sky background values due to the target’s proximity to the full-phase moon. The effects can be measured by comparing the scatter of the residuals in Figure 2. Scaled for exposure time, the magnitudes of the scatter for Transits 1 and 3 agree to within 15%, while the scatter of Transit 2 is  $\sim 50\%$  larger. The scatter of Transit 4 is similarly  $\sim 50\%$  larger, although Transit 4 was observed in a redder filter with less sky contamination.

### 3.7. Slit losses

We consider whether slit losses may have been significant during our observations. For Transits 1-3, the slits were 10" wide and 20" long, while for Transit 4 the slits were 10" wide and 10" long. We perform a least-squares fit of a Moffat profile to the point spread function of a bright but unsaturated star in each of the field acquisition images, then integrate to determine how much light would lie outside of the slit if placed over the star. For every night, the fraction of light lost is more than an order of magnitude lower than the Poisson noise.

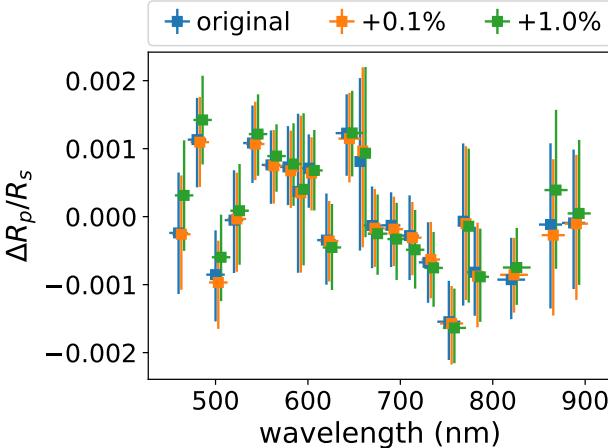
### 3.8. Detector linearity

We observed Transits 3 and 4 with longer integration times to reduce the relative readout overhead and improve our signal-to-noise ratios. However, this increased the maximum pixel values of the target spectrum from  $\sim 15,000$  to  $\sim 30,000$  ADU, nearly half of the full well value ( $\sim 65,000$  ADU). This raises the question of whether the detector’s response could have become non-linear (i.e. the gain was not uniform over the range of pixel values), which could bias the mean-subtracted transit spectra which we extract later in Section 4.

In principle, non-linearity should not be an issue for measuring the relatively small changes in transit depth from bin to bin, as the gain should not be expected to change over the small range of changing values. However, the pixel values at the peak of the target spectrum varied by up 15% from frame to frame and 40% over the course of each night, due mostly to seeing variations which changed the point spread function. Therefore it is worthwhile to determine the threshold beyond which non-linearity could significantly affect our results.

To do so, we modify the raw data for Transit 3 to reflect a 0.1% or 1% linear increase in the gain over the range of 0-35,000 ADU. We then reduce the modified data and extract the mean-subtracted transit spectra for both gain prescriptions, which we plot alongside the original spectrum in Figure 1. We find an average effect of  $0.07\sigma$  per bin for a 0.1% increase in gain, and  $0.23\sigma$  per bin for a 1% increase.

So long as the detector is linear to 0.1% from zero to half-well, non-linear effects should have a negligible impact on our results. Even in the case of 1% non-linearity, the effect



**Figure 1.** The mean-subtracted transit spectrum of Transit 3 with different gain modifications. We calculate and apply a gain enhancement to each pixel in the raw data, assuming the gain increases linearly by 0%, 0.1%, or 1% from 0–35,000 ADU. Provided the detector is linear to < 1% over this range, the mean-subtracted transit spectrum is largely unaffected.

is  $\ll 1\sigma$  except in a few bins. We are therefore confident that our results are robust to realistic levels of non-linearity, and we advise that future observers target a similar range of peak pixel values ( $\sim 35,000$  ADU) so as to optimally balance observational efficiency and detector linearity.

#### 4. LIGHT CURVE MODELING

We model our integrated and spectroscopic light curves using the analytic models introduced by Mandel & Agol (2002), and marginalize over the parameter space with a Markov-Chain Monte Carlo (MCMC) algorithm previously detailed in Rackham et al. (2017). As in previous works, we fit for the limb-darkening coefficients in order to account for any biases that might arise due to our imperfect knowledge of the intensity profile of the photosphere (Espinoza & Jordán 2015). As WASP-4b has been studied extensively through photometric observations, we hold the orbital parameters fixed to the mean values in Table 2.

The fitted parameters for the transit model include the planet-to-star radius ratio ( $R_p/R_s$ ), two parameters for a quadratic limb-darkening law which are sampled according to the method detailed in Kipping (2013), and the mid-transit time. The baseline flux is modeled as a second-order polynomial; this decision is discussed in more detail in Section 4.4.

We use the likelihood function of Carter & Winn (2009), Equation 41, in which the noise in the light curve is parameterized by two free noise parameters,  $\sigma_w$  and  $\sigma_r$ .  $\sigma_w$  describes the amplitude of uncorrelated (“white”) sources of noise (e.g. photon noise), while  $\sigma_r$  describes the amplitude of the correlated (“red”) noise, which is modeled as a superposition of time-localized oscillating signals known as wavelets.

The power spectral density of the red noise is modeled as  $S(f) \propto 1/f^\gamma$ ; following the authors’ example, we set  $\gamma = 1$ . We calculate the wavelet functions and likelihoods using our own Python module<sup>1</sup>.

We first fit a model to the white light curve of each transit to determine the mid-transit time, then fit each binned light curve independently with the mid-transit time fixed. The fitted white light curves for each night are shown in Figure 2, and the fitted binned light curves are presented in Appendix Figures A1 through A4.

Note that two of the transits exhibit spot-crossing features which have been excluded from the fit. To determine which data points to exclude, we first conservatively remove data near the spot and fit a model to the white light curve, then fit a one-dimensional Gaussian model to the shape of the spot in the residuals, and finally exclude only those data points which lie within two standard deviations of the fitted mean. For a more detailed analysis of these spot features, see Section 5.

The primary output of our fitting routine is a transmission spectrum for 19 independently fitted continuum bins, each approximately 20 nm in width, and three narrower bins centered on possible absorption features, including Na D, H $\alpha$ , and the K I doublet (767/770 nm). We select the bin width to be as small as possible while maintaining signal-to-noise ratios of at least a few hundred in each bin and data set. We attempt to keep the bin sizes consistent across the spectrum, but some bins have been adjusted to accommodate the three narrow bins and two chip gaps. The spectra for each transit are presented in Figure 3, covering wavelengths from 450–900 nm in the first three nights, and 570–900 nm for the fourth night. We have no data from 790–805 nm and 835–850 nm as these correspond to the location of the detector chip gaps in the target and reference star spectra.

##### 4.1. Combining data from separate nights

The average values of  $R_p/R_s$  in the spectrum of Transits 1–3 agree to within  $1.5\sigma$ , but Transit 4 disagrees at the  $\sim 5\sigma$  level. This mirrors the deviation in the fitted orbital parameters for Transit 4 discussed in Section 4.3, and the causes are likely the same.

Regardless of the reason for this disagreement, we are primarily interested in the change in  $R_p/R_s$  with wavelength. We subtract the weighted mean  $R_p/R_s$  from the individual spectra for Transits 2–4, then calculate the average values of  $\Delta R_p/R_s$  in each bin, weighted by the uncertainties for each night. The combined transmission spectrum is presented in Figure 4.

Due to the large uncertainties and incomplete baseline coverage for Transit 1, we do not include its spectrum in the com-

<sup>1</sup> <https://github.com/nespinoza/flicker-noise>

**Table 2.** Relevant previously measured properties of WASP-4 and its companion, with  $1\sigma$  uncertainties.

Parameter	Value	Reference
<b>WASP-4</b>		
$R_s (R_\odot)$	$0.873^{+0.036}_{-0.027}$	Gillon et al. (2009)
$M_s (M_\odot)$	$0.85^{+0.11}_{-0.07}$	...
[Fe/H]	$-0.03^{+0.09}_{-0.09}$	...
$\log(g)$ (cgs)	$4.487^{+0.019}_{-0.015}$	...
$T_s (K)$	$5540^{+55}_{-55}$	Maxted et al. (2011)
<b>WASP-4b</b>		
P (d)	1.33823204	Hoyer et al. (2013)
$R_p/R_s$	$0.15445^{+0.00025}_{-0.00025}$	...
$i$ (deg)	$88.52^{+0.39}_{-0.26}$	...
$a/R_s$	$5.463^{+0.025}_{-0.020}$	...
$e$	$\approx 0$	Beerer et al. (2011)
$R_p (R_J)$	$1.395^{+0.022}_{-0.022}$	Hoyer et al. (2013)
$M_p (M_J)$	$1.237^{+0.021}_{-0.021}$	Winn et al. (2009)

bined result; see Section 4.6. We also correct the spectrum of Transit 3 for known stellar contamination before incorporating it into the combined spectrum, as detailed in Section 4.7.

#### 4.2. Excluded bins

Two of the fitted bins have been excluded from our primary figures and results: the 450–470 nm bin of Transit 2, and the 875–900 nm bin of Transit 4. The values for both bins deviate by  $> 3\sigma$  from the average value of their respective spectra, and both bins are at the low S/N ends of the stellar spectrum. For reference, these bins are included in Appendix Figures A2 and A4 and Table B1.

The 450–470 nm bin of Transit 2 lies  $\sim 3\sigma$  below the average value of the spectrum; we can think of no astrophysical explanation for this effect. Furthermore, the likelihood of a statistical  $3\sigma$  outlier across all of our 82 bins is  $\sim 0.3\%$ . Most likely, our polynomial model is inadequate to describe the low S/N systematic trends in this bin.

The 875–900 nm bin of Transit 4 lies  $\sim 4\sigma$  above the weighted mean. While this wavelength range does correspond with water absorption, such an offset is not observed in the other spectra. Furthermore, the value of this bin depends strongly on whether a linear or quadratic polynomial is used to model the statistics, as discussed in Section 4.4. Again, it seems most likely that our systematics model is inadequate for this low S/N bin.

#### 4.3. Fitting orbital parameters

We also use our MCMC code to simultaneously fit for the transit depth, the inclination ( $i$ ) and the semi-major axis

**Table 3.** Fitted orbital parameters when modelling the planet radius, inclination, and semi-major axis jointly. For comparison we include the values derived by Hoyer et al. (2013).

	$R_p/R_s$	$a/R_s$	$i$ (deg)
Hoyer+13	$0.15445^{+0.00025}_{-0.00025}$	$5.463^{+0.025}_{-0.020}$	$88.52^{+0.39}_{-0.26}$
Transit 1	$0.15386^{+0.00135}_{-0.00174}$	$5.386^{+0.088}_{-0.157}$	$89.33^{+0.47}_{-0.73}$
Transit 2	$0.15617^{+0.00164}_{-0.00167}$	$5.439^{+0.070}_{-0.092}$	$87.90^{+1.24}_{-0.85}$
Transit 3	$0.15388^{+0.00045}_{-0.00047}$	$5.473^{+0.021}_{-0.038}$	$88.94^{+0.53}_{-0.52}$
Transit 4	$0.15654^{+0.00067}_{-0.00065}$	$5.520^{+0.026}_{-0.030}$	$89.46^{+0.37}_{-0.60}$

( $a/R_s$ ), while keeping the eccentricity and period fixed. The results of these fits are listed in Table 3. The parameters from Transits 1–3 are generally consistent with those measured by Hoyer et al. (2013), although the values from the partial Transit 1 are poorly constrained.

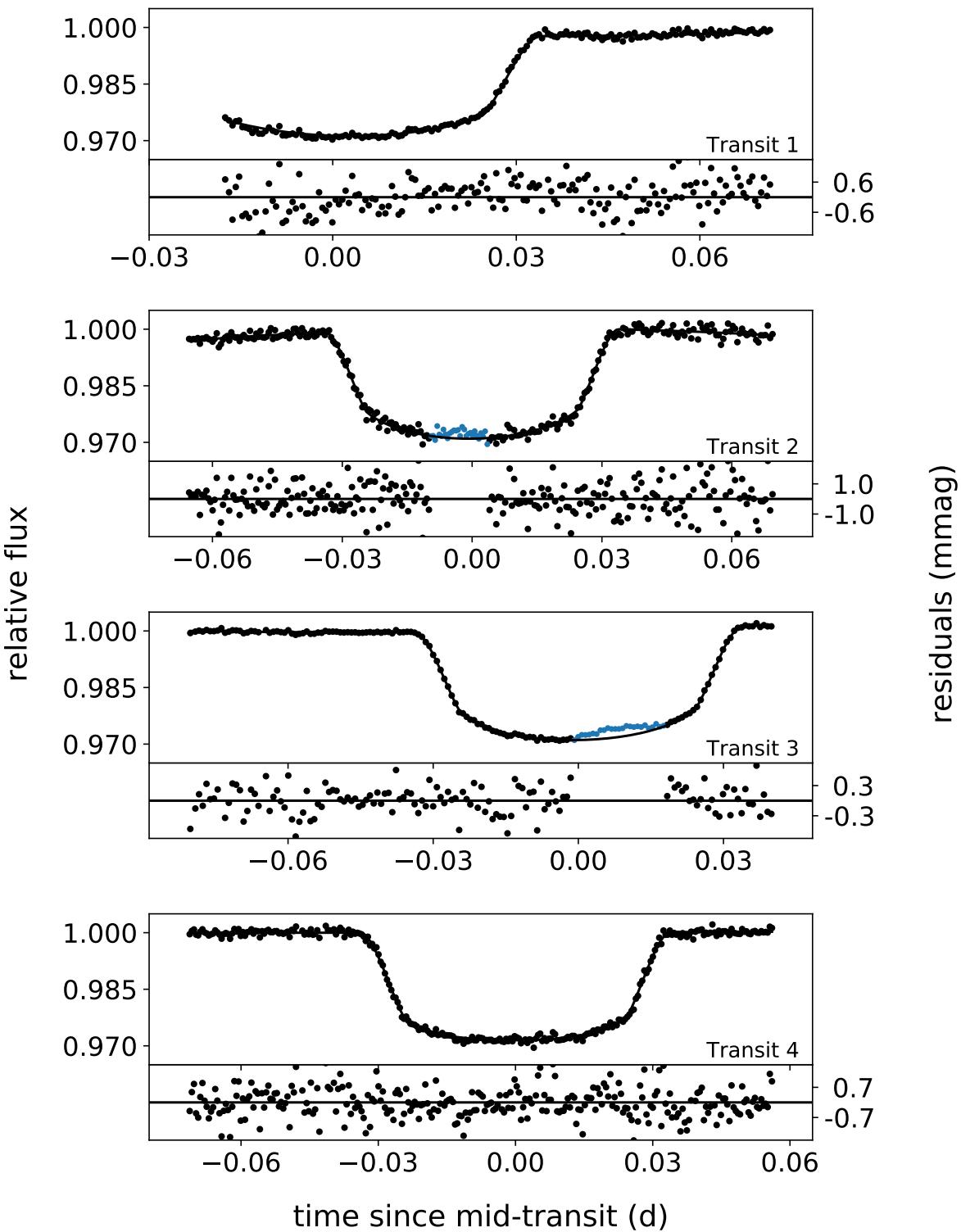
The parameters from Transit 4 deviate  $2–3\sigma$  from the previously measured values. It is not clear why this is the case. Stellar variability may be a partial contributor; as discussed in Section 6.2, WASP-4 is variable at the 6 mmag level, and this can lead to changes in the apparent transit depth. The deviation may also be related to losses due to the shorter slits employed during Transit 4, as discussed in Section 3.7.

Due to this this large deviation, we also try fitting the spectrum of Transit 4 by fixing the inclination and semi-major axis to the fitted values, rather than those in the literature. The result is only a  $\ll 1\sigma$  change in the binned values for  $\Delta R_p/R_s$ , so we believe the spectrum for this night is robust despite the disagreement in the orbital parameters. For consistency with the other data sets, we keep the parameters fixed to the literature values in the remaining sections.

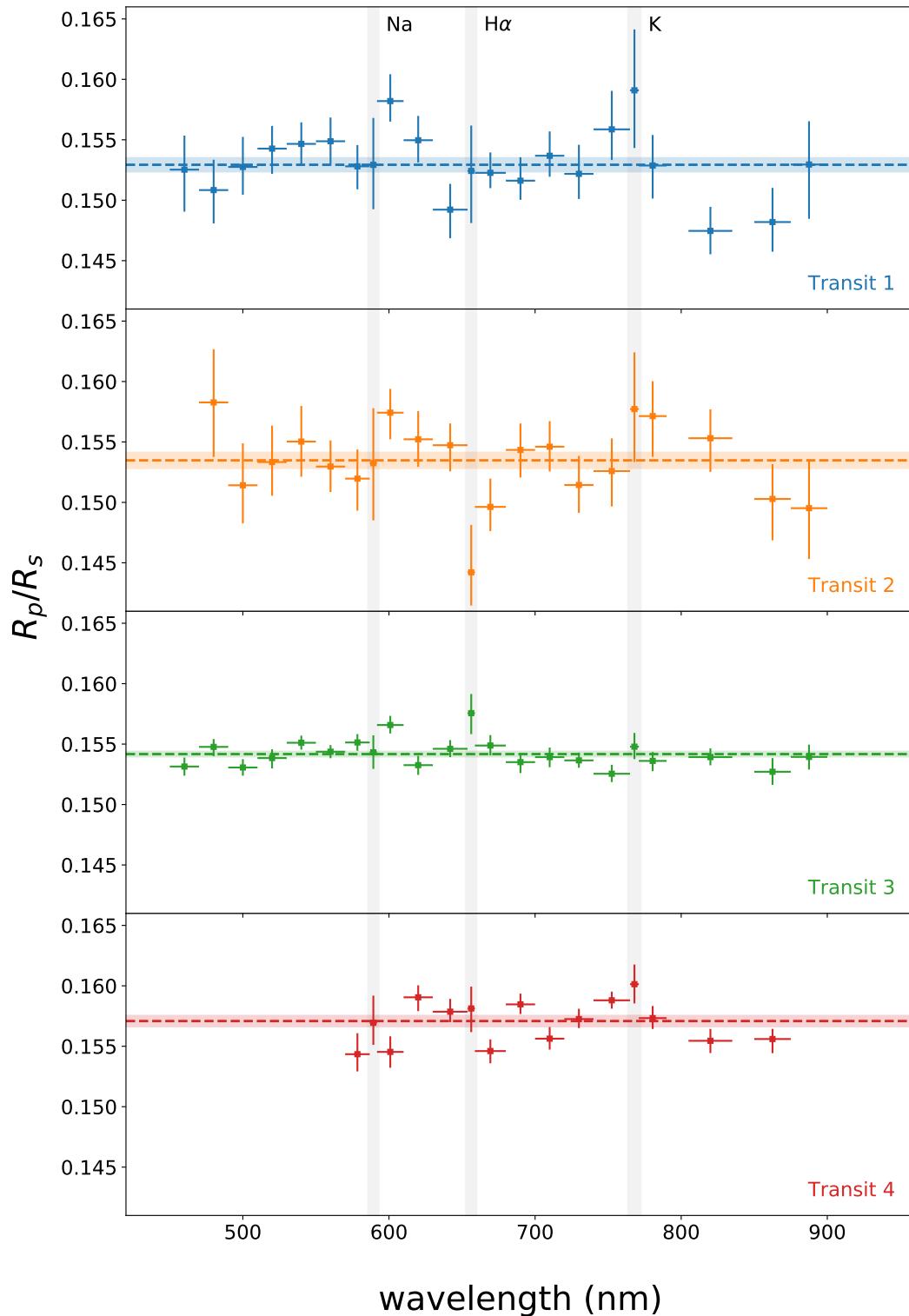
#### 4.4. Systematics model

We find that the trends which remain in the target light curves after dividing by the reference star light curves correlate with airmass, which tends to reach its minimum near the center of the transit. Nevertheless, no physically motivated (e.g., exponential) function of airmass is able to describe this trend better than a simple second-order polynomial in time, so we opt to use time-dependent polynomials to characterize our long-term systematics.

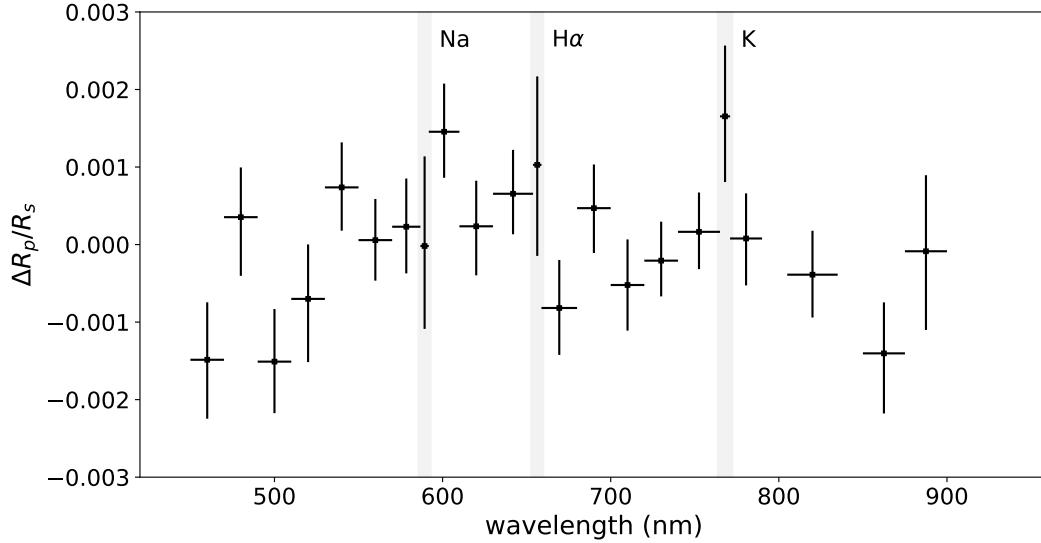
The long-term trends in many of our binned light curves cannot be adequately described by a linear polynomial in time, so we use a quadratic polynomial instead. In principle, any degeneracy between the transit depth and the polynomial value should be incorporated into our MCMC uncertainties. To measure the impact of the model choice on our results, we apply both linear and quadratic systematics models and compare the values of  $\Delta R_p/R_s$  for each.



**Figure 2.** White light curves with transit models and residuals ( $1\sigma$  values marked). Two of the light curves featured spot-crossing events (blue) which are not included in the fit. The long-term trend is modeled by a second-order polynomial.



**Figure 3.** Transmission spectra from each night, with no corrections or offsets applied. The dashed lines and shaded regions represent weighted average values with  $\pm 1\sigma$  uncertainties. The wavelengths of potential atomic features are highlighted.



**Figure 4.** Combined transmission spectrum from Transits 2-4, with the weighted average value subtracted from each night. Narrow bins centered on potential atomic features are highlighted, and the remaining (continuum) bins are black. The spectrum of Transit 3 is corrected for the effect of an occulted spot as described in Section 4.7 before it is incorporated into the combined result. Transit 1 is excluded due to poor phase coverage.

For Transits 1-3, the average absolute change in  $\Delta R_p/R_s$  between the linear and quadratic models is approximately  $0.5\sigma$  per bin, and no larger than  $1.6\sigma$  in any bin. For Transit 4 the average change is approximately  $1.5\sigma$  per bin, but the reddest bin (which has the lowest S/N) changes at the  $3\sigma$  level. Due to this large inconsistency between systematics models, we exclude the reddest bin from our results.

Some authors use the residuals to their white light curve models to calibrate out common (i.e., wavelength-independent) systematics from each bin (e.g., Sing et al. 2013; Nikolov et al. 2015; Huitson et al. 2017). In our case we find that such a correction is unwarranted. By eye, it is not apparent that our binned light curves share any common systematics (see Figures A1-A4 in the Appendix). Furthermore, when we apply the common mode correction, we find that the impact on the resulting transit spectra is negligible compared to the uncertainties.

#### 4.5. Red noise

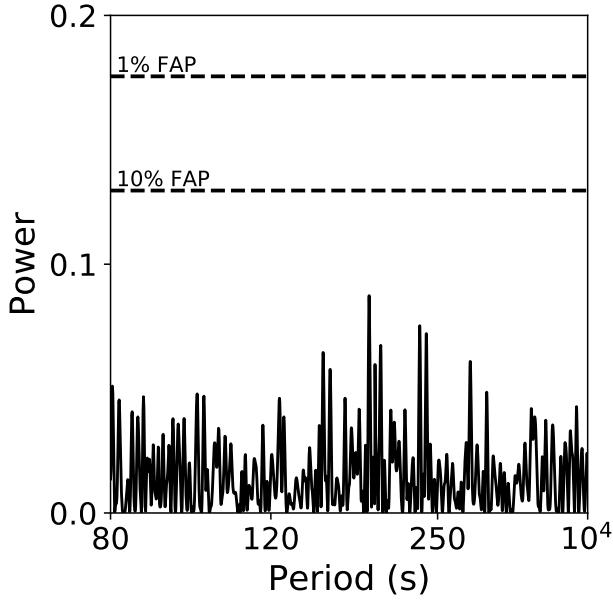
As discussed in Section 4, we fit a parameter  $\sigma_r$  to characterize the level of correlated (“red”) noise in our data. While the ratios in Table 1 suggest a large amount of systematic noise, this does not necessarily imply *correlated* noise; the systematic noise may still be captured by the white noise parameter.

To assess the level of red noise in our data, we construct Lomb-Scargle periodograms of the residuals to each binned light curve model, then use bootstrap resampling to calculate the false alarm probability (FAP; e.g., VanderPlas 2018) for signals with periods between the exposure times and the duration of our observations. Figure 5 shows an example periodogram for one of the bins of Transit 3. We find that only 6 out of 82 binned light curves possess signals with  $\text{FAP} < 10\%$ , and only 2 have  $\text{FAP} < 1\%$ . This suggests that there are few strong periodic signals in our light curves. Nevertheless, we conservatively choose to apply the red noise parameterization, which results in a  $\sim 15\%$  increase in our uncertainties.

#### 4.6. Effects of incomplete phase coverage

The light curves of Transits 1-3 all suffer from a partial lack of phase coverage, due either to poor weather, observing window constraints, or spot-crossing events which must be masked in the model fit. In principle, our MCMC approach to fitting the model should adequately incorporate this lack of information into the uncertainties on  $R_p/R_s$ . However, it is worthwhile to estimate the extent to which this missing information inflates our uncertainties.

The full light curve of Transit 4 allows us to investigate the effects of poor phase coverage in the previous three transits. Specifically, we remove data points from the binned light curves of Transit 4 to mimic the phase coverage of Transits



**Figure 5.** Lomb-Scargle periodogram of the residuals to the light curve model for the 610–630 nm bin of Transit 3. The lack of strong signals demonstrates that this light curve does not have strong periodic correlated noise across this range of frequencies.

1–3, then fit for the transit depth and calculate the resulting increase in the uncertainties on the binned values of  $R_p/R_s$ . Our results are summarized as follows:

- Mock Transit 1: We remove the pre-ingress baseline and most of the first half of the transit. This results in a 150% increase in uncertainties, as well as a  $4.5\sigma$  decrease in the wavelength-averaged radius. The  $\Delta R_p/R_s$  values change at the  $> 1\sigma$  level, and the change appears to be wavelength dependent, introducing a slope on the order of  $\sim 1\sigma$  from blue to red.
- Mock Transit 2: We remove the points near mid-transit which correspond to a spot-crossing event. This results in a 50% increase in uncertainties, and a  $0.5\sigma$  decrease in the average radius. The  $\Delta R_p/R_s$  values change at the  $\lesssim 0.2\sigma$  level.
- Mock Transit 3: First, we remove most of the post-egress baseline, which results in a 50% increase in uncertainties with no considerable change in the average radius, and  $\lesssim 0.5\sigma$  changes in  $\Delta R_p/R_s$ . Next, we exclude only those points which correspond to the spot crossing event in Transit 3, finding effects of similar magnitude. Removing both portions of the light curve, however, does not increase the uncertainties any further.

The results for Transit 1 lead to concerns over whether the lack of a pre-ingress baseline could introduce a bias into the shape of our transit spectrum. Furthermore, the large uncertainties mean that this transit contributes little to the combined spectrum. We therefore opt to exclude the spectrum of Transit 1 from the combined result.

Meanwhile, the results for Transits 2 and 3 suggest that little bias has been introduced into the transit spectrum due to the lack of phase coverage in each, while the uncertainties on  $\Delta R_p/R_s$  should be larger by  $\sim 50\%$  than they would be given full phase coverage.

#### 4.7. Correcting for occulted spots

The light curves of Transits 2 and 3 feature prominent spot occultation features which we exclude from the light curve model fit. While doing so allows us to fit the apparent transit depth without simultaneously modeling the spot, it does *not* eliminate the impact of the occulted spot on the resulting transit spectrum. Since the spot is cooler than the rest of the photosphere, it breaks the transit model’s assumption that the photosphere is azimuthally symmetric and amplifies the fitted transit depth in a wavelength-dependent manner (e.g., Deming et al. 2013). In previous works we have referred to this as the transit light source effect (TLSE); here we derive an approximate correction (ignoring limb-darkening) for the TLSE due to a single spot.

Assuming the photosphere to have specific flux  $F_{\lambda,\text{phot}}$  and radius  $R_s$ , the spot to have specific flux  $F_{\lambda,\text{spot}}$  and radius  $R_{\text{spot}}$ , and the planet to have apparent radius  $R_p$  and true radius  $R_{p,0}$ , then the observed relative transit depth will be

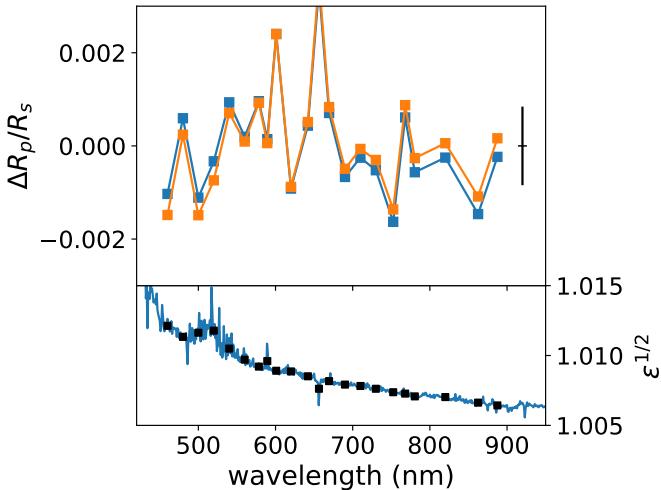
$$\left(\frac{R_p}{R_s}\right)^2 = \frac{F_{\lambda,\text{phot}}R_{p,0}^2}{F_{\lambda,\text{phot}}(R_s^2 - R_{\text{spot}}^2) + F_{\lambda,\text{spot}}R_{\text{spot}}^2} \quad (1)$$

By re-arranging this equation, we derive a corrective factor for the transit depth<sup>2</sup>:

$$\epsilon_\lambda \equiv \left(\frac{R_p}{R_{p,0}}\right)^2 = \left[1 - \left(1 - \frac{F_{\lambda,\text{spot}}}{F_{\lambda,\text{phot}}}\right) \left(\frac{R_{\text{spot}}}{R_s}\right)^2\right]^{-1} \quad (2)$$

Later, in Section 5, we find the feature in Transit 3 to be best described by a model of a single spot with radius  $R_{\text{spot}}/R_s = 0.27^{+0.03}_{-0.02}$  and a temperature contrast of  $T_s - T_{\text{spot}} = 100 \pm 5$  K. While this feature could in principle be due to a mixture of occulted spots and faculae, our data are insufficient to constrain more complex models, so we use the parameters of the single spot model for the purposes of correcting for the TLSE in this spectrum. We interpolate a PHOENIX (Husser et al. 2013) stellar photosphere model

<sup>2</sup> See also Equation 1 of McCullough et al. (2014), Equation 2 of Rackham et al. (2018).



**Figure 6.** (Top) The spectrum of Transit 3 before (blue) and after (orange) correcting for the large occulted spot. Both spectra have been mean-subtracted. The black error bar demonstrates the mean binned uncertainty, while the individual error bars are excluded for visibility. (Bottom) The radius contamination factor due to the spot, binned to 1 nm. The black squares are binned to match the spectrum above.

grid onto the values in Table 2 to compute  $F_{\lambda,\text{phot}}$  and  $F_{\lambda,\text{spot}}$  at their respective temperatures, then calculate  $\epsilon_\lambda$  from Equation 2.

In Figure 6 we plot  $\epsilon^{1/2}$ , which is the corrective factor for the planet *radius*, as well as the original and corrected spectra for Transit 3. The magnitude of the correction from 450 to 900 nm is approximately equal to the binned uncertainty. The corrective factor is listed in Table B1, binned to match our final spectrum. This corrective factor is not applied to the spectrum of Transit 3 in Figure 3, but is applied before creating the combined spectrum in Figure 4.

We do not derive a similar correction for Transit 2 because it is unclear (see Section 5) whether this feature is best modeled by a small or large spot, and the slope introduced by the large spot model is six times steeper across the wavelength range than that of the small spot. Furthermore, the large spot model would only introduce a  $\sim 0.3\sigma$  increase in the measured radius from the reddest to bluest bins.

Finally, it is worth noting that this effect is just as prominent for spots outside of the transit chord. Equation 2 can be further generalized by replacing  $(R_{\text{spot}}/R_s)^2$  with  $F_{\text{het}}$ , the areal spot covering fraction of the photosphere, and using an average spot temperature to compute  $F_{\lambda,\text{spot}}$ .

## 5. STARSPOTS

In Figure 2 we have identified spot-crossing features in the light curves of Transits 2 and 3, and we exclude these data points from our analysis of the transmission spectrum. Such features are common in transit light curves, and have

previously been found in transit observations of WASP-4b (Sanchez-Ojeda et al. 2011). The shape, central time, and magnitude of the feature carry information about the size, position, and contrast (or effective temperature) of the occulted spot, which are otherwise difficult to constrain. Transit spectroscopy allows us to measure the spot contrast at multiple wavelengths, permitting a more precise measurement of the spot’s effective temperature than is available from wavelength-integrated light curves.

The measurement of spot properties is worthwhile for a variety of reasons. In the context of this paper, knowing the properties of an occulted spot allows us to correct for its effect on the spectrum of the transit during which it is observed. More generally, observations of multiple spots across an extended period of time permit a better understanding of the makeup of the photospheres of other stars.

### 5.1. Spot modeling

We use SPOTROD (Béky et al. 2014) to produce light curves of transits with one or two spot-crossing events. In addition to the usual transit parameters, SPOTROD models the position and radius (in units of stellar radii) as well as the average spot-to-stellar flux contrast for each spot. We employ PyMultiNest (Buchner et al. 2014) to sample the parameter space and to calculate the Bayesian log-evidences  $\ln(Z)$  through which we can compare the different models, as discussed in Section 6.1. This MultiNest implementation<sup>3</sup> of SPOTROD has also been used to study spots observed during transits of WASP-19b (Espinoza et al. 2019).

We use a second-order polynomial to detrend the white light curve from each night, then fit the combined spot and star model. We limit the uniform prior on spot size to  $R_{\text{spot}}/R_s < 0.08$  or  $0 < R_{\text{spot}}/R_s < 1$  to probe for small and large spots. Using the effective temperature in Table 2 as a prior, the code then constructs and fits models for the flux contrast as a function of wavelength.

The optimal model parameters for each spot feature are presented in Table 4. The precision of the Transit 2 data only permits us to fit the spot feature and spectrum in a single, white-light bin. For this night, we find that our data are consistent with one- and two-spot solutions, as well as large and small spots. We opt for the simpler single-spot models and present the fit for both a single large and single cool spot. The data for Transit 3 favor a single large spot versus multiple or smaller spots ( $\Delta \ln(Z) > 2$ ). For this transit we are able to fit the spot spectrum to the same bins as our transmission spectrum, as shown in Figure 7.

The large spot model for Transit 3 seems inconsistent with smaller individual spots observed on the Sun; nevertheless, such wide features have been observed in transits before

<sup>3</sup> <https://github.com/nespinoza/spotnest>

**Table 4.** Best-fit parameters for the spot-crossing events in the light curves of Transits 2 and 3. The data for Transit 2 are consistent with both a small (2a) and large spot (2b). Median and  $1\sigma$  confidence intervals are reported.

	$R_{\text{spot}}/R_s$	$f_{\text{spot}}/f_s$	$T_s - T_{\text{spot}}$ (K)	$T_{\text{spot}}$ (K)
Transit 2a	$0.05^{+0.01}_{-0.01}$	$0.49^{+0.20}_{-0.28}$	$880 \pm 360$	$4670 \pm 360$
Transit 2b	$0.15^{+0.11}_{-0.07}$	$0.79^{+0.11}_{-0.18}$	$410 \pm 280$	$5140 \pm 280$
Transit 3	$0.27^{+0.03}_{-0.02}$	$0.91^{+0.01}_{-0.01}$	$100 \pm 5$	$5442 \pm 50$

(e.g., Espinoza et al. 2019). While our evidences do not favor a two-spot model, it is possible that the large, low-contrast spot is in fact a more complex active region consisting of multiple spots and/or faculae, as has been observed on the Sun. Regardless of the actual structure of the feature, its average contrast is well-described by a single spot with  $T_s - T_{\text{spot}} = 100$  K, and we choose this model to calculate the correction in Section 4.7.

### 5.2. Spot sizes and temperatures

Later, in Section 6, we assume a spot temperature in order to model the effects of unocculted spots on the transit spectrum. The temperature we assume reflects small, cool spots resembling those on the Sun, but the results for Transit 3 in Table 4 appear to suggest the existence of large, warm features on the photosphere of WASP-4. In short, we do not believe that the spots discovered in our transit light curves are necessarily representative of the spots present elsewhere in the photosphere or below our detection threshold. We offer these arguments as justification:

1. Sanchis-Ojeda et al. (2011) find evidence for a persistent small spot in multiple transit observations of WASP-4b. In this case, they find a spot with radius  $R_{\text{spot}}/R_s \approx 0.05$  and temperature contrast  $T_s - T_{\text{spot}} \approx 600$  K, suggesting that more Sun-like spots do exist on WASP-4.
2. Cool spots have previously been discovered through Doppler imaging of active G and K dwarfs (e.g., O’Neal et al. 1998, 2004; Strassmeier 2009). However, this method of spot detection is subject to biases regarding spot size and temperature.
3. The feature occulted during Transit 3 may indeed have a more complex structure consisting of cool spots and hot faculae which the precision of our data does not allow us to resolve.
4. While the precise radius and temperature of the spot observed during Transit 2 are largely unconstrained, we can say with at least  $1\sigma$  confidence that the spot

is smaller and cooler than the feature observed during Transit 3. Therefore, even our data suggest that small spots may be present in the photosphere of WASP-4.

It is nevertheless possible that the distribution of spot sizes on WASP-4 favors larger and hot spots, in which case the spectral contamination signal would be less severe. However, our use of the contamination model is not meant to accurately account for the actual surface heterogeneity, but rather to demonstrate that this aspect should not be ignored in determining the planet’s transmission spectrum.

## 6. ATMOSPHERIC RETRIEVAL

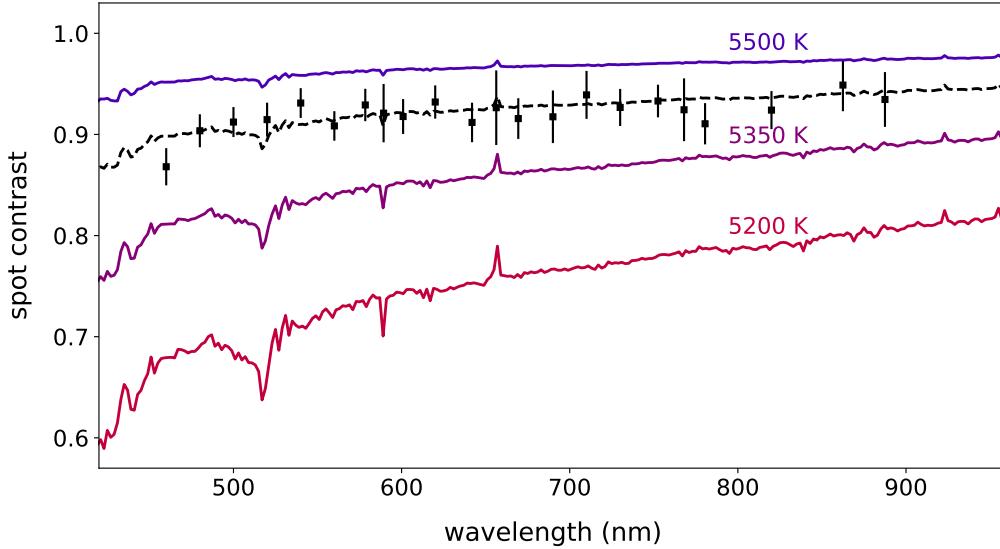
To interpret the combined transmission spectrum in Figure 4, we employ a Bayesian atmospheric retrieval code based on PyMultiNest which has previously been used to study WASP-19b (Espinoza et al. 2019). Following Bétrémieux & Swain (2017) and Heng & Kitzmann (2017), our atmosphere model includes two regions: an optically thick base region (which could also be interpreted as a cloud layer) with radius  $(R_p/R_s)_0$  and reference pressure  $P_0$ , and an isothermal optically thin region above with temperature  $T$ . The components of the optically thin region may include either or both of (i) a set of atomic and molecular species and (ii) a scattering haze defined by a cross-section power law  $\sigma = a\sigma_0 (\lambda/\lambda_0)^\gamma$ , where  $\sigma_0 = 5.31 \times 10^{-27} \text{ cm}^2$  is the Rayleigh scattering cross-section of H<sub>2</sub> at the reference wavelength  $\lambda_0 = 350$  nm (MacDonald & Madhusudhan 2017), and  $a$  is a dimensionless enhancement factor. We constrain  $\gamma$  to be between 0 (uniform opacity) and -4 (Rayleigh scattering), which should span the range of physical scattering processes.

As we have previously argued in Rackham et al. (2017, 2018), it is important that analyses of transmission spectra account for the transit light source effect (TLSE) which can be introduced by heterogeneous features (spots and faculae) on the stellar surface. We incorporate a three-parameter model for the stellar photosphere into our retrieval code, fitting this model simultaneously with that of the planet’s atmosphere. Section 6.2 discusses this model in more detail.

The parameters and priors for each component of the model are listed in Table 5. We consider eight models for the atmosphere, including all possible combinations of a cloud deck, a scattering haze, and Na and/or K absorption signals. Each of these is fitted independently or alongside a photosphere model (+1 free parameter,  $F_{\text{het}}$ ) for a total of sixteen models.

### 6.1. Model comparison

To compare between our models, we compute their relative Bayes factors. A thorough introduction to the use of Bayes factors in astronomy is given by Trotta (2008), and their use in transit spectroscopy is well-established (e.g., Benneke & Seager 2013). Here, we provide a brief overview.



**Figure 7.** Best-fit model for the occulted spot in the light curve of Transit 3. The data are best described by a single large, warm spot with a radius of  $0.27 R_s$  and a temperature contrast of 100 K.

Given a model  $M$  with parameters  $\theta$ , the optimal values of  $\theta$  to describe data values  $x$  are those which maximize Bayes' equation:

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

where  $P(x|\theta, M)$  is the likelihood function and  $P(\theta|M)$  is the prior distribution of  $\theta$ . Most Bayesian model fitting algorithms seek to maximize the likelihood function, but the maximum likelihoods of two different models cannot be directly compared.

To compare two models, we first compute their Bayesian evidences, defined as:

$$Z_M \equiv \int P(x|\theta, M)P(\theta|M)d\theta$$

The evidence for a model is high if the data are well-described by a large region of the prior parameter space. The Bayes factor between two models is defined as the ratio of their evidences:

$$\ln(B) \equiv \ln(Z_2) - \ln(Z_1)$$

In general, model 2 is weakly favored over model 1 if  $0 < \ln(B) < 2$ , and strongly favored if  $\ln(B) > 5$ . If  $\ln(B) < 0$  then model 2 is disfavored to model 1. Complex models will be favored over simpler models only if the evidence justifies the additional parameters. However, a preference towards simpler models does not necessarily mean that the complex

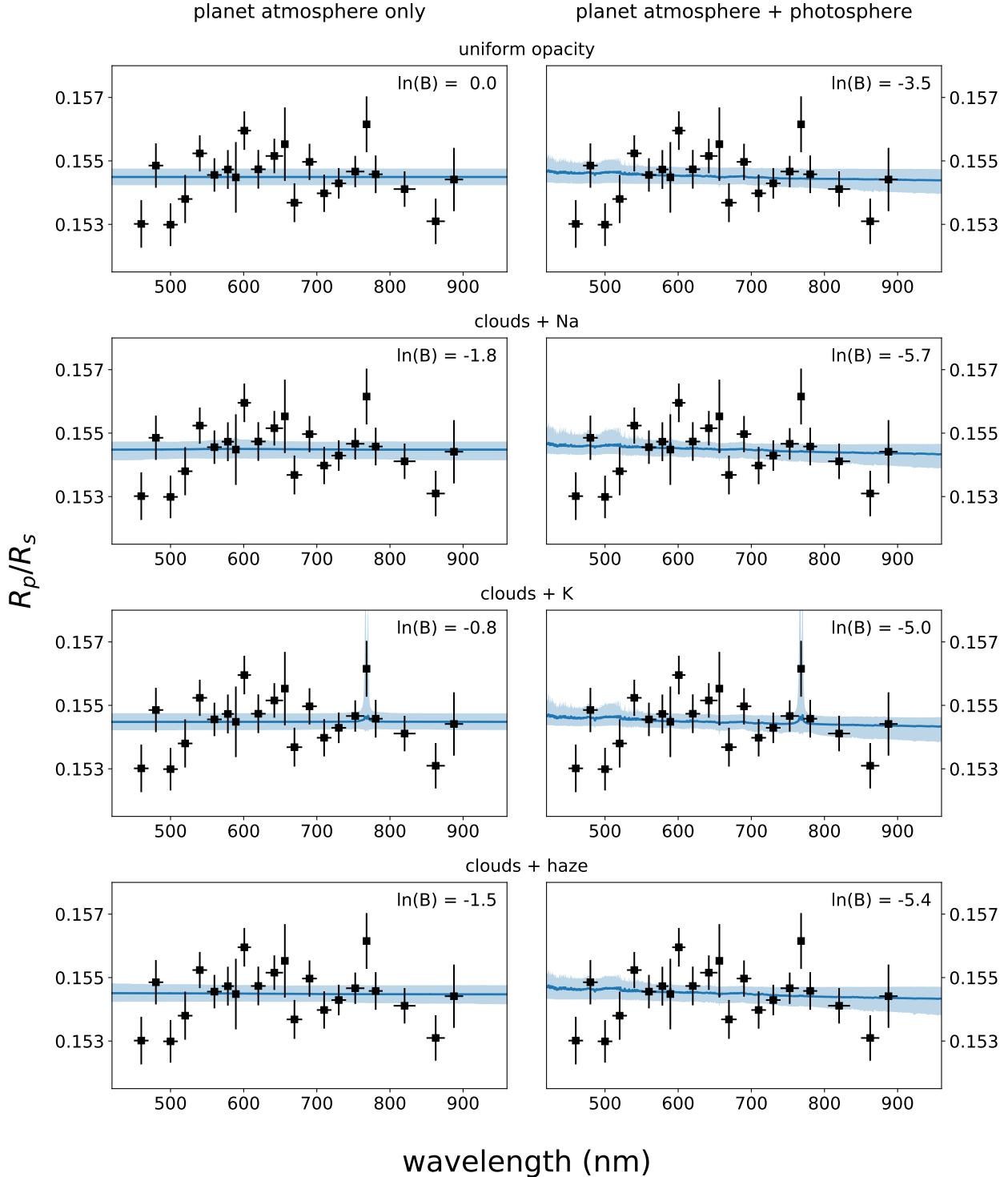
models are incorrect, but rather that they are not justified by the data at hand.

Our retrieval algorithm computes the Bayesian evidence parameter for each of the fitted models. In the following sections, we cite the Bayes factor  $\ln(B)$  for each model with respect to the model for a uniform opacity atmosphere with no contamination from the photosphere (i.e., a flat line). In general, we find that the more complex models have  $\ln(B) < 0$ , meaning that they are *disfavored* versus a flat line.

## 6.2. Stellar heterogeneity

Assessing the level of TLSE contamination is critical to correctly interpreting high-precision transmission spectra of transiting exoplanets (Apai et al. 2018). This is particularly important for investigating features that could have a planetary or stellar origin, such as the He 10,833 Å line (Spake et al. 2018) or the Na and K alkali lines (Cauley et al. 2018). We have previously demonstrated this effect in the spectrum of GJ 1214b (Rackham et al. 2017), which has a featureless near-IR spectrum but a visible spectrum that is apparently offset below the near-IR transit depths. In this study, we found that for reasonable assumptions about the star's activity, the planetary transmission spectrum of GJ 1214b can be shown to be flat in the visible as well. We also detect the signature of stellar activity in one of our *Magellan/IMACS* transit data sets for WASP-19b (Espinoza et al. 2019).

WASP-4 is of slightly later spectral type than the Sun, and multiple spot-crossing events have been observed during pre-



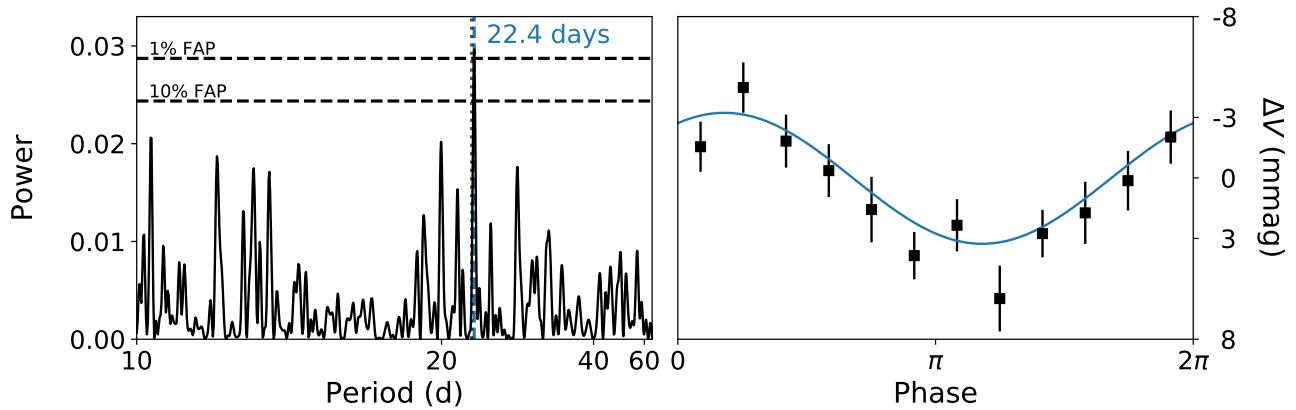
**Figure 8.** A subset of fitted models for the observed transmission spectrum of WASP-4b; four separate models for the planetary atmosphere are presented alone (left) or combined with a model for the heterogeneous stellar photosphere (right). The atmosphere models, described in more detail in Section 6, include (top) a uniform opacity atmosphere, (middle) a cloud deck and *Na* absorption feature, a cloud deck and *K* absorption feature, and (bottom) a cloud deck and scattering haze. The shaded region represents the 95% confidence interval around the mean model fit. Bayes factors are given for each model relative to the flat line model (upper left). A constant offset of +0.1545 has been added to the combined spectrum.

**Table 5.** Parameters for our combined photosphere and atmosphere models. Not all parameters are incorporated in all models; rather, Bayesian log-evidences are used to compare separate models.

Model component	Parameter	Units	Description	Prior distribution
Base	$(R_p/R_s)_0$	-	Radius corresponding to the top of the cloud layer or $\tau \gg 1$	Uniform(0.1,0.2)
	$P_0^\dagger$	bar	Reference pressure at $(R_p/R_s)_0$	Log-uniform( $10^{-6}$ , $10^6$ )
	$T^\dagger$	K	Average temperature of the optically-thin region	Uniform(1000,2400)
Atomic features	$X$	-	Mixing ratio of species X	Log-uniform( $10^{-30}$ ,1)
Haze	$a$	-	Amplitude of the haze cross-section power law	Log-uniform( $10^{-20}$ , $10^{10}$ )
	$\gamma$	-	Index of the haze cross-section power law	Uniform(-4,0)
Stellar photosphere	$T_{\text{occ}}$	K	Average temperature of the transit chord <sup>‡</sup>	Fixed(5540)
	$T_{\text{het}}$	K	Average temperature of the heterogeneous surface features	Fixed(4200)
	$F_{\text{het}}$	-	Fraction of the unocculted photosphere covered by spots	Uniform(0,1)

<sup>†</sup>The temperature and reference pressure are not modeled in the case of a uniform opacity atmosphere.

<sup>‡</sup>This excludes any occulted features which can be identified in the light curve.



**Figure 9.** (Left) Lomb-Scargle periodogram for the V-band magnitude of WASP-4 from ASAS-SN photometry. The largest peak, marked with a blue line, corresponds to the previously measured rotation period, marked with a black dotted line. The horizontal lines denote the false alarm probability calculated through the bootstrapping method. (Right) Fitted model for the phase-folded light curve, assuming a period of 22.4 days. The data are binned for visibility.

vious transit observations of WASP-4b (Sanchis-Ojeda et al. 2011) as well as in two of our data sets. To assess the photometric variability, we retrieve four years of photometric monitoring data from the ASAS-SN online database<sup>4</sup> (Shappee et al. 2014; Kochanek et al. 2017). We remove outliers with a  $5\sigma$  cut and create a Lomb-Scargle periodogram to search for periodic signals between 1 and 100 days, calculating the false alarm probability as in Section 4.5. As shown in Figure 9, we find the largest peak at 22.4 days, and while the significance of the detection is only moderate, it corresponds well with the rotational period of 22.2 days which Sanchis-Ojeda et al. (2011) measure from consecutive spot-crossing events. This is to be expected if the photometric variability is dominated by the rotation of spots and faculae in and out of view. Modeling the variability as a sine curve and assuming a period of 22.4 days, we perform a least-squares fit to the phase folded data to find a peak-to-trough variability amplitude of  $\sim 6$  mmag (0.6%). These results suggest that the surface of WASP-4 is moderately heterogeneous, so any analysis of the transmission spectrum of WASP-4b which ignores stellar contamination may be overly optimistic. Furthermore, since the spot covering fraction can not be determined from the variability alone (Rackham et al. 2018), the photosphere heterogeneity should be modeled alongside the atmosphere.

In the retrieval code described above, we characterize the heterogeneity with three parameters:  $T_{occ}$ , the effective temperature of the transit chord,  $T_{het}$ , the mean effective temperature of the heterogeneous features, and  $F_{het}$ , the fraction of the unocculted photosphere which is covered in spots. We then compute the effect on the observed transit depth following Section 4.7 and Equation 2, but replacing the area ratio  $(R_{\text{spot}}/R_s)^2$  with a covering fraction  $F_{het}$ .

We test more complex parameterizations which include multiple types of unocculted heterogeneities (e.g., Zhang et al. 2018), but determine from the evidences that they are not warranted by the data.

Given the quality of our data, we find it necessary to fix some parameters of the heterogeneity model to reasonable values. In Table 5 we fix the spot temperature to match typical spots on the Sun (e.g., Solanki 2003, and references therein), and the transit chord temperature to the measured effective temperature of the photosphere.

$F_{het}$  is allowed to vary over all possible values, and serves as a simple measure of the level of stellar contamination in the transmission spectrum. However, we note that  $F_{het}$  represents an *enhancement* over the effect of the large occulted spot in Transit 3, which has already been corrected for in Section 4.7.

### 6.3. Retrieval results

In Figure 8 we present four of our atmosphere models with and without photosphere contamination. In Table 6 we list the Bayes factors for all sixteen of the models tested.

#### 6.3.1. Favored models

The evidences favor a uniform opacity model for the atmosphere, with other atmosphere models being disfavored by  $\ln(B) = -1$  to  $-2$ . Even when the haze is included, the amplitude  $a$  of the haze opacity is small; the posterior distributions do not converge well enough from the log-uniform prior to offer a meaningful upper limit, but the mode of the distribution for  $a$  lies between  $10^{-18}$  and  $10^{-12}$ , which is lower than in the prior distribution.

Atmosphere models including Na and/or K are disfavored versus uniform opacity models by  $\ln(B) = -1$  to  $-2$ . Even though the narrow K bin is elevated above the continuum by  $\sim 1\sigma$ , ultimately the presence of potassium is not warranted by the data. We note that we cannot place upper limits on the abundances of Na or K, as these are degenerate with the reference pressure.

The models which include a heterogeneous photosphere are strongly disfavored by  $\ln(B) = -5$  to  $-6$  versus their counterparts with a homogeneous photosphere. While it is quite likely that  $F_{het} > 0$ , the spectrum *alone* does not reveal evidence for it beyond what we have already corrected for in Section 4.7. The effect of even a low spot covering fraction ( $< 3\%$ ) is apparent in Figure 8, and is degenerate with the effect of a scattering haze. For this reason, we recommend the joint modeling of the photosphere and atmosphere in future analyses of optical transmission spectra, and the use of Bayesian evidences or likelihood parameters for model selection (see also Pinhas et al. 2018).

#### 6.3.2. Constraints on spot covering fraction

The parameters in the clear and hazy atmosphere models do not converge enough from their priors to offer meaningful constraints. However, in the special case of a uniform opacity atmosphere with stellar contamination, we can place a 95% upper limit on the spot covering fraction,  $F_{het} < 3.4\%$ . We caution that this is under the assumption of 4200 K spots and does not factor in the likely diversity of spot and faculae temperatures on the stellar surface. Instead, it should be taken as a limit on the net spectral effect of contamination.

### 6.4. Correcting vs. fitting the contamination in Transit 3

In Section 4.7 we detail a method for correcting the contamination due to the cool feature which was occulted by the planet during Transit 3. In this retrieval, however, we also include a varying parameter  $F_{het}$  and fixed parameter  $T_{het}$  to characterize the heterogeneity of the photosphere. We offer the following argument to justify our decision to visit the heterogeneity twice.

<sup>4</sup> <https://asas-sn.osu.edu/>

**Table 6.** Bayes factors for the full suite of atmosphere models, some of which are shown in Figure 8. The two columns represent atmosphere-only (A) and combined atmosphere-photosphere (A+P) models.  $\ln(B)$  is calculated relative to the flat line model; since  $\ln(B) < 0$  for the more complex models, they are all disfavored versus a flat line.

Model	$\ln(B)_A$	$\ln(B)_{A+P}$
uniform opacity	0.0	-3.5
clouds + Na	-1.8	-5.7
clouds + K	-0.8	-5.0
clouds + Na,K	-1.0	-5.4
clouds + haze	-1.5	-5.4
clouds + haze + Na	-1.9	-6.2
clouds + haze + K	-1.9	-5.4
clouds + haze + Na,K	-1.9	-6.2

First, while the detailed structure of the occulted feature from Transit 3 is unclear, it is well-described by a circular spot model with a constant temperature, and this is the same model which we use to correct for its contamination. By excluding this correction, we would not be leveraging all of the available information from our light curve. However,  $F_{het}$  must still be non-zero to describe the remaining unocculted features.

Second, we demonstrate that this correction is accounted for during the atmospheric retrieval. As a test, we fit the model of a uniform opacity atmosphere with a heterogeneous photosphere (Figure 8, top-right), this time fixing  $T_{het} = 5442$  K to match the temperature of the occulted feature. When we fit the uncorrected combined spectrum, we find that the median of the posterior distribution for  $F_{het}$  is larger than for the corrected combined spectrum. The difference is consistent with a single spot of size  $R_{spot}/R_s \approx 0.22$ . Similarly, when the retrieval is applied to the corrected and uncorrected spectra from Transit 3 only, the difference in the median value of  $F_{het}$  is consistent with a single spot of size  $R_{spot}/R_s \approx 0.33$ . Both of these are similar to the spot size assumed in the correction,  $R_{spot}/R_s = 0.27^{+0.03}_{-0.02}$ . This indicates that the effect of the correction is carried forward into the retrieval; we are not over-correcting for the effect of the spot.

## 7. COMPARISON TO PUBLISHED RESULTS

### 7.1. Gemini/GMOS

Huitson et al. (2017) (hereafter H17) have previously published an optical transmission spectrum of WASP-4b from four combined transit observations with *Gemini*/GMOS, and find that their data favor a uniform opacity model. In this

section, we assess the agreement of their results with our data from IMACS. In Figure 10 we compare the red and blue GMOS spectra with the combined spectrum from this work, and find that the slopes of our spectra are generally in agreement.

#### 7.1.1. Absorption features

The H17 data reveal tentative evidence for Na absorption at 589 nm. Our data reveal no evidence for such a feature, but we concede that the quality of our data in the blue is not sufficient to definitively rule out an atomic feature of small optical depth.

H17 exclude bins in their spectra which are coincident with terrestrial telluric  $O_2$  absorption, finding that their correction for differential atmospheric refraction was not reliable in this wavelength range. Magellan is equipped with an ADC so our data do not require this correction (see Section 3.4). As a result, we are able to test for a K I feature in the same region, and find little evidence as discussed in Section 6.

## 7.2. Magellan/IMACS

May et al. (2018) (hereafter M18) used *Magellan*/IMACS to study this target with the same grism as we used for Transits 1-3, and they also find no evidence for features in the transit spectrum. Figure 10 demonstrates that our results are in agreement.

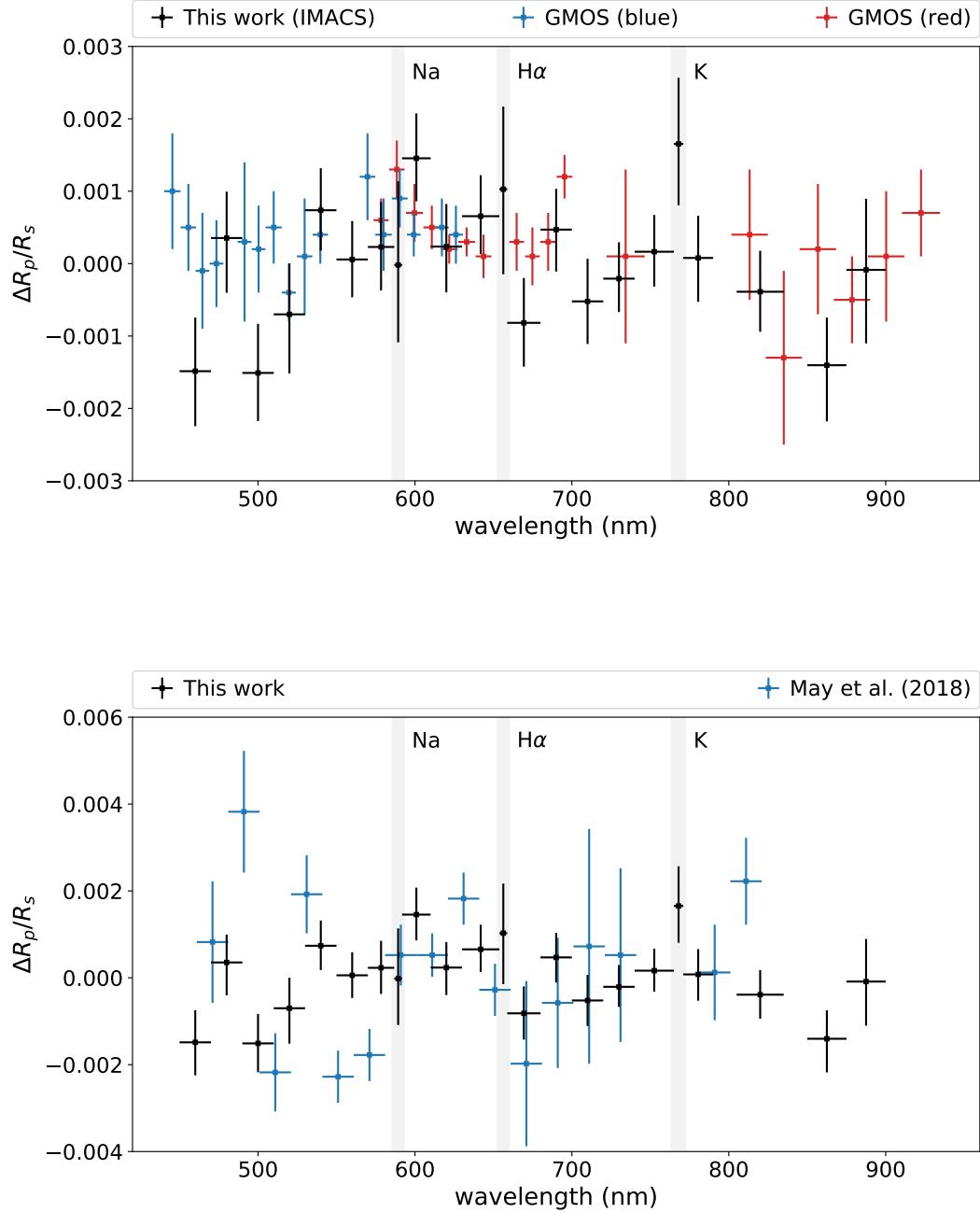
#### 7.2.1. Lack of a spot crossing event

The transit observed by M18 occurred on 19 August 2015, approximately 5.4 days after our Transit 3, though their light curves show no evidence for a large occulted photosphere feature. This is not unexpected: the planet's period is prograde and nearly equatorial, and the stellar rotation period is  $\sim 22.2$  days (Triaud et al. 2010; Sanchis-Ojeda et al. 2011). This implies that the feature observed during Transit 3, which was occulted in the second half of the transit, would have rotated  $\sim 90$  degrees and off the projected stellar disk within the period of time between the transits.

## 7.3. Combined analysis

We combine the data from H17 and M18 with our own by subtracting the weighted mean from each spectrum, then repeat the retrieval analysis of Section 6. Table 7 displays the Bayes factor relative to a flat line for each of the 2x8 models. As before, we find that most atmosphere models are disfavored versus a uniform opacity model by  $\ln(B) = -1$  to  $-2$ , but models including Na are slightly less disfavored than in Table 6. This indicates that there is *more* evidence for Na absorption in the combined data set than in ours alone, but still not enough to justify its inclusion.

Models which include the photosphere are disfavored by  $\ln(B) \sim -5.5$ . In the case of a uniform opacity atmosphere,



**Figure 10.** A comparison of our combined transmission spectrum to the spectrum of (top) Huitson et al. (2017) (Table 4) as observed with *Gemini*/GMOS and (bottom) May et al. (2018) (Table 5) as observed with IMACS. The wavelengths of potential atomic features are highlighted. The weighted mean is subtracted from each of the three data sets.

**Table 7.** Bayes factors for the full suite of atmosphere models (similar to Table 6) where we have included data from H17 and M18.

Model	$\ln(B)_A$	$\ln(B)_{A+P}$
uniform opacity	0.0	-3.8
clouds + Na	-0.2	-5.5
clouds + K	-1.3	-5.7
clouds + Na,K	-0.7	-4.9
clouds + haze	-1.7	-6.4
clouds + haze + Na	-1.2	-6.0
clouds + haze + K	-1.6	-6.3
clouds + haze + Na,K	-1.6	-5.9

the 95% upper limit on the spot covering fraction is  $F_{\text{het}} < 1.8\%$ , which is smaller than the value we report in Section 6.3.2. However, we caution that the red and blue GMOS spectra were observed during separate transits, which may mask the slope introduced by TLSE contamination.

## 8. FEATURELESS ATMOSPHERE

Our transit spectra, in combination with the spectra by H17 and M18, suggest that WASP-4b displays a mostly featureless (uniform opacity) spectrum without a strong optical spectral slope or alkaline absorption (Na I or K I lines). Although more observations will be required to place tighter constraints on the optical spectrum to verify this, in the following we will explore what a uniform opacity spectrum would suggest for WASP-4b.

Cloud-free, broadly solar-composition atmospheres are predicted to display strong absorption at visible wavelengths by alkali metal atoms (Na I or K I doublets, Seager et al. 2000; Sudarsky et al. 2000; Seager & Sasselov 2000). These prominent (deep and broad) features have been observed in the spectra of multiple transiting hot Jupiters (e.g., Charbonneau et al. 2002). The depths of the truncated alkali features are often used as proxy for the atmospheric pressures probed (e.g., Sing et al. 2015). The emerging evidence argues that transit sightlines are often limited by cloud decks, which then lead to truncated alkali line cores or, for very low-pressure particles, even the absence of detectable alkali absorption.

Another marked deviation from a flat (featureless, zero-slope) visible spectrum would arise in a clear atmosphere from Rayleigh scattering of the starlight by molecules or very small particles. Rayleigh scattering is more efficient at smaller wavelengths, resulting in apparently larger planet diameters (i.e., a lower pressure level  $\tau = 1$  surface) at shorter wavelengths (e.g., Pont et al. 2008). The actual wavelength-dependence of the Rayleigh-scattering in a given transiting exoplanet atmosphere will depend on the particle size dis-

tribution in its upper atmosphere and can vary by orders of magnitude (e.g., Heng & Kitzmann 2017). However, visual-wavelength slopes in transiting exoplanets may also be introduced by the transit light source effect described in Section 4.7; given this consideration, the simultaneous presence of visual slopes and alkali line absorption are considered to be the most robust indicators of clear (particle-free) upper atmospheres.

Therefore, with multiple transit spectra suggesting the lack of alkali absorption *and* the lack of a strong visual slope for WASP-4b, it is important to consider the possibility that WASP-4b's upper atmosphere is not clear but contains particles at high-altitude. Given this strong possibility, in the following we explore the possible nature of these particles and the mechanisms that may inject particles into the upper atmosphere. Based on *Spitzer*/IRAC eclipse (dayside emission) measurements in the [3.6] and [4.6] filters, Beerer et al. (2011) found that the best-fit blackbody temperature of WASP-4b's dayside is 1,700 K. This temperature is close to the radiative equilibrium estimate of 1,650 K (assuming zero albedo) and significantly lower than 2,000 K, the temperature that would correspond to a zero-albedo, dayside-emission-only case.

The dayside temperatures of WASP-4b are hot enough to vaporize less refractory grains and – in a smaller fraction of the hemisphere centered on the sub-stellar point – it is likely hot enough to vaporize metal-oxide and silicate grains. Cloud formation through the evaporation and condensation of refractory grains is common and relatively well-studied in non-irradiated brown dwarfs (e.g., Apai et al. 2013; Buenzli et al. 2014; Metchev et al. 2015) and directly imaged exoplanets (Biller et al. 2015; Zhou et al. 2016) of similar temperatures. Equilibrium condensation models coupled to global circulation models show that refractory species (e.g., CaTiO<sub>3</sub>, MgSiO<sub>3</sub>, MnS, Na<sub>2</sub>S) will also form clouds in the upper atmospheres (typically 10–100 mbar) of hot Jupiters (e.g., Parmentier et al. 2016; Showman et al. 2009; Kataria et al. 2016). Therefore, our observations suggesting the lack of evidence for a clear atmosphere are fully consistent with the general expectations set by observations and models of brown dwarf and hot Jupiter atmospheres of similar temperatures.

## 9. CONCLUSIONS

We have extracted a combined optical transmission spectrum from three transit observations of WASP-4b using *Magellan*/IMACS, and use a MultiNest-based retrieval code to test different atmospheric models for the data. We find that a uniform opacity model for the atmosphere is weakly favored over alternatives with hazes, Na, and/or K. In particular, no meaningful evidence for potassium is found despite the elevated radius in the narrow bin centered on the K doublet.

Our results, in addition with those of Huitson et al. (2017) and May et al. (2018), suggest that no strong signals exist in the optical transit spectrum. Nevertheless, higher quality data may yet reveal evidence for scattering or atomic absorption; for example, Huitson et al. (2017) find inconclusive evidence for Na absorption.

We are also able to fit the size and contrast of the star spots occulted by the planet during Transits 2 and 3. Assuming a single spot model, the quality of the data from Transit 3 allows us to tightly constrain the spot size and contrast, which suggest a spot that is much larger and warmer than is typical for spots on the Sun. More complex models which include several spots and faculae could be consistent with the data as well, but are not justified by the evidences. We use this spot model to correct for the wavelength-dependent effect on the transmission spectrum from Transit 3 before averaging the individual nights' spectra.

Further space-based or larger aperture ground-based observations should be conducted to search for low amplitude signatures of scattering or absorption. However, we note that the presence and strength of a scattering haze is particularly degenerate with the presence of spots and faculae on the star. Since WASP-4 is known to be variable, the stellar

photosphere and planet atmosphere should be modeled simultaneously in future analyses of this planet's transmission spectrum.

The results reported herein benefited from collaborations and/or information exchange within NASA's Nexus for Exoplanet System Science (NExSS) research coordination network sponsored by NASA's Science Mission Directorate. This paper includes data gathered with the 6.5 meter Magellan-Baade Telescope located at Las Campanas Observatory, Chile. This research has made use of NASA's Astrophysics Data System.

A.B. acknowledges support from the NASA Earth and Space Science Fellowship Program under grant No. 80NSSC17K0470. B.R. acknowledges support from the National Science Foundation Graduate Research Fellowship Program under grant No. DGE-1143953.

**Software:** Astropy (Astropy Collaboration et al. 2018), corner (Foreman-Mackey 2016), Matplotlib (Hunter 2007), NumPy (Oliphant 2006), PyMC (Salvatier et al. 2016), PyMultiNest (Buchner et al. 2014), SciPy (Jones et al. 2001), and SPOTROD (Béky et al. 2014).

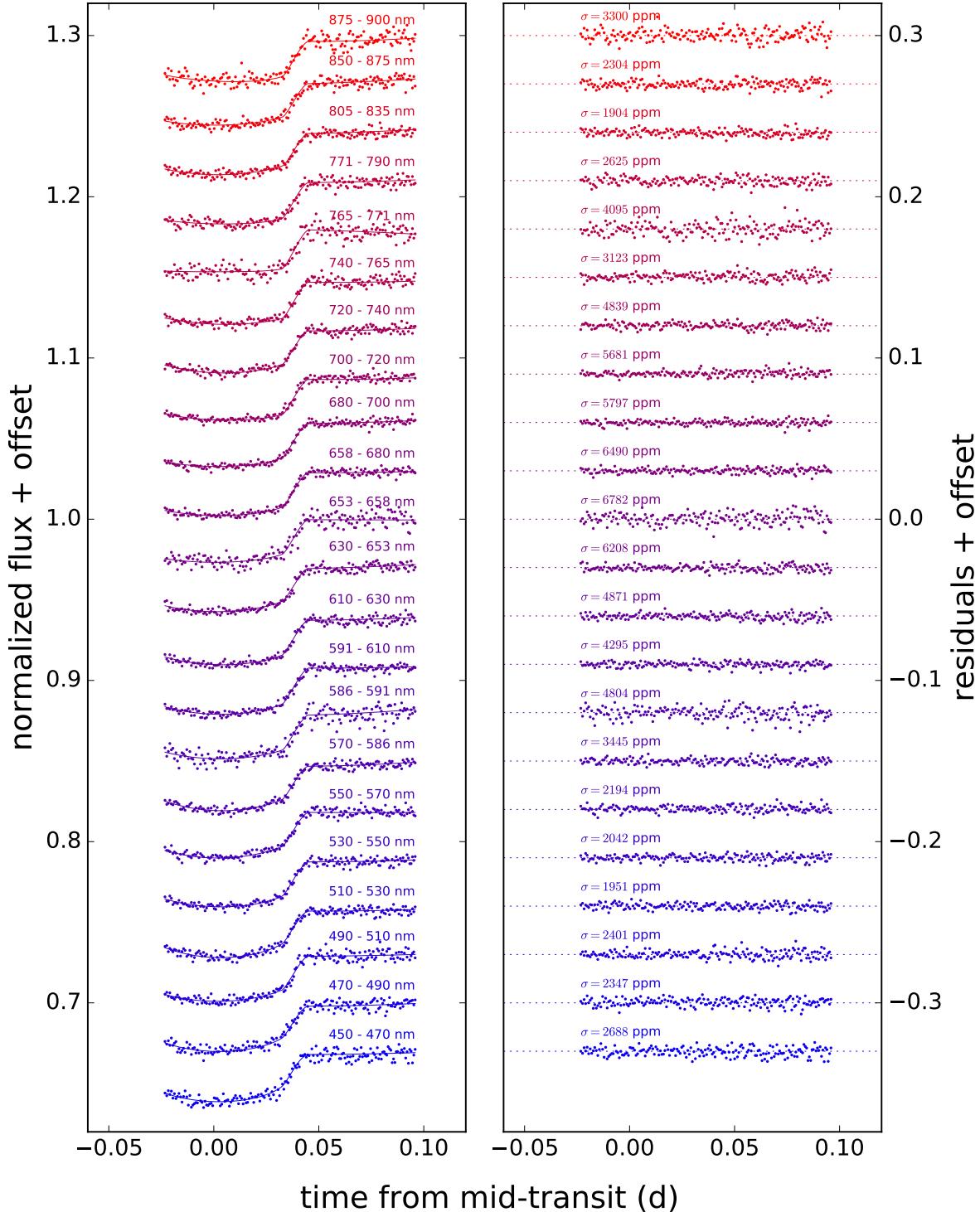
## REFERENCES

- Apai, D., Radigan, J., Buenzli, E., et al. 2013, ApJ, 768, 121, doi: [10.1088/0004-637X/768/2/121](https://doi.org/10.1088/0004-637X/768/2/121)
- Apai, D., Rackham, B. V., Giampapa, M. S., et al. 2018, ArXiv e-prints. <https://arxiv.org/abs/1803.08708>
- Astropy Collaboration, Price-Whelan, A. M., Sipócz, B. M., et al. 2018, AJ, 156, 123, doi: [10.3847/1538-3881/aabc4f](https://doi.org/10.3847/1538-3881/aabc4f)
- Beerer, I. M., Knutson, H. A., Burrows, A., et al. 2011, ApJ, 727, 23, doi: [10.1088/0004-637X/727/1/23](https://doi.org/10.1088/0004-637X/727/1/23)
- Béky, B., Kipping, D. M., & Holman, M. J. 2014, MNRAS, 442, 3686, doi: [10.1093/mnras/stu1061](https://doi.org/10.1093/mnras/stu1061)
- Benneke, B., & Seager, S. 2013, ApJ, 778, 153, doi: [10.1088/0004-637X/778/2/153](https://doi.org/10.1088/0004-637X/778/2/153)
- Bétrémieux, Y., & Swain, M. R. 2017, MNRAS, 467, 2834, doi: [10.1093/mnras/stx257](https://doi.org/10.1093/mnras/stx257)
- Biller, B. A., Vos, J., Bonavita, M., et al. 2015, ApJL, 813, L23, doi: [10.1088/2041-8205/813/2/L23](https://doi.org/10.1088/2041-8205/813/2/L23)
- Brown, T. M. 2001, ApJ, 553, 1006, doi: [10.1086/320950](https://doi.org/10.1086/320950)
- Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125, doi: [10.1051/0004-6361/201322971](https://doi.org/10.1051/0004-6361/201322971)
- Buenzli, E., Apai, D., Radigan, J., Reid, I. N., & Flateau, D. 2014, ApJ, 782, 77, doi: [10.1088/0004-637X/782/2/77](https://doi.org/10.1088/0004-637X/782/2/77)
- Carter, J. A., & Winn, J. N. 2009, ApJ, 704, 51, doi: [10.1088/0004-637X/704/1/51](https://doi.org/10.1088/0004-637X/704/1/51)
- Cauley, P. W., Kuckein, C., Redfield, S., et al. 2018, AJ, 156, 189, doi: [10.3847/1538-3881/aaddf9](https://doi.org/10.3847/1538-3881/aaddf9)
- Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L. 2002, ApJ, 568, 377, doi: [10.1086/338770](https://doi.org/10.1086/338770)
- Deming, D., Wilkins, A., McCullough, P., et al. 2013, ApJ, 774, 95, doi: [10.1088/0004-637X/774/2/95](https://doi.org/10.1088/0004-637X/774/2/95)
- Diamond-Lowe, H., Stevenson, K. B., Bean, J. L., Line, M. R., & Fortney, J. J. 2014, ApJ, 796, 66, doi: [10.1088/0004-637X/796/1/66](https://doi.org/10.1088/0004-637X/796/1/66)
- Dressler, A., Bigelow, B., Hare, T., et al. 2011, PASP, 123, 288, doi: [10.1086/658908](https://doi.org/10.1086/658908)
- Espinoza, N., & Jordán, A. 2015, MNRAS, 450, 1879, doi: [10.1093/mnras/stv744](https://doi.org/10.1093/mnras/stv744)
- Espinoza, N., Rackham, B. V., Jordán, A., et al. 2019, MNRAS, 482, 2065, doi: [10.1093/mnras/sty2691](https://doi.org/10.1093/mnras/sty2691)
- Foreman-Mackey, D. 2016, The Journal of Open Source Software, 24, doi: [10.21105/joss.00024](https://doi.org/10.21105/joss.00024)
- Gaia Collaboration. 2018, VizieR Online Data Catalog, I/345
- Gillon, M., Smalley, B., Hebb, L., et al. 2009, A&A, 496, 259, doi: [10.1051/0004-6361:200810929](https://doi.org/10.1051/0004-6361:200810929)
- Heng, K., & Kitzmann, D. 2017, MNRAS, 470, 2972, doi: [10.1093/mnras/stx1453](https://doi.org/10.1093/mnras/stx1453)
- Hoyer, S., López-Morales, M., Rojo, P., et al. 2013, MNRAS, 434, 46, doi: [10.1093/mnras/stt962](https://doi.org/10.1093/mnras/stt962)
- Hubbard, W. B., Fortney, J. J., Lunine, J. I., et al. 2001, ApJ, 560, 413, doi: [10.1086/322490](https://doi.org/10.1086/322490)

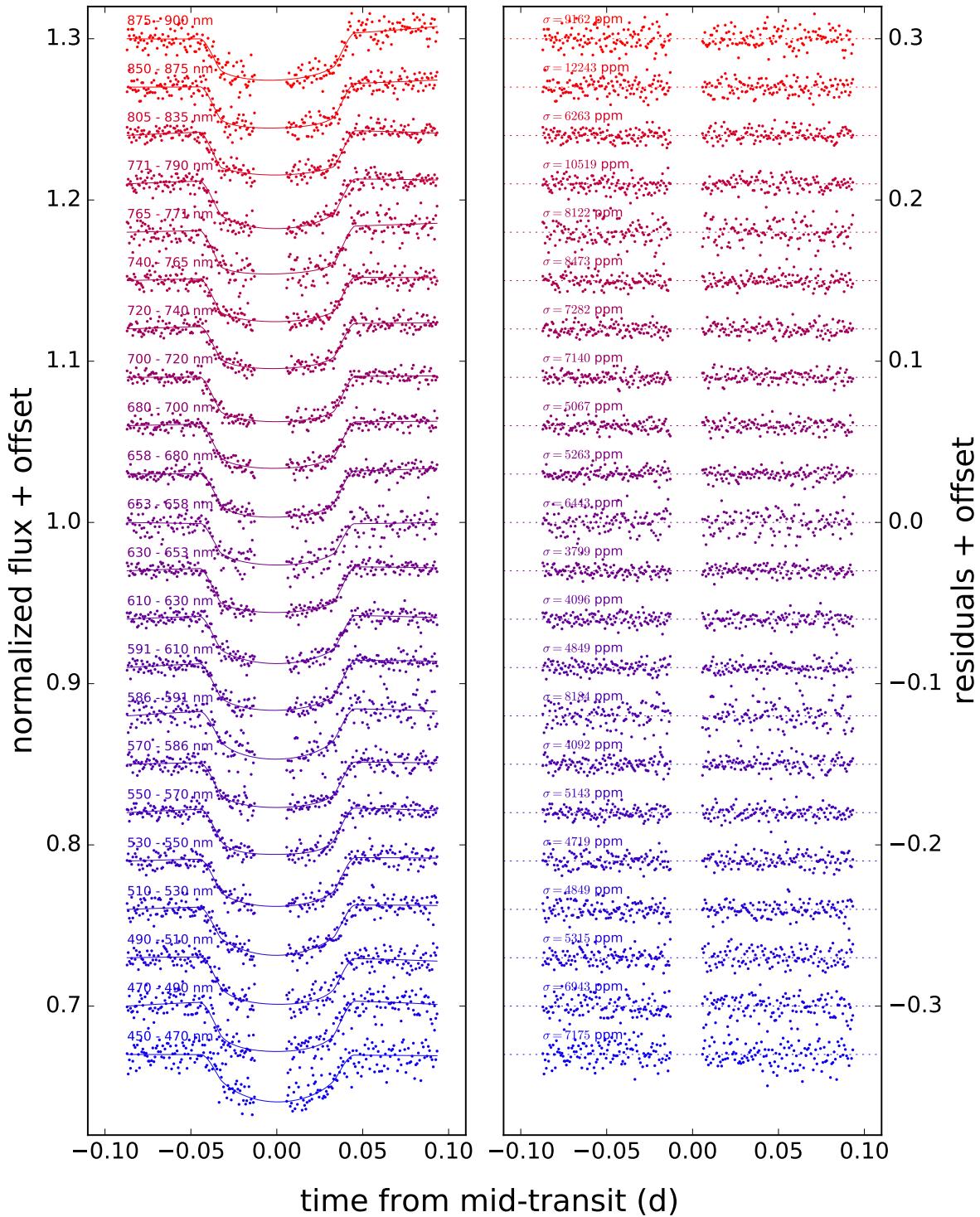
- Huitson, C. M., Désert, J. M., Bean, J. L., et al. 2017, AJ, 154, 95, doi: [10.3847/1538-3881/aa7f72](https://doi.org/10.3847/1538-3881/aa7f72)
- Hunter, J. D. 2007, Computing In Science & Engineering, 9, 90, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- Husser, T.-O., Wende-von Berg, S., Dreizler, S., et al. 2013, A&A, 553, A6, doi: [10.1051/0004-6361/201219058](https://doi.org/10.1051/0004-6361/201219058)
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python. <http://www.scipy.org/>
- Jordán, A., Espinoza, N., Rabus, M., et al. 2013, ApJ, 778, 184, doi: [10.1088/0004-637X/778/2/184](https://doi.org/10.1088/0004-637X/778/2/184)
- Kataria, T., Sing, D. K., Lewis, N. K., et al. 2016, ApJ, 821, 9, doi: [10.3847/0004-637X/821/1/9](https://doi.org/10.3847/0004-637X/821/1/9)
- Kipping, D. M. 2013, MNRAS, 435, 2152, doi: [10.1093/mnras/stt1435](https://doi.org/10.1093/mnras/stt1435)
- Kochanek, C. S., Shappee, B. J., Stanek, K. Z., et al. 2017, PASP, 129, 104502, doi: [10.1088/1538-3873/aa80d9](https://doi.org/10.1088/1538-3873/aa80d9)
- MacDonald, R. J., & Madhusudhan, N. 2017, MNRAS, 469, 1979, doi: [10.1093/mnras/stx804](https://doi.org/10.1093/mnras/stx804)
- Mandel, K., & Agol, E. 2002, ApJL, 580, L171, doi: [10.1086/345520](https://doi.org/10.1086/345520)
- Marsh, T. R. 1989, PASP, 101, 1032, doi: [10.1086/132570](https://doi.org/10.1086/132570)
- Maxted, P. F. L., Koen, C., & Smalley, B. 2011, MNRAS, 418, 1039, doi: [10.1111/j.1365-2966.2011.19554.x](https://doi.org/10.1111/j.1365-2966.2011.19554.x)
- May, E. M., Zhao, M., Haidar, M., Rauscher, E., & Monnier, J. D. 2018, AJ, 156, 122, doi: [10.3847/1538-3881/aad4a8](https://doi.org/10.3847/1538-3881/aad4a8)
- McCullough, P. R., Crouzet, N., Deming, D., & Madhusudhan, N. 2014, ApJ, 791, 55, doi: [10.1088/0004-637X/791/1/55](https://doi.org/10.1088/0004-637X/791/1/55)
- Metchev, S. A., Heinze, A., Apai, D., et al. 2015, ApJ, 799, 154, doi: [10.1088/0004-637X/799/2/154](https://doi.org/10.1088/0004-637X/799/2/154)
- Nikolov, N., Henning, T., Koppenhoefer, J., et al. 2012, A&A, 539, A159, doi: [10.1051/0004-6361/201118336](https://doi.org/10.1051/0004-6361/201118336)
- Nikolov, N., Sing, D. K., Burrows, A. S., et al. 2015, MNRAS, 447, 463, doi: [10.1093/mnras/stu2433](https://doi.org/10.1093/mnras/stu2433)
- Oliphant, T. E. 2006, Guide to NumPy, Provo, UT
- O’Neal, D., Neff, J. E., & Saar, S. H. 1998, ApJ, 507, 919, doi: [10.1086/306340](https://doi.org/10.1086/306340)
- O’Neal, D., Neff, J. E., Saar, S. H., & Cuntz, M. 2004, AJ, 128, 1802, doi: [10.1086/423438](https://doi.org/10.1086/423438)
- Parmentier, V., Fortney, J. J., Showman, A. P., Morley, C., & Marley, M. S. 2016, ApJ, 828, 22, doi: [10.3847/0004-637X/828/1/22](https://doi.org/10.3847/0004-637X/828/1/22)
- Pinhas, A., Rackham, B. V., Madhusudhan, N., & Apai, D. 2018, MNRAS, 480, 5314, doi: [10.1093/mnras/sty2209](https://doi.org/10.1093/mnras/sty2209)
- Pont, F., Knutson, H., Gilliland, R. L., Moutou, C., & Charbonneau, D. 2008, MNRAS, 385, 109, doi: [10.1111/j.1365-2966.2008.12852.x](https://doi.org/10.1111/j.1365-2966.2008.12852.x)
- Pont, F., Sing, D. K., Gibson, N. P., et al. 2013, MNRAS, 432, 2917, doi: [10.1093/mnras/stt651](https://doi.org/10.1093/mnras/stt651)
- Rackham, B., Espinoza, N., Apai, D., et al. 2017, ApJ, 834, 151, doi: [10.3847/1538-4357/aa4f6c](https://doi.org/10.3847/1538-4357/aa4f6c)
- Rackham, B. V., Apai, D., & Giampapa, M. S. 2018, ApJ, 853, 122, doi: [10.3847/1538-4357/aaa08c](https://doi.org/10.3847/1538-4357/aaa08c)
- Ranjan, S., Charbonneau, D., Désert, J.-M., et al. 2014, ApJ, 785, 148, doi: [10.1088/0004-637X/785/2/148](https://doi.org/10.1088/0004-637X/785/2/148)
- Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, PeerJ Computer Science, 2, e55, doi: [10.7717/peerj-cs.55](https://doi.org/10.7717/peerj-cs.55)
- Sanchis-Ojeda, R., Winn, J. N., Holman, M. J., et al. 2011, ApJ, 733, 127, doi: [10.1088/0004-637X/733/2/127](https://doi.org/10.1088/0004-637X/733/2/127)
- Seager, S., & Sasselov, D. D. 2000, ApJ, 537, 916, doi: [10.1086/309088](https://doi.org/10.1086/309088)
- Seager, S., Whitney, B. A., & Sasselov, D. D. 2000, ApJ, 540, 504, doi: [10.1086/309292](https://doi.org/10.1086/309292)
- Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, ApJ, 788, 48, doi: [10.1088/0004-637X/788/1/48](https://doi.org/10.1088/0004-637X/788/1/48)
- Showman, A. P., Fortney, J. J., Lian, Y., et al. 2009, ApJ, 699, 564, doi: [10.1088/0004-637X/699/1/564](https://doi.org/10.1088/0004-637X/699/1/564)
- Sing, D. K., Pont, F., Aigrain, S., et al. 2011, MNRAS, 416, 1443, doi: [10.1111/j.1365-2966.2011.19142.x](https://doi.org/10.1111/j.1365-2966.2011.19142.x)
- Sing, D. K., Leconte, J., Lecavelier des Etangs, A., Fortney, J. J., et al. 2013, MNRAS, 436, 2956, doi: [10.1093/mnras/stt1782](https://doi.org/10.1093/mnras/stt1782)
- Sing, D. K., Wakeford, H. R., Showman, A. P., et al. 2015, MNRAS, 446, 2428, doi: [10.1093/mnras/stu2279](https://doi.org/10.1093/mnras/stu2279)
- Solanki, S. K. 2003, Astronomy and Astrophysics Review, 11, 153, doi: [10.1007/s00159-003-0018-4](https://doi.org/10.1007/s00159-003-0018-4)
- Spake, J. J., Sing, D. K., Evans, T. M., et al. 2018, Nature, 557, 68, doi: [10.1038/s41586-018-0067-5](https://doi.org/10.1038/s41586-018-0067-5)
- Strassmeier, K. G. 2009, Astronomy and Astrophysics Review, 17, 251, doi: [10.1007/s00159-009-0020-6](https://doi.org/10.1007/s00159-009-0020-6)
- Sudarsky, D., Burrows, A., & Pinto, P. 2000, ApJ, 538, 885, doi: [10.1086/309160](https://doi.org/10.1086/309160)
- Triaud, A. H. M. J., Collier Cameron, A., Queloz, D., et al. 2010, A&A, 524, A25, doi: [10.1051/0004-6361/201014525](https://doi.org/10.1051/0004-6361/201014525)
- Trotta, R. 2008, Contemporary Physics, 49, 71, doi: [10.1080/00107510802066753](https://doi.org/10.1080/00107510802066753)
- VanderPlas, J. T. 2018, ApJS, 236, 16, doi: [10.3847/1538-4365/aab766](https://doi.org/10.3847/1538-4365/aab766)
- Wilson, D. M., Gillon, M., Hellier, C., et al. 2008, ApJL, 675, L113, doi: [10.1086/586735](https://doi.org/10.1086/586735)
- Winn, J. N., Holman, M. J., Carter, J. A., et al. 2009, AJ, 137, 3826, doi: [10.1088/0004-6256/137/4/3826](https://doi.org/10.1088/0004-6256/137/4/3826)
- Zhang, Z., Zhou, Y., Rackham, B. V., & Apai, D. 2018, AJ, 156, 178, doi: [10.3847/1538-3881/aade4f](https://doi.org/10.3847/1538-3881/aade4f)
- Zhou, Y., Apai, D., Schneider, G. H., Marley, M. S., & Showman, A. P. 2016, ApJ, 818, 176, doi: [10.3847/0004-637X/818/2/176](https://doi.org/10.3847/0004-637X/818/2/176)

## APPENDIX

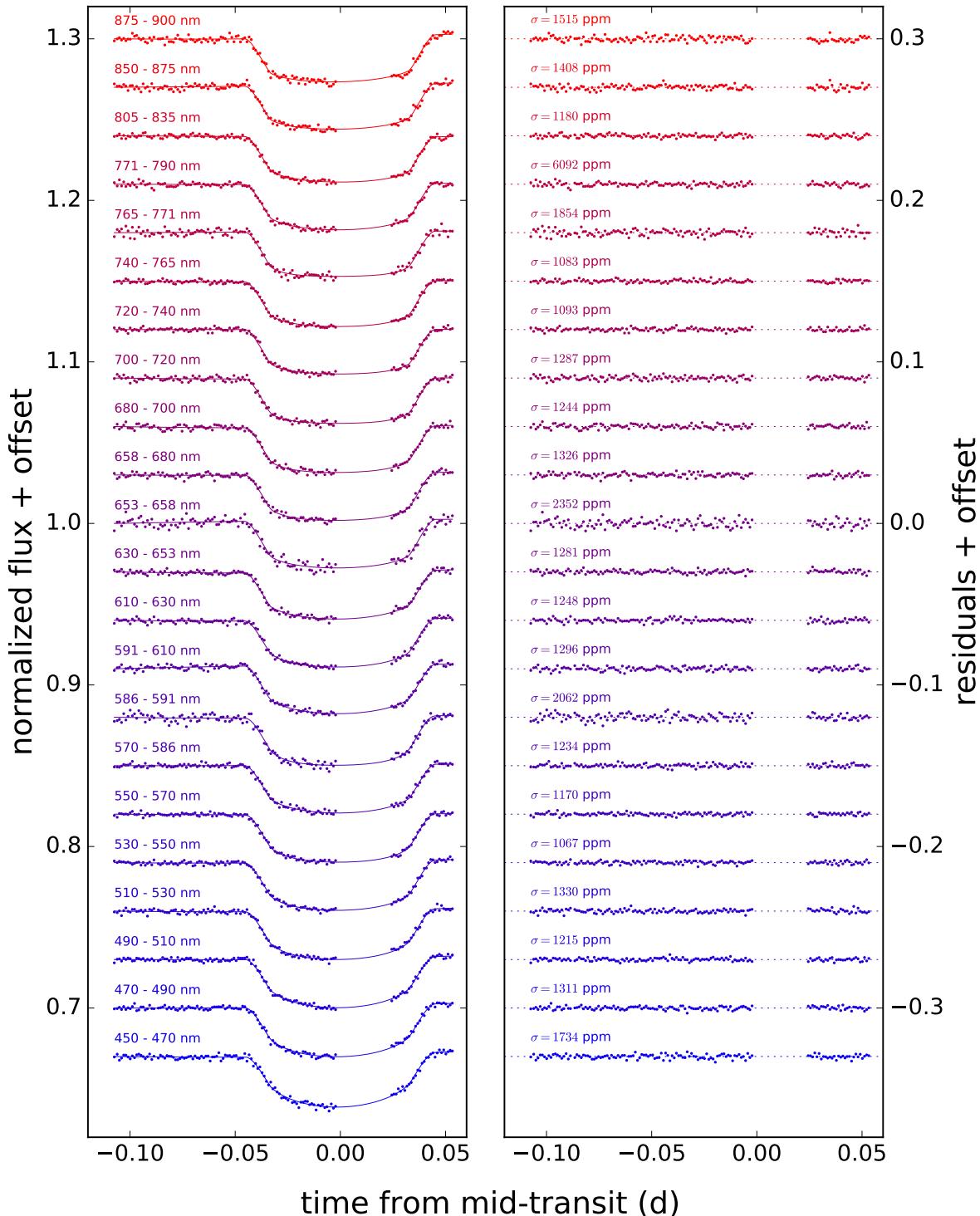
## A. BINNED LIGHT CURVES



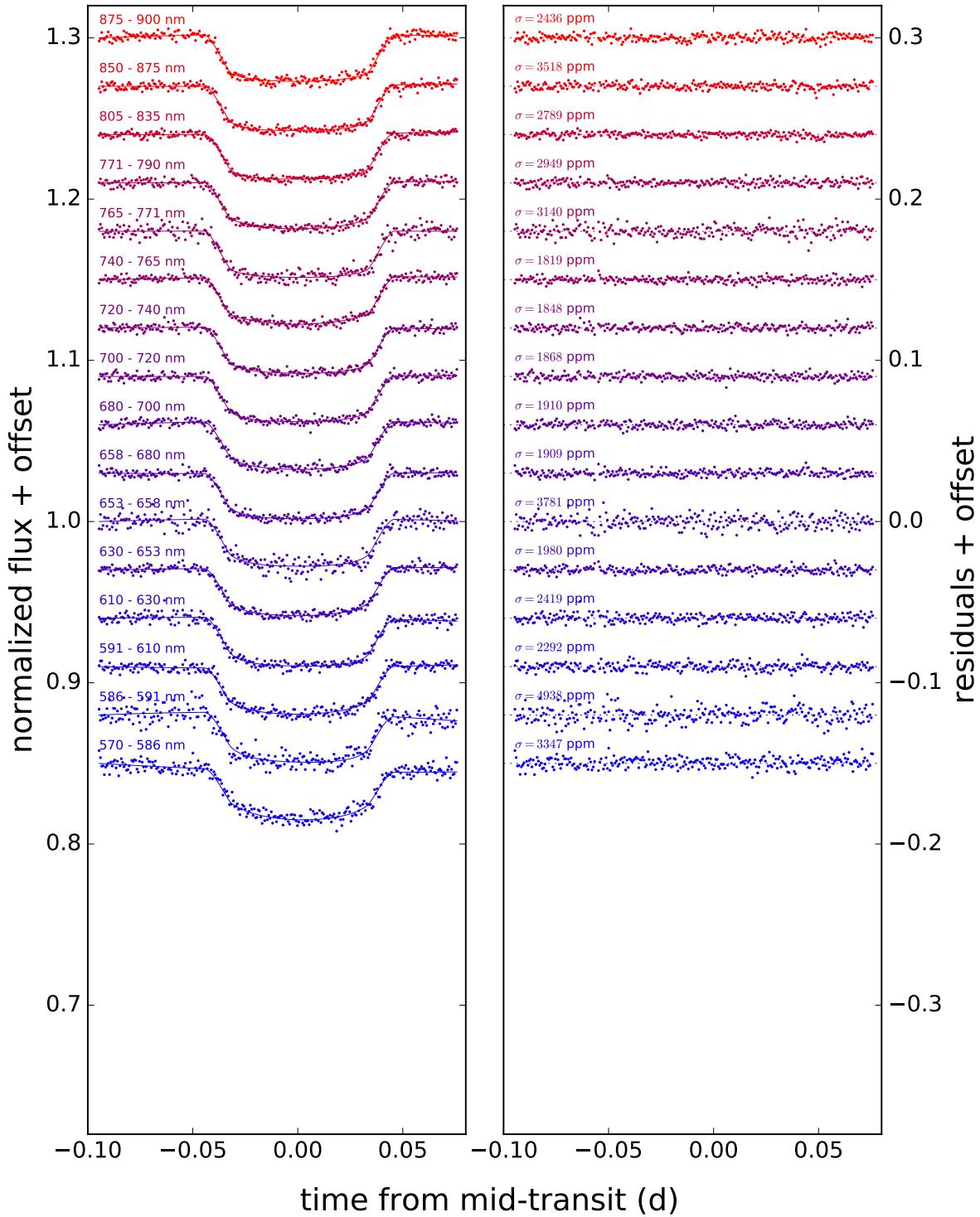
**Figure A1.** Binned light curve fits from the night of 2013-09-24. The models in the left panel include the linear systematics component.



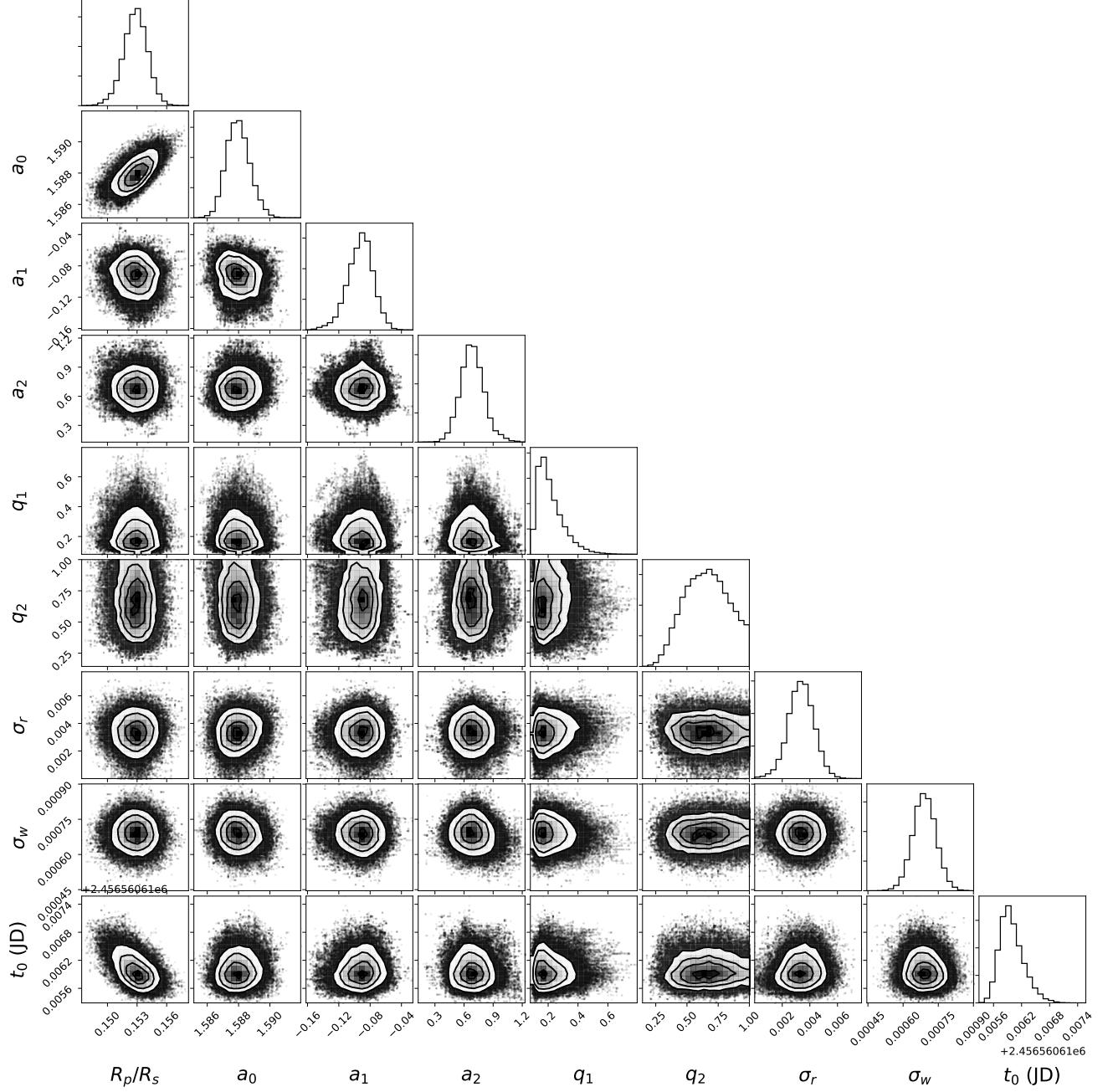
**Figure A2.** Binned light curve fits from the night of 2013-10-17. The models in the left panel include the linear systematics component. A potential spot-crossing event is excluded from the fit.



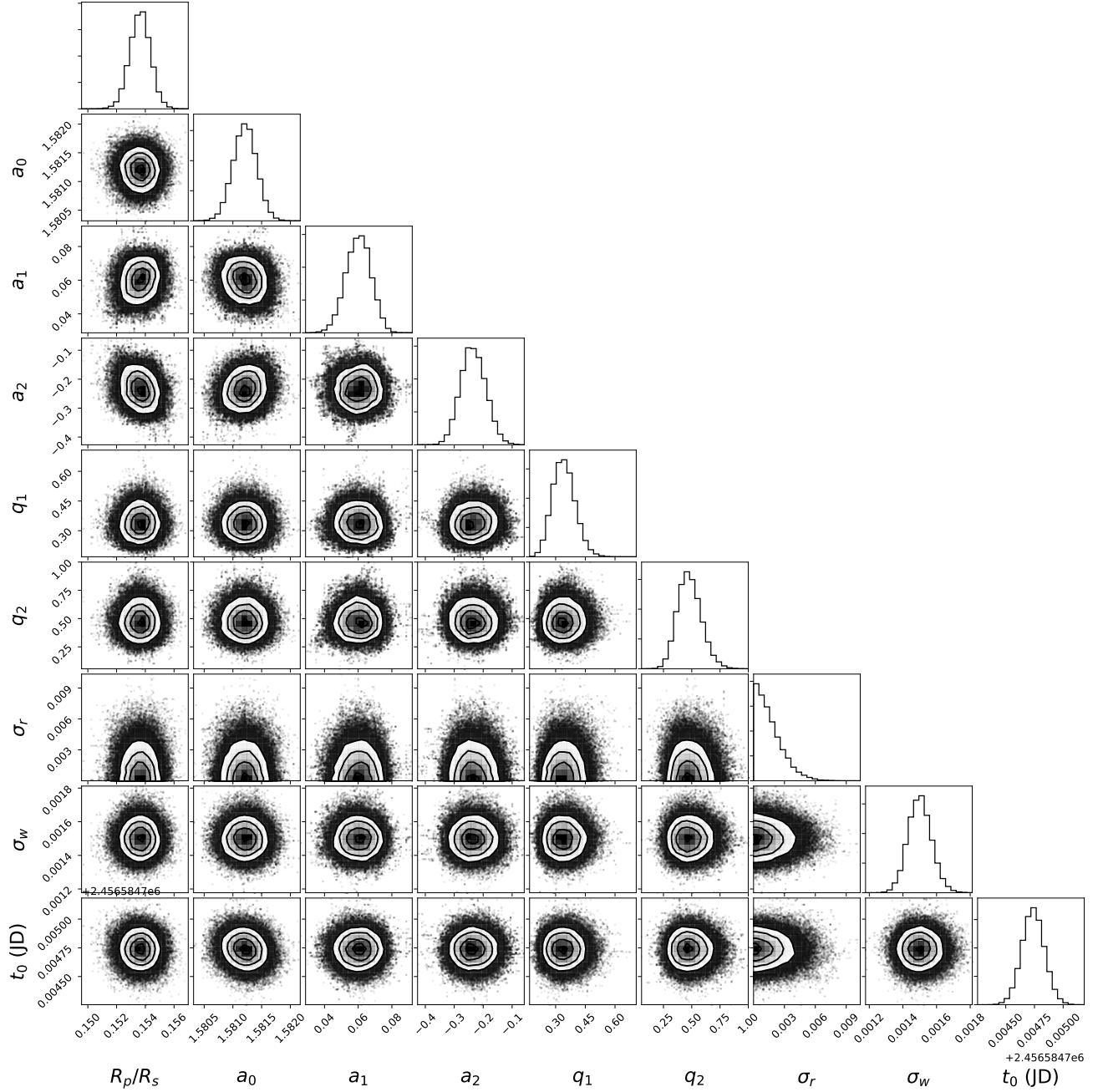
**Figure A3.** Binned light curve fits from the night of 2015-08-14. The models in the left panel include the linear systematics component. The spot-crossing event is excluded.



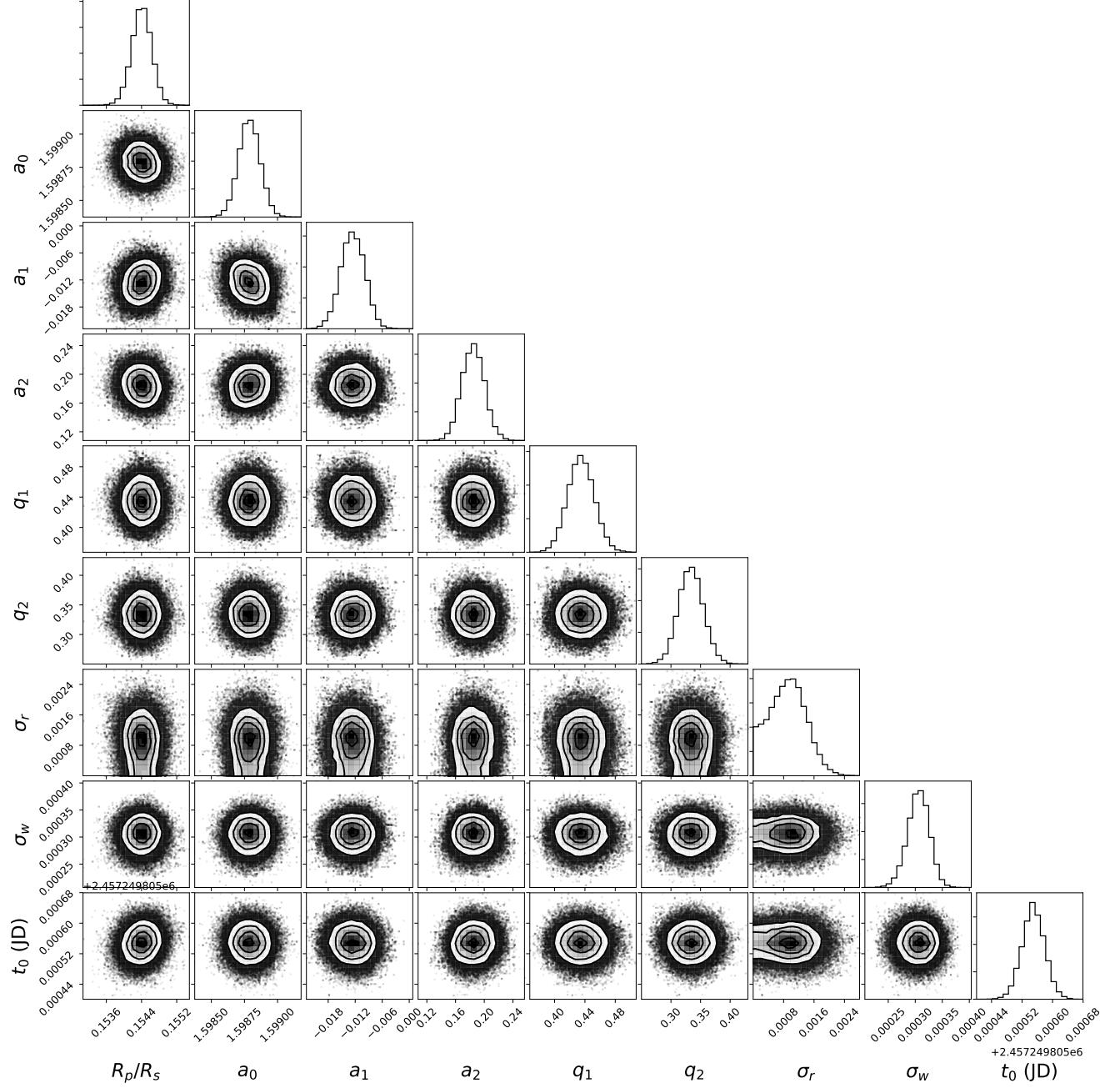
**Figure A4.** Binned light curve fits from the night of 2015-09-26. The models in the left panel include the linear systematics component. Six of the bins are excluded for lack of data due to the narrower filter.



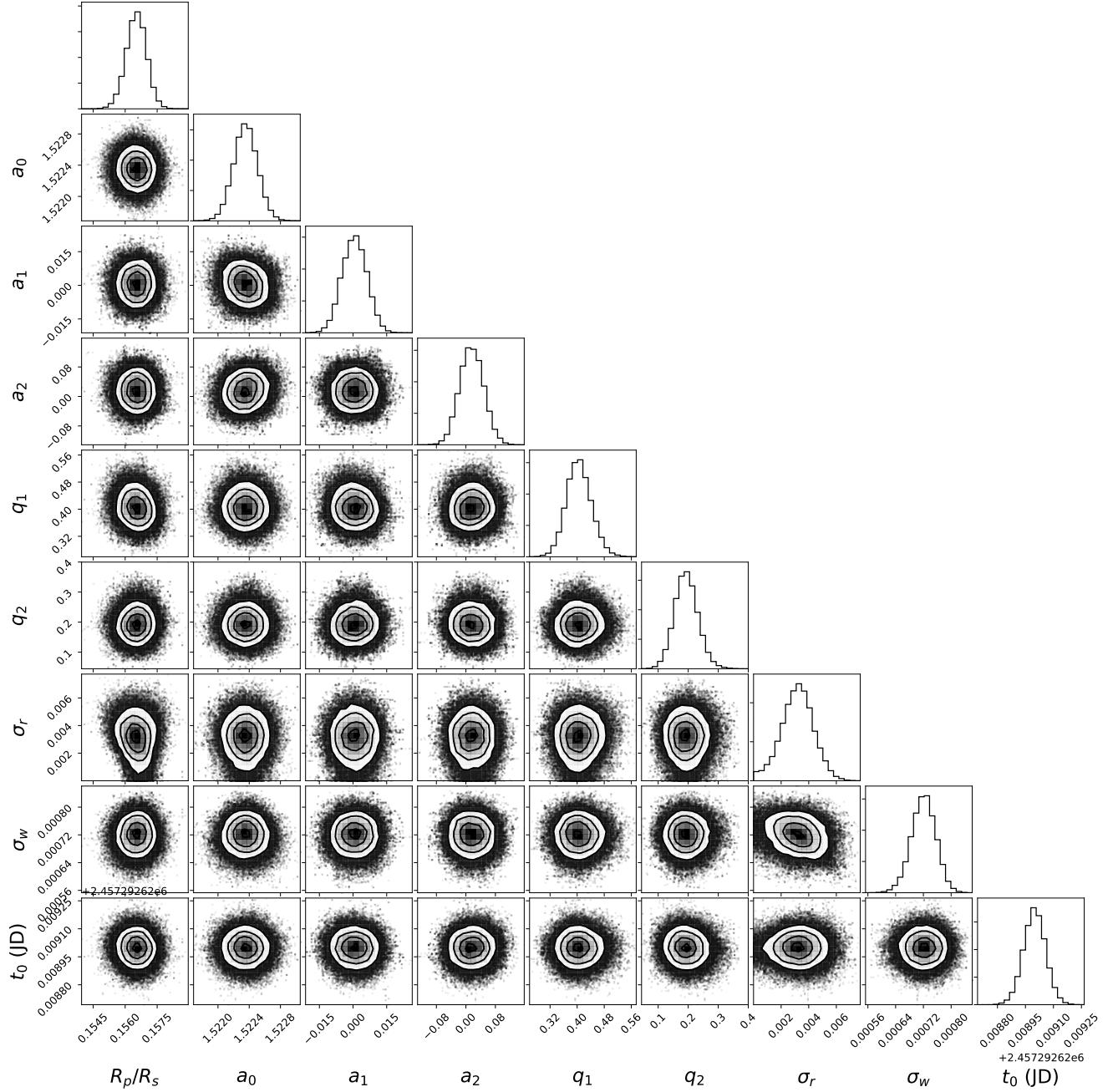
**Figure A5.** Joint posterior distributions for the white light curve model for Transit 1, including the planet-to-star radius ratio ( $R_p/R_s$ ), the correlated and uncorrelated noise parameters ( $\sigma_r$ ,  $\sigma_w$ ), the mid-transit time ( $t_0$ ), three parameters for the quadratic polynomial systematics model ( $a_0$ ,  $a_1$ ,  $a_2$ ), and two parameters for the quadratic limb-darkening law ( $q_1$ ,  $q_2$ ). The light curve of this transit is incomplete, so we exclude its transit spectrum from our combined result.



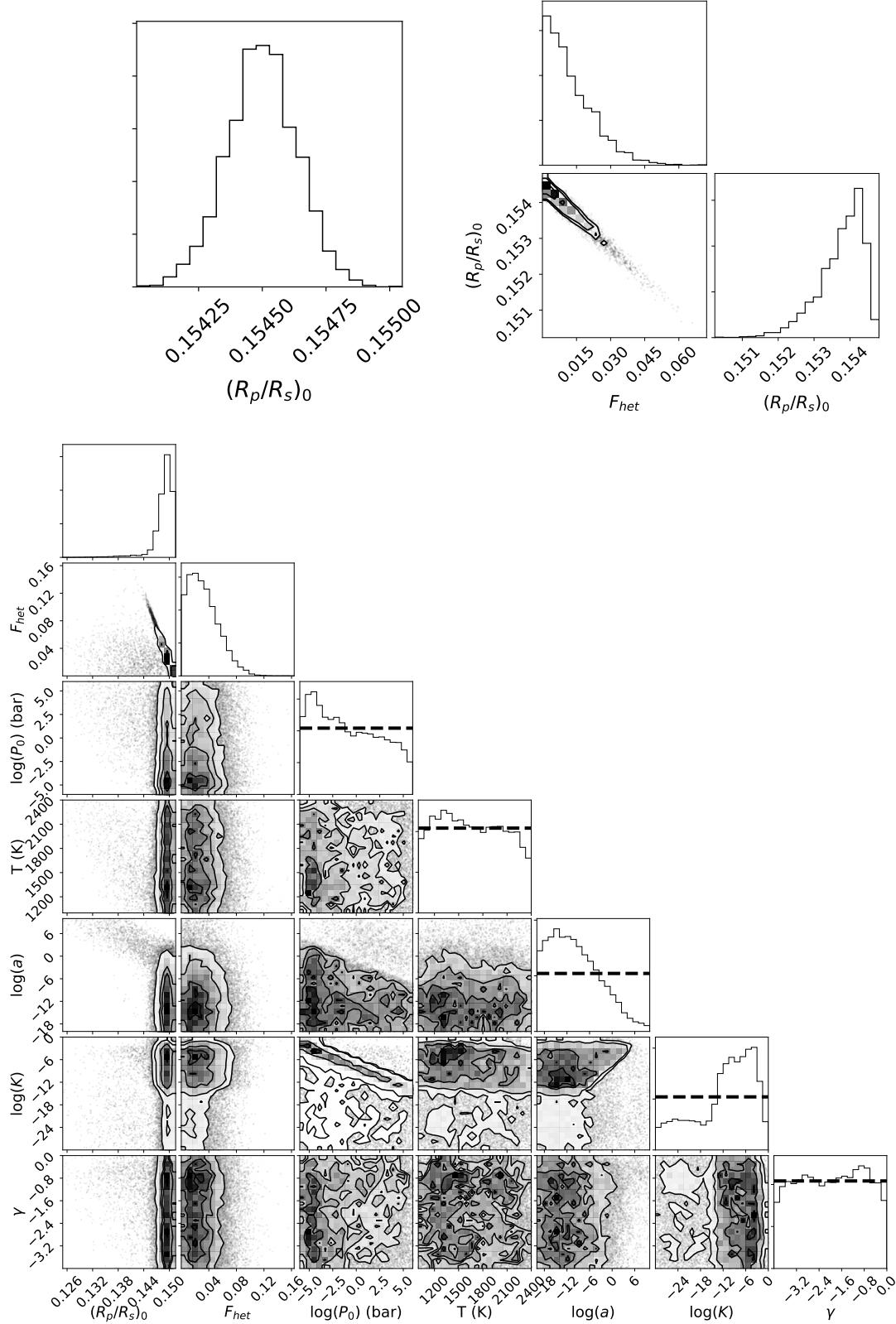
**Figure A6.** Joint posterior distributions for the white light curve model for Transit 2.



**Figure A7.** Joint posterior distributions for the white light curve model for Transit 3.



**Figure A8.** Joint posterior distributions for the white light curve model for Transit 4.



**Figure A9.** Joint posterior distributions for a subset of the atmospheric retrieval models shown in Figure 8. The parameters and prior distributions are described in Table 5. (Top) Models for a uniform opacity atmosphere with (right) and without (left) the contamination features from the photosphere. Here the posterior distributions have converged, and the degeneracy between the planet's radius and the spot covering fraction is apparent. (Bottom) Model for a hazy atmosphere with K absorption and contamination from the photosphere. Some of the parameters do not converge enough from their priors to yield meaningful constraints. In these cases, the prior distributions are marked with a dashed line.

## B. TABULATED TRANSMISSION SPECTRA

**Table B1.** Data for the combined and individual transmission spectra shown in Figures 3 and 4 with  $1\sigma$  uncertainties. The second column is the weighted mean of the mean-subtracted values from Transits 2–4, where the Transit 3 values have first been corrected for the presence of an occulted spot. The last column is the multiplicative effect on the fitted radii for Transit 3 due to the occulted spot, where the true radius  $R_{p,0} = R_p/\epsilon^{1/2}$ .

Bin (nm)	$\Delta R_p/R_s$	Transit 1	Transit 2	Transit 3	Transit 4	$\epsilon^{1/2}$
450.0 - 470.0	-0.0015 ± 0.0008	0.1525 ± 0.0031	0.1430 ± 0.0028 <sup>‡</sup>	0.1531 ± 0.0008		1.0121
470.0 - 490.0	0.0004 ± 0.0007	0.1508 ± 0.0026	0.1583 ± 0.0045	0.1548 ± 0.0007		1.0113
490.0 - 510.0	-0.0015 ± 0.0007	0.1528 ± 0.0024	0.1514 ± 0.0033	0.1531 ± 0.0007		1.0116
510.0 - 530.0	-0.0007 ± 0.0008	0.1543 ± 0.0020	0.1533 ± 0.0029	0.1538 ± 0.0008		1.0118
530.0 - 550.0	0.0007 ± 0.0006	0.1547 ± 0.0018	0.1550 ± 0.0029	0.1551 ± 0.0006		1.0105
550.0 - 570.0	0.0001 ± 0.0005	0.1549 ± 0.0019	0.1530 ± 0.0021	0.1544 ± 0.0005		1.0097
570.0 - 586.8	0.0002 ± 0.0006	0.1528 ± 0.0018	0.1520 ± 0.0025	0.1551 ± 0.0007	0.1543 ± 0.0016	1.0092
586.8 - 591.8 <sup>†</sup>	-0.0000 ± 0.0011	0.1529 ± 0.0038	0.1532 ± 0.0046	0.1543 ± 0.0014	0.1569 ± 0.0020	1.0096
591.8 - 610.0	0.0015 ± 0.0006	0.1582 ± 0.0020	0.1574 ± 0.0021	0.1566 ± 0.0007	0.1545 ± 0.0013	1.0089
610.0 - 630.0	0.0002 ± 0.0006	0.1550 ± 0.0019	0.1552 ± 0.0023	0.1533 ± 0.0008	0.1591 ± 0.0011	1.0089
630.0 - 653.8	0.0007 ± 0.0005	0.1492 ± 0.0022	0.1547 ± 0.0020	0.1546 ± 0.0007	0.1579 ± 0.0010	1.0085
653.8 - 658.8 <sup>†</sup>	0.0010 ± 0.0012	0.1524 ± 0.0040	0.1442 ± 0.0033	0.1576 ± 0.0017	0.1581 ± 0.0019	1.0076
658.8 - 680.0	-0.0008 ± 0.0006	0.1523 ± 0.0015	0.1496 ± 0.0022	0.1549 ± 0.0008	0.1546 ± 0.0010	1.0082
680.0 - 700.0	0.0005 ± 0.0006	0.1516 ± 0.0018	0.1543 ± 0.0022	0.1535 ± 0.0008	0.1585 ± 0.0008	1.0079
700.0 - 720.0	-0.0005 ± 0.0006	0.1537 ± 0.0019	0.1546 ± 0.0021	0.1539 ± 0.0008	0.1556 ± 0.0009	1.0078
720.0 - 740.0	-0.0002 ± 0.0005	0.1522 ± 0.0022	0.1514 ± 0.0024	0.1537 ± 0.0006	0.1572 ± 0.0008	1.0076
740.0 - 765.0	0.0002 ± 0.0005	0.1559 ± 0.0029	0.1526 ± 0.0028	0.1525 ± 0.0007	0.1588 ± 0.0007	1.0074
765.0 - 771.0 <sup>†</sup>	0.0017 ± 0.0009	0.1591 ± 0.0049	0.1577 ± 0.0045	0.1548 ± 0.0011	0.1601 ± 0.0016	1.0073
771.0 - 790.0	0.0001 ± 0.0006	0.1529 ± 0.0026	0.1571 ± 0.0031	0.1536 ± 0.0008	0.1573 ± 0.0009	1.0071
805.0 - 835.0	-0.0004 ± 0.0006	0.1475 ± 0.0020	0.1553 ± 0.0026	0.1539 ± 0.0007	0.1555 ± 0.0010	1.0070
850.0 - 875.0	-0.0014 ± 0.0007	0.1482 ± 0.0026	0.1503 ± 0.0032	0.1527 ± 0.0011	0.1556 ± 0.0010	1.0066
875.0 - 900.0	-0.0001 ± 0.0010	0.1530 ± 0.0040	0.1495 ± 0.0040	0.1539 ± 0.0010	0.1608 ± 0.0011 <sup>‡</sup>	1.0064

<sup>†</sup> Denotes bins centered on Na, H $\alpha$ , and K absorption lines.

<sup>‡</sup> These outliers are excluded from the analysis; see Section 4.2