REPEATABILITY OF *SPITZER*/IRAC EXOPLANETARY ECLIPSES WITH INDEPENDENT COMPONENT ANALYSIS

G. MORELLO, I. P. WALDMANN, AND G. TINETTI
Department of Physics & Astronomy, University College London, Gower Street, WC1E6BT, UK; giuseppe.morello.11@ucl.ac.uk

ABSTRACT

The research of effective and reliable detrending methods for *Spitzer* data is of paramount importance for the characterization of exoplanetary atmospheres. To date, the totality of exoplanetary observations in the mid- and far-infrared, at wavelengths $>3\,\mu$m, have been taken with *Spitzer*. In some cases, in past years, repeated observations and multiple reanalyses of the same data sets led to discrepant results, raising questions about the accuracy and reproducibility of such measurements. Morello et al. (2014, 2015) proposed a blind-source separation method based on the Independent Component Analysis of pixel time series (pixel-ICA) to analyze InfraRed Array Camera (IRAC) data, obtaining coherent results when applied to repeated transit observations previously debated in the literature. Here we introduce a variant to the pixel-ICA through the use of wavelet transform, wavelet pixel-ICA, which extends its applicability to low-signal-to-noise-ratio cases. We describe the method and discuss the results obtained over 12 eclipses of the exoplanet XO3b observed during the "Warm *Spitzer*" era in the 4.5 μm band. The final results are reported, in part, also in Ingalls et al. (2016), together with results obtained with other detrending methods, and over 10 synthetic eclipses that were analyzed for the "IRAC Data Challenge 2015." Our results are consistent within 1σ with the ones reported in Wong et al. (2014) and with most of the results reported in Ingalls et al. (2016), which appeared on arXiv while this paper was under review. Based on many statistical tests discussed in Ingalls et al. (2016), the wavelet pixel-ICA method performs as well as or better than other state-of-art methods recently developed by other teams to analyze *Spitzer*/IRAC data, and, in particular, it appears to be the most repeatable and the most reliable, while reaching the photon noise limit, at least for the particular data set analyzed. Another strength of the ICA approach is its highest objectivity, as it does not use prior information about the instrument systematics, making it a promising method to analyze data from other observatories. The self-consistency of individual measurements of eclipse depth and phase curve slope over a span of more than three years proves the stability of Warm *Spitzer*/IRAC photometry within the error bars, at the level of 1 part in 10^4 in stellar flux.

Key words: methods: data analysis – planets and satellites: atmospheres – planets and satellites: individual (XO3b) – techniques: photometric

1. INTRODUCTION

Observations of transits and eclipses of exoplanets have been used, in the last decade, to characterize the nature of more than 100 of these alien worlds with unprecedented detail. Molecular, atomic, and ionic signatures have been detected in the atmospheres of exoplanets through transmission spectroscopy, i.e., multiwavelength measurements of the transit depth, showing differential absorption/scatter of the stellar light through the planetary limb (e.g., Charbonneau et al. 2002; Tinetti et al. 2010; Deming et al. 2013). For the brightest targets, emission spectra have been measured during (planetary) eclipses to constrain the atmospheric chemistry, pressure-temperature profile, albedo, and global circulation patterns (e.g., Charbonneau et al. 2005; Swain et al. 2009). Many data sets were obtained using the InfraRed Array Camera (IRAC; Fazio et al. 2004) on board the *Spitzer Space Telescope*. Since depletion of the telescope's cryogen in 2009, IRAC continues to operate in the 3.6 and 4.5 μm bands, as the "*Spitzer*" Warm Mission."

A precision of 0.01% is required to study the atmospheres of giant exoplanets through transmission and/or emission spectroscopy (Brown 2001; Tessenyi et al. 2013; Waldmann et al. 2015a, 2015b). Detrending instrumental systematics in raw data is necessary to achieve the target precision. It is not always obvious how to decorrelate the data using auxiliary information of the instrument and, in some cases, different methods lead to

significantly different results (see e.g., Waldmann 2012; Morello et al. 2015). The majority of exoplanets' transit and eclipse multi-band photometric data adopted in the literature are obtained by combining data at different wavelengths from separate epochs years apart. Repeated observations are necessary to estimate the overall level of variability, due to astrophysical variations and possible uncorrected instrument effects. If consistent, compared to single observations, repeated observations can provide more accurate parameter values, leading to higher signal-to-noise-ratio (S/N) atmospheric spectra for exoplanets, and potentially allowing the characterization of smaller exoplanets around fainter stars.

The main instrumental effect affecting *Spitzer*/IRAC data at 3.6 and 4.5 μm is due to intra-pixel gain variations and spacecraft-induced motion, the so-called pixel-phase effect. The measured flux from the star is correlated with its position on the detector, hence the idea of correcting the data with a polynomial function of the stellar centroid as proposed in the literature (Charbonneau et al. 2005; Morales-Caldéron et al. 2006; Stevenson et al. 2010; Beaulieu et al. 2011). Multiple reanalyses of the same data sets with the polynomial method show that, in some cases, results can be sensitive to some specific options/variants, such as the degree of the polynomial adopted, partial data rejection, including temporal or other decorrelations (e.g., Beaulieu et al. 2008, 2011; Désert et al. 2009; Stevenson et al. 2010; Knutson et al. 2011).

Recently, several alternative methods have been proposed to decorrelate *Spitzer*/IRAC data: gain mapping (Ballard et al. 2010; Cowan et al. 2012; Knutson et al. 2012; Lewis et al. 2013; Zellem et al. 2014), bilinearly interpolated sub-pixel sensitivity mapping (BLISS, Stevenson et al. 2012), Independent Component Analysis using pixel time series (pixel-ICA; Morello et al. 2014, 2015), pixel-level decorrelation (PLD; Deming et al. 2015), and Gaussian Process models (Evans et al. 2015). A comparison between pixel-ICA and PLD methods is reported in Morello (2015). The discussion about the detrending methods, their performances, reliability, and potential biases for *Spitzer*/IRAC data is a hot topic. In the "IRAC Data Challenge 2015" different methods have been tested over synthetic data created by the IRAC team, which contain 10 simulated eclipse observations of the exoplanet XO3b (Ingalls et al. 2016). Researchers were also encouraged to reanalyze a similar set of real observations obtained in the 4.5 μm band.

In this paper, we describe an evolution of the pixel-ICA method proposed in Morello et al. (2014, 2015) and Morello (2015), and present the results obtained by applying the method to the analysis of 12 eclipses of the exoplanet XO3b taken with Warm *Spitzer*/IRAC in the 4.5 μm band. The pixel-ICA method differs from the other detrending methods proposed in the literature, as it is an unsupervised machine learning algorithm. The lack of any prior assumptions about the instrument systematics and astrophysical signals ensures a high degree of objectivity, and indicates that the same method could be applied to detrend data taken with different instruments. The pixel-ICA method gave coherent results when applied over multiple transit observations of the exoplanets HD189733b and GJ436b, for which the previous literature reported discrepant results (Morello et al. 2014, 2015), and over simulated observations with a variety of instrumental systematics (Morello 2015). Similar techniques have been used in the literature to detrend *Spitzer*/IRS and *Hubble*/NICMOS data, the main difference being in the choice of the input time series (Waldmann 2012, 2014; Waldmann et al. 2013). The ability of ICA to decorrelate non-Gaussian signals is inherently limited to a low Gaussian white noise amplitude relative to the non-Gaussian signals. In this paper, we propose a wavelet pixel-ICA algorithm, which outperforms the traditional pixel-ICA algorithm in low-S/N observations, extending applicability to planetary eclipses taken during the "Warm *Spitzer*" era.

XO3b is a hot Jupiter ($M_p = 11.7 \pm 0.5\,M_{\mathrm{Jup}}$, Johns-Krull et al. 2008; Hirano et al. 2011) in an eccentric orbit ($e = 0.283 \pm 0.003$, Knutson et al. 2014) with a period of 3.19 days and orbital semimajor axis of $a = 0.045$ au (Winn et al. 2008). The host star is F5V with $T_* = 6760 \pm 80$ K, and $\log g = 4.24 \pm 0.03$ (Winn et al. 2008; Torres et al. 2012). A previous analysis of the 12 eclipses reported an average eclipse depth of $1.58^{+0.03}_{-0.04} \times 10^{-3}$ relative to the out-of-eclipse flux (star +planet), and a phase curve slope of $6.0^{+1.3}_{-1.6} \times 10^{-4}$ days^{-1} (Wong et al. 2014). Here we compare our results with the ones reported in Wong et al. (2014).

2. DATA ANALYSIS

2.1. Observations

We analyze 12 eclipse observations of XO3b taken with *Spitzer*/IRAC in the 4.5 μm band. Ten individual eclipses were

Table 1
Eclipse Observations Dates and Orbit Numbers of XO3b

Obs. Number	UT Date	Orbit Number
1	2010 Apr 8	0
2	2012 Nov 11	297
3	2012 Nov 17	299
4	2012 Nov 20	300
5	2012 Nov 23	301
6	2012 Dec 2	304
7	2012 Dec 9	306
8	2013 Apr 22	348
9	2013 May 5	352
10	2013 May 18	356
11	2013 May 21	357
12	2013 May 24	358

observed over six months (2012 November 11–2013 May 24), including two sets of three consecutive eclipses, and another eclipse is contained within a full-orbit observation on 2013 May 5 (PID: 90032). Each individual observation consists of 14,912 frames over 8.4 hr using IRAC's sub-array readout mode with 2.0 s integration time. In sub-array mode 64 frames are taken consecutively, the reset time is ∼1 s. We extracted 14,912 frames from the full-orbit observation to analyze the light-curve of the eclipse over a time interval similar to other observations. The last eclipse was extracted out of a 66 hr observation on 2010 April 8 (PID: 60058). Table 1 reports the dates in which the eclipses were observed.

2.2. The Eclipse Model

In our model the stellar flux is constant in time, and normalized to 1. We adopt the formalism of Mandel & Agol (2002) for the eclipse model, accounting for the planet being occulted by the star. We approximate the planet's phase curve in the region of the eclipse as a linear function of the time, the slope is called "phase constant," adopting the same terminology of Wong et al. (2014). The slope is only due to planet's flux variations. While the planet is completely occulted by the star, the flux is constantly 1. The eclipse depth is defined as the (unseen) planet's flux at the center of eclipse, in units of stellar flux (see Figure 1). When fitting the eclipse models, the orbital parameters are fixed to the values reported in Table 2, taken from Wong et al. (2014), while the center of eclipse, eclipse depth, and phase constant are free parameters to determine. We also considered models with zero phase curve's slope (see Appendix A.5).

2.3. Wavelet ICA

2.3.1. Continuous Wavelet Transform (CWT)

The wavelet transform (WT) decomposes a given signal, $x(t)$, into its frequency components. This is done by convolving the time signal with a basis of highly localized impulses or "wavelets." To fix the ideas, we assume that $x(t)$ is a time series, although this is not necessary, as the Discrete Wavelet Transform (DWT) can be applied to different kinds of signals. The individual wavelet functions are derived from a single "mother wavelet," $\psi(t)$, through translation and dilation of the mother wavelet. The mathematical definition of the CWT is

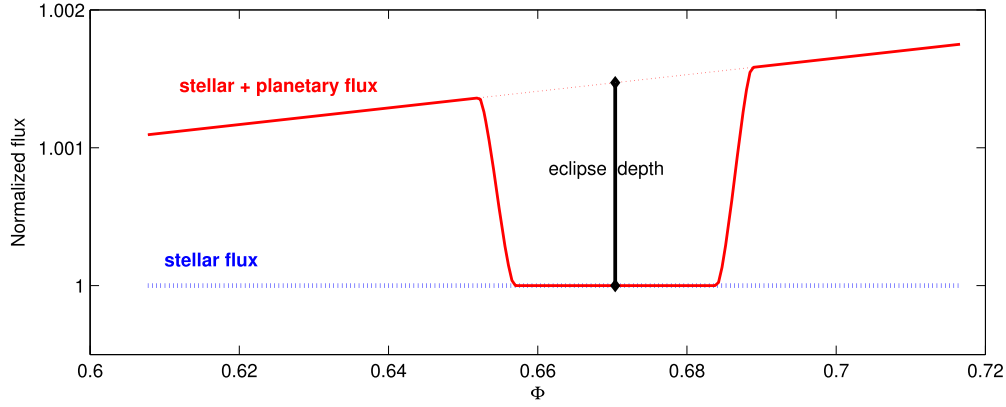$$c_{\tau,\varphi} = \int_{\mathbb{R}} x(t)\psi_{\tau,\varphi}(t)\,dt \qquad (1)$$

Figure 1. Scheme of the eclipse model adopted: the stellar flux is constant and normalized to 1 (blue dashed line), the planetary flux is a linear function of time, and disappears during the eclipse (red line). The eclipse depth is the extrapolated planetary flux, in units of stellar flux, at the eclipse center.

Table 2
Values of the Parameters Fixed while Generating the Eclipse Models

Orbital period, P (days)	3.19153285
Scaled semimajor axis, a/R_*	7.052
Inclination, i (degree)	84.11
Eccentricity, e	0.2833
Argument of periastron, ω (degree)	346.8

$$\psi_{\tau,\varphi}(t) = \frac{1}{\sqrt{2}} \, \psi\left(\frac{t-\tau}{\varphi}\right) \qquad (2)$$

where $\psi_{\tau,\varphi}$ is the mother wavelet for a given scaling φ and translation τ, and $c_{\tau,\varphi}$ is the wavelet coefficient with respect to τ and φ.

If the wavelet basis is orthogonal, the inverse wavelet transform can be used to reconstruct the original time series:

$$x(t) = \sum_{\varphi \in \mathbb{Z}} \sum_{\tau \in \mathbb{Z}} c_{\tau,\varphi} \psi_{\tau,\varphi}(t). \qquad (3)$$

The mother wavelet can be chosen among a variety of wavelet families with different properties. For more details we refer to the relevant literature, e.g., Daubechies (1992) and Percival & Walden (2000).

2.3.2. Discrete Wavelet Transform (DWT)

Astronomical data are usually in the form of discrete time series. For the DWT the mother wavelet is denoted by $h(t)$, and the scaling function, also called "father wavelet," is denoted by $g(t)$. The mother and father wavelets act as high-pass and low-pass frequency filters, respectively. They are related by

$$g(L-1-t) = (-1)^t \, h(t), \qquad (4)$$

where L is the filter length and corresponds to the number of points in the time series $x(t)$.

The one-level DWT is defined by

$$cA_1(\tau) = (x * g)(t) \downarrow 2$$
$$= \sum_{t=-\infty}^{+\infty} x(t) \, g(2\tau - t) \qquad (5)$$

$$cD_1(\tau) = (x * h)(t) \downarrow 2$$
$$= \sum_{t=-\infty}^{+\infty} x(t) \, h(2\tau - t). \qquad (6)$$

The cA_1 time series approximates the underlying low-frequency trend of $x(t)$ (average coefficients), while the cD_1 time series represents a higher frequency component (detail coefficients). They are down-sampled by a factor of two ("$\downarrow 2$" in Equations (5) and (6)) with respect to the original time series, because of the Nyquist theorem.

It is possible to apply the g and h filters to the cA_1 time series, then obtaining new sets of coefficients, cA_2 and cD_2, and iterate the process. The n-level DWT includes the cA_n series of average coefficients, down-sampled by a factor of 2^n, and n series cD_1–cD_n of detail coefficients, representing bands of higher frequencies. The original data can be reconstructed by reversing the process:

$$x(t) = cA_n(\tau) \, g(-t + 2\tau)$$
$$+ \sum_{i=1}^{n} \sum_{\tau=-\infty}^{+\infty} (cD_i \, h(-t + 2\tau)). \qquad (7)$$

2.3.3. Wavelet ICA

In this section we describe the wavelet ICA algorithm, which is used in a variety of contexts, such as medical sciences (e.g., Inuso et al. 2007; La Foresta et al. 2012; Mammone et al. 2012), engineering (e.g., Lin & Zhang 2005), acoustic (e.g., Moussaoui et al. 2006; Zhao et al. 2006), image denoising (e.g., Karande & Talbar 2014), and astrophysics (e.g., Waldmann 2014).

Be $\boldsymbol{x} = (x_1, x_2, ..., x_m)^T$ the column vector of observed signals. In this paper, $\boldsymbol{x}_k$ are individual pixel time series, so-called pixel light curves, which are mixtures of different source signals, astrophysical or instrumental in nature, and Gaussian noise. The formalism adopted in this subsection is valid in a more general context, where the x_k can be any kind of mixed signals. ICA is a linear transformation of the observed (mixed) signals, which minimizes the mutual information to decorrelate the independent components:

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}, \; \boldsymbol{s} = \boldsymbol{W}\boldsymbol{x} \qquad (8)$$

where $\boldsymbol{s} = (s_1, s_2, ..., s_m)^T$ is the column vector of the original source signals, $\boldsymbol{A}$ is the matrix of mixing coefficients, and $\boldsymbol{W} = \boldsymbol{A}^{-1}$. We refer the reader to Waldmann (2012) and Morello et al. (2014, 2015) for additional details.

The ability of ICA to decorrelate non-Gaussian signals is inherently limited to a low Gaussian noise amplitude relative to
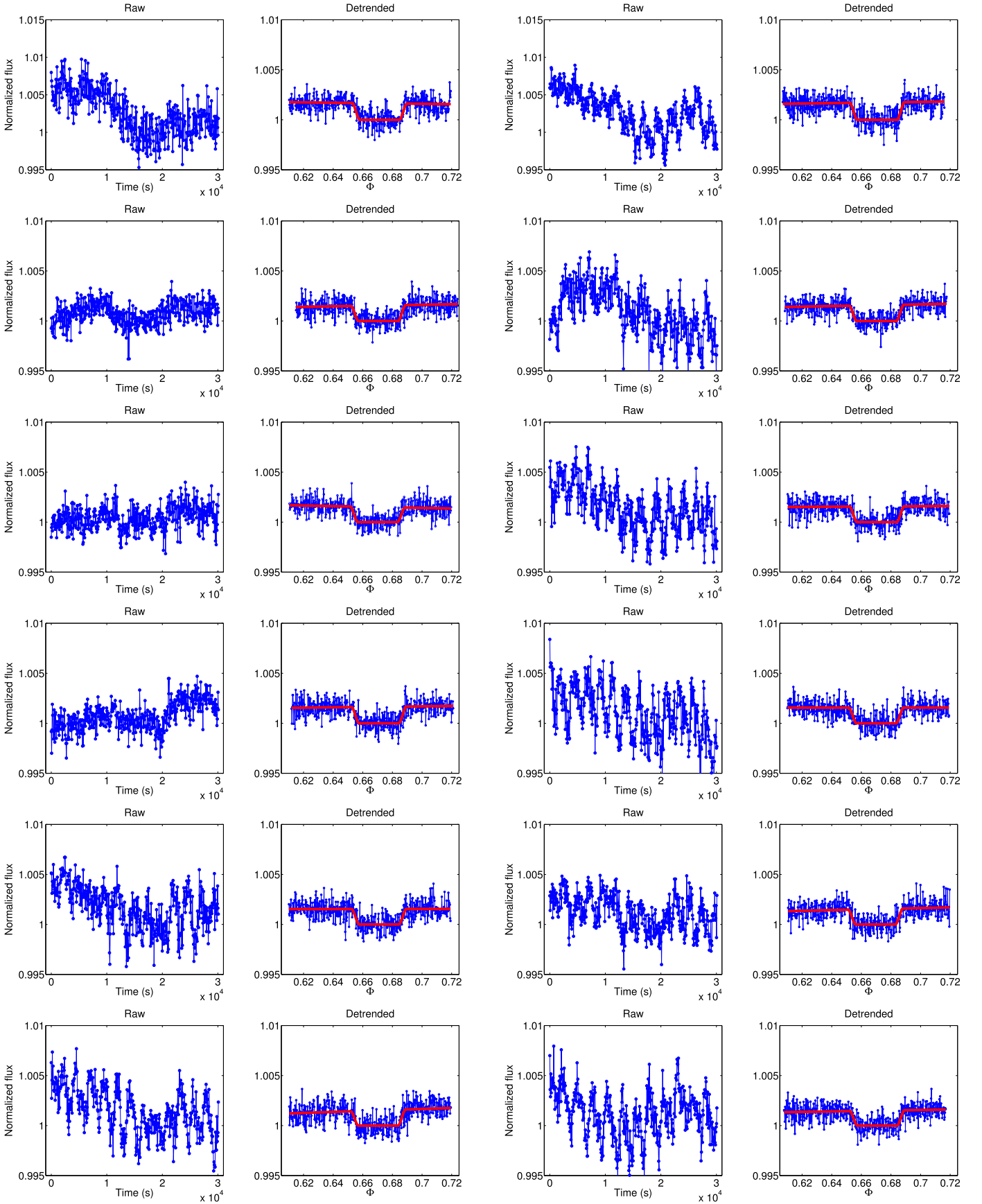
Figure 2. Left panels: (blue) raw light curves obtained from 5 × 5 array of pixels. Right panels: (blue) detrended eclipse light curves obtained with the wavelet pixel-ICA method, and (red) best eclipse models. All the light curves are binned over 32 frames, i.e., ~64 s.

the non-Gaussian signals. Wavelet ICA algorithms are designed to be less sensitive to white noise compared to the simple ICA separation, described above. In wavelet ICA algorithms, the DWT is applied to the observed signals:

$$x_k(t) \rightarrow \hat{x}_k = (cA_{k,n}, cD_{k,n}, ..., cD_{k,1}) \tag{9}$$

$$\boldsymbol{x}(t) \rightarrow \hat{\boldsymbol{x}} = (\hat{x}_1(t), \hat{x}_2(t), ..., \hat{x}_m(t))^T \tag{10}$$

The ICA separation is performed with the series of coefficients:

$$\hat{\boldsymbol{s}} = \hat{\boldsymbol{W}}\hat{\boldsymbol{x}} \tag{11}$$

The independent components series of coefficients are

$$\hat{s}_l = (cA_{l,n}, cD_{l,n}, ..., cD_{l,1}) \tag{12}$$

They can be converted into the time domain through inverse DWT (Equation (7)).

The DWT preliminarly separates the high-frequency components from the low-frequency trend, enhancing the ability of ICA to disentangle the low-frequency independent components. This step is particularly important in cases where the Gaussian noise is dominant. Additional processing options/variants have been proposed in the literature to further improve the ICA performances in specific contexts, e.g., coefficients' thresholding (Stein 1981; Donoho 1995), suppression of some frequency ranges (Lin & Zhang 2005), taking individual levels as input to ICA (Inuso et al. 2007; Mammone et al. 2012). In this paper, we aim to provide the most objective analysis of the data sets with minimal prior assumptions, hence those variants are not considered. The impact of those variants and other operations to the data will be carefully investigated in future studies.

2.4. Detrending Method, Light-curve Fitting, and Error Bars

In this section we list the main steps of the wavelet pixel-ICA method, followed by a more accurate description and comments:

1. Selecting an array of pixels. The raw light-curve is the sum of the individual pixel time series within the selected array.
2. Removing outliers.
3. Subtracting the background from the raw light curve.
4. Computing the wavelet transforms of the time series from the pixels within the selected array, hereafter called pixel light curves.
5. Performing ICA decomposition of the wavelet-transformed pixel light curves.
6. Computing the inverse wavelet transforms of the independent components.
7. Simultaneous fitting of the components (except the eclipse one) and astrophysical model on the raw light curve.
8. Estimating parameter error bars.

2.4.1. Selecting the Pixel-array

We use squared arrays of pixels as photometric apertures; in this paper, we tested 5×5 and 7×7 arrays with the stellar centroid at their centers. By default, all pixel light curves within the selected array are also used to decorrelate the instrument systematics through ICA. Our previous analyses of transit observations indicated the 5×5 and 7×7 arrays to be

optimal choices, and, in general, results were very little affected by the choice of different arrays (Morello et al. 2014, 2015).

2.4.2. Outlier Rejection

We flag and correct outliers in the flux time series. First, we calculate the standard deviations of any set of five consecutive points and take the median value as the representative standard deviation. We define the expected value in one point as the median of the four closest points, i.e., two before and two after. Points differing from their expected values by more than five times the standard deviation are flagged as outliers. They are then replaced with the mean value of the points immediately before and after, or, in case of two consecutive outliers, with a linear interpolation between the closest points which are not outliers. We checked that the outliers removed after this process are coincident with outliers that would have been spotted by eye. They are less than 0.35% the number of points in each observation.

2.4.3. Background Subtraction

The background is estimated by taking the mean flux over four arrays of pixels with the same size of the selected array (5×5 or 7×7) near the corners of the sub-array area. In Appendix A.3 we discuss how this preliminary step improves the results. Here we anticipate that the impact on individual measurements of the eclipse depth is at the level of $\sim 10^{-5}$, well below the error bars. However, this difference might become significant when averaging results from multiple observations, and the error bars are reduced.

2.4.4. Wavelet Transforms

The main novelty of the algorithm proposed in this paper is that the pixel light curves are wavelet transformed before performing the ICA separation. More specifically, we adopt one-level DWT with mother wavelet Daubechies-4 (Daubechies 1992). We found that different choices of the mother wavelet, among different families and numbers, lead to exactly the same results, and higher-level DWTs are not useful. We also investigated the effect of binning the time series (see Appendices A.1 and A.2).

2.4.5. ICA Decomposition

It is performed on the wavelet-transformed pixel light curves (see Equation (11)).

2.4.6. Inverse Wavelet Transforms

The independent components are transformed in the time domain through inverse DWTs (see Equation (7)).

2.4.7. MCMC Fitting

The raw light curves are linear combinations of the independent components. One of the components is the eclipse signal (with some residual systematics), other components may be instrumental systematics and/or other astrophysical signals. Instead of fitting an eclipse model to the eclipse component, more robust estimates of the eclipse parameters are obtained by fitting a linear combination of the eclipse model and the non-eclipse components, hereafter full models, to the raw light curves. The free parameters of the eclipse model are fitted
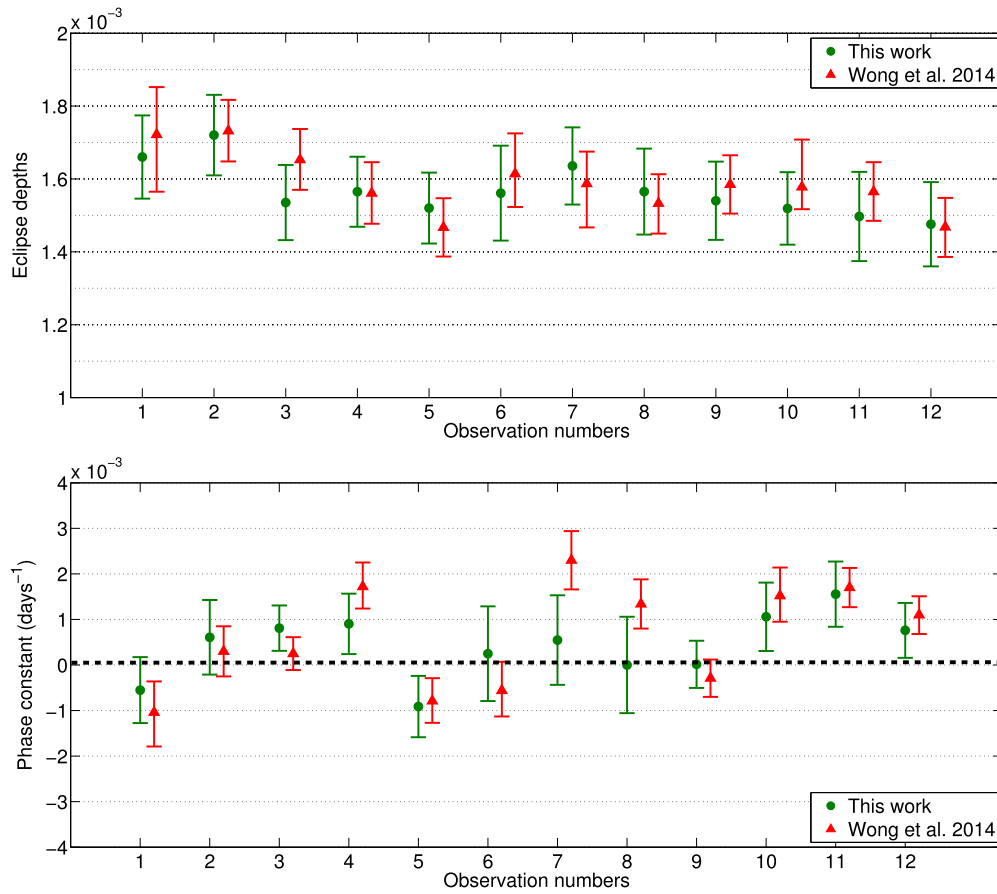
Figure 3. Top panel: (green circles) individual best eclipse depth measurements obtained in this work , and (red triangles) results from Wong et al. (2014). Bottom panel: the same for individual measurements of the phase constant.

Table 3
Individual Best Parameters Results Obtained in this Work

Obs. Number	Depth (10^{-3})	1σ error	Phase Constant (10^{-4} days^{-1})	1σ Error
1	1.66	0.11	−6	7
2	1.72	0.11	6	8
3	1.54	0.10	8	5
4	1.56	0.10	9	7
5	1.52	0.10	−9	7
6	1.56	0.13	2	10
7	1.64	0.11	5	10
8	1.57	0.12	0	11
9	1.54	0.11	0	5
10	1.52	0.10	11	8
11	1.50	0.12	16	7
12	1.48	0.12	8	6

together with the scaling coefficients of the independent components. First estimates of the parameters and scaling coefficients are obtained through a Nelder–Mead optimization algorithm (Lagarias et al. 1998); they are then used as optimal starting points for an Adaptive Metropolis algorithm with delayed rejection (Haario et al. 2006), generating chains of 300,000 values. The output chains are samples of the posterior (Gaussian) distributions. We adopt the mean values of the chains as final best estimates of the parameters, and the standard deviations as zero-order error bars, $\sigma_{par,0}$.

2.4.8. Final Error Bars

The final parameter error bars are

$$\sigma_{par} = \sigma_{par,0} \sqrt{\frac{\sigma_0^2 + \sigma_{ICA}^2}{\sigma_0^2}} \qquad (13)$$

σ_0^2 is the sampled likelihood variance, approximately equal to the variance of the residuals for the best transit model; σ_{ICA}^2 is a term accounting for the potential bias of the components obtained with ICA.

$$\sigma_{ICA}^2 = \sum_j o_j^2 \, \mathrm{ISR}_j \qquad (14)$$

where ISR is the so-called Interference-to-Signal-Ratio matrix, and o_j are the coefficients of the non-eclipse components. The term relative to the goodness of the fit of the components does not appear in Equation (14), as it is automatically included in σ_0 when the components' coefficients and astrophysical parameters are fitted simultaneously.

2.5. Results

Figure 2 shows the raw light curves for the 12 observations, the correspondent detrended eclipses and best models, obtained using the 5 × 5 array and binning over 32 frames, i.e., ∼64 s. The individual measurements of eclipse depth and phase constant are reported in Figure 3 and Table 3. The results from
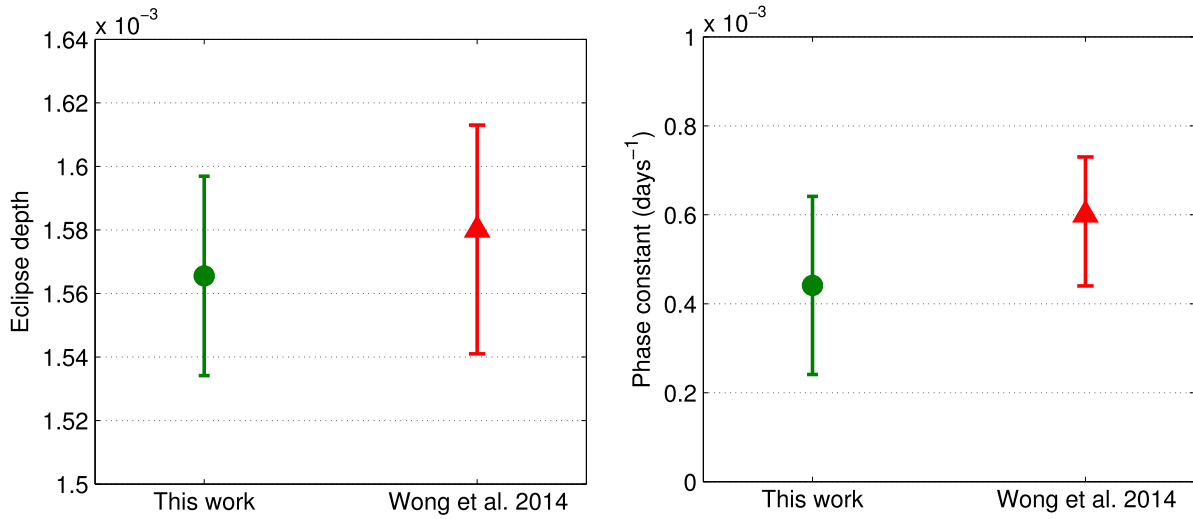
Figure 4. Left panel: (green circle) best global eclipse depth estimate obtained in this work, and (red triangle) in Wong et al. (2014). Right panel: the same for the global phase constant.

Table 4
Weighted Mean Parameter Results, Dispersions, and Reduced Chi Squared Values Obtained in this Paper and Reported in Wong et al. (2014)

Eclipse Depth	This Work	Wong et al. (2014)
Best estimate	$(1.57 \pm 0.03) \times 10^{-3}$	$1.580^{+0.033}_{-0.039} \times 10^{-3}$
Dispersion	7.2×10^{-5}	8.4×10^{-5}
χ_0^2	0.42	0.86

Phase Constant (days^{-1})	This work	Wong et al. (2014)
Best estimate	$(4.4 \pm 2.0) \times 10^{-4}$	$6.0^{+1.3}_{-1.6} \times 10^{-4}$
Dispersion	7.0×10^{-4}	11.3×10^{-4}
χ_0^2	1.0	4.3

all epochs are consistent within the error bars, suggesting the lack of any detectable astrophysical variability for this system, and residual instrument variability. By taking the weighted means of the individual measurements, we obtain global best estimates of $(1.57 \pm 0.03) \times 10^{-3}$ for the eclipse depth, and $(4.4 \pm 2.0) \times 10^{-4}$ days^{-1} for the phase constant.

3. DISCUSSION

3.1. Reduced Chi-squared Tests

The underlying assumption for the weighted mean to be a valid parameter estimate is that individual measurements of that parameter are normally distributed around the same mean value with variances σ_i^2, and there are no systematic errors. The reduced chi-squared can be used to test, in part, this hypothesis:

$$\chi_0^2 = \frac{1}{n-k} \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{\sigma_i^2} \quad (15)$$

where $x_i \pm \sigma_i$ are the individual measurements, $\bar{x}$ is the weighted mean value, $n = 12$ is the number of measurements, and $k = 1$ is the number of calculated parameters. Ideally, if the assumption is valid, we should expect $\chi_0^2 \lesssim 1$. Conventionally, the hypotesis is rejected if $\chi_0^2 > M_{n,k}$, where $M_{n,k}$ is the critical value corresponding to a probability of less than

5% for the hypothesis to be valid. We found $\chi_0^2 = 0.42$ for the eclipse depth, and $\chi_0^2 = 1.0$ for the phase constant, confirming the non-detection of any inter-epoch variability. $\chi_0^2 = 0.42$ may suggest that the error bars for the eclipse depth are overestimated, but this is not totally surprising, given that we actively increased them to account for potential uncorrected systematics and biases in the detrending method. Note that the reduced chi-squared tests whether the actual dispersion in the measurements is consistent within their error bars, but it is not sensitive to a uniform bias for all measurements, e.g., a constant shift. Hence, it is not sufficient alone to justify the use of the weighted mean as global estimate of a parameter. Additional tests, reported in the appendices, show that the weighted mean result is very stable for the eclipse depth. The phase constant appears to be more dependent on certain detrending options, in particular background subtraction. In this case, the adopted weighted mean error bar of 2.0×10^{-4} days^{-1} is a lower limit, valid under some caveats. In the worst-case scenario, the maximum error bar, calculated without scaling when combining multiple measurements, is 7×10^{-4} days^{-1}.

3.2. Comparison with a Previous Analysis of the Same Observations

Our results are consistent within 1σ with the ones from a previous analysis reported in Wong et al. (2014) (see Figures 3 and 4). Our error bars are generally larger by a factor 0.8–1.5 for the eclipse depth (smaller in 1 case) and 1.0–2.0 for the phase constant compared to the ones in the literature. The factors for the weighted mean eclipse depth and phase constant are 0.9 and 1.4, respectively. Slightly larger error bars are a worthwhile trade-off for much higher objectivity, which derives from the lack of assumptions about the origin of instrument systematics and their functional forms in our detrending method. We also note that, despite the larger nominal error bars, the dispersions in our best parameter estimates are slightly smaller than the ones calculated from the results reported in Wong et al. (2014; see Table 4).
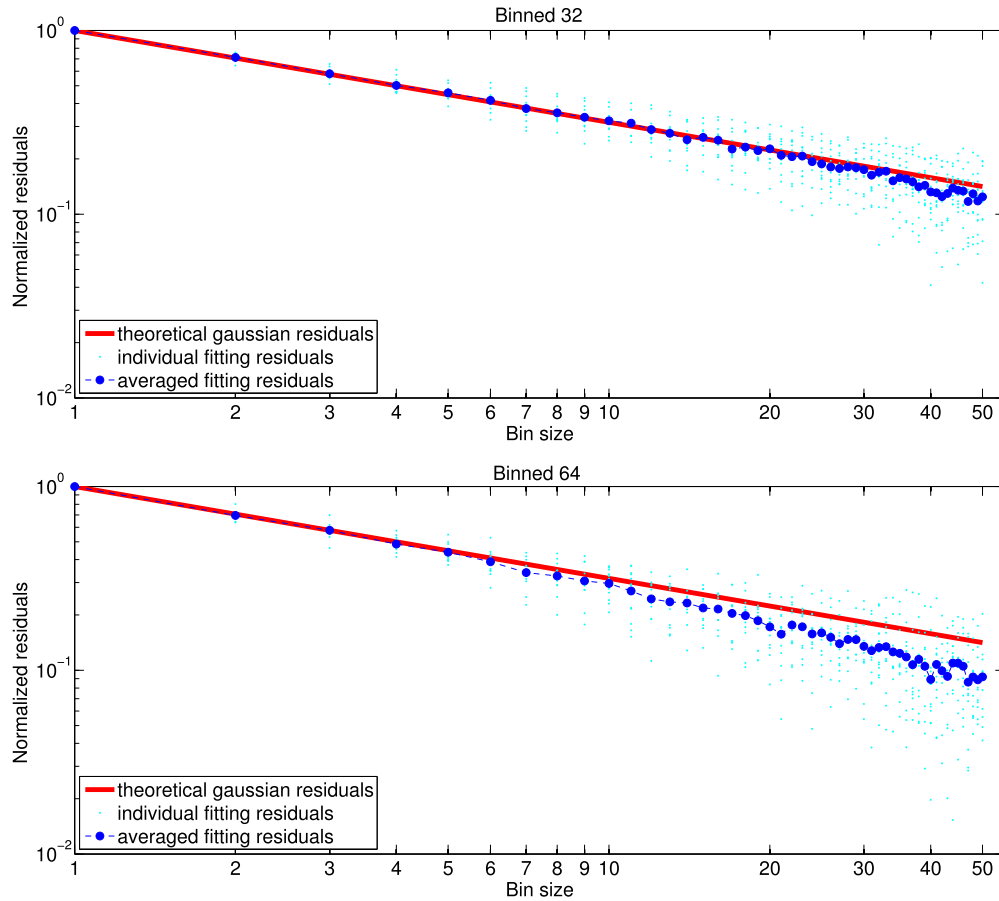
Figure 5. Top panel: rms of residuals as a function of bin size for (red line) Gaussian white noise, (cyan dots) individual observations analyzed with bin size of 32 frames, and (blue circles) values averaged over the 12 observations. Bottom panel: the same for individual observations analyzed with a bin size of 64 frames.

The reduced chi-squared values inferred from their individual parameter estimates are $\chi_0^2 = 0.86$ for the eclipse depth, and $\chi_0^2 = 4.3$ for the phase constant. While the first χ_0^2 value is consistent with the hypothesis of a constant transit depth within the quoted error bars, the second χ_0^2 value indicates that the analogous hypothesis for the phase constant can be rejected (less than 0.1% probability of being true). This may suggest either that they were able to detect some astrophysical variability of the phase curve's slope, or that their individual error bars are underestimated by a factor ~2. Given that the astrophysical slope is degenerate with other instrumental trends, such as long-term position drift of the telescope and possible thermal heating, it is possible that their individual error bars do not fully accounts for these degeneracies, as the authors themselves state. If this is the case, we observe that their final error bar on the phase constant, derived by a joint fit of all eclipses, could be equally underestimated, because it is not guaranteed that residual systematic errors cancel out over multiple observations as if they were random errors. Note that the joint fit approach is theoretically valid under the same assumptions for which the weighted mean is valid, and the two approaches are expected to lead to very similar results (we checked that this happens in this case). In conclusion, our reanalysis confirms the results reported in Wong et al. (2014) for the eclipse depth, and relative inter-epoch variability, but not the 4σ detection of a non-flat phase curve's slope during the eclipse, as larger error bars are needed to account for the possible residual systematics.

4. CONCLUSIONS

We have applied a blind signal-source separation method to analyze twelve photometric observations of the eclipse of the exoplanet XO3b obtained with Warm *Spitzer*/IRAC at 4.5 μm. The method is an evolution of the pixel-ICA method proposed and successfully used by our team to analyze real and synthetic transit observations. By adding a wavelet transform of the time series, we extend the applicability of pixel-ICA to detrend low-S/N observations with instrumental systematics stronger than the astrophysical signal. Wavelet pixel-ICA results are consistent within 1σ with results reported in the literature. They also have smaller dispersions in the eclipse parameters measurements, even including the most recent results that appeared on arXiv while this paper was under review. While the individual error bars are usually more conservative, as they fully accounts for the possible uncertainties, the final error bar on the eclipse depth is equal or smaller than the ones obtained with other methods discussed in the literature.

No significant inter-epoch variations are detected over twelve repeated observations in 3 years interval. This is convincing evidence that, with appropriate data detrending methods, transit, and eclipse measurements based on *Spitzer*/IRAC observations can achieve this level of precision and reproducibility, and therefore are useful to characterize the atmospheres of exoplanets. Also, the lack of any detectable astrophysical variability, for the XO3b system, allows us to combine multiple observations to increase the accuracy in stellar and planetary parameters.
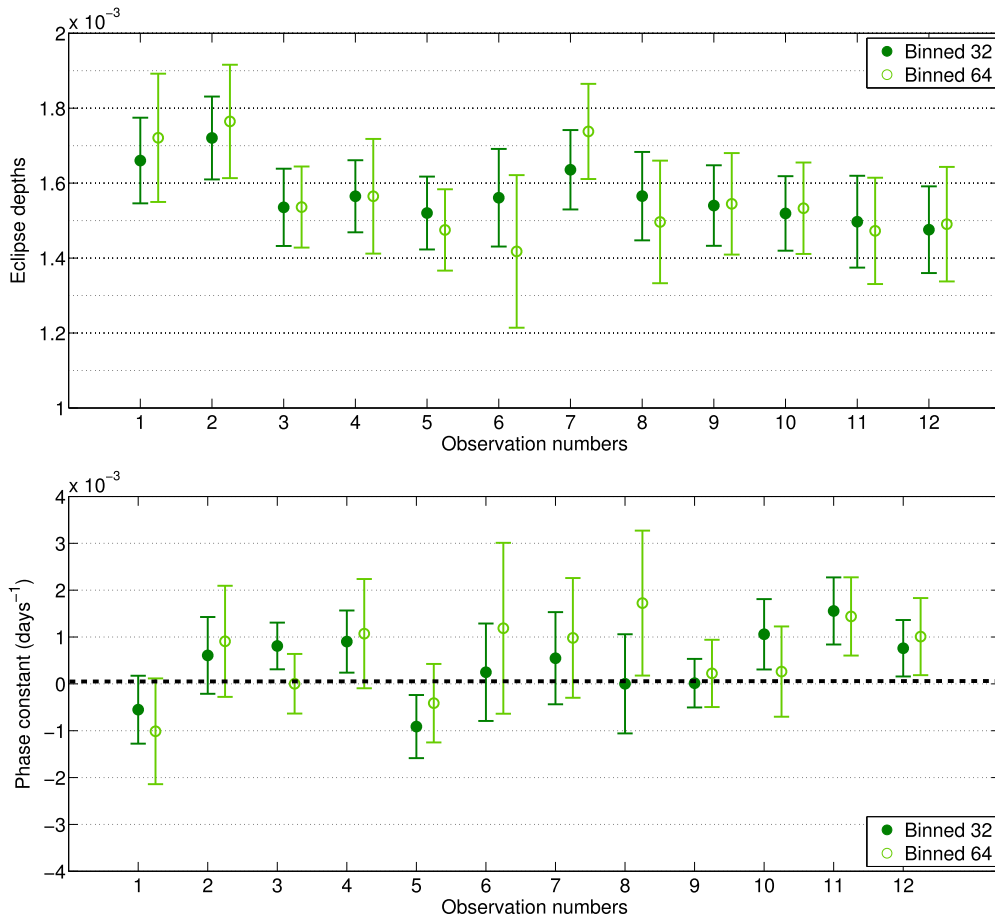
Figure 6. Top panel: individual eclipse depth measurements obtained with 5 × 5 array, background subtraction, and (dark green, full circles) binning over 32 frames, and (light green, empty circles) binning over 64 frames. Bottom panel: the same for individual measurements of the phase constant.

APPENDIX
TESTING THE ROBUSTNESS OF RESULTS

A.1. The Effect of Binning

Given the large number of frames (14,912) for each observation, binning data is very useful to decrease the computational time needed to run the MCMCs for the eclipse parameters and scaling coefficients for the independent components (see Section 2.4.7). Some authors suggest that an optimal choice of the binning size can be useful to reduce the noise on the timescale of interest (Deming et al. 2015; Kammer et al. 2015), provided that the theoretical curve is similarly binned as necessary to eliminate bias, and the bin size is not too long to cause significant loss of astrophysical information (Kipping 2010).

An additional choice for the pixel-ICA algorithm is whether to bin the pixel time series prior the ICA separation, or to bin the independent components extracted from unbinned pixel time series. We found that the two options are almost equivalent, as the eclipse signals obtained after removing the systematic components from the raw light-curve are identical (discrepancies of one to two orders of magnitudes smaller than the fitting residuals), except in cases for which the unbinned ICA separation fails to retrieve an eclipse component. It is worth noting that for the unbinned case, the amplitude of the total noise plus systematics is higher than the the eclipse depth. Thus we decided to bin the individual pixel light curves prior ICA retrieval.

We compare the results obtained for all the observations with bin sizes of 32 and 64 frames, i.e., 64 and 128 s, respectively. First, we test the gaussianity of fitting residuals by calculating their rms as a function of the bin size, b. If fitting residuals are white noise, the rms would scale as $1/\sqrt{b}$. Figure 5 shows that, in both cases, the rms of fitting residuals slightly deviates from the expected behavior of white noise. Those deviations are smaller for the analysis with a bin size of 32 frames. The parameter results obtained with the two binning choices are all consistent within 0.5σ, and on average within 0.16σ (see also Figures 6 and 12). Both the error bars and the overall dispersions are smaller for the cases with bin size of 32 frames by factors of ~1.4. We consider the results obtained with bin size of 32 frames as our best results.
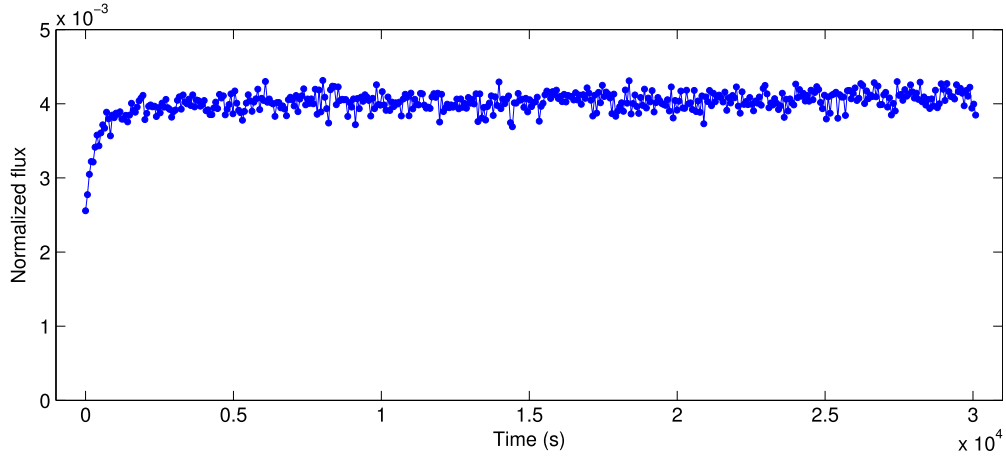
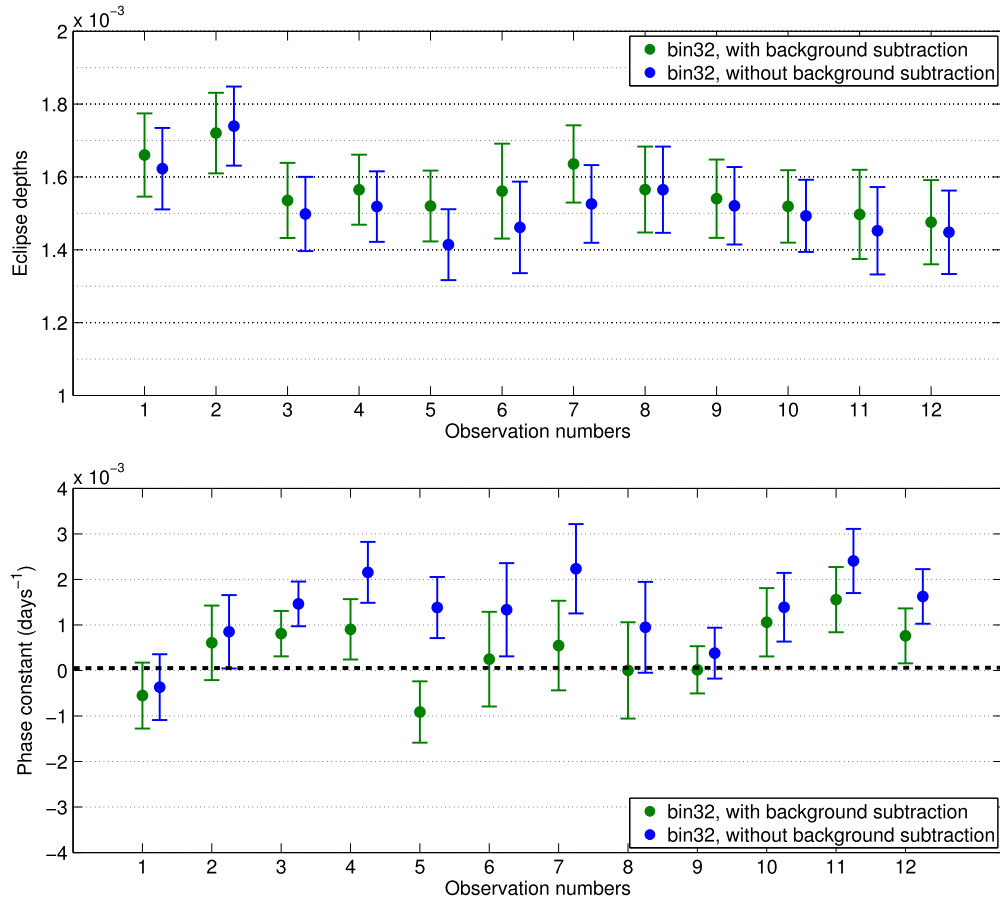**Figure 7.** Example of background time series binned over 32 frames.

Figure 8. Top panel: individual eclipse depth measurements obtained with 5 × 5 array, time series binned over 32 frames (green circles) with background subtraction, and (blue circles) without background subtraction. Bottom panel: the same for individual measurements of the phase constant.

A.2. Specific Wavelet ICA Options

The DWT of a time series is specified by the choice of a mother wavelet and the number of levels (see Section 2.3.2). We adopted a Daubechies-4 mother wavelet, and one-level decomposition. For a few observations we tested multiple choices of the mother wavelet, among the Daubechies, Biorthogonal, Symlets, and Coiflets families, and number of decomposition levels. We refer to Daubechies (1992), Percival & Walden (2000) for details about the wavelet properties. We

found that different choices of the mother wavelet are not significant, e.g., discrepancies in the eclipse signals are one to two orders of magnitudes smaller than the fitting residuals, while level decompositions higher than 1 usually appear to make it impossible for ICA to retrieve the eclipse. The difficulties with higher-level DWTs may arise from subsampling the average coefficients, and the fact that some of the low-frequency non-Gaussian components may be smeared over higher levels of detail coefficients.
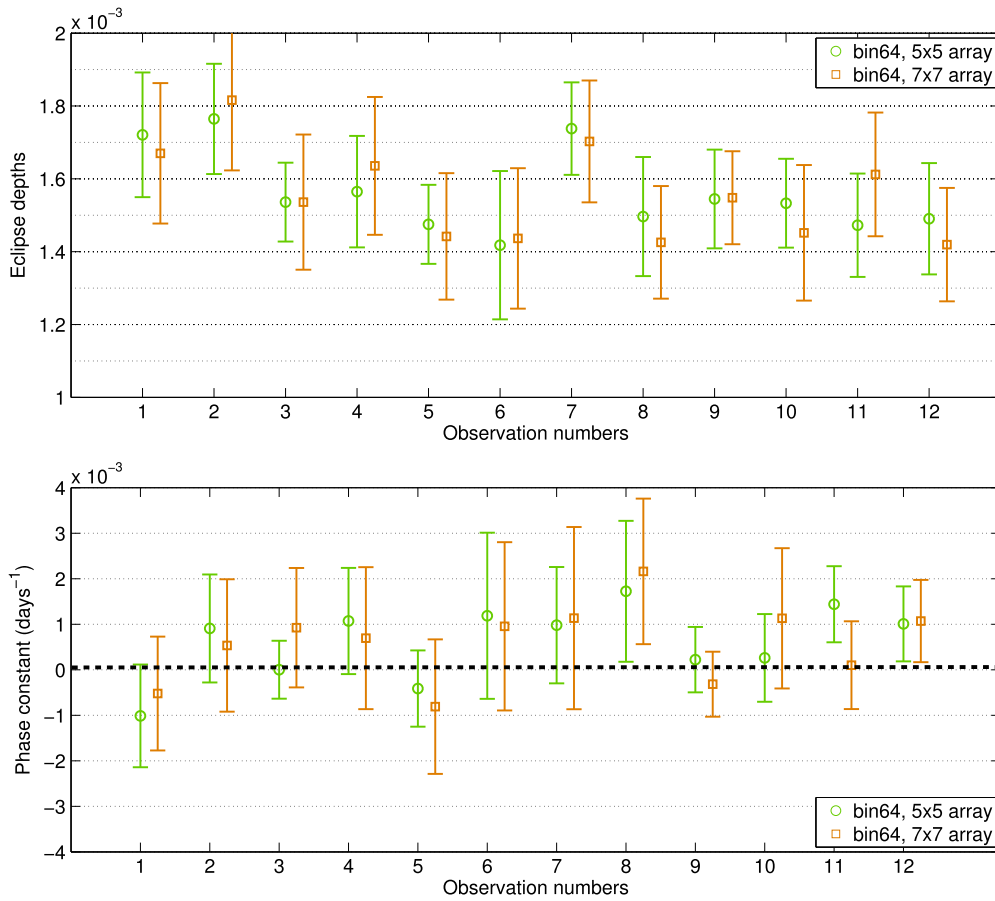
Figure 9. Top panel: individual eclipse depth measurements obtained with (green circles) 5×5 array, and (brown squares) 7×7 array; both are obtained with time series binned over 64 frames, and background subtraction. Bottom panel: the same for individual measurements of the phase constant.

A.3. About Background Subtraction

Uncorrected background may bias the normalized amplitude of the eclipse depth, as well as the phase curve's slope, if the background is not constant over time. The typical morphology of a background time series is either a constant function or a slow monotonic drift. The lack of a distinct temporal structure and non-gaussianity makes it difficult to disentangle with ICA, as well as other statistical methods. For this reason, we performed ad hoc background subtractions before ICA detrending (see Section 2.4.3). In this section, we discuss the impact of this step in the analyses, by comparing results obtained with and without background subtraction. These are also used to infer the maximum parameter errors that can be caused by an inappropriate background correction.

Figure 7 shows an example of background time series estimated for one observation. The measured mean background level slightly varies from one observation to the other, but it is always less than 0.6% of the total flux from the system, hence potentially affecting the eclipse depth by less than 10^{-5}, well below the error bars. The background is also not constant during one observation: it has a small ramp for the first ~20 minutes (except for the eclipses extracted from longer observations), then continues to slowly increase.

Figure 8 shows the parameter results obtained with and without background subtraction. The greatest discrepancies are observed for the phase constant, which is systematically higher by $2\text{–}20 \times 10^{-4}\,\mathrm{days}^{-1}$ for the cases without background subtraction, suggesting the presence of uncorrected

systematics. It is quite remarkable that our individual error bars automatically account for those systematics, but, given their non-random nature, the weighted mean error bars for the phase constant cannot be compared. It is difficult to prove the superiority of the results obtained with background subtraction over the others, as the residuals between the full models and the relevant raw light curves (see Section 2.4.7) are very similar for the two cases, e.g., similar amplitudes and time correlations, leading to similar error bars. Slightly smaller parameter dispersions are obtained with background subtraction rather than without, in particular 7×10^{-5} versus 9×10^{-5} for the eclipse depth, and 7×10^{-4} versus $8 \times 10^{-4}\,\mathrm{days}^{-1}$ for the phase constant. The higher mean value of the phase constant obtained without background subtraction, i.e., $\sim\!13 \times 10^{-4}\,\mathrm{days}^{-1}$, appears to be less likely, as it would require a higher than expected increase in the atmospheric temperature due to stellar irradiation during the eclipse, and/or strong horizontal disomogeneities either in temperature, chemical composition and/or clouds (Cowan & Agol 2011; Kataria et al. 2013; Agúndez et al. 2014). If we assume that systematics are removed with background subtraction, we can take the weigthed mean as the best estimate for the phase constant.

The discrepancies between eclipse depth measurements with and without background subtraction are in the range $10^{-5}\text{–}10^{-4}$, and, on average, the eclipse depth is smaller by $\sim\!4 \times 10^{-5}$ for the case without background subtraction. Although this is more than the 10^{-5} difference expected from the mean background level relative to the mean stellar flux (see
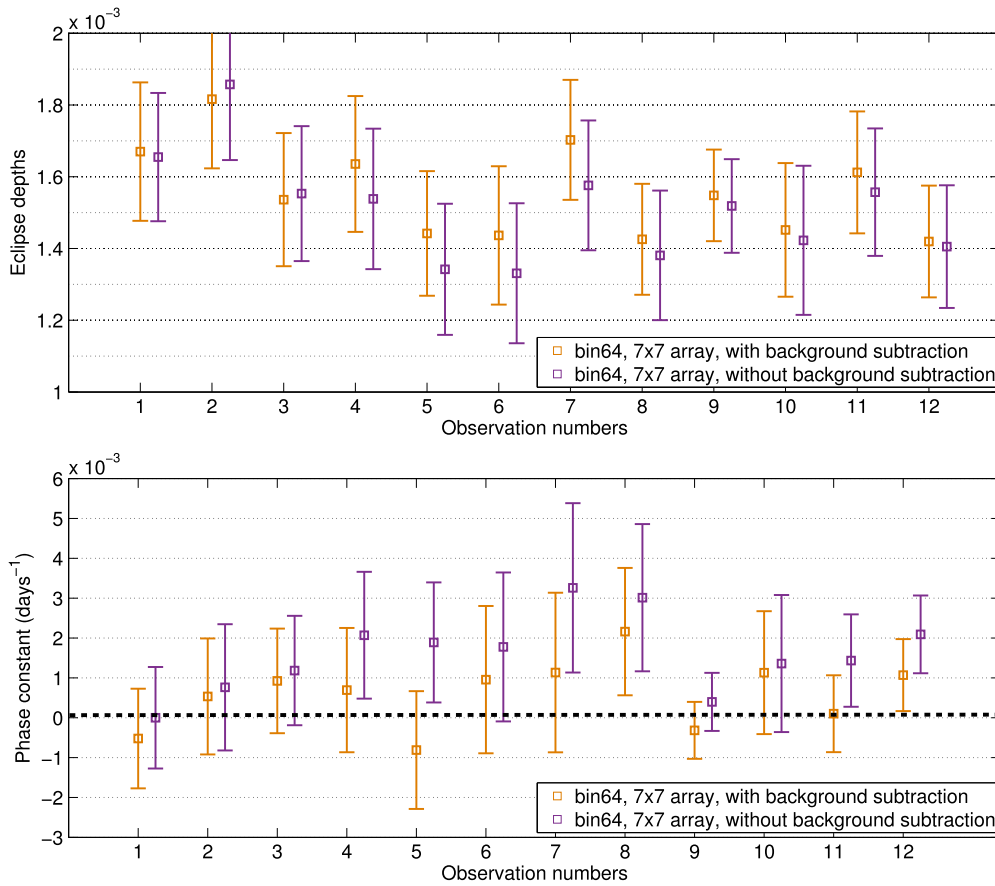
Figure 10. Top panel: individual eclipse depth measurements obtained with 7×7 array, time series binned over 64 frames (brown squares) with background subtraction, and (purple squares) without background subtraction. Bottom panel: the same for individual measurements of the phase constant.

the discussion in the previous paragraph), in this case, the two weighted means are consistent within 0.5σ. We found that the main effect of background on the eclipse depth is due to correlations between the measured eclipse depth at the eclipse center and the phase constant, in terms of Pearson correlation coefficients between the relevant MCMCs. The details of this study are beyond the scope of this paper.

A.4. Using Different Arrays of Pixels

In previous studies, we found that, for high-S/N transit observations, pixel-ICA performances are only slightly dependent on the choice of the array (Morello et al. 2014, 2015). The 3×3 array is too narrow compared to the *Spitzer*/IRAC Point Spread Function, and for this reason, it is not photometrically stable (although it usually leads to consistent results). Larger arrays are more photometrically stable, but they include more noise. Also, noisier pixel light curves, appear to increase the uncertainties in the ICA decomposition (σ_{ICA}, see Section 2.4.8), so that the smallest error bars were usually obtained using the 5×5 array.

In this work, we analyzed all data sets with two different choices of the pixel-array, i.e., 5×5 and 7×7. For the analyses with the 7×7 array, we only adopted the 64 frames bin size, for which the MCMC fitting is faster. Figure 9 compares the results obtained with the two different arrays and the same bin size. The two sets of results are consistent well within 1σ, but error bars obtained with the 7×7 array are 1–1.5 times larger than the ones obtained with the 5×5 array,

despite the residuals in the fits are similar and often smaller for the 7×7 cases. This confirms the conclusions obtained from previous analyses, in particular:

1. The 5×5 array leads to smaller error bars than other squared arrays of pixels;
2. Parameter results from different arrays are consistent well within 1σ.

In this case, the choice of a less optimal array increases the error bars more significantly than in our previous analyses, most likely because of the lower S/N of the observations analyzed here.

Figure 10 compares the results obtained with and without background subtraction for the case of 7×7 array. The same considerations discussed in Appendix A.3 for the 5×5 array are valid for the 7×7 array.

A.5. Breaking Degeneracy with the Phase Constant

At the end of Appendix A.3 we revealed the existence of a correlation between the eclipse depth and phase constant parameter, which is stronger for data with a higher slope, such as the cases without background subtraction. This may affect the eclipse depth estimate in case of residual systematics with a slope, e.g., uncorrected background. In our analyses, the maximum bias on the eclipse depth due to correlations with this kind of systematics is $\sim 4 \times 10^{-5}$, provided the systematic slope is not larger than our individual error bars.
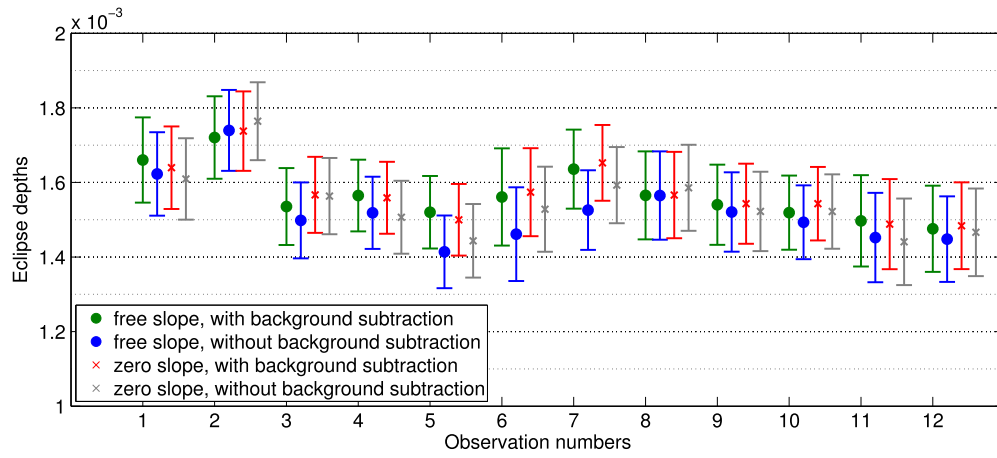
Figure 11. Individual eclipse depth measurements obtained with 5×5 array, time series binned over 32 frames (green circles) with fitted phase constant and background subtraction, (blue circles) with fitted phase constant and without background subtraction, (red "x") with zero phase constant and background subtraction, and (gray "x") with zero phase constant and without background subtraction.
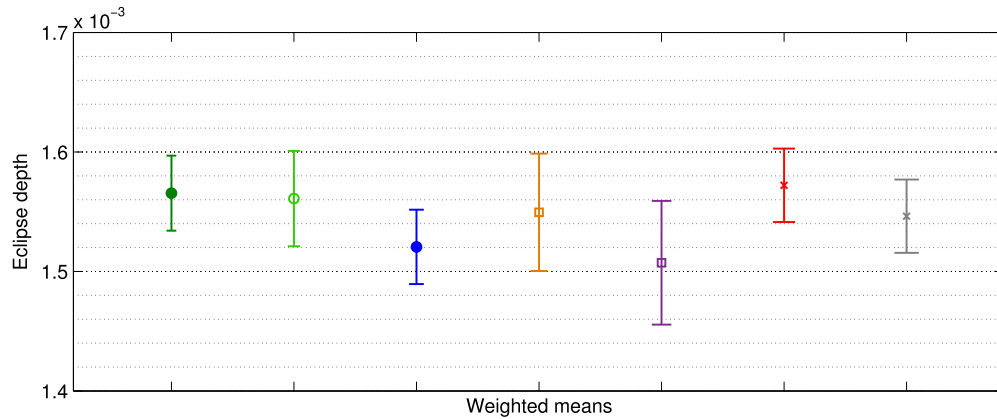

Figure 12. Weighted mean eclipse depth values obtained with different tests: (dark green, full circle) 5×5 array, with background subtraction, binned over 32 frames; (light green, empty circle) 5×5 array, with background subtraction, binned over 64 frames; (blue, full circle) 5×5 array, without background subtraction, binned over 32 frames; (brown, empty square) 7×7 array, with background subtraction, binned over 64 frames; (purple, empty square) 7×7 array, without background subtraction, binned over 64 frames; (red "x") 5×5 array, with background subtraction, binned over 32 frames, and zero phase constant; (gray "x") 5×5 array, without background subtraction, binned over 32 frames, and zero phase constant.

Here we test the consequences of adopting zero phase constant while fitting for the eclipse depth. This is equivalent to the assumption that the phase curve is flat before and after the eclipse, and the slope is entirely due to instrumental effects. Compared to the cases with the phase constant as free parameter, fitting residuals with zero phase constant are not significantly larger. Figure 11 compares results for the eclipse depth with free and zero phase constant, with and without background subtraction. The weighted mean results are reported in Figure 12. We note that differences in eclipse depth measurements with and without background subtraction are smaller for the zero phase constant models, suggesting that eclipse depth measured with zero phase constant models is less affected by residual systematics with a slope. Free phase constant models are valid in a more general context, as, differently from the zero phase constant models, they approximate a planet's flux variability during the observation. The consistency between the results obtained with the two classes of models indicates that, in this case, the planet's flux variability in the proximity of the eclipse is smaller than the error bars.

A.6. Robustness of the Eclipse Depth Measurement

Figure 12 reports the weighted mean eclipse depth estimated for all the tests discussed in the appendices. Note that they are mutually consistent at the 0.5σ level, and the range is 6×10^{-5}.

REFERENCES

Agúndez, M., Parmentier, V., Venot, O., Hersant, F., & Selsis, F. 2014, A&A, 564, A73
Ballard, S., Charbonneau, D., Deming, D., et al. 2010, PASP, 122, 1341
Beaulieu, J. P., Carey, S., Ribas, I., & Tinetti, G. 2008, ApJ, 677, 1343
Beaulieu, J. P., Tinetti, G., Kipping, D. M., et al. 2011, ApJ, 731, 16
Brown, T. M. 2001, ApJ, 553, 1006
Charbonneau, D., Allen, L. E., Megeath, S. T., et al. 2005, ApJ, 626, 523
Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L. 2002, ApJ, 568, 377
Cowan, N. B., & Agol, E. 2011, ApJ, 726, 82
Cowan, N. B., Machalek, P., Croll, B., et al. 2012, ApJ, 747, 82
Daubechies, I. 1992, Ten Lectures on Wavelets (Philadelphia, PA: Society for Industrial and Applied Mathematics)
Deming, D., Knutson, H. A., Kammer, J., et al. 2015, ApJ, 805, 132
Deming, D., Wilkins, A., McCullough, P., et al. 2013, ApJ, 774, 95
Désert, J. M., Lecavelier des Etangs, A., Hébrard, G., et al. 2009, ApJ, 699, 478

Donoho, D. L. 1995, ITIT, 41, 613
Evans, T. M., Aigrain, S., Gibson, N., et al. 2015, arXiv:1504.05942v1
Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, ApJS, 154, 10
Haario, H., Laine, M., Mira, A., & Saksman, E. 2006, Statistics and Computing, 16, 339
Hirano, R., Narita, N., Sato, B., et al. 2011, PASJ, 63, L57
Ingalls, J. G., Krick, J. E., Carey, S. J., et al. 2016, arXiv:1601.05101v1
Inuso, G., La Foresta, F., Mammone, M., & Morabito, F. C. 2007, Int. J. Conf. Neural Netw., 20, 1524
Johns-Krull, C. M., McCullough, P. R., Burke, C. J., et al. 2008, ApJ, 677, 657
Kammer, J. A., Knutson, H. A., Line, M. R., et al. 2015, ApJ, 810, 118
Karande, K. J., & Talbar, S. 2014, Independent Component Analysis of Edge Information for Face Recognition (New Delhi: Springer India)
Kataria, T., Showman, A. P., Lewis, N. K., et al. 2013, ApJ, 767, 76
Kipping, D. K. 2010, MNRAS, 408, 1758
Knutson, H. A., Fulton, B. J., Montet, B. T., et al. 2014, ApJ, 785, 126
Knutson, H. A., Lewis, N., Fortney, J. J., et al. 2012, ApJ, 754, 22
Knutson, H. A., Madhusudhan, N., Cowan, N. B., et al. 2011, ApJ, 735, 27
La Foresta, F., Mammone, N., & Morabito, F. C. 2012, in Neural Nets, Vol. 3931, ed. B. Apolloni (Berlin: Springer), 78
Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. 1998, SJCO, 9, 112
Lewis, N. K., Knutson, H. A., Showman, A. P., et al. 2013, ApJ, 766, 95
Lin, J., & Zhang, A. 2005, NDTI, 38, 421
Mammone, N., La Foresta, F., & Morabito, F. C. 2012, IEEE Sensors Journal, 12, 533
Mandel, K., & Agol, E. 2002, ApJL, 580, L171

Morales-Caldéron, M., Stauffer, J. R., Davy Kirkpatrick, J., et al. 2006, ApJ, 653, 1454
Morello, G. 2015, ApJ, 808, 56
Morello, G., Waldmann, I. P., Tinetti, G., et al. 2014, ApJ, 786, 22
Morello, G., Waldmann, I. P., Tinetti, G., et al. 2015, ApJ, 802, 117
Moussaoui, R., Rouat, J., & Lefebvre, R. 2006, in Proc. IEEE Int. Conf. Acoustic, Speech and Signal, Vol. 5 (Piscataway, NJ: IEEE), V645
Percival, D. B., & Walden, A. T. 2000, Wavelet Methods for Time Series Analysis (Cambridge: Cambridge Univ. Press)
Stein, C. M. 1981, AnSta, 9, 1135
Stevenson, K. B., Harrington, J., Fortney, J. J., et al. 2012, ApJ, 754, 136
Stevenson, K. B., Harrington, J., Nymeyer, S., et al. 2010, Natur, 464, 1161
Swain, M. R., Tinetti, G., Vasisht, G., et al. 2009, ApJL, 690, L114
Tessenyi, M., Tinetti, G., Savini, G., & Pascale, E. 2013, Icar, 226, 1654
Tinetti, G., Deroo, P., Swain, M. R., et al. 2010, ApJL, 712, L139
Torres, G., Fischer, D. A., Sozzetti, A., et al. 2012, ApJ, 757, 161
Waldmann, I. P. 2012, ApJ, 747, 12
Waldmann, I. P. 2014, ApJ, 780, 23
Waldmann, I. P., Rocchetto, M., Tinetti, G., et al. 2015a, ApJ, 813, 13
Waldmann, I. P., Tinetti, G., Deroo, P., et al. 2013, ApJ, 766, 7
Waldmann, I. P., Tinetti, G., Rocchetto, M., et al. 2015b, ApJ, 802, 107
Winn, J. N., Holman, M. J., Torres, G., et al. 2008, ApJ, 683, 1076
Wong, I., Knutson, H. A., Cowan, N. B., et al. 2014, ApJ, 794, 134
Zellem, R. T., Lewis, N. K., Knutson, H. A., et al. 2014, ApJ, 790, 5
Zhao, C. H., Liu, J., & Sun, J. D. 2006, Journal of Electronics and Information Technology, 28, 1565