The Astrophysical Journal, 819:10 (19pp), 2016 March 1 doi:10.3847/0004-637X/819/1/10

MARGINALIZING INSTRUMENT SYSTEMATICS IN *HST* WFC3 TRANSIT LIGHT CURVES

H. R. Wakeford[1,2], D. K. Sing[2], T. Evans[2], D. Deming[3], and A. Mandell[1]
[1] NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA; hannah.wakeford@nasa.gov
[2] University of Exeter, Exeter, Devon, EX4 4QL, UK
[3] University of Maryland, College Park, MD 20742, USA

Received 2015 September 25; accepted 2016 January 8; published 2016 February 23

ABSTRACT

Hubble Space Telescope (*HST*) Wide Field Camera 3 (WFC3) infrared observations at 1.1–1.7 μm probe primarily the H_2O absorption band at 1.4 μm, and have provided low-resolution transmission spectra for a wide range of exoplanets. We present the application of marginalization based on Gibson to analyze exoplanet transit light curves obtained from *HST* WFC3 to better determine important transit parameters such as R_p/R_*, which are important for accurate detections of H_2O. We approximate the evidence, often referred to as the marginal likelihood, for a grid of systematic models using the Akaike Information Criterion. We then calculate the evidence-based weight assigned to each systematic model and use the information from all tested models to calculate the final marginalized transit parameters for both the band-integrated and spectroscopic light curves to construct the transmission spectrum. We find that a majority of the highest weight models contain a correction for a linear trend in time as well as corrections related to *HST* orbital phase. We additionally test the dependence on the shift in spectral wavelength position over the course of the observations and find that spectroscopic wavelength shifts $\delta_\lambda(\lambda)$ best describe the associated systematic in the spectroscopic light curves for most targets while fast scan rate observations of bright targets require an additional level of processing to produce a robust transmission spectrum. The use of marginalization allows for transparent interpretation and understanding of the instrument and the impact of each systematic evaluated statistically for each data set, expanding the ability to make true and comprehensive comparisons between exoplanet atmospheres.

Key words: methods: data analysis – planets and satellites: atmospheres – techniques: spectroscopic

1. INTRODUCTION

H_2O is the most spectroscopically dominant species expected in hot Jupiter atmospheres and a key molecule for constraining atmospheric compositions. In most lower-atmosphere models of hot Jupiters H_2O is well-mixed throughout the atmosphere, a majority of absorption features fall between 0.7 and 2.5 μm, and are the result of H_2O vibration-rotation bands (Brown 2001). *Hubble Space Telescope* (*HST*) Wide Field Camera 3 (WFC3) infrared (IR) observations at 1.1–1.7 μm probe primarily the H_2O absorption band at 1.4 μm and have provided low-resolution transmission spectra for a wide range of exoplanets (e.g., Berta et al. 2012; Gibson et al. 2012; Deming et al. 2013; Wakeford et al. 2013; Knutson et al. 2014; Stevenson et al. 2014a; Sing et al. 2015). Using both transmission and emission spectra, Kreidberg et al. (2014b) measured the relative H_2O abundance in the atmosphere of WASP-43b. They compare the measured H_2O abundance with the solar system's giant planets using the metallicity expected for a "broadly" solar case, indicating that the trend observed in the metal abundance of the solar system giant planet atmospheres, i.e., decreasing metal enhancement with increasing mass, extends to exoplanet atmospheres. Wakeford et al. (2013) were also able to obtain a strong detection of H_2O in the atmosphere of the hot Jupiter HAT-P-1b. Their analysis made use of a common spectral type star in the field of view of the WFC3 detector and performed differential spectrophotometry using the companion spectrum which is similar to methods used from the ground. The robust nature of the transmission spectral shape shown for analysis methods with and without the use of a companion spectrum shows that the emphasis needs to be on methods that not only can effectively reduce the data but that can also be successfully applied to multiple data sets simultaneously for a true comparative study.

A wide range of reduction and analysis methods have been used on *HST* WFC3 transit data sets, making comparisons between data sets and planetary atmospheres difficult. Studies of multiple WFC3 data sets have been conducted across multiple programs and observation modes. These studies have attempted to define a common de-trending technique applicable to all data sets (Deming et al. 2013; Mandell et al. 2013; Kreidberg et al. 2014a, 2014b; Ranjan et al. 2014; Haynes et al. 2015). WFC3 has two main observation modes commonly used for transiting exoplanet spectra: stare mode and spatial scan mode. Stare mode has been used for a majority of *HST* observations and is useful when observing dimmer target stars where the photon counts pixel^{-1} s^{-1} is low; observing brighter targets in this mode leads to saturation. Stare mode maintains a constant pointing of the telescope throughout the observation and maintaining the same pixel position on the detector. Spatial scanning mode was made available on WFC3 in Cycle 19 (2012) and is now implemented as the main mode of observation for transiting exoplanets, as targets observed for atmospheric follow-up with such instruments often orbit brighter target stars (*V* magnitudes brighter than 11). WFC3 spatial scanning involves nodding the telescope during an exposure to spread the light along the cross-dispersion axis, resulting in a higher number of photons by a factor of ten per exposure while considerably reducing overheads. This also increases the time of saturation of the brightest pixels, allowing for longer exposure times (McCullough 2011).

Mandell et al. (2013) conducted the first re-analysis test of WFC3 data for WASP-19b and WASP-12b with the addition of

the first analysis of WASP-17b. The re-analysis incorporated a wavelength-dependent systematic correction over the methods used in the previous analysis. While this study produced almost photon-limited results in individual spectral bins, the spectral features observed in the transmission spectra were degenerate with various models of temperature and composition making interpretation difficult. Ranjan et al. (2014) conducted a study of four stare mode WFC3 transit light curves from the large *HST* program, GO 12181. However, they were unable to resolve any features in the transmission spectrum for three of the planetary atmospheres and were unable to extract a robust transmission spectrum for one of the data sets, as different treatments to the data gave moderately different results.

Additionally, the analysis of the very hot Jupiter WASP-12b that was observed as part of GO 12230 (P. I. M. Swain) is a good example of differences caused by analysis techniques. WASP-12 was observed in stare mode using WFC3 G141 slitless grism, which contained the spectrum of both the target planetary host star, and the M-dwarf binary companions to WASP-12, which overlapped in the spectral response on the detector. The most recent re-analysis of this data explores the effect of systematic model correction on the absolute transit level measured for the atmospheric transmission spectrum and finds that the absolute transit depth is sensitive to the systematic model applied to the data (Stevenson et al. 2014b).

Spectroscopically resolving transmission absorption features in-transit light curves is important for determining the compositional information of exoplanet atmospheres. Unlike *Spitzer*'s IRAC instrument, which is able to obtain photometric points in H_2O absorption bands, *HST* WFC3 is vitally able to spectroscopically resolve features in the atmospheres of exoplanets. Accurately determining the parameters of a transiting planet via timeseries observations is dependent on robustly accounting for any systematic effects from the telescope and the detector for both ground- and space-based instruments. Each of the models used in the literature attempt to correct the array of systematics observed independently in each data set. However, not all data sets display the same combination of systematics, which appear to be highly dependent on the observational mode and setup.

In this paper we apply the marginalization method proposed by Gibson (2014) to not only incorporate the analysis of multiple systematic treatments on the light curve, but also to make it applicable to multiple data sets and allow for a cross-comparison between different transmission spectra. In Section 2 we outline the current methods used in the literature to reduce and analyze current WFC3 transit data sets. In Section 3 we put forward and discuss the marginalization analysis technique, which incorporates a number of previously published methods to produce robust transit parameters. We then discuss marginalization in the context of our observations in Section 4 and highlight the key aspects of marginalization for transit data sets. We then contrast this technique with others in Section 5 to assess the impact of different analysis methods and on the computed transit parameters.

2. SYSTEMATIC MODEL CORRECTIONS

A transit light curve consists of N stellar flux measurements observed at time t, collectively referred to as the data D or the light curve. To model each of the light curves we calculate a Mandel & Agol (2002) transit model $T(t, \lambda)$, following a nonlinear limb-darkening law as defined in Claret (2000). The model is then fit to the data by allowing the baseline flux F_*, R_p/R_*, and the center of transit time to vary. We fix the other planetary system parameters such as the inclination and a/R_* from previously published values because the phase coverage of *HST* light curves do not well constrain these parameters on their own. The final form of the model fit to the data becomes

$$F(\lambda, t) = F_* \times T(t, \lambda) \times S(t, \lambda), \quad (1)$$

where $S(t, \lambda)$ is the sytematics model normalized to unity.

One of the issues encountered when analyzing observational data sets is determining the impact that instrumental systematics have on the resultant measurements. Since the advent of WFC3's application to transiting exoplanets, a number of systematic models have been used to reduce G141 spectroscopic data (e.g., Berta et al. 2012; Gibson et al. 2012; Deming et al. 2013; Line et al. 2013; Wakeford et al. 2013; Stevenson et al. 2014b).

Figure 1 shows three examples of systematic trends observed in WFC3 transit light curves: "*HST* breathing" effects, visit-long slopes, and the "ramp" effect. The "*HST* breathing" effect shown in Figure 1(a) displays a periodic systematic across each orbit of data. This is attributed to the known thermal variations that occur during the orbit of *HST* as it passes into and out of the Earth's shadow, causing expansion and contraction of *HST*. This can be most easily seen in the middle panel in relation to the *HST* orbital phase. The "*HST* breathing" effect systematic has been noted and corrected for in multiple data sets with a variety of parametrized models (e.g., Line et al. 2013; Wakeford et al. 2013; Stevenson et al. 2014b) which remove systematics based on functions of the *HST* orbital time period and phase.

Many groups have also reported a visit-long trend in WFC3 light curves. This trend can be seen clearly in the raw band-integrated light curve of HD 209458 shown in Figure 1(b), which displays a significant slope across the entire observation period. This can be seen in relation to both the planetary phase and the exposure number. This systematic trend has not been correlated with any other physical parameter of WFC3 observations. However, it has been shown to significantly affect the resultant system parameters obtained from the light curve.

In addition to orbital phase trends in both planetary and *HST* space, a number of light curves has been dominated by a systematic increase in intensity during each group of exposures obtained between buffer dumps which is referred to as the "ramp" or "hook" effect (e.g., Berta et al. 2012; Mandell et al. 2013). The data set of WASP-17 in Figure 1(c) clearly displays this effect and the bottom residual plot in terms of exposure number shows the highly repeatable aspect of the systematic. This is thought to be caused by charge trapping on the detector and it has been found that the "ramp" is on average zero when the count rate is less than about 30,000 electrons per pixel (Wilkins et al. 2014). This is commonly seen in stare mode observations where the count rapidly builds up over a small pixel range.

Strong wavelength-dependent shifts of the stellar spectrum across the detector throughout the course of the observation can also affect the spectroscopic light curves and therefore the measured transit parameters (Deming et al. 2013). Large shifts

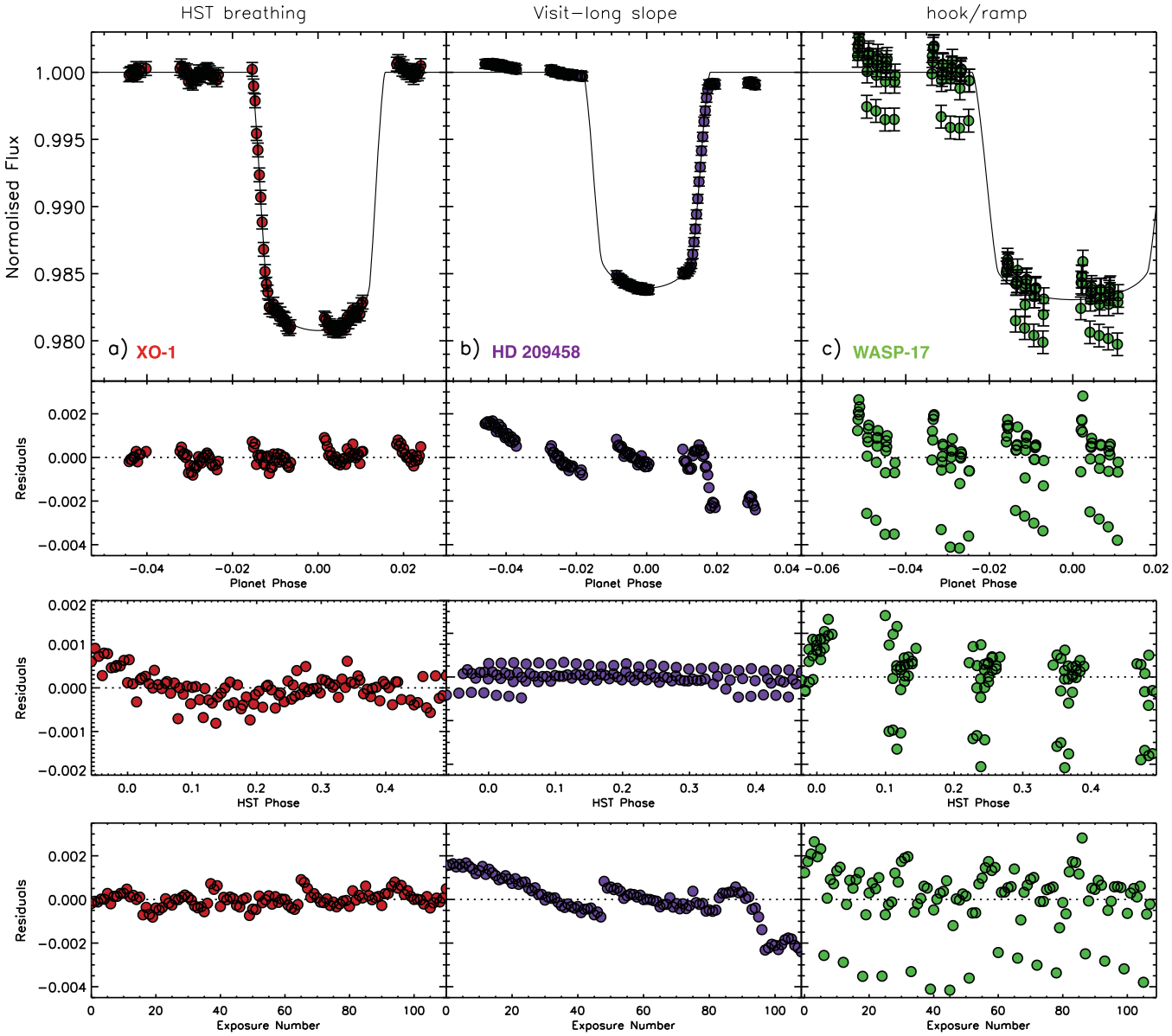

Figure 1. Three of the main systematic effects observed in *HST* WFC3 transit data sets: (a) "*HST* breathing" effect caused by the temperature variations in the orbital period of *HST*. (b) Visit-long slope, which is a linear trend observed across the entire observing period for all transit light curves observed with *HST* WFC3. (c) The "ramp" effect which is thought to be caused by charge trapping between buffer dumps. Lower panels show the residuals of each data set with respect to different timeseries parameters. Top: residuals in terms of planetary phase. Middle: residuals in terms of *HST* orbital phase, in which each *HST* orbit of data is overplotted on subsequent orbits. Bottom: residuals in terms of exposure number.

in the wavelength direction on the WFC3 detector may introduce subpixel- to pixel-sized variations in each spectroscopic bin when dividing the spectrum into wavelength channels. This is likely the result of fast scanning rates used for very bright targets where additional motion during a rapid scan rate can introduce variations between spectral exposures, as fast scan rates spread the stellar spectrum across a larger range of the detector and make it much harder for the fine guidance system to hold a fixed wavelength position on the detector.

A combination of observational and phase-dependent instrument systematics has been observed in all WFC3 transit data sets. We outline the main systematic models used to correct for these effects with the full list of published systematic correction models presently used outlined in Table 1.

2.1. Exponential Model Ramp

The first method put forward to correct WFC3 systematics, as an analytical model whose parameters attempt to represent the physical processes of the instrument, was outlined by Berta et al. (2012) defined as the exponential model-ramp. The exponential models apply an exponential ramp over sets of exposures, corrects for orbit-long and visit-long slopes, and is intended to model the charge trapping. Line et al. (2013) show the exponential model-ramp in the form of the equation

$$\frac{F_{\rm orb}}{F_{\rm cor}} = (C + V\theta + B\phi)(1 - Re^{\psi/\tau}), \quad (2)$$

where $F_{\rm orb}/F_{\rm cor}$ are the light curve residuals, θ is the planetary phase, ϕ is the *HST* orbital phase, and ψ is the phase over which the ramp feature occurs, which accounts for the visit-

Table 1
Table of Exoplanets Observed Using WFC3 G141 Grism, the Mode in Which the Data is Observed, the Systematic Model Used to De-trend the Transmission Spectral Data, Author, *HST* Program, and Year of Observation

HST Program	PI	Cycle	Planet	Year	Mode	Model	Paper
GO 11740	F. Pont	Cycle 17	HD 189733	2010	stare	$\mathcal{GP}$	Gibson et al. (2012)
GO 12181	D. Deming	Cycle 18	WASP-17	2011	stare	θ + doot + δ_λ	Mandell et al. (2013)
			HD 209458	2012	scan	$\theta + \delta_\lambda(\lambda).\ A(\lambda)$	Deming et al. (2013)
			XO-1	2011	scan	$\theta + \delta_\lambda(\lambda).\ A(\lambda)$	Deming et al. (2013)
			HAT-P-12	2011	stare	$\theta + e^{\psi/\tau} + \phi$	Line et al. (2013)
			WASP-19	2011	stare	θ + doot	Huitson et al. (2013)
						θ + doot + δ_λ	Mandell et al. (2013)
			WASP-4	2010	stare	doot	Ranjan et al. (2014)
			TrES-2	2010	stare	doot	Ranjan et al. (2014)
			TrES-4	2010	stare	θ + doot	Ranjan et al. (2014)
			CoRoT-1	2012	stare	θ + doot	Ranjan et al. (2014)
GO 12230	M. Swain	Cycle 18	WASP-12	2011	stare	θ	Swain et al. (2013)
						θ + doot + δ_λ	Mandell et al. (2013)
						$\theta + \phi^N$	Sing et al. (2013)
						$\theta + e^{\psi/\tau} + \phi$	Stevenson et al. (2014c)
GO 12251	Z. Berta	Cycle 18	GJ 1214	2010	stare	doot	Berta et al. (2012)
GO 12449	D. Deming	Cycle 19	HAT-P-11	2012	scan	$\theta + \delta_\lambda(\lambda).\ A(\lambda)$	Fraine et al. (2014)
GO 12473	D. Sing	Cycle 19	WASP-31	2012	scan	$\theta + \phi^N + \delta_\lambda^N$	Sing et al. (2015)
			HAT-P-1	2012	scan	$\theta + \phi^N$	Wakeford et al. (2013)
GO 12881	P. McCullough	Cycle 20	HD 189733	2013	scan	θ	McCullough et al. (2014)
GO 13021	J. Bean	Cycle 20	GJ 1214	2012–13	scan	$\theta + e^{\psi/\tau} + \phi$	Kreidberg et al. (2014a)
GO 13064	D. Ehrenreich	Cycle 19	GJ 3470	2013	stare	doot	Ehrenreich et al. (2014)
GO 13338	K. Stevenson	Cycle 21	GJ 436	2013	scan	$\theta^2 + e^{\phi} + \phi$	Stevenson et al. (2014a)
GO 13467	J. Bean	Cycle 21	WASP-43	2013	scan	$\theta + e^{\psi/\tau} + \phi$	Kreidberg et al. (2014b)
GO 13501	H. Knutson	Cycle 20	HD 97658	2014	scan	$\theta + e^{\psi/\tau}$	Knutson et al. (2014)

Note. $\mathcal{GP}$—Gaussian Process; doot—Divide-oot; θ—Visit-long slope; ϕ—*HST* phase; δ_λ—wavelength shift; $e^{\psi/\tau}$—model-ramp.

long slope V, the orbit-long slope B, and a vertical offset C applied to the whole light curve. The exponential model for the ramp has an additional two parameters: the ramp amplitude R and the ramp timescale τ. This method is used by a number of groups when a ramp is observed in the raw band-integrated light curve (Knutson et al. 2014; Kreidberg et al. 2014a, 2014b).

In cases where the orbital timescale matches the phase over which the ramp occurs, a simplification of the ramp-model can be used as seen in Stevenson et al. (2014a),

$$S(t,\ \lambda) = [1 + r_0\theta + r_1\theta^2] \times [1 - e^{r_2\phi + r_3} + r_4\phi], \quad (3)$$

where r_{0-4} are free parameters and the phase ψ over which the ramp feature occurs is now equal to the *HST* orbital phase ϕ. This is displayed as "$\theta + e^{\phi} + \phi$" in Table 1. Fraine et al. (2014) find that a simple linear visit-long correction and a single exponential ramp in *HST* phase can correct for the systematics observed in the transit light curve of the hot Neptune HAT-P-11 without the need for a squared term in time.

Both of these methods rely on the timescales of each ramp to be the same in each orbit and for each orbit to have the same repeating systematic. This, therefore, depends heavily on the scheduling of each exposure within an *HST* orbit, which a number of *HST* WFC3 data sets do not meet.

2.2. Polynomial Models

An additional polynomial method to correct for "*HST* breathing" effects in the data is discussed in Wakeford et al. (2013), which assumes that the effects fit a polynomial function rather than being exponential in nature. This method also seeks to remove the visit-long slope observed in each WFC3 transit data set using a linear time trend in the planetary phase in addition to the *HST* phase corrections,

$$S(t,\ \lambda) = T_1\,\theta + \sum_{i=1}^{n} p_i\phi^i, \quad (4)$$

where θ is the planetary phase representing a linear slope over the whole visit when multiplied by the free or fixed parameter T_1, ϕ is the *HST* phase accounting for "*HST* breathing" effects when multiplied by p_{1-n}, which are either free parameters or fixed to zero to fit the model to the data.

In addition to "*HST* breathing" trends, functions in wavelength shift on the detector were needed to correct for systematics seen in the light curve of WASP-12 (Sing et al. 2013). The systematic model then takes the form

$$S(t,\ \lambda) = T_1\,\theta + \sum_{i=1}^{n} p_i\phi^i + \sum_{j=1}^{n} l_j\delta_\lambda^j, \quad (5)$$

where δ_λ is the array of the shift in the wavelength (x) direction on the detector over the visit and l_{1-n} are fixed to zero or free parameters similar to that used for the "*HST* breathing" correction. δ_λ is created by cross-correlating the stellar spectrum for each exposure with a template stellar spectrum across the whole wavelength range. This is shown in Table 1 by a combination of the different systematics corrected for e.g., $\theta + \phi^N + \delta_\lambda^N$.

Importantly, this model does not require each of the orbits to have the same number of exposures, or consistent repeating systematics in each orbit or between each buffer-dump making it a robust method to apply to any transit data set from WFC3. However, due to the large number of potential free parameters in each systematic model and the potential to go to high orders of polynomial for each systematic, using this method can potentially introduce additional systematics if the correct model is not initially chosen.

2.3. Divide Out-of-transit

Berta et al. (2012) also suggested a separate method called divide-oot for correcting the systematic "ramp" or "hook" observed in a number of data sets. The divide out-of-transit method (divide-oot) relies on the hook systematic being "extremely" repeatable between orbits during the *HST* phase in a visit.

Divide-oot uses the out-of-transit orbits to compute a weighted average of the flux evaluated at each exposure within an orbit and divides the in-transit orbits by the template created. This requires each of the in-transit exposures to be equally spaced in time with the out-of-transit exposures being used to correct them such that each corresponding image has the same *HST* phase and that additional systematic effects are not introduced. While this does not rely on knowing the relationship between measured photometry and the physical state of the camera, it does require there to be an even number of exposures equally spaced from orbit to orbit where the exposures occupy the same *HST* phase space. This is so that the systematics induced by the known "*HST* breathing" trend caused by temperature variations in its orbit can be effectively eliminated. The divide-oot method relies on the cancellation of common-mode wavelength-independent systematic errors by operating only on the data themselves through the use of simple linear procedures and by relying on trends to be similar in the time domain over a number of orbits. This is listed as "doot" in Table 1.

2.4. Spectral Template Corrections

A somewhat similar technique was adopted by Deming et al. (2013) and Mandell et al. (2013) for their analysis of WFC3 data relying on common trends in the wavelength domain. To account for this subpixel change in the spectral wavelength solution for the stellar spectrum across each exposure, Deming et al. (2013) introduced a spectral template technique to the extracted spectrum of each image. The template spectrum is constructed using the average observed spectrum from exposures within one hour of the first and fourth contacts. The template is then used to fit the wavelength solution to each exposure spectrum by stretching and shifting it in wavelength, stepping through small increments to determine the best fit solution. The individual spectra are then divided by the template spectrum to create residuals, effectively removing any additional background contributions and the wavelength shift on the detector. This technique ("stretch and shift") also allows for the cancellation of common-mode systematics, similar to the divide-oot method; however, this requires the common-mode systematics to be consistent in wavelength rather than in time. This technique is labeled as "$\delta_\lambda(\lambda)$" in Table 1. However, this method only produces a relative depth measurement for the atmospheric signatures and does not provide absolute planet-to-star radius ratio values needed for comparative studies and combined data sets.

3. MARGINALIZATION

Thus far it is not clear which of the systematic models that have been employed to correct WFC3 transit light curves is the best for each individual data set. However, it is clear that none of these simple systematic models will work for all data sets. We attempt to rectify this by performing a marginalization over a grid of systematic models which incorporate corrections for the different systematics observed across the exoplanet transit data sets. In this way, model selection does not have such a large influence on the final result.

Use of the Bayesian Information Criterion (BIC) to select between different systematic models, which takes Occam's Razor into effect by penalizing models with increasing complexity, has been previously used for WFC3 observations (e.g., Sing et al. 2013, 2015; Wakeford et al. 2013; Haynes et al. 2015) and has been applied to a range of data sets (e.g., Crossfield et al. 2013; Huitson et al. 2013; Nikolov et al. 2014). Here we take this one step further by adopting the methodology proposed by Gibson (2014) of marginalizing over multiple systematic models to calculate robust transit parameters. This allows us to quantify the degeneracy between our physical parameters of interest and our choice of systematic model. In this case we want to determine the value and associated uncertainty of the planet-to-star radius ratio (R_p/R_*) for each of our light curves after correcting for the systematics inherent in the data. For each systematic model used to correct the data we calculate the evidence of fit, defined as the probability that the data would be produced given the systematic model, which is then used to apply a weight to the parameter of interest measured using that model. A weighted average of R_p/R_* is then calculated which takes into account the individual weights of each fit and statistical likelihood of each model. This ensures that a variety of systematic models is taken into account when measuring the R_p/R_* without having to choose between models.

Before marginalization the overall systematic models that are going to contribute to the final weighting must be decided upon. We use a grid of 49 polynomial models which incorporate all known identified systematic trends in the data (see Table 2) with the addition of the two exponential models outlined by Stevenson et al. (2014a) and the uncorrected light curve. All 52 models are then placed into an array of varying free parameters to be fit or fixed in turn and looped over for each light curve fit to calculate the weighting assigned to each in turn. We assume equal priors on all the systematic models tested where no model is preferentially preferred over another.

It is also important to note that marginalization relies on the fact that at least one of the models being marginalized over is a good representation of the systematics in the data. Our grid of parametrized models includes all combinations of factors up to the fourth order in both *HST* phase to correct for "*HST* breathing" effects and up to the fourth order in wavelength shift, in addition to the visit-long linear trend noted by all groups. By also incorporating the Stevenson et al. (2014a) exponential *HST* phase models with a linear and squared planetary phase trend, we make the assumption that this condition of marginalization is satisfied. Under this approach we effectively average the results obtained from a suite of systematic models in a principled manner. In doing so we

Table 2
Table of All Parametrized Systematic Models Applied to the Light Curves Showing the Combination of Visit-long Trends as a Function of Planetary Phase (θ), Functions of *HST* Orbital Phase (ϕ), and Functions Dependent on Wavelength Shifts (δ_λ) in the Data

No.	θ	ϕ	ϕ^2	ϕ^3	ϕ^4	δ_λ	δ_λ^2	δ_λ^3	δ_λ^4	No.	θ	ϕ	ϕ^2	ϕ^3	ϕ^4	δ_λ	δ_λ^2	δ_λ^3	δ_λ^4
0	…	…	…	…	…	…	…	…	…	25	√	…	…	…	…	…	…	…	…
1	…	…	…	…	…	√	…	…	…	26	√	…	…	…	…	√	…	…	…
2	…	…	…	…	…	√	√	…	…	27	√	…	…	…	…	√	√	…	…
3	…	…	…	…	…	√	√	√	…	28	√	…	…	…	…	√	√	√	…
4	…	…	…	…	…	√	√	√	√	29	√	…	…	…	…	√	√	√	√
5	…	√	…	…	…	…	…	…	…	30	√	√	…	…	…	…	…	…	…
6	…	√	…	…	…	√	…	…	…	31	√	√	…	…	…	√	…	…	…
7	…	√	…	…	…	√	√	…	…	32	√	√	…	…	…	√	√	…	…
8	…	√	…	…	…	√	√	√	…	33	√	√	…	…	…	√	√	√	…
9	…	√	…	…	…	√	√	√	√	34	√	√	…	…	…	√	√	√	√
10	…	√	√	…	…	…	…	…	…	35	√	√	√	…	…	…	…	…	…
11	…	√	√	…	…	√	…	…	…	36	√	√	√	…	…	√	…	…	…
12	…	√	√	…	…	√	√	…	…	37	√	√	√	…	…	√	√	…	…
13	…	√	√	…	…	√	√	√	…	38	√	√	√	…	…	√	√	√	…
14	…	√	√	…	…	√	√	√	√	39	√	√	√	…	…	√	√	√	√
15	…	√	√	√	…	…	…	…	…	40	√	√	√	√	…	…	…	…	…
16	…	√	√	√	…	√	…	…	…	41	√	√	√	√	…	√	…	…	…
17	…	√	√	√	…	√	√	…	…	42	√	√	√	√	…	√	√	…	
18	…	√	√	√	…	√	√	√	…	43	√	√	√	√	…	√	√	√	…
19	…	√	√	√	…	√	√	√	√	44	√	√	√	√	…	√	√	√	√
20	…	√	√	√	√	…	…	…	…	45	√	√	√	√	√	…	…	…	…
21	…	√	√	√	√	√	…	…	…	46	√	√	√	√	√	√	…	…	…
22	…	√	√	√	√	√	√	…	…	47	√	√	√	√	√	√	√	…	…
23	…	√	√	√	√	√	√	√	…	48	√	√	√	√	√	√	√	√	…
24	…	√	√	√	√	√	√	√	√	49	√	√	√	√	√	√	√	√	√
50	$\theta \times (1 - e^{\phi}) + \phi$									51	$(\theta + \theta^2) \times (1 - e^{\phi}) + \phi$								

In Addition to These Models we Apply the Two Exponential Orbital Phase Models Outlined in Stevenson et al. (2014a)

marginalize over our uncertainty as to which systematic model is actually the "correct" one. This is especially important when several models have equally well fitting systematics as is often the case.

3.1. Evidence and Weight

To calculate the weighting assigned to each of the systematic models and subsequently the final marginalized parameter for the planet-to-star radius ratio for each planetary transit, we first have to determine the evidence that each systematic model has when fit to the data. The evidence E_q of fit assigned to each systematic model S_q is given by the probability of the data D given the model q and is often referred to as the marginal likelihood. In the absence of accurate priors on which to place a likelihood we can use an approximation for the evidence (Gibson 2014), such as,

$$\ln E_q = \ln \mathcal{P}(D|S_q) \approx -\frac{1}{2}\mathrm{BIC} = \ln\left[\mathcal{P}(D|\alpha_*, S_q)\right] - \frac{1}{2}M \ln N, \quad (6)$$

where the BIC is the Bayesian Information Criterion which is equated to the logarithmic probability of the data given the parameter and systematic model ($\ln[\mathcal{P}(D|\alpha_*, S_q)]$) minus the number of free parameters M multiplied by the log number of data points being fit N.

The BIC is the most commonly used criterion to select between models in the current exoplanet literature. Alternatively to the BIC, the Akaike Information Criterion (AIC) can be calculated, which does not penalize the model as strongly for added complexity given a large number of data points,

$$\ln E_q = \ln \mathcal{P}(D|S_q) \approx -\frac{1}{2}\mathrm{AIC} = \ln\left[\mathcal{P}(D|\alpha_*, S_q)\right] - M. \quad (7)$$

Both the AIC and BIC have strong theoretical foundations and can be used for model selection (Burnham & Anderson 2004). As the number of data points in each data set greatly exceeds that of the number of free parameters in our most complex model, we choose to minimize the AIC to give our best-fitting model the largest evidence. This is also favored in Gibson (2014) as it allows for more flexible models into the likelihood, which typically leads to more conservative error estimates.

The evidence calculated for each model additionally relies upon the uncertainty placed on the data (σ), which is dominated by photon noise in spectral extraction pipelines. To ensure appropriate uncertainties, we start by applying photon noise error bars to each point and include our systematics models when running *MPFIT* (Markwardt 2009), which uses L–M least-squares minimization to fit the data. We then determine the best systematic model used based on the AIC and rescale the light curve uncertainties such that it has a reduced χ^2 of 1. We then re-run each of the systematic models with the light curves prior to to marginalization. Typically this rescales the errors by a factor of ~1.1 to ~1.2 times the theoretical photon noise limit of WFC3. Once we have applied this inflation to our error bars the approximated evidence is modified to incorporate the likelihood function.

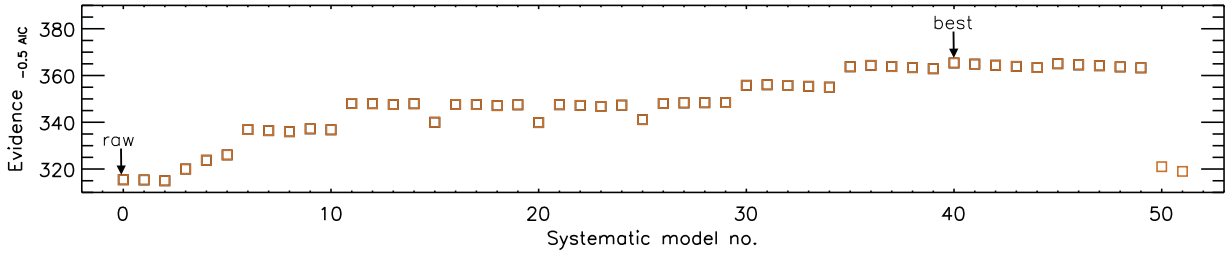

Figure 2. Evidence based on the AIC plotted against the systematic model number corresponding to Table 2. This is an example of the evidence computed for WASP-31 where the best-fitting systematic model and the raw light curve evidence are indicated with arrows.

To apply this to our data set we expand the above likelihood function,

$$\ln\left[\mathcal{P}(D|\alpha_*, S_q)\right] = \ln\left[\prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{r_i^2}{\sigma^2}}\right], \quad (8)$$

where σ is the uncertainty on the data and r_i represents the model residual for the ith data point.

$$\ln\left[\mathcal{P}(D|\alpha_*, S_q)\right] = \sum_{i=1}^{N} \ln\left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{r_i^2}{2\sigma^2}}\right], \quad (9)$$

$$= \sum_{i=1}^{N} \ln\left[\frac{1}{\sigma\sqrt{2\pi}}\right] - \frac{1}{2}\left(\frac{r_i}{\sigma}\right)^2, \quad (10)$$

$$= -N \ln\left(\sigma (2\pi)^{\frac{1}{2}}\right) - \frac{1}{2}\chi^2, \text{ leading to} \quad (11)$$

$$\ln\left[\mathcal{P}(D|\alpha_*, S_q)\right] = -N \ln\sigma - \frac{N}{2}\ln 2\pi - \frac{1}{2}\chi^2. \quad (12)$$

Substituting Equation (11) into (6) we arrive at

$$\ln E_q = -N \ln\sigma - \frac{1}{2} N \ln 2\pi - \frac{1}{2}\chi^2 - M. \quad (13)$$

This gives us the final form of the evidence function for each of our systematic models to which the data was applied. This now needs to be transformed into a weighting so that each of the systematic models is assigned a percentage of the overall probability and when normalized, $\sum_q \mathcal{P}(S_q|D) = 1$.

The individual weight (W_q) for each systematic model is calculated by

$$W_q = \mathcal{P}(S_q|D) = E_q / \sum_{q=0}^{N_q} E_q, \quad (14)$$

where N_q is the number of models fit, α_m is the marginalized parameter, and α_q is the measured parameter for each model.

The weighting assigned to each model due to the evidence parameter can then be used to calculate the weighted mean of all the parameters (α) of interest:

$$\alpha_m = \sum_{q=0}^{N_q} (W_q \times \alpha_q), \quad (15)$$

and the uncertainty (σ_α) on that parameter can be determined from σ_{α_q} i.e., the uncertainty on the parameter α determined from the qth model,

$$\sigma(\alpha) = \sqrt{\sum_{q=0}^{N} (W_q [(\alpha_q - \alpha_m)^2 + \sigma_{\alpha_q}^2])}. \quad (16)$$

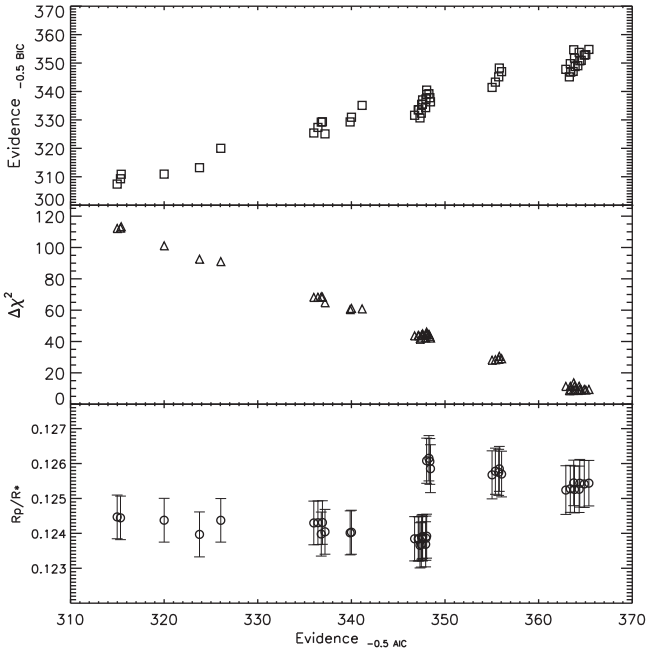

Figure 3. Evidence based on the AIC plotted against the evidence based on the BIC (top squares), the χ^2 (middle triangles), and the R_p/R_* (bottom circles) for the 52 models computed for WASP-31b. The best-fitting model is that with the highest AIC evidence parameter; while a similar trend is seen by reducing the $\Delta\chi^2$, the AIC penalizes the models for added complexity.

This assumes that the AIC is a good approximation for the evidence calculated for each systematic model and uses a Gaussian approximation for the posterior distributions (Frühwirth-Schnatter 2006).

From the marginalization over all 52 systematic models on the band-integrated light curve, the best-fitting model can reveal the dominant contributing systematics to each data set. Figure 2 shows the calculated evidence based on the AIC approximation for all 52 models when fit to an example data set where the systematic model with the highest overall weighting for the marginalization is indicated by an arrow along with the raw light curve fit. This clearly shows the sample of systematic models that are favored when correcting this data set and emphasizes the need for thorough systematic model analysis. It is sometimes the case that a number of models will have a strong weighting on the band-integrated light curve where their weight >10%. Often these models correct for the same combination of systematic trends assigning a different order to the polynomial used for correction to within one order.

In Figure 3 we demonstrate the statistical correlation between different factors used to select models, comparing the evidence based on the AIC approximation to that of the

Table 3
Observation Parameters for the Five Planetary Transits Measured with WFC3

	HAT-P-1	WASP-31	XO-1	HD 209458	WASP-17
GO Program	12473	12473	12181	12181	12181
Date	2012 Jul 05	2012 May 13	2011 Sep 30	2012 Sep 25	2011 Jul 08
Mode	Scan	Scan	Scan	Scan	Stare
NSAMP	4	8	9	5	16
Subarray size	512	256	128	256	512
Exposure Time (s)	46.695	134.35	50.382	22.317	12.795
Peak Pixel Count	25,000	38,000	38,000	44,000	64,000
No. Exposures	111	74	128	125	131
Scan Rate (pix s^{-1})	1.07	0.15	0.43	7.43	...
R_p (R_J)	1.319	1.55	1.209	1.38	1.932
M_p (M_J)	0.525	0.48	0.942	0.714	0.477
$T_{\rm eff}$ (K)	1322	1570	1210	1459	1755
Period (days)	4.46529976	3.405886001	3.94163	3.52474859	3.7354833692
Epoch (MJD)	53979.43202	56060.69042	55834.3419	56195.7595	55750.2973239
Inclination (°)	85.634	84.670	88.92	86.59	86.917160
a/R_*	9.91	8.19	11.24	8.859	7.03
T_* (K)	5980	6250	5750	6117	6550
[Fe/H]*	0.130	−0.2	0.02	−0.02	−0.25
Best-fit Band-integrated Limb-darkening coefficients					
c1	0.58115522	0.49271309	0.57505270	0.55407671	0.46913939
c2	0.08416305	0.32553801	0.07068500	0.15318674	0.38874433
c3	−0.1886652	−0.5299505	−0.1131334	−0.2741835	−0.6396845
c4	0.05957531	0.21191308	0.01649965	0.09583456	0.26591443

Note. The planetary system parameters used for each of the data sets is also listed along with the band-integrated limb-darkening parameters calculated using 1D Kurucz stellar models.

evidence based on the BIC approximation and to the $\Delta\chi^2$ for all 52 systematic models. Additionally we show the difference in the measured R_p/R_* computed for each model relative to the weight applied to that fit, again demonstrating the importance of using the correct systematic model when correcting transit data.

4. OBSERVATIONS

Observations of single transit events are conducted using *HST*'s WFC3 in the IR with the G141 spectroscopic grism. Our observations span two *HST* large programs and two observation modes, stare mode and spatial scan mode, acquired between 2011 and 2012 over 25 *HST* orbits.

Table 3 outlines the observational and planetary system parameters for each of the exoplanet hot Jupiters studied here: HAT-P-1b, WASP-31b, XO-1b, HD 209458b, and WASP-17b. Each of these exoplanet transmission spectra has been previously analyzed and published: HAT-P-1b (Wakeford et al. 2013), WASP-31b (Sing et al. 2015), XO-1b (Deming et al. 2013), HD 209458b (Deming et al. 2013), and WASP-17b (Mandell et al. 2013). Across these five exoplanet transmission spectra there are a variety of measured features over the expected H_2O absorption bands, from full amplitude features extending several scale heights in the atmosphere, to muted and absent features. The study here is intended to demonstrate the most uniform analysis and comparison to date, such that differences between spectra can be more easily tied to the planets themselves and changes between different reduction techniques can be minimized.

We use the "*ima*" outputs from WFC3's *Calwf3* pipeline. For each exposure, *Calwf3* conducts the following processes: bad pixel flagging, reference pixel subtraction, zero-read subtraction, dark current subtraction, nonlinearity correction, flat-field correction, and gain and photometric calibration. The resultant images are in units of electrons per second. A single exposure for each of the five targets is shown in Figure 4 with the target spectrum outlined. This figure already demonstrates the difference in the individual observation strategies used for WFC3. The larger subarrays used for HAT-P-1 and WASP-17 include both the zero and first order spectrum, while the smaller subarrays only contain the first order spectrum. There is also a clear difference between the brightest target, HD 209458 (V mag = 7.7), and dimmer targets (V mag > 10.3) where a much larger scan area is needed for bright targets so that the detector is not saturated during the exposure. We also note that each HAT-P-1 exposure includes the spectral trace of the companion star to the target exoplanet host star.

For each of our five exoplanetary transit data sets we extract each exposure spectrum with our custom *IDL* routine *spextract*, which optimizes the aperture over which the target spectrum is exposed on each image (see Figure 5). We then compute the band-integrated light curve by summing the flux over all exposed wavelengths to obtain the planet-to-star radius ratio (R_p/R_*) and center of transit time by fitting a Mandel & Agol (2002) transit model, created with nonlinear limb-darkening parameters, to our data using the *IDL* code *MPFIT*. Due to the limited phase coverage of the *HST* transit observations, when fitting the band-integrated and spectroscopic light curves for R_p/R_* we fix the system parameters such as inclination and a/R_* from a joint fit between *HST* STIS, WFC3, and *Spitzer* IRAC using a Markov Chain Monte Carlo (MCMC) analysis (Sing et al. 2016), as further constraints cannot be placed with these individual data sets.

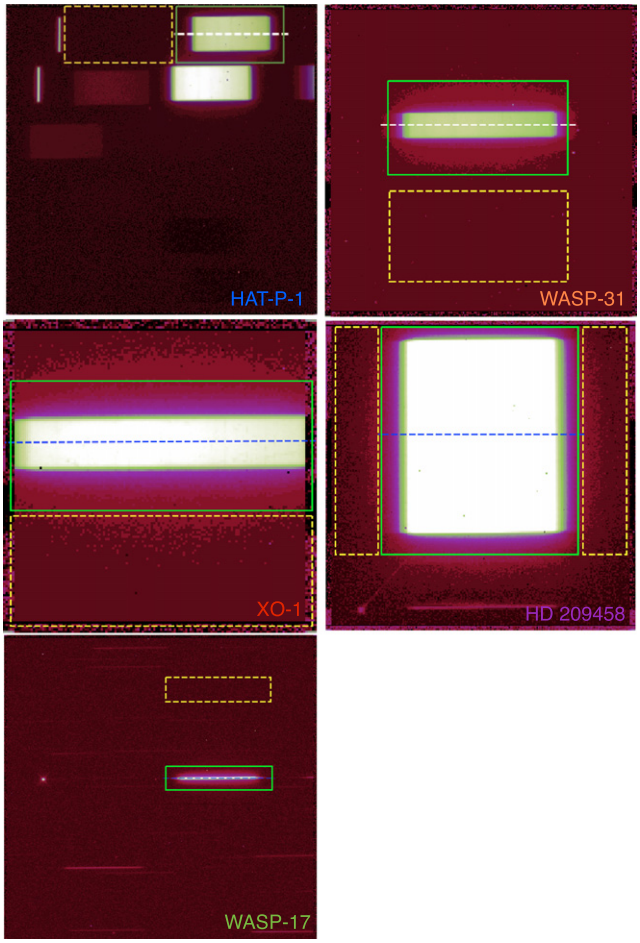

Figure 4. Single exposure ima output image for each observed exoplanet host star. Each spectral exposure is outlined in green with the center of the spectrum marked with a dotted line. The region used for background subtraction is boxed by a dashed yellow line.

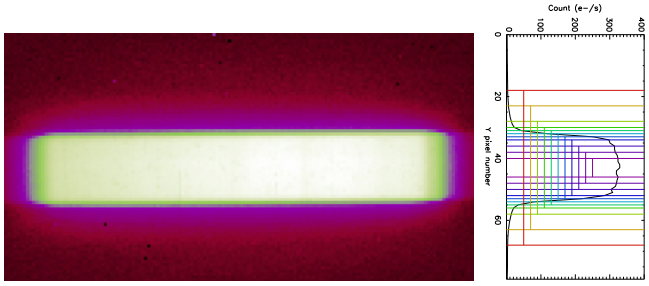

Figure 5. Left: example exposure in spatial scan mode from the "ima" output. Right: the spectral trace in the cross-dispersion direction from the one-pixel column is shown in black with a series of aperture sizes marked out by the colored brackets. The best aperture size is determined by minimizing the standard deviation of the light curve residuals when fit with a standard Mandel & Agol (2002) transit model from the base planetary parameters.

For each of the spectroscopic transit observations we compute the band-integrated light curve and transmission spectrum from 1.1–1.7 μm. Here we present the results from each of the hot Jupiter exoplanet transit light curves together, discussing the data with respect to the analysis technique and individual observation strategies.

4.1. Band-integrated Light Curve Marginalization

We use our systematic grid to determine the impacting systematics for each transit data set. The grid consists of 52 systematic models using a combination of systematics in the polynomial model and the two different exponential models (see Table 2). We then perform a marginalization on the computed planet-to-star radius ratio and center of transit time, and determine the best-fitting systematic model with the maximum weight from the band-integrated light curve.

For each of our data sets we use marginalization as a tool to compute robust transit depths from the band-integrated light curve. This also allows us to determine the main systematics impacting the light curves from the different spectral targets in a common and directly comparative manner. Figure 6 shows the evidence based on the AIC parameter for each systematic model used to correct the band-integrated light curve following the model numbers in Table 2. We highlight the model favored by this criterion corresponding to the systematic corrections with the highest weighting. From each of the fitting statistics it becomes clearer to see the differences across the data sets. For HAT-P-1 there is little difference between a large number of the systematic models with distinct groups of models which correct for only wavelength shifts and no "*HST* breathing" effects being disfavored in the final marginalization. Interestingly, the band-integrated light curve of WASP-31 does not favor systematic models that only correct for the known "*HST* breathing" effect, while the introduction of an additional visit-long linear correction makes these models most favorable. The systematic model correction for XO-1 favors higher-order polynomials across *HST* phase and wavelength shift with little difference incurred with or without the visit-long slope. HD 209458 strongly disfavors corrections with only the "*HST* breathing" effect, with the evidence based on the AIC approximation several hundred points below the other systematic correction fits (these points have been artificially shifted on the scale shown with their original values listed above the shown points). This makes the other systematic models relatively stable in that no model appears to be greatly favored above any others.

For WASP-17b it can clearly be seen in the raw light curve that there is a strong hook feature present (Figure 1) with an orbit-to-orbit repeating pattern in the residuals. The grid of systematic models used to correct the white light curve do not accurately account for this hook. To effectively correct for this systematic in the white light curve we use the divide-oot routine discussed in Section 2.3, which removes common-mode time-dependent systematics. We then additionally apply a marginalization over the systematic grid to remove any additional time-independent systematics and determine the planet-to-star radius ratio in which most models incorporating θ, ϕ, and δ_λ show some favorability. Those only correcting for "*HST* breathing" trends are again the least-favored in each group of models applied to the data.

We also note that the two exponential models for each of the light curves are not heavily favored and likely contribute negligibly to the resultant transit parameters. This may be the result of equal priors being placed on all the models tested. While this assumption holds when considering models from within the same family (i.e., polynomial expansions), separate priors may need to be placed when combining information from two different families of models. However, given that the

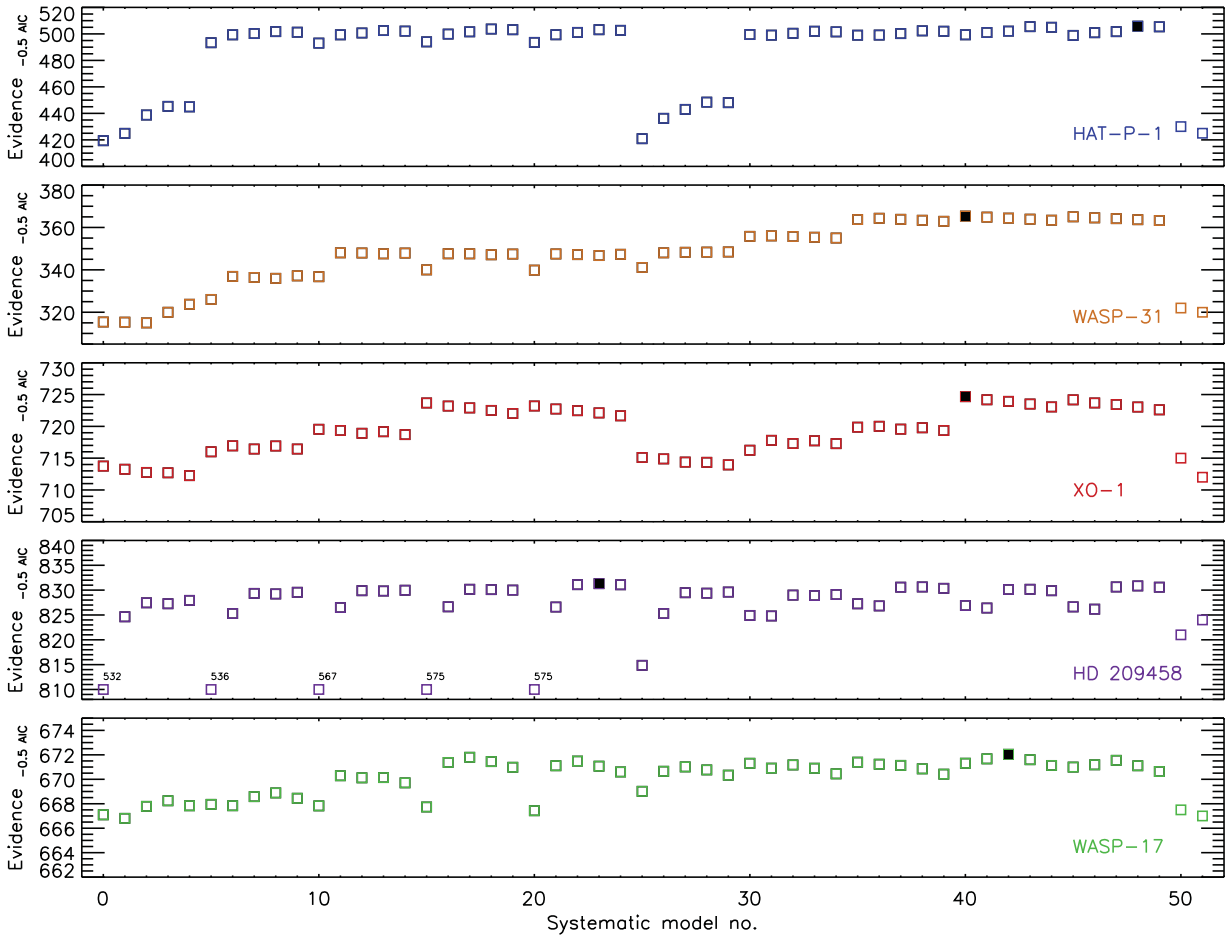

Figure 6. Evidence based on the AIC approximation for each systematic model applied to the band-integrated light curve for each exoplanet transit. Top to Bottom: HAT-P-1 (blue), WASP-31 (orange), XO-1 (red), HD 209458 (purple), and WASP-17 (green). The best-fitting systematic model for the band-integrated light curve is filled in with a black square for each planet in each plot. A number of points on the HD 209458 plot have been artificially shifted on the scale shown with their original values listed above the shifted points as they are strongly disfavored by the data.

exponential models were always found to give negligible weights, they are not likely sensitive to the priors.

Marginalization offers a transparent analysis method in which each stage can be deconstructed and used to understand both the data set and the instrument. Figure 6 shows the top 10 models favored for each of the transit data sets analyzed and the resultant planet-to-star radius ratio following the L–M least-squares fitting (L–M) with the Mandel & Agol (2002) transit model. This shows more clearly the impact that the different systematic models have on the desired transit parameters. When the weighting assigned to the systematic model becomes negligible, the impact on the final values also reduces, meaning only the best models become relevant in the final parameter determination. From this it is also possible to determine the main source of corrections associated with different models. For example, the absolute radius ratio computed from the WASP-17b data indicates that the visit-long correction impacts both the calculated radius ratio and the uncertainty associated with the fit. The visit-long correction is especially important for this data set to account for the absence of post-transit data used to determine the stellar baseline flux. By using the results computed from all of the systematic models that may best represent the data, marginalization produces a robust and accurate description of the specific data set and thus the planetary parameter as can be seen in the final band-integrated radius ratio and the associated uncertainty which is shown by the colored lines in each plot.

Similar to our previously published work (e.g., Wakeford et al. 2013; Nikolov et al. 2014; Sing et al. 2015), to corroborate the uncertainties estimated by L–M using MPFIT, we also applied a MCMC analysis using the IDL routine EXOFAST (Eastman et al. 2012) to the band-integrated light curves for the five most favored systematic models in each planetary data set. While MPFIT uses L–M least-squares fitting the parameter errors from the covariance matrix calculated using numerical derivatives, the MCMC evaluates the posterior probability distribution for each parameter of the model. Using the MCMC analysis we find that the uncertainties are within 10% of those calculated using our analysis with L–M, as they are largely Gaussian in nature. We calculate the mean percentage difference between the MCMC and L–M analysis for the uncertainties associated with the measured R_p/R_*, calculated using the top 5 systematic models on the band-integrated light curve for each data set, HAT-P-1b, WASP-31b, XO-1b, HD 209458b, and WASP-17b, respectively, 1.6%, 7.0%, 6.1%, 4.8%, and −16.4%. In the case of WASP-17b, the MCMC solution produces slightly larger uncertainties on average compared with MPFIT. This slight overestimation in the MCMC is likely the result of separately fitting the residual

systematics with the model and divid-oot whereas when using MPFIT we fit these simultaneously.

Marginalization is therefore applicable to a range of data sets where the general systematics that may impact the data are known but not well understood for each observation. By computing the desired parameter for each systematic model and marginalizing over all results according to their likelihood, information is preserved. This is important for transit light curves where the absolute radius ratio is important for combining data from different instruments (i.e., STIS, WFC3, and Spitzer) and when conducting comparative studies between different exoplanet atmospheres.

4.2. Spectroscopic Light Curve Marginalization

In addition to performing marginalization on the band-integrated light curves, we also compute the evidence and weighting for each systematic model applied to the individual spectroscopic light curves to test the impact on the calculated transmission spectrum.

We compute the spectroscopic transmission spectrum by binning up each exposure spectrum into a number of wavelength bins. We perform sigma clipping on each of the data sets using the best-fitting band-integrated light curve systematic model from the grid shown in Table 2 to remove wavelength-dependent outliers that deviate from the residual scatter by greater than 4σ. We then fit the transit light curve in each bin with our systematic grid to determine any wavelength-dependent systematics in each bin. We then compute the marginalized transit parameters for each spectroscopic light curve and thus the final transmission spectrum.

The grid of systematic models applied to each data set is based on three main systematics observed in WFC3 timeseries data. Of these three systematics, visit-long slope, *HST* thermal variations, and wavelength shift, only the shift in wavelength on the detector is based on a wavelength-dependent array. As outlined in Section 2.2 the array for the shift in wavelength over the course of the observation is calculated by cross-correlating the stellar spectrum with a template spectrum. This array can be calculated for the whole stellar spectral band or calculated for each spectroscopic bin using a wavelength-defined section of the spectral trace. In the following sections we analyze the transmission spectrum of our five targets using the band-integrated wavelength shifts (δ_λ) and the individual spectroscopic wavelength shifts ($\delta_\lambda(\lambda)$) through use of marginalization to evaluate the impact of the systematics. We then apply a spectroscopic shift and stretch ($\delta_\lambda(\lambda)$. $A(\lambda)$) method to determine the robust nature of transmission spectral structures in each planetary data set (see Figure 8) and present marginalization as the best technique for measuring planetary transmission spectra.

4.2.1. Band-integrated Wavelength Shifts

As shown by the highest weight models from the band-integrated light curves (Figure 7), shifts in the wavelength position of the spectrum across the course of the observation can play a significant role in the light curve systematics. To determine the impact of the physical shift on the detector across the whole spectral trace we compare the wavelength shift to the raw flux residuals computed from the band-integrated light curve (δ_λ) and calculate the correlation coefficient and false alarm probability for each data set as measured by the Spearman's correlation coefficient. Figure 9 shows that there is no generic correlation between the shift in wavelength on the detector and the systematics observed in the band-integrated light curve for almost all data sets considered, although some do show significant trends. This highlights the need to investigate the correlation between wavelength shift on the detector and the systematics present in the band-integrated light curve as the effect could be significant in order to detrend the data.

The strong correlation between wavelength shifts and the raw data measured for HD 209458, which was also noted by Deming et al. (2013), suggests that wavelength-dependent systematics are dominant in these observations. This correlation is also shown by the most favored systematic model from the band-integrated light curve based on the weighting calculated prior to marginalization. The spectrum of HD 209458b differs significantly from the other targets in this sample as it is a much brighter target, roughly 3 mag. This requires a much higher scan rate over a larger range, which likely leads to the larger shifts measured. We investigate the impact this large wavelength shift has on the computed transmission spectrum and the use of different techniques related to the wavelength shift to account for this noted systematic effect.

Similar to the analysis conducted on the band-integrated light curve, we calculate the R_p/R_* and uncertainty by fitting each spectroscopic light curve with all 52 systematic models using the δ_λ array. We then marginalize over all models in each spectroscopic bin using the MLE based on the AIC to produce the atmospheric transmission spectrum for each planet. This assumes that the shift in wavelength position on the detector is constant across the whole spectral trace with no additional systematic shifts in each individual wavelength regime selected by binning the data. Again, by using marginalization in each bin separately, we allow the data to define the systematic corrections being applied to each light curve and then use all of the information to calculate the final transmission spectrum. In each case this produces a marginal amount of additional scatter in the computed R_p/R_* for each bin while maintaining a robust transmission spectral shape across the whole spectral range.

4.2.2. Spectroscopic Wavelength Shifts

To determine if there is a wavelength dependence to the shifts on the detector, we calculate an array of shifts for each spectral bin, $\delta_\lambda(\lambda)$. Each $\delta_\lambda(\lambda)$ array is calculated in the individual wavelength channels by cross-correlating each portion of the spectrum with a template of the same region of the stellar spectrum for that spectroscopic channel. A similar method is used in the analysis of WASP-33b emission spectra by Haynes et al. (2015), which showed additional wavelength-dependent systematics associated with the shift of the spectral trace in the wavelength position on the detector.

Figure 10 shows the wavelength shift calculated for each exposure by cross-correlating the spectra for bins spanning 10 pixels compared with the wavelength shift across the whole spectrum shown as a bold black line. This shows that in a majority of the data sets the shift in wavelength is consistent across the spectrum and across each exposure, as shown from the correlation in Figure 9. However, the spectrum of HD 209458b shows significant shifts in each wavelength bin similar to the shifts in the full spectrum with additional changes between the different spectral channels. Again this may be the

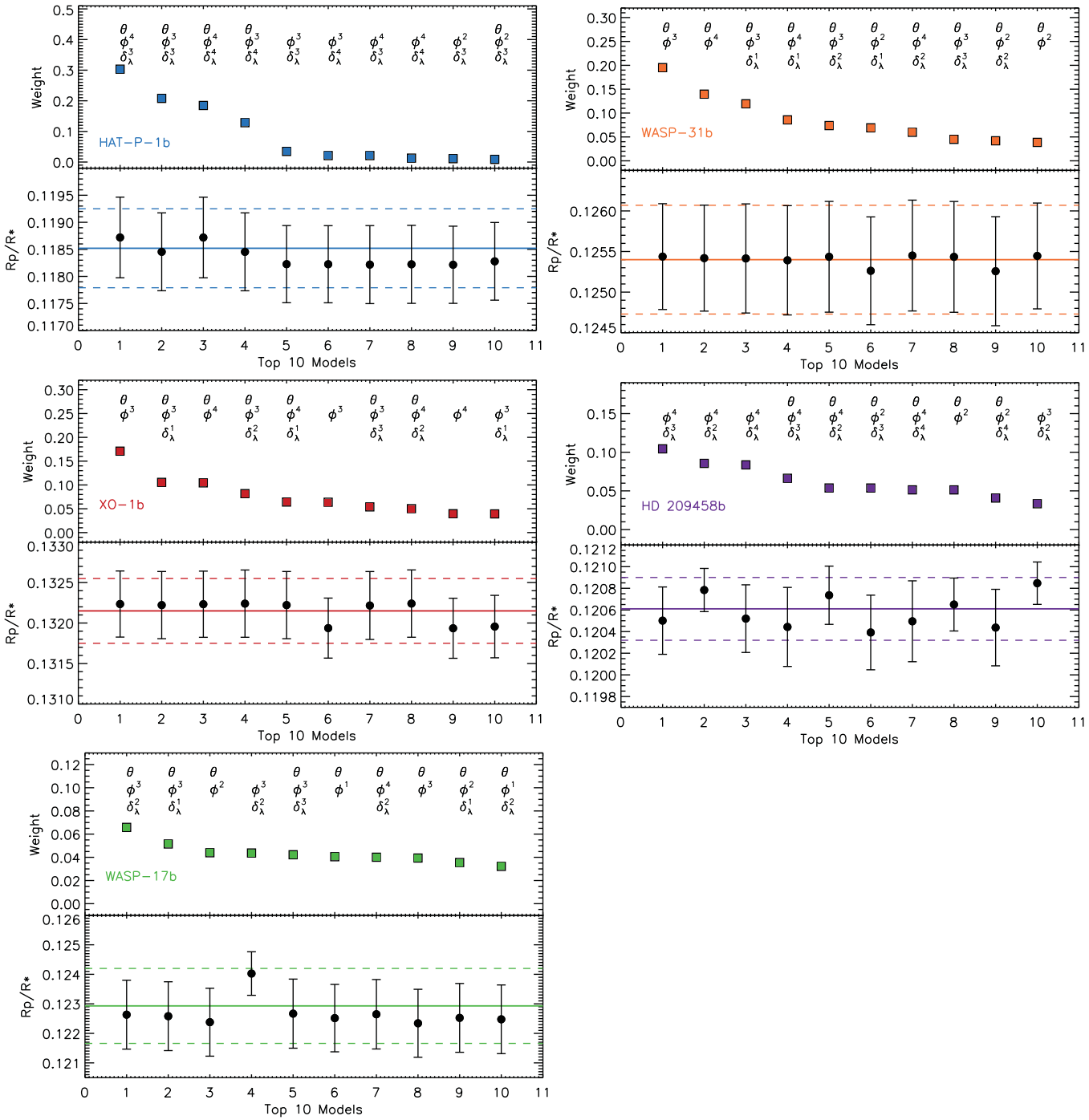

Figure 7. For each planetary data set: HAT-P-1b (blue, top left), WASP-31b (orange, top right), XO-1b (red, middle left), HD 209458b (purple, middle right), and WASP-17b (green, bottom left). Top: weighting for the top 10 models fit to the band-integrated light curve based on the AIC approximation. The model parameters are outlined below each model with best to worst from left to right. θ corrects a visit-long slope, ϕ^N the *HST* orbital phase, where N is the order of the polynomial used and δ_λ^N is the wavelength shift polynomial correction applied. Bottom: The computed R_p/R_* and uncertainty for each of the models as in the top plot. The solid horizontal line represents the final marginalized radius ratio with the dashed lines marking the uncertainty range.

result of the fast scan rate used in the observations of HD 209458 as it is a much brighter star.

To test the effect these have on the resultant transmission spectrum we used the wavelength shift for each channel for the polynomial models in our systematic grid and applied it to each bin (i.e., replacing our δ_λ array with an array of $\delta_\lambda(\lambda)$ for each wavelength bin). We then marginalized over the whole of our systematic grid to produce the final transmission spectrum across the spectral range.

In Figure 11 we show the transmission spectra of all five targets for each systematic model using the $\delta_\lambda(\lambda)$ arrays. Similar to the analysis conducted using the δ_λ arrays, the shape of each transmission spectrum appears robust across most systematic models with variations in the scatter across each of the data sets. In the bottom of each plot we display the marginalized transmission spectrum (triangles) calculated using the whole systematic grid. This demonstrates how each of the systematic models affect the final R_p/R_* value in each bin.

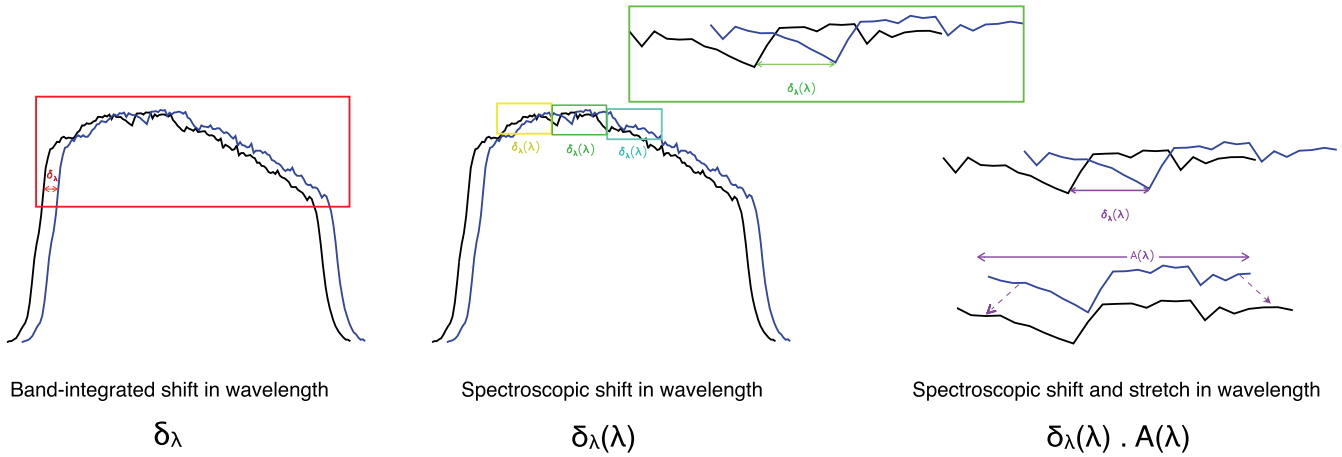

Figure 8. Diagram showing the three different methods for addressing the shift in the spectral wavelength position on the detector over the course of the observations. Band-integrated shift in wavelength, δ_λ, calculated using the whole spectral range. Spectroscopic shifts in wavelength, $\delta_\lambda(\lambda)$, calculated for each wavelength bin. Spectroscopic shift and stretch in wavelength, $\delta_\lambda(\lambda)$. $A(\lambda)$, calculated using both a shift and stretch in the spectral template to correct systematic shifts on the detector during the observation.

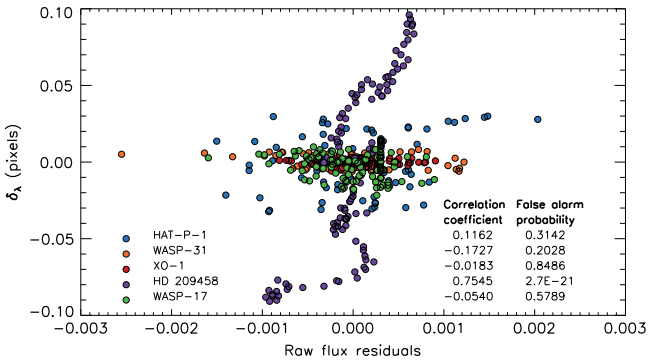

Figure 9. Trends in band-integrated wavelength shift (δ_λ) in correlation to the observed stellar flux, where the transit is removed. The correlation coefficients show that the δ_λ is weakly correlated with all of our data sets apart from HD 209458b, which shows a strong correlation (0.7545) with a negligible false alarm probability.

Larger scatter represents larger uncertainties on the final points, and where small uncertainties are observed it is likely the highest weighted models lie in a specific portion of the systematic grid for each bin. In this analysis we also show the transmission spectrum calculated using the highest weighted systematic model, defined by the band-integrated light curve in Section 4.1 (asterix). In three out of five cases this matches very closely with the marginalized transmission spectrum, suggesting in these cases that additional wavelength-dependent systematics do not greatly affect the spectroscopic light curves. However, in the case of WASP-17b the transmission spectrum calculated using a single systematic model results in a significant difference in the the shape and uncertainties associated with the spectrum. As seen in the band-integrated light curve analysis (Figure 7) this is likely due the small final weighting assigned to any particular systematic model from the grid, so no particular model is adequately able to define the systematics in the light curves.

In contrast with the small changes observed between the computed transmission spectra of WASP-17, the transmission spectra computed for HD 209458b show distinct differences between the two spectra with increased scatter in over 20% of the spectral bins and a number of points falling outside the range of the plot. In the case of HD 209458b's transmission spectrum, using the most favored band-integrated systematic model to fit the spectroscopic data clearly no-longer applies to the individual bins when using the $\delta_\lambda(\lambda)$ array. This further emphasizes the power of the marginalization technique where the systematics associated with the observations are not well known over a single systematic model fit based on general light curve statistics.

4.2.3. Wavelength Stretch and Shift Template

To further test the impact of wavelength shift treatment on the resultant transmission spectrum we additionally apply the wavelength "stretch and shift" template method employed by Deming et al. (2013) as described in Section 1.1. In each bin the spectrum is smoothed and fit with a template that is shifted in wavelength position and stretched in amplitude ($\delta_\lambda(\lambda)$. $A(\lambda)$), as an averaged spectrum will likely be slightly broadened by the wavelength shifts. This produces a residual array for each wavelength channel with both common-mode band-integrated and wavelength shift systematic corrections removed.

This method, which is similar to divide-oot, relies on canceling out common-mode systematics and assumes that *HST* breathing effects and linear trends in time are both wavelength-independent systematics. To account for any residual wavelength-dependent systematics in the spectroscopic light curves, we modify the "stretch and shift" technique to run with our systematic grid to determine if any additional systematics are present in the spectroscopic light curve. We then marginalize over the whole systematic grid to calculate the final transmission spectrum for each planet.

4.3. Marginalized Transmission Spectra

In Figure 12 we show the marginalized transmission spectrum from all three of our wavelength shift treatments: δ_λ (circles), $\delta_\lambda(\lambda)$ (triangles), and $\delta_\lambda(\lambda)$. $A(\lambda)$ (squares). The transmission spectra for HAT-P-1b, WASP-31b, XO-1b, and WASP-17b show no statistically significant differences between each of the three treatments. Importantly, this demonstrates that the more intensive treatments to the

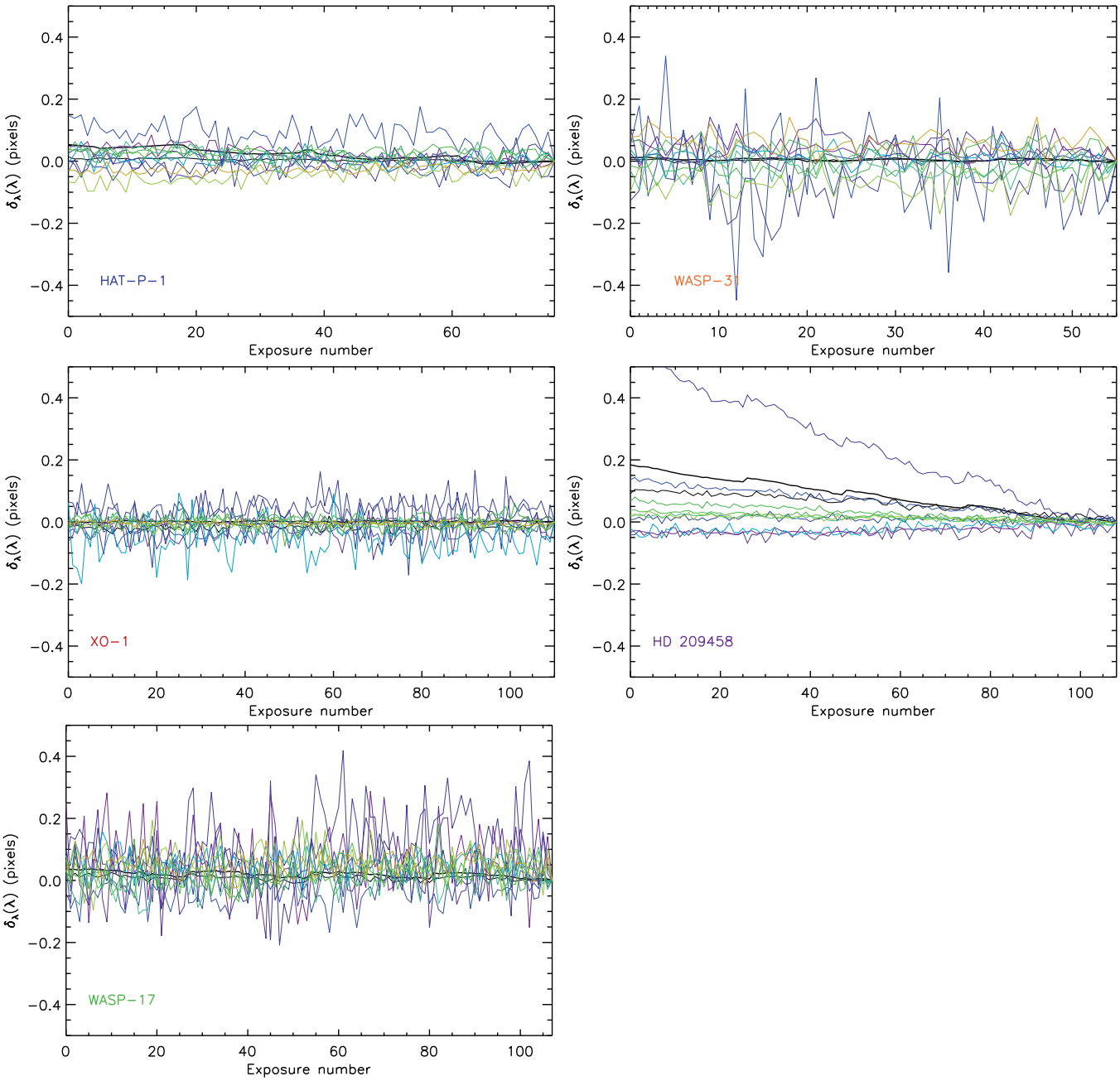

Figure 10. Plot of the wavelength shift measured in each spectroscopic bin of the observed stellar spectrum, $\delta_\lambda(\lambda)$, when cross-correlated with a template spectrum for each bin. The wavelength shift measured using the whole spectrum is shown as a thick black line. Each planetary data set is represented in a separate plot. The wavelength shifts are shown for spectroscopic bins spanning 10 pixels.

wavelength shift systematics are not producing additional non-physical absorption signatures in the transmission spectrum.

The transmission spectrum of HD 209458b, however, shows large differences between all three of the methods employed to correct the significant wavelength shifts that occurred over the course of the observations. This strongly suggests that the shifts in wavelength observed at the band-integrated level and at the spectroscopic level have a large effect on the systematic treatment and therefore on the interpretation that can be placed on the calculated transmission spectrum. Using the δ_λ correction produces a flat transmission spectrum with large uncertainties, while structure emerges with reduced uncertainties using both the $\delta_\lambda(\lambda)$ and the "stretch and shift" $\delta_\lambda(\lambda)$. $A(\lambda)$ techniques. This suggests that there are wavelength-dependent systematics associated with the wavelength shifts in the spectroscopic light curves which are used to calculate the atmospheric transmission spectrum.

In this study we present marginalization as a method of systematic correction that can be applied to a large series of data sets simultaneously using the individual data sets to define the corrections applied. As such, the systematic models being marginalized over need to represent the diversity in data sets and observation strategies. Here we present marginalization along with the $\delta_\lambda(\lambda)$ treatment as a main result of our analysis, as it treats the wavelength dependence of the observational systematics in a consistent manner as defined by the individual

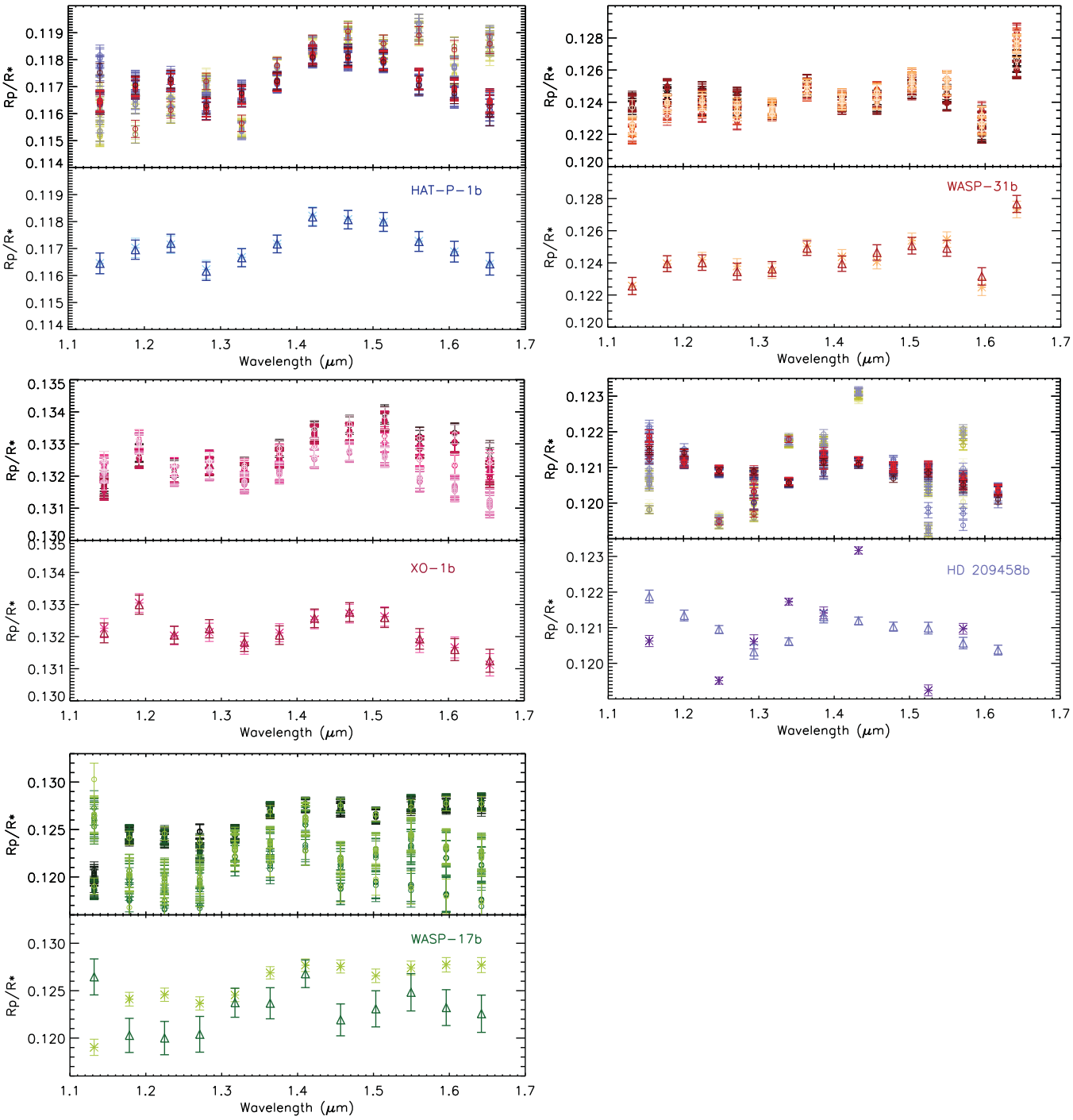

Figure 11. Marginalized transmission spectrum for each of our five exoplanet atmospheres using spectroscopic wavelength shifts, $\delta_\lambda(\lambda)$. Top of each plot: the transmission spectrum for each planet calculated using our systematic grid. Bottom of each plot: the marginalized transmission spectrum calculated using systematic wavelength shifts, $\delta_\lambda(\lambda)$ (triangles), and the transmission spectrum calculated using spectroscopic wavelength shifts and the best-fitting systematic model as determined from the band-integrated light curve analysis (stars). This shows that in most cases there is little significant difference between the two choices of treatment to the wavelength shifts. However, when the difference between shifts in each spectroscopic bin is large as seen in HD 209458b, the spectroscopic treatment of the systematics using the best-fitting model from the band-integrated fit is not effective.

data set while still allowing for a direct comparison between each planetary atmosphere by maintaining the relative absolute depth of the observed transit. The $\delta_\lambda(\lambda)$ correction is favored over the stretch and shift method, as stretch and shift uses smoothing and common-mode techniques which adjust the absolute baseline of the transmission spectrum and make interpretation of low amplitude spectra difficult. The $\delta_\lambda(\lambda)$ correction is also favored over the broad-band δ_λ correction as it accounts for the small variations in the wavelength shifts measured in each bin which impact the data rather than using the average, as the deviation from the average can vary from observation to observation (see Figure 10). We quote the computed transmission spectra for each of the planetary data sets following marginalization of our systematic grid with the $\delta_\lambda(\lambda)$ correction in Tables 4–7. In the case of HD 209458 where fast scan rates were used as a result of a bright target star

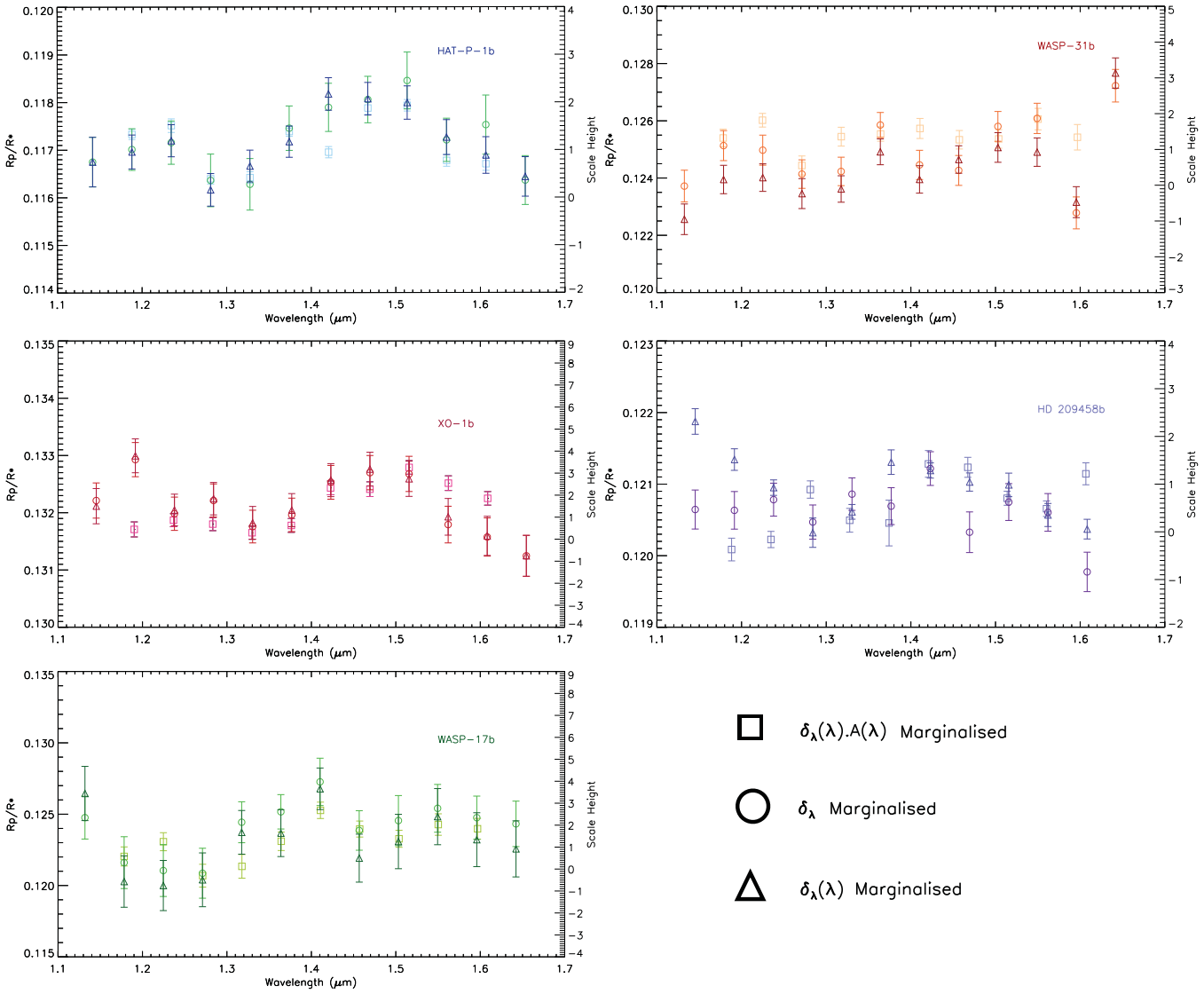

Figure 12. Marginalized transmission spectrum calculated using stretch and shift $\delta_\lambda(\lambda)$. $A(\lambda)$ (squares), the marginalized transmission spectrum calculated using spectroscopic wavelength shifts $\delta_\lambda(\lambda)$ (triangles), and the marginalized transmission spectrum calculated using the band-integrated wavelength shift δ_λ (circles). Here we present the $\delta_\lambda(\lambda)$ transmission spectrum as the most robust and consistent systematic treatment of WFC3 data sets, to enable large cross-comparisons between multiple observations, and therefore atmospheric detections, when related to the specific planetary scale height (right y-axis).

Table 4
Table of Marginalized $\delta_\lambda(\lambda)$ Transmission Spectral Properties of HAT-P-1b Binned to 46.5 nm

Wavelength	R_p/R_*	σ
band-integrated	0.118520	0.000730
1.1416	0.116748	0.000520
1.1881	0.116959	0.000356
1.2346	0.117194	0.000336
1.2811	0.116162	0.000342
1.3276	0.116660	0.000334
1.3741	0.117171	0.000327
1.4206	0.118171	0.000341
1.4671	0.118070	0.000342
1.5136	0.117989	0.000350
1.5601	0.117262	0.000366
1.6066	0.116885	0.000384
1.6531	0.116434	0.000414

we present the "stretch and shift" method, $\delta_\lambda(\lambda.A(\lambda)$, as brighter targets require an additional level of processing to produce robust and accurate transmission spectra (Table 8).

Figure 13 shows the marginalized transmission spectra computed for all of our targets compared with the previously published transmission spectral values. Where appropriate we have attempted to match the resolution of the previous measurements. In the case of WASP-17b this was not possible due to the quality of the data and the lower-resolution spectrum is shown. In the cases of XO-1b and HD 209458b the published transmission spectra have been shifted by −0.0016 and −0.00022, respectively, as the published spectra include a common-mode technique which affects the absolute baseline of the calculated transmission spectra. This shows that the marginalization technique can be used both on the band-integrated light curve and the spectroscopic light curves while maintaining the absolute depth and shape of transit. For HD 209458b we find larger uncertainties associated with

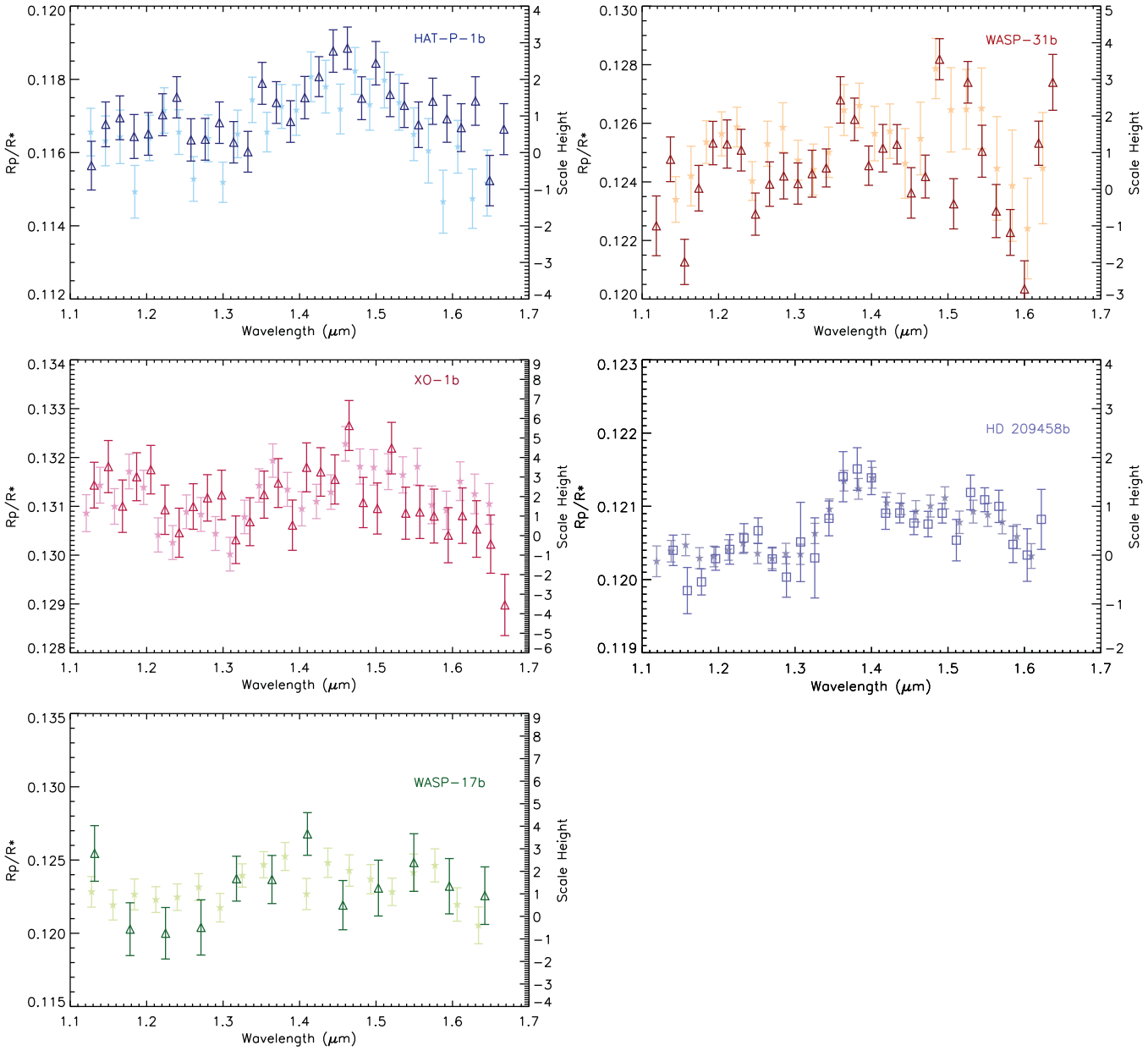

Figure 13. Marginalized transmission spectrum calculated using spectroscopic wavelength shifts $\delta_\lambda(\lambda)$ (triangles) plotted against the published transmission spectra (stars). HAT-P-1b (Wakeford et al. 2013), WASP-31b (Sing et al. 2015), XO-1b, and WASP-17b (Mandell et al. 2013), and the transmission spectrum computed using "stretch and shift" (squares) for HD 209458b (Deming et al. 2013).

several bins as marginalization incorporates the information from all models in the systematic grid while the published data is fit only with a linear trend in time.

5. CONCLUSION

We have applied the systematic marginalization technique proposed by Gibson (2014) to *HST* WFC3 light curves to determine robust transit parameters across multiple data sets. This has allowed us to incorporate all of our current knowledge of *HST* WFC3 into a single systematic treatment when estimating the planet parameters. We examine the marginalization technique in relation to the current field and present an outline of the methods used. We compute the maximum likelihood estimation for each systematic model when fit to the data based on the AIC. We then calculate the weight assigned to each systematic model and use the information from all tested models to calculate the final marginalized transit parameters for both the band-integrated light curve and the spectroscopic light curves to construct the atmospheric transmission spectrum. This allows for a more thorough exploration of the degeneracies between the planet signal and the affecting systematics to derive more realistic uncertainties on the data.

We demonstrate that marginalization is able to more accurately describe the data producing realistic uncertainties as well as provide valuable insight into the instrument and impacting systematics. We find that the systematic models with the highest weight all include a higher order (third or fourth order) polynomial correction in *HST* phase to account for the thermal variations over the course of each orbit. We also find that a linear correction in time across the whole visit is required in over five out of the top 10 systematic models selected for

Table 5
Table of Marginalized $\delta_\lambda(\lambda)$ Transmission Spectral Properties of WASP-31b Binned to 46.3 nm

Wavelength	R_p/R_*	σ
Band-integrated	0.125400	0.000670
1.1324	0.122559	0.000534
1.1787	0.123949	0.000495
1.2250	0.124015	0.000481
1.2713	0.123456	0.000523
1.3176	0.123617	0.000461
1.3639	0.124915	0.000451
1.4102	0.123952	0.000479
1.4565	0.124648	0.000479
1.5028	0.125068	0.000517
1.5491	0.124905	0.000496
1.5954	0.123157	0.000540
1.6417	0.127668	0.000529

Table 6
Table of Marginalized $\delta_\lambda(\lambda)$ Transmission Spectral Properties of XO-1b Binned to 46.3 nm

Wavelength	R_p/R_*	σ
Band-integrated	0.132151	0.000401
1.1452	0.132115	0.000307
1.1915	0.132994	0.000296
1.2378	0.132048	0.000276
1.2841	0.132245	0.000283
1.3304	0.131824	0.000285
1.3767	0.132045	0.000289
1.4230	0.132569	0.000288
1.4693	0.132762	0.000295
1.5156	0.132593	0.000307
1.5619	0.131929	0.000320
1.6082	0.131596	0.000341
1.6545	0.131248	0.000356

Table 7
Table of Marginalized $\delta_\lambda(\lambda)$ Transmission Spectral Properties of WASP-17b Binned to 46.4 nm

Wavelength	R_p/R_*	σ
Band-integrated	0.122931	0.001270
1.1318	0.125451	0.001899
1.1782	0.120278	0.001802
1.2246	0.119998	0.001758
1.2710	0.120399	0.001885
1.3174	0.123727	0.001530
1.3638	0.123671	0.001641
1.4102	0.126780	0.001456
1.4566	0.121919	0.001681
1.5030	0.123084	0.001905
1.5494	0.124827	0.001967
1.5958	0.123212	0.001887
1.6422	0.122568	0.001965

each data set, which suggests that it has a high impact on the observational systematics in WFC3 data. We additionally test the dependence on the shift in spectral wavelength position over the course of the observations and find that spectroscopic wavelength shifts $\delta_\lambda(\lambda)$ best describe the associated systematic when applying the full systematic grid to each spectroscopic light curve to produce the final transmission spectrum. However, in the case of bright targets where scan rates are large such as HD 209458b, additional treatment to the spectral position systematic are required to produce a robust and accurate transmission spectrum.

Table 8
Table of Marginalized "Stretch and Shift" ($\delta_\lambda(\lambda)$. $A(\lambda)$) Transmission Spectral Properties of HD 209458b Binned to 46.5 nm

Wavelength	R_p/R_*	σ
Band-integrated	0.120610	0.000291
1.1683	0.120086	0.000158
1.2100	0.120226	0.000113
1.2517	0.120924	0.000120
1.3934	0.120494	0.000166
1.3335	0.120456	0.000319
1.3767	0.121282	0.000186
1.4184	0.121234	0.000139
1.4600	0.120802	0.000098
1.5017	0.120656	0.000108
1.5434	0.121144	0.000153

The use of marginalization accounts for all of the information from each systematic model tested, compared with relying solely on the BIC to select a particular systematic model for the data. Marginalization across a systematic grid is shown to have the greatest effect on data sets where the impacting systematics are not well-defined but can be estimated by a family of similar models. Using marginalization to correct for instrument systematics in both the band-integrated and spectroscopic light curves we show the measured transmission spectrum of five exoplanet atmospheres which maintain the absolute transit depth. In most cases we find similar overall shapes and baseline depths using marginalization. Compared with previously published results, marginalization can increases the error bar values on each measurement by up to $\sim$30 ppm in the most extreme cases when the systematics cannot be well described by known systematics.

Critically, we contrast this technique with current analysis methods used in the field and demonstrate the uses of marginalization to make large comparative studies of different exoplanet atmospheres possible. Previous studies often use the most favored systematic model alone to correct transit light curves, we show that using marginalization across a systematic grid of models can more accurately represent data where multiple systematic models fit equally well. The use of marginalization over a grid of systematic models related to the instrument where observations strategy and instrument modes may have a varied effect allows for a direct comparison between multiple planetary atmospheres by allowing the data to define the analysis while keeping the characterization consistent, which cannot be done when varied singular analysis methods are applied.

Marginalization can be applied to a multitude of transit observations, expanding the ability to make a robust comparison of exoplanet atmospheres as more favorable targets are observed. Additionally, marginalizing over a series of systematic models will be important for new generations of instruments to more easily determine the impact individual systematics have on the data and to aid in optimized observing strategies. This will be particularly important for *James Webb Space Telescope* observations where there is potential to observe

smaller and colder worlds, as the observing strategies and treatment of systematics will likely define the measurements.

As can be seen in Figure 12 the transmission spectra show an overall similarity around the H_2O absorption band centered at 1.4 μm with varied depressed amplitude features. In a follow-up paper we use marginalization across a wider range of hot Jupiter transit light curves from WFC3 in a comparative study of their atmospheres and discuss the interpretation with a series of theoretical models and T-P profiles.

The authors would like to thank N. Gibson for useful comments and discussions on this paper and the analysis technique presented. H.R. Wakeford acknowledges support by an appointment to the NASA Postdoctoral Program at Goddard Space Flight Center, administered by Oak Ridge Associated Universities through a contract with NASA. H.R. Wakeford, D. K. Sing, and T. Evans acknowledge funding from the European Research Council under the European Unions Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement no. 336792. This work is based on observations made with the NASA/ESA *Hubble Space Telescope*. This research has made use of NASAs Astrophysics Data System and components of the IDL astronomy library.

Facilities: *HST* (WFC3).

REFERENCES

Berta, Z. K., Charbonneau, D., Désert, J.-M., et al. 2012, ApJ, 747, 35
Brown, T. M. 2001, ApJ, 553, 1006
Burnham, K. P., & Anderson, D. R. 2004, Sociological Methods & Research, 33, 261
Claret, A. 2000, A&A, 363, 1081
Crossfield, I. J. M., Barman, T., Hansen, B. M. S., & Howard, A. W. 2013, A&A, 559, A33
Deming, D., Wilkins, A., McCullough, P., et al. 2013, ApJ, 774, 95
Eastman, J., Gaudi, B. S., & Agol, E. 2012, PASP, 125, 83
Ehrenreich, D., Bonfils, X., Lovis, C., et al. 2014, A&A, 570, A89
Fraine, J., Deming, D., Benneke, B., et al. 2014, Natur, 513, 526
Frühwirth-Schnatter, S. 2006, Finite Mixture and Markov Switching Models (Berlin: Springer)
Gibson, N. P. 2014, MNRAS, 445, 3401
Gibson, N. P., Aigrain, S., Pont, F., et al. 2012, MNRAS, 422, 753
Haynes, K., Mandell, A. M., Madhusudhan, N., Deming, D., & Knutson, H. 2015, ApJ, 806, 146
Huitson, C. M., Sing, D. K., Pont, F., et al. 2013, MNRAS, 434, 3252
Knutson, H. A., Dragomir, D., Kreidberg, L., et al. 2014, ApJ, 794, 155
Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014a, Natur, 505, 69
Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014b, ApJL, 793, L27
Line, M. R., Knutson, H., Deming, D., Wilkins, A., & Desert, J.-M. 2013, ApJ, 778, 183
Mandel, K., & Agol, E. 2002, ApJL, 580, L171
Mandell, A. M., Haynes, K., Sinukoff, E., et al. 2013, ApJ, 779, 128
Markwardt, C. B. 2009, in Astronomical Data Analysis Software and Systems XVIII ASP Conf. Ser. 411, Nonlinear Least-squares Fitting in IDL with MPFIT, ed. D. A. Bohlender, D. Durand, & P. Dowler (San Francisco, CA: ASP), 251
McCullough, P. 2011, WFC Space Telescope Analysis Newsletter 6
McCullough, P. R., Crouzet, N., Deming, D., & Madhusudhan, N. 2014, ApJ, 791, 55
Nikolov, N., Sing, D. K., Pont, F., et al. 2014, MNRAS, 437, 46
Ranjan, S., Charbonneau, D., Désert, J.-M., et al. 2014, ApJ, 785, 148
Sing, D. K., Fortney, J. J., Nikolov, N., et al. 2016, Natur, 529, 59
Sing, D. K., Lecavelier des Etangs, A., Fortney, J. J., et al. 2013, MNRAS, 436, 2956
Sing, D. K., Wakeford, H. R., Showman, A. P., et al. 2015, MNRAS, 446, 2428
Stevenson, K. B., Bean, J. L., Fabrycky, D., & Kreidberg, L. 2014a, ApJ, 796, 32
Stevenson, K. B., Bean, J. L., Seifahrt, A., et al. 2014b, AJ, 147, 161
Stevenson, K. B., Désert, J.-M., Line, M. R., et al. 2014c, Sci, 346, 838
Swain, M., Deroo, P., Tinetti, G., et al. 2013, Icar, 225, 432
Wakeford, H., Sing, D., Deming, D., et al. 2013, MNRAS, 435, 3481
Wilkins, A. N., Deming, D., Madhusudhan, N., et al. 2014, ApJ, 783, 113