MNRAS **000**, 1–20 (2019) Preprint 23 May 2019 Compiled using MNRAS LATEX style file v3.0

Over-1000 transit timings for 17 exoplanets homogeneously derived: empirical calibration of stellar limb-darkening laws and reassessed TTV trends for WASP-12 and WASP-4

R.V. Baluev[1,2⋆], E.N. Sokov[1], H.R.A. Jones[3], V.Sh. Shaidulin[2], I.A. Sokova[1], L.D. Nielsen[4], P. Benni[5], M.E. Schneiter[6], C.S. Villarreal D'Angelo[6], A.Yu. Burdanov[7], E. Fernández-Lajús[8,9], R.P. Di Sisto[8,9], Ö. Baştürk[10], V.-P. Hentunen[11], S. Shadick[12], Y. Jongen[13], W. Kang[14], T. Kim[14,15], E. Pakstiene[16], J.K.T. Qvam[17], C.R. Knight[18], P. Guerra[19], A. Marchini[20], F. Salvaggio[20], P. Evans[21], M. Salisbury[22], F. Garcia[23], D. Molina[24], J. Garlitz[25], N. Esseiva[26], Y. Ogmen[27], Yu. Karavaev[28], S. Rusov[1], M.A. Ibrahimov[29], R.G. Karimov[30], and I.M. Asfandiyarov[30]

Accepted 2019 .... ??. Received 2019 .... ??; in original form 2019 May 31

ABSTRACT

We homogeneously analyse $\sim 3 \times 10^5$ photometric measurements for ~ 1100 transit lightcurves belonging to 17 selected exoplanet hosts. The photometric data cover 16 years 2004–2019 and include amateur and professional observations. Old archival lightcurves were reprocessed in view of the up-to-date exoplanetary parameters and empirically debiased limb-darkening models. We also derive self-consistent transit+RV fits for 13 targets. We confirm the nonlinear TTV trend in the WASP-12 data at a high significance, and with a consistent magnitude. However, Doppler data reveal hints of a radial acceleration about −7.5 m/s/yr, indicating the presence of unseen distant companions, and suggesting that roughly 10 per cent of the observed TTV was induced via the Roemer effect (variable light arrival delay). For WASP-4, a similar TTV trend suspected after the recent TESS observations appears controversial and model-dependent, given the very heterogeneous nature of the TTV data available. It is not supported by our homogeneus TTV sample, including 10 ground-based EXPANSION lightcurves obtained in 2018 simultaneously with TESS. Even if the TTV trend itself does exist in WASP-4, its magnitude and tidal nature are uncertain. Doppler data cannot entirely rule out the Roemer effect induced by possible distant companions.

Key words: planetary systems - techniques: photometric - techniques: radial velocities - methods: data analysis - methods: statistical - surveys

AUTHORS' AFFILIATIONS

[1]Central Astronomical Observatory at Pulkovo of Russian Academy of Sciences, Pulkovskoje sh. 65/1, St Petersburg 196140, Russia
[2]Saint Petersburg State University, Faculty of Mathematics & Mechanics, Universitetskij pr. 28, Petrodvorets, St Petersburg 198504, Russia
[3]University of Hertfordshire, Centre for Astrophysics Research, STRI, College Lane, Hatfield AL10 9AB, UK
[4]Geneva Observatory, University of Geneva, Chemin des Mailettes 51, 1290 Versoix, Switzerland
[5]Acton Sky Portal (Private Observatory), Acton, MA, USA
[6]Instituto de Astronomía Teoríca y Experimental, Universidad Nacional de Córdoba, Laprida 854, Córdoba X5000BGR, Argentina
[7]Space sciences, Technologies and Astrophysics Research (STAR) Institute, Universite de Liege, Allee du 6 Aout 17, B-4000 Liege, Belgium
[8]Facultad de Ciencias Astronómicas y Geofísicas - Universidad Nacional de La Plata, Paseo del Bosque S/N - 1900 La Plata, Argentina

⋆ E-mail: r.baluev@spbu.ru

© 2019 The Authors

[9]Instituto de Astrofísica de La Plata (CCT La Plata - CONICET/UNLP), Argentina
[10]Ankara University, Faculty of Science, Department of Astronomy and Space Science, TR-06100, Tandogan, Ankara, Turkey
[11]Taurus Hill Observatory, Warkauden Kassiopeia ry., Härkämäentie 88, 79480 Kangaslampi, Finland
[12]Physics and Engineering Physics Department, University of Saskatchewan, 116 Science Place, Saskatoon, Saskatchewan, S7N 5E2, Canada
[13]Observatoire de Vaison la Romaine, 1075 RD 51, Le Palis, 84110 Vaison-la-Romaine, France
[14]National Youth Space Center, Goheung, Jeollanam-do, 59567, S.Korea
[15]Department of Astronomy and Space Science, Chungbuk National University, Cheongju-City, 28644, S.Korea
[16]Institute of Theoretical Physics and Astronomy, Vilnius University, Sauletekio al. 3, Vilnius 10257, Lithuania
[17]Horten Videregaende Skole, Bekkegata 2, 3181 Horten, Norway
[18]Ngileah Observatory, 144 Kilkern Road, RD 1. Bulls 4894, New Zealand
[19]Observatori Astronomic Albanya, Cami de Bassegoda s/n, 17733 Albanya, Spain
[20]Astronomical Observatory–DSFTA, University of Siena, Via Roma 56, I-53100 Siena, Italy
[21]El Sauce Observatory, Coquimbo Province, Chile
[22]School of Physical Sciences, The Open University, Milton Keynes, MK7 6AA, UK
[23]La Vara, Valdes Observatory, 33784 Munas de Arriba, Valdes, Asturias, Spain
[24]Anunaki Observatory, Calle de los Llanos, 28410 Manzanares el Real, Spain
[25]Private Observatory, 1155, Hartford St, Elgin, OR 97827, USA
[26]Observatory Saint Martin, code k27, Amathay Vesigneux, France
[27]Green Island Observatory, Code B34, Gecitkale, Famagusta, North Cyprus
[28]Institute of Solar-Terrestrial Physics (ISTP), Russian Academy of Sciences (Siberian Branch), Irkutsk 664033, p.b. 291, Lermontov street 126a, Russia
[29]Institute of Astronomy of Russian Academy of Sciences, Pyatnitskaya Str. 48, Moscow 119017, Russia
[30]Ulugh Beg Astronomical Institute of Uzbek Academy of Sciences, Astronomicheskaya Str. 33, Tashkent 100052, Uzbekistan

1 INTRODUCTION

No doubts that transit photometry is now one of the primary exoplanets detection tools. This method has a very promising descedant branch — transit timing variations, or TTVs. The outstanding value of the TTV method comes from its ability to derive independent hints of N-body interactions in a planetary system. This method is even capable to detect previously unknown planets (Agol & Fabrycky 2017). Moreover, it is capable to reveal unique objects directly demonstrating tidal interactions with the star, like now famous example of WASP-12 b (Bailey & Goodman 2019). This star demonstrates subtle period drift, as if the planet was slowly spiraling down onto its host star. Such a physical phenomenon brings us unique opportunities to test the theories of tidal planet-star interaction and even put some constrains on the interior structure of this particular exoplanet (Patra et al. 2017). Recently, hints of an analogous TTV drift were also reported for WASP-4 (Bouma et al. 2019), based on the first TESS observations.

Therefore, our present work is devoted to a further development of the TTV method. Basically, this paper presents results of a revised analyis following (Baluev et al. 2015) but includes additional observation targets, expanded photometric data, and improved processing algorithms. However, if the primary goal of Baluev et al. (2015) was to demonstrate the potential of amateur observations in the TTV field, the goal of the present work is to highlight the importance of using homogeneously derived TTV data.

The exoplanetary transit times published in literature are dervied by a vaste amount of various independent teams that used very different methods and models. For example, some works assume linear limb-darkening law, some quadratic. Concerning the limb-darkening coefficients, they may be set to fixed theoretically predicted values, or fitted as free parameters of the lightcurve. What concerns the photometric noise, it can be modelled differently as well: while some early measurements did not yet take into account the red noise, the other did, but in a different fashion. Some try to reduce systematic effects by decorrelating them with airmass, some use more complicated models, and some just fit the systematics by a deterministic model (e.g. trends plus multiple oscillations).

Moreover, the transit lightcurve fit also depend on the exoplanetary parameters which have an obvious tendency to improve their accuracy with time. While many earlier transit observations were rather inaccurate because they could not rely on accurate exoplanetary parameters, later ones could use bigger records of observations to derive more accurate results.

As such, the transit times published in the literature appear very heterogeneous: they may have subtle systematic biases in both the times and in their uncertainties. These biases are difficult to deal with, because they vary from one team to another in an impredictable manner. Hence, it might appear too dangerous to analyse such merged TTV data as if they were homogeneous: this may ultimately result in false detections of spurious variations and too unreliable conclusions.

This work presents an attempt to carefully reprocess the archival and new observation in a homogeneous way, relying on the same analysis protocol, including the use of the same methods and of the same transit and noise models. Now we can reprocess the entire photometry avaialble for each target in a self-consistent manner, i.e. we should not necessarily fit all the exoplanetary parameters in each lightcurve independently. This approach was already tested in (Baluev et al. 2015), and it allows us to reduce drastically the number of degrees of freedom of the fit, thus improving a lot the usability of lower-quality observations like the amateur ones.

Such a goal naturally infers a big work on systematizing the photometric data available, careful identification of possibly outlying measurements or even entire lightcuirves, as well as those lightcurves that possibly included erratic (but correctable) processing. Such a work necessarily implies an investigation of the models involved, in particular the limb-darkening models and noise models. This also includes a systematic analysis of the photometric noise, and seeking of

improved data-processing methods based on results of this investigation.

Moreover, we now aim to undertake a multimethod study not relying on just the photometric observations. We performed a self-consistent analysis of our homogeneously processed photometry jointly with Doppler data, since the combination of the transit and Doppler methods allows for a much more comprehensive characterization of a planetary system. This is especially important for sevelar unique exoplanets, like e.g. WASP-12 or WASP-4 which were either suspected or even confirmed to demonstrate a systematic departure from linear ephemeris (Maciejewski et al. 2018b; Bouma et al. 2019).

Finally, this work presents results of the first big practical test of the EXPANSION project (EXoPlanetary trANsit Search with an International Observational Network), grown on the basis of the Exoplanet Transit Database that was used by Baluev et al. (2015). Now EXPANSION is a standalone international project joining a network of several dozens of relatively small-aperture telescopes, aimed to monitor the exoplanetary transits (Sokov et al. 2018). This network covers amateur as well as professional observatories spreaded over the world in the both hemispheres.

The structure of the paper is as follows. In Sect. 2 we provide a detailed description of all the photometric and radial velocity data that we include in our analysis. In Sect. 3 we introduce the algorithms used to process the photometric data. In Sect. 4 we present results of empirical debiasing for the available theoretic models of the limb-darkening coefficients. In Sect. 5 we present the TTV data derived from the photometry of our 17 targets and results of their analysis, including a detailed discussion of possible TTV trends in WASP-12 and WASP-4. In Sect. 6 we present results of self-cosistent fits using both the transit and radial velocity data, available for 13 targets. In Sect. 7 we discuss in yet more detail the case of WASP-12, deriving a purely tidal part in its observed TTV trend.

2 PHOTOMETRIC AND DOPPLER DATA

An international project EXPANSION (EXoPlanetary trANsit Search with an Internation Observational Network) was organized for a long-term monitoring of exoplanetary transits. Amateur and professional observatories from Russia, Europe, North and South America with relatively small telescopes from 25-cm to 2-m are used in the photometric observations (Sokov et al. 2018). In this work we used photometric data from this same network. They also included all the data from ETD (Exoplanet Transit Database) that were used in (Baluev et al. 2015). Additionally, we used lightcurves published in the literature or kindly provided by the observers, as listed in Table 1. Most of them are available in the VIZIER database.

We expanded our targets list by seven exoplanets: Qatar-2, WASP-3, -6, -12, HAT-P-3, -13, and XO-5, thus increasing their number to 17. The total amount of the input data has grown considerably. This time we had $\sim 3\times10^5$ photometric measurements in $\sim$ 1000 lightcurves, compared to $\sim 8\times10^4$ measurements in $\sim$ 300 lightcurves processed by Baluev et al. (2015).

Whenever necessary, the timestamps in the photometric series were transformed to the $\mathrm{BJD_{TDB}}$ system by means of the public IDL software developed by Eastman et al. (2010). To perform this reduction, we used ICRS coordinates through the SIMBAD database which originate from GAIA DR2 (Brown et al. 2018). We did not apply any correction to these coordinates due to proper motion, since this would imply only a negligible correction to the time (below $\sim$ 0.1 sec).

Additionally, we used the precision radial velocity (RV) measurements obtained from the archival spectra of the HARPS, HARPS-N, SOPHIE, Keck/HIRES spectrographs. This involves the following targets from our photometry sample: Corot-2, GJ 436, TrES-1, WASP-2, -4, -5, -6, -12, HD 189733, XO-2N. The spectra were processed with the HARPS–TERRA pipeline (Anglada-Escudé & Butler 2012). Some of these data represent reprocessed versions of the RV data available in the literature, e.g. some were published in (Baluev et al. 2015) but now reprocessed, and some are new. Whenever performing a self-consistent transit and radial velocity analysis we transform all the Doppler time staps to the $\mathrm{BJD_{TDB}}$ system consistent with the photometry. However, the RV data that we release here correspond to the UTC rather than TDB system (as traditionally adopted for this type of the data).

Since Wilson et al. (2008) additional 31 RV measurements have been obtained for WASP-4 with the high resolution spectrograph CORALIE on the Swiss 1.2 m Euler telescope at La Silla Observatory, Chilie (Queloz et al. 2001). RVs were re-computed for the new data and the dataset presented in Wilson et al. (2008), for 45 measurements in total, by cross-correlating each spectrum with a G2 binary mask, using the standard CORALIE data-reduction pipeline.

For WASP-2, WASP-3, WASP-4, WASP-12, and XO-5 additional Keck/HIRES observations were presented by Knutson et al. (2014), which we included in the analysis in the published form. The Keck RV data from (Knutson et al. 2014) for XO-2N and GJ 436, and from (Albrecht et al. 2012) for GJ 436 were not used as they were found in our TERRA-processed sample. The HAT-P-13 data available in (Knutson et al. 2014) mysteriously appeared older and much less complete than RV data set by Winn et al. (2010), so we used the latter one. Some more in-transit RV data for WASP-12 are also mentioned in (Albrecht et al. 2012) but not published, and we currently did not acquire this set of raw Keck spectra to reprocess with TERRA.

The data files containing the photometic and radial-velocity measurements are attached to this paper as the online-only material. The format of the files follows that of (Baluev et al. 2015). Concerning the RV data, we currently release only a partial set, since we still plan to seek more RV data and perform their more detailed analysis in a future work.

We also notice that some of the TERRA-processed RV data published in (Baluev et al. 2015) appeared partly erratic. First, the HARPSN data for HD 189733 appeared entirely wrong, suspectively because they belong to another star (possibly, to the known companion B). Secondly, the difference between the new and old HARPS data for GJ 436 revealed a clear systematic trend indicating some processing error in the old data set. This trend was even highly significant in the previous RV release, but now it appeared spurious.

Table 1. Sources of the photometric data (except for the EXPANSION project).

Target	References	Note
CoRoT-2	Gillon et al. (2010)	
GJ 436	Gillon et al. (2007)	
	Bean et al. (2008)	HST Fine Guidance Sensor
	Shporer et al. (2009)	
	Cáceres et al. (2009)	Very high cadence; we binned these data to 10 sec chunks
HAT-P-3	Torres (2007)	
	Chan et al. (2011)	
	Nascimbeni et al. (2011a)	Data initially uploaded to VIZIER were not actually in BJD system as claimed (priv. comm.); correct data uploaded in 2017
	Mancini et al. (2018)	
HAT-P-13	Bakos et al. (2009)	
	Szabó et al. (2010)	
	Nascimbeni et al. (2011b)	
	Fulton et al. (2011)	
	Southworth et al. (2012)	
HD 189733	Bakos et al. (2006)	
	Winn et al. (2007a)	T10APT data involve *double* HJD correction by mistake (priv. comm.)
	Pont et al. (2007)	HST Advanced Camera for Surveys
	McCullough et al. (2014)	HST Wide Field Camera 3
	Kasper et al. (2019)	Multi-band transmission spectroscopy; very high accuracy data
Kelt-1	Siverd et al. (2012)	
	Maciejewski et al. (2018b)	
Qatar-2	Bryan et al. (2012)	It is not fully clear, whether the "BJD" times are given in UTC or TDB system. We assume BJD TDB, because the TTV residuals look bad otherwise.
	Mancini et al. (2014)	
TrES-1	Winn et al. (2007b)	
WASP-2	Southworth et al. (2010)	Danish telescope timings might be unreliable (Nikolov et al. 2012; Petrucci et al. 2013)
WASP-3	Tripathi et al. (2010)	
	Nascimbeni et al. (2013)	
WASP-4	Wilson et al. (2008)	
	Gillon et al. (2009a)	
	Winn et al. (2009)	Superseded by Sanchis-Ojeda et al. (2011)
	Southworth et al. (2009b)	Danish telescope timings might be unreliable (Nikolov et al. 2012; Petrucci et al. 2013)
	Sanchis-Ojeda et al. (2011)	
	Nikolov et al. (2012)	
	Petrucci et al. (2013)	These data were kindly provided by the authors
WASP-5	Southworth et al. (2009a)	Danish telescope timings might be unreliable (Nikolov et al. 2012; Petrucci et al. 2013)
WASP-6	Gillon et al. (2009b)	
	Tregloan-Reed et al. (2015)	
WASP-12	Hebb et al. (2009)	These data were kindly provided by the authors
	Chan et al. (2011)	
	Maciejewski et al. (2013)	Partly superseded by Maciejewski et al. (2016)
	Stevenson et al. (2014)	Multi-band transmission spectroscopy; very high accuracy data
	Maciejewski et al. (2016)	
	Maciejewski et al. (2018b)	
WASP-52	Chen et al. (2017)	Multi-band transmission spectroscopy; very high accuracy data
	Mancini et al. (2017)	
XO-2N	Fernandez et al. (2009)	
	Kundurthy et al. (2013)	
	Damasso et al. (2015)	
XO-5	None	

3 DERIVING TRANSIT TIMES FROM PHOTOMETRY

Our derivation of transit timing variations from photometry uses a similar algorithm to that of Baluev et al. (2015), though we augmented it somewhat in order that we can follow the processing stages below.

(i) Fit the raw transit photometry, followed by the fit of the resulting transit timings by a reference TTV model (linear ephemeris plus a possible quadratic trend, see eq. (5)).
(ii) Clean TTV outliers (bad lightcurves) by verifying the normality of the TTV residuals and then reprocess the remaining data.
(iii) Clean photometry outliers in the remaining lightcurves in a similar way and then reprocess the data.
(iv) Reprocess the data using semi-empirical limb-darkening coefficients for lightcurves in which the limb-darkening was ill-fitted or had poor accuracy. Those semi-empirical values represent theoretical predictions by Claret & Bloemen (2011) corrected by systematic biases derived in Sect. 4.
(v) Among the remaining lightcurves, identify higher-quality (HQ) ones, and reprocess them separately.

Our previous work followed only Stages 1 and 3. Stage 2 could not be completed due to a relative lack of TTV data. Stage 4 was not previously performed due to a simplistic limb-darkening treatment, which is now revised, and Stage 5 was absent. Most of the analysis is performed by means of an improved version 3 of the dedicated PlanetPack software (Baluev 2013, 2018). We now consider each stage in more detail.

3.1 Stage 1: lightcurve fitting

The light curve fitting is based on maximum-likelihood fitting with a dedicated model of the photometric noise and follows (Baluev et al. 2015). The procedure is self-consistent: we jointly fit the entire set of the lightcurves available for a given target. In particular, we fit all the data with a single parameters for transit duration, impact parameter, and planet radius. In other words, we bind each of these parameters across different lightcurves, rather than fit them separately for each lightcurve. Only the transit mid-times were fitted independently, to enable possible TTV detection. Such an approach allows to increase the overall statistical accuracy of the fit. In particular, this improved the TTV accuracy of the amateur observations, which otherwise would not always allow the robust determination of all the transit parameters. This approach may possibly have some side effects, e.g. if the planet demonstrates transit duration variations (TDV) in addition to TTVs, or variations in the remaining bound parameters like the impact factor. But in practice, transit duration and other parameters are determined with a much worse accuracy than mid-times, so it is reasonable to seek a relible TTV detection first, and only after that consider more complicated models.

As in (Baluev et al. 2015), we use circular model of the curved transiter motion. Most of our targets do not have a detectable orbital eccentricity, except for GJ436b. However, the photometric data for GJ436 appeared mostly of a too low quality. Except for a very few space-based HST observations,

they do not justify the use of a general Keplerian model. In any case, we include non-zero orbital eccentricities in the joint transit+Doppler analysis below.

The initial steps of the algorithm involve a set of preliminary fits, needed to avoid pathological solutions and fitting traps:

(i) Fit the data with a fixed transit impact parameter, fixed limb darkening coefficients and with a strictly quadratic TTV ephemeris by fixing TTV residuals to zero. Here we have the first difference from (Baluev et al. 2015), who adopted a linear ephemeris. Now we decided to use a quadratic one because now we have two candidates with a quadratic TTV trend (WASP-12 and WASP-4), and all other targets should be processed homogeneously.
(ii) Refit after releasing the transit impact parameter and mid-times.
(iii) Refit after releasing limb darkening coefficients (except for those that are fixed at the corrected theoretical values at Stage 4 or 5).
(iv) Determine very high-quality lightcurves that allow independent fitting of the limb darkening coefficients and if such lightcurves exist, refit the model yet again.

After that, the red noise auto-detection sequence is invoked. The photometric red noise is modelled by a Gaussian random process with the exponential correlation function and fittable noise parameters (white noise jitter $\sigma_{\rm w}$, red noise jitter $\sigma_{\rm r}$, and red noise timescale τ, see all the details in Baluev et al. 2015 and in Baluev 2018).

On the first substep each lightcurve was considered individually: we kept the red noise term in the final model only if it appeared at least mildly significant. We did not preserve the red noise terms, if they appeared ill-fitted with a given lightcurve. The criterion of a robust red noise detection was: (i) the log-likelihood ratio statistic Z should be at least 2, implying the asymptotic false detection probability $\chi^2_2(Z) = \exp(-Z) \sim 14$ per cent, (ii) the uncertainty of the red jitter $\sigma_{\rm r}$ is at least the estimated value (iii) the uncertainty of the red noise correlation timescale τ is at least twice the estimated value. These criteria were worked out empirically, and appear very mild (even more mild than in Baluev et al. 2015). In fact, they imply that most of the lightcurves *must* contain some red noise by default, adopting the purely white noise model only if the red noise could not be modelled reliably.

Also, in this work we used 3 starting initial values for τ, thus running up to 3 probe red-noise fits for each lightcurve. These initial values were spreaded logarithmically in the range from T/N to T (where T is the total time span of the lightcurve, and N is the number of its photometric measurements). In the previous work (Baluev et al. 2015) just a single initial value $\tau = T/\sqrt{N}$ was used, with a single probe fit per a lightcurve. It appeared that among our $\sim$ 1000 lightcurves, almost all reveal their red noise after just this very first trial fit with intermediary initial τ. However, in a few cases it appeared that the first fit did not converge to a robust solution because the actual best fitting value of τ was too far from the starting approximation $\tau = T/\sqrt{N}$. By adding two more probe fits starting from τ closer to the low and upper limits of the range, we could robustly detect the red noise in several lightcurves additionally.

But even with these very mild detection criteria and multiple trial fits it appeared that only 1/4 to 1/3 of our lightcurves revealed an individually fittable red noise. This result is basically in agreement with Baluev et al. (2015), however such a low fraction of the red-noised lightcurves still appears surprising. The red noise may exist in the rest of the lightcurves too, even though the individual parameters appear ill-fitted. Therefore, leaving the noise models in such a partial state might mean to make the resulting TTV data less homogeneous, because some of them would be obtained with a white noise model and some with a red noise one. For example, the uncertainties in the white-noise portion of TTV data may appear systematically smaller than in the red-noise one. To soften this possible effect we tried to fit the red noise in the remaining lightcurves in an averaged sense by using a restricted model. Since the most uncertain and poorly determinable red noise parameter is τ, we assumed that this τ is the same among all the lightcurves that did not reveal an individually detectable red noise. While binding the τ at such a shared 'average' value, the value of $\sigma_{\rm r}$ was still assumed individually fittable for each lighcurve to allow an adaptive match of the red noise magnitude. In this way, if this derived shared τ appeared inconsistent with the actual observations in a given lightcurve then this τ could be just ignored by reducing $\sigma_{\rm r}$ to zero.

After that the fraction of lightcurves enclosed by a red-noise model was raised to $50-80$ per cent, depending on the target. The rest of the data had the best fitting $\sigma_{\rm r} = 0$, implying that they contradict to either the derived shared τ, or to the red-noise hypothesis itself, formally suggesting the presence of a *blue* noise instead (or $\sigma_{\rm r}^2 < 0$). In the latter case, the uncertainty of the derived TTV measurement would formally appear even *smaller* than what is implied by the white noise model. However, we consider this to be unrealistic in general and do not allow the TTV uncertainties to decrease below their white-noise estimations.

Since we have a large set of red noise estimations for numerous lighcurves, it is now possible to consider some statistics. In Fig. 1 we show the histograms of the derived red noise parameters τ and $\sigma_{\rm r}$, and of the ratio $\sigma_{\rm r}/({\rm RMS})$, which reflects the relative contribution of the red noise in the total error budget. We can see that τ reveals a wide range from $\sim$ 10 sec to $\sim$ 50 min, but primarily located in the range $1-5$ min. The typical magintude of the red noise is $\sim$ 1 mmag, but also can deviate a lot from this peak value. The relative red noise contrubution is typically above 30 per cent (smaller values typically can not be detected or estimated reliably, so they were simply ignored in these histograms).

Yet another major difference with the previous work (Baluev et al. 2015) is a more careful treatment of the limb darkening. As before, we adopted a quadratic limb darkening model:

$$I(\rho)/I(0) = 1 - A(1-\mu) - B(1-\mu)^2, \quad \mu = \sqrt{1-\rho^2}, \quad (1)$$

where ρ is the projected distance from the disk center, and A and B should satisfy the constraints

$$A + B \leq 1, \quad A + 2B \geq 0, \quad A \geq 0, \quad (2)$$

which guarantee that $I(\rho)$ never turns negative and always remains monotonically decreasing (no limb brightening allowed), see (Baluev et al. 2015; Kipping 2013).

In (Baluev et al. 2015) the limb darkening coefficients A

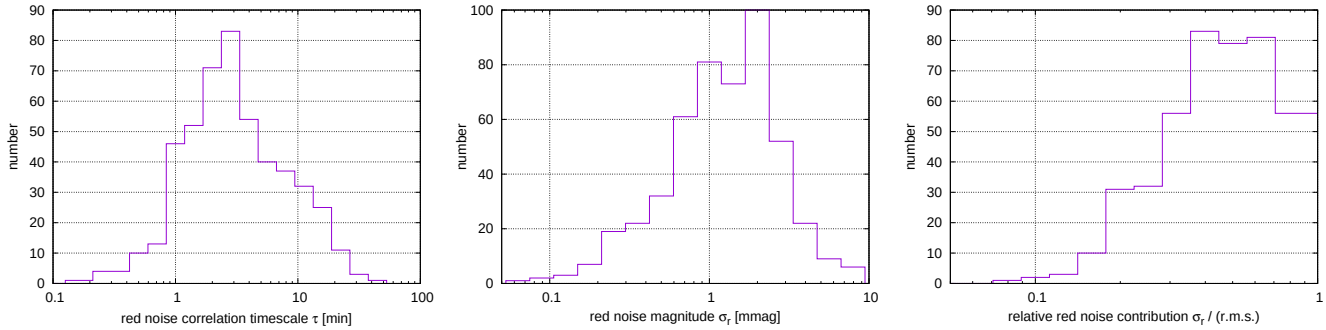


Figure 1. Histograms of the estimated photometric red noise parameters, τ and $\sigma_{\rm r}$, and of the relative red noise contribution in the total RMS. We used only those estimations of τ and $\sigma_{\rm r}$ on Stage 4 that exceeded their respective uncertainty.

and B were assumed the same for the most of the lightcurves, even regardless of their spectral band. In the present work we considered it as an inadmissibly rough assumption. Although a fully independent fit of these coefficients for every lightcurve is unnecessary (and even practically impossible), we need at least to fit them independently for different spectral filters.

We split all available lightcurves into several sets that correspond to the same or similar spectral band. For example, we combined in a single set the Johnson R_J and Cousins R_C filters, as well as the Sloan r or r' ones, treating them as the same "generic R" filter. The limb darkening coefficients appear almost equal in all these filters: the differences are smaller than e.g. those implied by different models of stellar atmosphere in (Claret 2000, 2004; Claret & Bloemen 2011).

Thus we sorted all our data into 8 classes, corresponding to the following "generic" spectral ranges: U, B, V, G, R, I, Z, K. Many lightcurves (mostly amateur ones) were obtained without any filter at all, or using a wide-band IR-UV cut-off filter, and we joined all such data under yet another class labelled "clear".

We fitted the data assuming that limb darkening coefficients are the same for all lightcurves belonging to the same band-class. After that, we carried out a systematic comparison of the resulting "observed" coefficients A and B with their theoreticalally predicted values based on (Claret 2000, 2004; Claret & Bloemen 2011). This comparison revealed certain systematic biases, discussed in Section 4 below.

A few lightcurves could not be assigned to any of the above band-classes, because they were obtained in another spectral band or with a different technique. Most of that data appeared of an exceptional quality, so we always fitted their limb darkening laws independently. These special cases include observations from Hubble Space Telescope (with different equipment), "white" lightcurves from a transmission spectroscopy, and data from some other specialized instruments.

3.2 Stages 2 and 3: cleaning the outliers

The cleaning of outliers is performed by means of inspecting the Gaussian quantile-quantile (QQ) plots of the residuals. The QQ plot is a non-linearly re-scaled graph of the empirical cumulative distribution $F_{\rm emp}(\varepsilon)$ of the normalized residuals $\varepsilon_i = r_i/\sigma_i$, where r_i is the best-fit residual and σ_i is its modelled standard deviation.[1] If all input data were good and all models (including the noise model for σ_i) were correct and correctly fitted, then the cumulative distribution of ε_i should be standard Gaussian, $F(\varepsilon) = \Phi(\varepsilon)$. Hence $F_{\rm emp}(\varepsilon)$ should be close to $\Phi(\varepsilon)$ as well, and the quantile function $Q(\varepsilon) = \Phi^{-1}(F_{\rm emp}(\varepsilon))$ should be close to $Q(\varepsilon) = \varepsilon$.

The graph of $Q(\varepsilon)$ is the QQ plot that we examine. If it lied close to the diagonal, the data would be practically Gaussian. These plots are given in the online-only Fig. 2, 1st row. We can see that the empirical curves are indeed close to the diagonal, suggesting almost Gaussian behaviour, but there is some number of outliers that deviate in the tails much more than a normal distribution would allow. Therefore, the outliers can be identified as points that reside in these tails.

In our work there is no difference between removing outliers from the raw photometry or from the derived TTV data: we again plot the relevant residuals and cut off the tails. The corresponding QQ plots are shown in the online-only Fig. 2, 2nd row.

We thouroughly reviewed the list of potential ~ 20 TTV outliers, and decided to manually 'whitelist' two lightcurves looking like outliers. Namely, this is one lightcurve for HAT-P-13 from (Szabó et al. 2010) and one for HD 189733 from (Kasper et al. 2019), both with $\varepsilon \simeq 4$. Concerning HAT-P-13, it demonstrated inconclusive hints of a TTV in the past, and it might appear to be the case that (Szabó et al. 2010) measurements actually reveal a true TTV, rather than a statistical outlier (e.g. induced by known nontransiting companions, see Winn et al. 2010). However, after that we noticed that this large residual was finally corrected on Stage 4, thanks to using corrected limb-darkening coefficients which appeared ill-fitted on Stage 3. Concerning the HD 189733 lightcurve by Kasper et al. (2019), it belongs to a homogeneous set of high-quality and homogeneous transmission spectroscopy observations. Therefore, we should remove either all the data by Kasper et al. (2019) together or remove none of them. We opted the latter solution, even though the particular lightcurve remains rather anomalous till the final Stage 5. Possible reasons of such anomalies in the Kasper et al. (2019) data and their consequencies are discussed in more details in Sect. 5.1.

[1] These σ_i were derived based on the multiplicative noise model without red noise, i.e. they are the estimated fit uncertainties normalized per each lightcurve so to have a unit reduced χ^2 statistic.

3.3 Stage 4: applying empirically corrected limb-darkening

Many lightcurves appear to have relatively poor quality, so they do not allow to fit the two-parametric limb-darkening law (1) reliably. Therefore, on Stage 3 several estimates of the limb-darkening coefficients appear to have large uncertainties about unity, and the limb-darkening coefficients themselves lie on the boundary of their admissible domain (Kipping 2013), indicating a poor fit.

To overcome potential side effects of this outcome, we perform one more processing of the data, fixing the limb-darkening coefficients with poor accuracy at certain semi-empirical values. Those semi-empirical values represent the theoretical predictions by Claret (2000, 2004) corrected by systematic biases obtained from the transit fits on Stage 3. The limb-darkening estimates are considered too poor if the uncertainty of either A or B appeared above 0.2. See more detailed discussion and motivation in Sect. 4.

It is notable that potentially poorly fit lightcurve by Szabó et al. (2010), mentioned in the previous subsection due to its large timing residual ($\sim$ 4-sigma), was finally corrected on Stage 4. This lightcurve was assigned a fixed semi-empirical limb-darkening coefficients that appeared very different from the fitted ones (see the discussion of Sect. 4). With this lightcurve model, its timing residual was reduced, while the uncertainty increased due to a detected red-noise term. Then the point becomes a more tolerable $\sim$ 2.4-sigma outlier. We therefore believe that problems with this point appeared due to photometric red noise that mimiced by a bias in the limb-darkening coefficients.

3.4 Stage 5: determining high quality lightcurves

To identify transit times of a higher quality, we first introduce the following "quality characteristic" of a lightcurve:

$$\mathcal{Q} = \frac{\sqrt{\text{measurements density}}}{\text{residuals RMS}} \tag{3}$$

The quantity $1/\mathcal{Q}$ determines the uncertainty offered by a "standard" chunk of the lightcurve of a unit length. The uncertainty of an arbitrary chunk of length t scales as $1/(\mathcal{Q}\sqrt{t})$, so $\mathcal{Q}\sqrt{t}$ can be accepted as a rough "quality characteristic" of the chunk. [2]

This characteristic is not yet indicative concerning the particular exoplanet. Let τ be the transit duration, and $r' = r/R_\star$ be the planet/star radii ratio. Then the uncertainty of the in-transit piece of the lightcurve would be $1/(\mathcal{Q}\sqrt{\tau})$, and it should be compared to the transit depth r'^2. That is, the following normalized parameter:

$$\mathcal{Q}' = \mathcal{Q}\sqrt{\tau}/r'^2 \tag{4}$$

can serve as our idealized quality characteristic. Say, $\mathcal{Q}' = 100$, then the transit depth can be measured with an accuracy of 1 per cent, while $\mathcal{Q}' = 10$ implies relative accuracy of merely 10 per cent.

Now let us plot the empirical disribution of $\mathcal{Q}'$ computed for all our lightcurves (online-only Fig. 2, bottom). We can see that $\mathcal{Q}'$ varies in an extremely wide range from a few tens to a few thousands. We choose a threshold $Q' > 100$ to select the HQ lightcurves. Such a threshold keeps about 2/3 of the entire sample, so it is a relatively mild filter. Our goal was mainly to filter out only very inaccurate and probably useless data, rather than to select a minor portion of highly accurate ones.

Note that whenever a lightcurve has low $\mathcal{Q}'$, this does not necessarily mean that it must be immediately removed from the analysis as unreliable. Such a lightcurve just has a poor overall accuracy, but it already survived the normality tests of the previous processing stages. Statistically, the derived timing value remains quite admissible and usable (within its uncertainty).

Besides, some of our targets like GJ436 and Kelt-1 have very small transit depth, and would be left with very few TTV measurements after passing through Stage 5. By this reasoning, below we consider results of Stage 4 and Stage 5 simultaneously and that the reader can compare them.

4 EMPIRICAL CALIBRATION OF THE LIMB-DARKENING COEFFICIENTS

Using the techique presented above, we performed a per-target and per-band fit of the limb darkening coefficients A and B in the quadrating model (1). After that, we compared these empirical estimates with their theoretically predicted values from (Claret 2000, 2004) and their newer update from (Claret & Bloemen 2011) band-by-band. The coefficients for the "clear" lightcurve class were compared with the bolometric estimates by Claret. We utilized the JKTLD code by Southworth (2015) that offers a convenient interface for interpolating the original tables by Claret.[3] The necessary stellar parameters ($T_{\rm eff}$, $\log g$, [Fe/H]) were taken mainly from the SWEET-Cat (Santos et al. 2013), and from (Siverd et al. 2012) for Kelt-1. In almost all cases we used the coefficients corresponding to the ATLAS models, except for GJ 436, for which only the PHOENIX-based coefficients were available. We assumed microturbulence velocity v_t of 2 km/s in all the cases.

We found that many of our A and B estimations, even the most accurate ones, significantly deviate from their predicted values by Claret. In itself, this is not very surprising, because the theoretical values are expected to have some bias (Heyrovský 2007). Even if the theoretical brightness profile was entirely perfect, the two-parameteric models like (1) cannot approximate it everywhere equally well. The resulting "theoretical" coefficients A and B depend on how we fit this profile: they may appear biased to better fit one its portion or another. These coefficients should not necessarily coincide with the empirical ones obtained from transit fitting (even if the latters had no significant errors at all).

In online-only Fig. 3, some worst-case discrepancies are

[2] Here we neglect possible red noise, so even neighbouring measurements are assumed uncorrelated.

[3] See http://www.astro.keele.ac.uk/jkt/codes/jktld.html for download; we actually augmented this code to make it capable of processing the newer tables by Claret & Bloemen (2011), and applied an additional post-interpolation with respect to the metallicity parameter [M/H], which is merely selected rather than interpolated by JKTLD.

demonstrated. The empirical A and B estimations correspond to the processing Stage 3, i.e. they were obtained after removal of both the timing and photometry outliers. Here, the theoretical values were derived from (Claret 2000, 2004), and one can see that they systematically deviate by $\sim 0.1 - 0.2$ from the empirical estimates.

We tried to compute the universal shifts ΔA and ΔB, necessary to minimize the differences between the observed and theoretical coefficients. The weighted least squares fit yielded the biases $\Delta A = 0.059 \pm 0.008$ and $\Delta B = -0.172 \pm 0.014$ for the quadratic law, and $\Delta A = -0.112 \pm 0.002$ for the linear law. These shifts refer to the older tables by Claret (2000, 2004), and take into account only the UBVGRIZK filters.

By fitting the newer models by Claret & Bloemen (2011), corresponding to the flux-conservation method (FCM), and for the same spectral filters as above, we obtained the following biases: $\Delta A = 0.004 \pm 0.008$ and $\Delta B = -0.099 \pm 0.014$ for the quadratic law, and $\Delta A = -0.035 \pm 0.002$ for the linear law. The newer tables are clearly better, although some minor bias still remains in B.

By adding these best-fitting corrections to the theoretical A and B values the agreement can be improved remarkably. This is especially obvious in several high-accuracy cases (e.g. WASP-4, Qatar-2), see online-only Fig. 4.

The coefficients from Claret & Bloemen (2011) obtained with the LF method (least-square fit of the brightness profile) appear less accurate than the FCM ones and more similar to those from (Claret 2000, 2004). The corrections determined above refer mainly to the ATLAS models and UBVGRIZK filters.

The coefficients given in the older works (Claret 2000, 2004) differ from the newer and more accurate ones from (Claret & Bloemen 2011) remarkably, as well as coefficients obtained with the FCM and LF method. But the TTV data published in the literature may correspond to either of two systems, or can be fitted from a particular transit lightcurve. We caution the reader against direct mixing of different timing measurements, especially those released before 2011 and after 2011, as these timings may have different biases due to different limb-darkening coefficients. This highlights the value of a homogeneous TTV analysis, based on simultaneous fitting of all raw lightcurves at once and using the same analysis pipeline.

In this work, we adopt a hybrid approach to model the limb-darkening profile and so we illustrate some key aspects below.

(i) If at Stage 3 both A and B had the fitting uncertainty of better than 0.2 and simultaneously did not reside on either boundary of (2) then we did not rely on the theoretical values. Even the corrected ones may still appear to be biased for an individual star (in particular, due to possible inaccuracies in stellar parameters, $T_{\rm eff}$, $\log g$, metallicity, or turbulence velocity). In such a case we allowed the coefficients to be fitted from the transit curves as free parameters, taking into account the common binding contraints within each spectral filter.

(ii) The limb-darkening coefficients corresponding to filters other than UBVGRIZK, were always fitted, including the no-filter ("Clear") cases, regardless of their resulting accuracy. Notice that in the online-only Figs 3, 4 we compare the "Clear" band with the predicted bolometric values only for a reference: we do not actually rely on the bolometric coefficients in our processing.

(iii) If our data-based limb-darkening estimations appeared too uncertain (the uncertainty of either A or B above 0.2 on Stage 3) or ill-fitted (the best fitting solution resided at the boundary of the domain (2)) then we fixed such coefficients at their theoretical FCM values from (Claret & Bloemen 2011), corrected by the biases derived above. This refers to only the UBVGRIZK filters. The motivation here was mainly to get rid of ill-fitted and definitely unrealistic solutions rather than to improve the resulting accuracy of the derived limb-darkening law by incorporating prior assumptions, as e.g. the Bayesian approach would assume. The semi-empirical estimates of A and B are in fact posterior quantities, since they include a posterior correction, but whenever the data-derived uncertainty of A and B becomes so large as 0.2, the fit (and fit uncertainties) practically does not depend on the adopted A and B values, except if they become completely unrealistic. Therefore, by fixing A and B in such a case we only cut off that unrealistic solutions.

The graphs of these limb-darkening coefficients are presented in the online-only Fig. 5.

5 RESULTS OF THE TRANSITS ANALYSIS

5.1 Verifying the quality of the derived timings

Before presenting our main TTV analysis results, we need to discuss the quality of the derived transit timing data. In fact, our transit analysis pipeline is still experimental to a certain extent, and it has some important differences from the methods applied usually. In particular, we treat the red noise using a parametric model by Gaussian processes, while many other works may use different techniques, e.g. ascending to (Pont et al. 2006) or (Foreman-Mackey et al. 2017). Also, we used different statistical treatment paradigms and different software. Finally, and probably most importantly, we analyse jointly lightcurves of a very different quality, from amateur ones to professional ground-based and even space-based HST data. Although we undertook multiple efforts to handle such a heterogeneity, its side effects may still exist.

Therefore, we need to perform some benchmarking of the accuracy and quality of our TTV data. This can be done by comparing them with analogous TTV data from other published works. However, most of the published TTV data were derived by different teams who used different techniques and different assumptions (e.g. concerning the limb-darkening). Hence, such data may have inconsistent systematic errors or statistical properties, and their mixed compilations cannot serve as reliable benchmarks. We need to identify such a target for which a long record of TTV data is available, but all the timings were obtained mostly by the same single team. In other words, we need to compare our results with another homogeneously-derived set of timings for heterogeneous photometric data.

In our target list, only WASP-12 perfectly suits our needs. It has $\sim$ 200 observed transit lightcurves in total, and about half of them were processed by the same group (Maciejewski et al. 2013, 2016, 2018b). Simultane-

ously, these data were obtained at quite different telescopes located in different astroclimate conditions. Therefore, they have different quality characteristics, offering the necessary degree of physical heterogeneity.

The next potential comparison benchmark would be HD 189733 ($\sim$ 100 transits available), but this star was observed by multiple independent groups that applied different processing algorithms, so any comparison with such mixed TTV series would remain inconclusive and potentially doubtable.

Here we used the transit times from (Maciejewski et al. 2018b) and (Maciejewski et al. 2016) that included most reprocessed lightcurves of (Maciejewski et al. 2013). We did not include the timings from the 2013 paper not reprocessed in the 2016 one (they would be statistically different). After that we sampled the same transits from our homogeneous data release. Thus we obtained two similar TTV time series to be compared, containing 73 data points each at equal epochs. We fitted both these data sets with a quadratic trend model (5). The best fitting models appeared practically identical. We then computed the resulting RMS: 22.1 sec for the Maciejewski et al. data and 24.9 sec for our data release. The Maciejewski et al. data winning with a slightly smaller scatter of the residuals, though this 12 per cent difference is comparable to the probable statistical uncertainty ($1/\sqrt{N}$ for $N = 73$ yields the same 12 per cent). Therefore, the intrinsic statistical accuracy of our processing pipeline appears similar to the best TTV data sets available in the literature.

However, our data reveal important difference in another aspect. The value of the reduced χ^2 for the quadratic TTV model is 0.79 for Maciejewski et al., implying that they tend to overestimate their TTV uncertainties by 12 per cent on average. On the contrary, our data imply the reduced χ^2 of 1.49, which means that our TTV uncertainties appear underestimated by the factor $\sqrt{1.49}$, or by $\sim$ 22 per cent on average.

We notice that it is quite frequent that the uncertainties reported for some measurements have a remarkable systematic bias. In fact, we should never expect that any uncertainties are accurate, because there are always subtle physical effects that were missed, or shortcomings of the adopted models, or hidden inaccuracies of the statistical processing. All this may lead to a systematically wrong uncertainties in the derived data. The same phenomenon was known long ago in the precision Doppler data (Wright 2005; Baluev 2009), and a generally similar effect should be expected in TTVs.

Then just like the RV data in Doppler programmes, the TTV data released by different groups may have quite a different level of a hidden noise. Some groups may overestimate their average TTV errors, some may underestimate them, and each team would have an individual correction factor. We therefore caution the reader against simplistic joining of TTV data coming from different sources. Such a merging should be made in an adaptive manner instead, taking into account possibly different relative weights of heterogeneous data. One way of such an adaptive treatment is demonstrated below in the WASP-4 case.

Concerning the reason why the Maciejewski et al. data appeared to have overestimated uncertainties, i.e. a smaller actual scatter than expected, this remains unclear. This is relatively non-physical situation in fact, only indicating that there are some unclassified inaccuracies in the statistical processing or models adopted. Concerning our TTV data which have underestimated uncertainties, this can be explained in a bit more physical manner via the effect of an additional noise source, not taken into account when performing the processing.

We believe a good candidate here might be the effect of starspot transit events. Initially, we expected that such transit curve anomalies might be partly taken into account by a red noise model, however it appeared that lightcurves with obvious spot-transit anomalies usually do not have a detectable or even fittable red noise.

Moreover, in practice it sometimes appeared that such transit curves demonstrated hints of a *blue* noise with $\sigma_{\rm r}^2 < 0$. After some thinkering one may even argue that such a behaviour is reasonable. What is the difference between the white, red, or blue noise? The type of the noise is basically determined by its rate of decrease whenever it is averaged: either $N^{-\frac{1}{2}}$ (white), or slower than that (red), or quicker than that (blue). A single spot-transit perturbation in the lightcurve is actually not a noise in the strict meaning: it is a determenistic anomaly. The noise-like effect here appears only because these anomalies change randomly from one transit to another. However, since for a given lightcurve a spot-transit anomaly is determenistic, it behaves as a deterministic function, e.g. it is averaged out at the rate $1/T$, where T is the length of the observation sequence. This corresponds to the decay rate of $1/N$, as N is accumulated with T linearly. Therefore, such an anomaly can be interpreted as a blue noise rather than red or white one.

The activity-related nature of the excessive TTV jitter is indirectly confirmed by the following observation: the values of the χ^2 for the derived transit times (see Table 2 below) are systematically larger for our HQ sample. This might be deemed as a confirmation that the unaccounted source of TTV errors is diluted in the lower-quality lightcurves.

We also notice that the value of the TTV scatter is very target-dependent. While for the most targets the value of the $\sqrt{\chi^2_{\rm TTV}}$ is about 1.3-1.5, it rises to $\sim$ 2 or even above for only three targets: HD 189733, XO-5, and Kelt-1. This might indicate that these three targets have a spotting activity larger than average. While for XO-5, and possibly for Kelt-1, this effect can be also explained via the small number statistics or other purely statistical or data-analysis concerns owing to the deficient data, the star HD 189733 has very rich set of HQ photometry. There is no way to explain its very high TTV scatter other than by the activity effects.

In particular, we notice that some HQ observations by Kasper et al. (2019) may be affected by hidden starspot transit or other activity-related phenomena (even though they are not obvious from the lightcurve). This would explain why one of them was identified as an outlier deviating by $\sim$ 2 min (see Sect. 3.2). Note that our estimation of this transit time is essentially consistent with the original Kasper et al. (2019) value (the shift by 6 sec, our uncertaity is 12 sec compared to the original uncertainty of 11 sec).

5.2 Analysis of the TTV

We processed our TTV data in the homogeneous manner, using the same protocol for each target. For the first step, we tested the existence of a possible long-term nonlinear trend

in the TTV time series, expressing it as a quadratic model:

$$\mathrm{TTV}(n) = T_0 + P(n-n_0) + \frac{1}{2}\frac{dP}{dn}(n-n_0)^2, \qquad (5)$$

where n is the transit count, P orbital period, and dP/dn is the small quadratic coefficient (treated as a solid fittable coefficient). By setting a temporal variable $t = nP$, we can alternatively use a more intuitive quantity $\dot{P}/P = P^{-2}dP/dn$ and rewrite (5) as:

$$\begin{aligned} \mathrm{TTV}(t) &= T_0 + (t-t_0) + \frac{1}{2}\frac{\dot{P}}{P}(t-t_0)^2 \\ &= T_0 + (t-t_0) - \frac{(t-t_0)^2}{2T_\mathrm{d}}. \end{aligned} \qquad (6)$$

In such a way we can seek for a long-term drift in an orbital period. In this model the quantity $T_\mathrm{d} = -P/\dot{P}$ represents a characteristic time of orbital decay (the time when the apparent period would turn zero in the linear approximation). Since it has an intuitive interpretation, we often use this quantity below as a reference fit parameter (rather than the quadratic coefficient itself).

However, we emphasize that considerably different physical phenomena may be approximated by mathematically the same formula (5): tidal orbital decay, tidal apsidal drift, or even non-tidal effect of a perturbation from a distant companion (causing the TTV via the variable light arrival delay).

The TTV residuals themselves are plotted in the online-only Fig. 6 and 7 for all our targets. They correspond to a linear TTV ephemeris and are given separately for Stage 4 (all data) and Stage 5 (HQ data).

We were able to easily confirm the TTV trend of WASP-12 (Maciejewski et al. 2016) at this stage. This case is discussed in details below in a separate section.

Based on observations prior to 2018, we first obtained promising hints of a similar trend for WASP-4. But this trend was not confirmed based on a set of 10 transit observations obtained with the EXPANSION network in 2018. Surprisingly, the recent publication of the TESS timing data (Bouma et al. 2019) claimed this trend to exist, so we investigated this target more closely in view of the TESS data from 2018. The results are given below in a separate section.

The other targets did not demonstrate detectable hints of nonlinear trends.

Furthermore, we performed a search for periodic TTV signals. We constructed a periodogram z_3 from (Baluev 2008), shown in the online-only Fig. 8 (for Stage 4) and Fig. 9 (for Stage 5). The base model for this periodogram always included a quadratic trend.

We could not find any periodic TTV for any of the targets. Peridograms did not revel hints of significant periodicities. In particular, we do not detect any hints of previously claimed controversial TTV for HAT-P-13 (Nascimbeni et al. 2011b; Pál et al. 2011) or for WASP-3 (Maciejewski et al. 2010; Montalto et al. 2012; Maciejewski et al. 2018a). Concerning the HAT-P-13 target, it has a second companion HAT-P-13 *c*, and also reveals hints of additional long-period companions appearing as a linear RV trend (Winn et al. 2010). These additional companions would clearly impose a variable light travel delay effect on the inner tight subsystem, causing therefore a TTV. However, this type of TTV is not detectable in HAT-P-13 due to a small magnitude (e.g. ~ 7 sec from HAT-P-13 *c*). In this work we did not investigate the TTVs possibly coming from gravitational perturbations of the planet *b* orbital motion.

For WASP-4 and using only its HQ data, we find that multiple peaks rise above the two-sigma significance level in the short-period range. However, these peaks look more like a noise rather than a systematic variation. Moreover, they disappear if we remove just a single timing measurements, namely the one derived from the lightcurve by Sanchis-Ojeda et al. (2011), dated by 02 Aug 2009. We believe that this lightcurve could be slightly affected by a subtle residual systematic effect or by a spot-transit event, even though it was not classified as an outlier and looks visually reasonable. Similar notice holds for HD 189733, which involves at least one lightcurve by Kasper et al. (2019) with anomalous timing.

We note that in (Baluev et al. 2015) we claimed for WASP-4 inconclusive hints of periodic TTVs in the range of a few days. However those hints disappeared after we adopted a more careful treatment of the limb-darkening coefficients. This highlights the practical value of an accurate limb-darkening model, even if the limb-darkening apparently does not seem so important for TTV studies.

5.3 Updated planetary transit fits

The Table 2 contains fitted transit parameters for all our 17 exoplanets, both for the Stage 4 and Stage 5 data. We give only rather raw parameters, while the complete set can be determined only from the transit+RV fits (considered below). In addition, we give the number of red-noise lightcurves per each target (fitted individually or with shared τ), the qumulative quality characteristic for each target, the maximum and mean absolute correlation of the derived transit times (which appears between different transits through the shared planetary parameters), and the reduced χ^2 for the derived transit times (relatively to a quadratic TTV model).

We also performed alternative fits assuming that all the transit times follow a quadratic model, with an independent red-noise detection sequence (expectedly resulting in a slightly larger number of red-noised lightcurves). For these alternative fits we only give the three parameters of the best fitting quadratic ephemeris (5).

Concerning the correlations between transit times, we find that most of them a practically zero. For GJ 436 and WASP-6 there are pairs of transits that generated a large correlation, up to 0.68. This is because contrary to (Baluev et al. 2015), in this work we included several partial transits in the analysis. The timing value derived for such transits relies a lot on the common transit parameters like the transit duration, and therefore depends on the other lightcurve through such shared parameters. Nonetheless, in the average the effect of correlations becomes negligible, so we decided to keep such transits, because transit data for these two targets are in a deficit.

For Kelt-1, we obtain the estimation of the impact parameter $b = 0.05 \pm 0.64$. This is obviously a formal and non-informative value since the parameter b becomes severely nonlinear and hence nongaussian whenever it becomes comparable to the uncertainty. In this case, a considerably more linear parameter might be $a = \sqrt{1-b^2}$ with $\sigma_a = (b/a)\sigma_b$

(if b is the distance of the transit trajectory from the star disk center, a is its distance from the star limb). This corresponds to $a = 0.999 \pm 0.032$, implying the 1σ low limit on a of 0.967, hence a more realistic upper 1σ limit on b of 0.25 (rather than 0.64). Notice that in (Baluev et al. 2015) the Kelt-1 best fit would formally correspond to an $a > 1$, i.e. imaginary b, so it was set to the least physically sound value $b = 0$. Clearly, the true value of b is almost zero in any case, but its uncertainty is large, so the best fit may move to a non-physical domain $a > 1$ because of the noise.[4] To avoid this mathematical pecularity near $b = 0$, one could use e.g. b^2 as a primary fit parameter, however we keep using b as it is more traditional and intuitive.

Finally, the most important observation from Table 2 is that all values of $\chi^2_{\rm TTV}$ are significantly above unit. This indicates, most probably, that our algorithm does not take into account all the noise sources in full. As we already noticed above in Sect. 5.1, the main such escaped noise source is likely the effect of spotting activity causing random anomalies in transit lightcurves.

For our goals it is important that this activity effect, whatever physical source it has, can be easily modelled on the TTV processing stage. This can be done by adding the additive noise to the derived transit timing uncertainties, or by multiplying them by a constant factor (we did not find definite hints clearly favouring either of these models). These noise models are discussed in details in (Baluev 2009, 2015).

Yet another important consequence of this phenomenon is that all self-consistent fits that avoided the TTV processing stage may appear to have underestimated uncertainties. This refers, in particular, to the quadratic ephemeris given in Table 2. For example, for WASP-12 the relative uncertainty of dP/dn from the table is 6.7 per cent, while after processing the transit times with an adaptive noise model (see Section 5.4 below) we obtain the more realistic relative uncertainty about 9.2 per cent. The ratio of these uncertainties is almost equal to the value of $\sqrt{\chi^2_{\rm TTV}}$ from the table.

We expect that the same effect of underestimated uncertainty affects the values of T_0 and P from Table 2, and also the values of P and $T_{\rm d}$ from Table 6 containing the self-consistent transit+RV fit. Their uncertainties following from a self-consistent fit should be increased by the factor of $\sqrt{\chi^2_{\rm TTV}}$ to become more realistic. Concerning the other fitted parameters, they may also have an underestimated uncertainty, although the correction factor is currently difficult to guess (it is not necessarily related to $\chi^2_{\rm TTV}$, if the parameter has no direct relationship with transit times).

5.4 WASP-12: a nonlinear TTV trend

Our analysis yielded above 9-sigma significance of the WASP-12 quadratic TTV term. This appears convincing, and the trend itself can be easily distinguished in online-only Fig. 6. We obtained the characteristic orbit decay time $T_{\rm d} = P/|\dot{P}| = 3.57 \pm 0.33$ Myr (or 3.60 ± 0.34 Myr for the HQ subsample). This is consistent with the recent estimations by

[4] This formally means that the planet radius following from the apparent duration of the transit ingress/egress phases appeared smaller than the radius following from the transit depth.

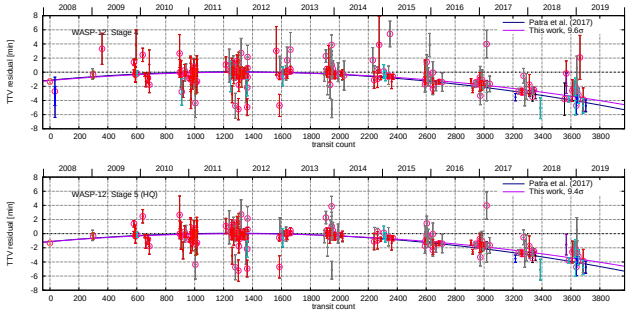


Figure 2. Transit times of WASP-12 derived in this work. Top panel is for all the TTV data (stage 4), bottom panel is for only HQ ones (stage 5). The models of the quadratic TTV trend are also plotted (for the multiplicative noise model).

Patra et al. (2017) and Maciejewski et al. (2018b). This estimate is based on the multiplicative noise model (Baluev 2015), which assumes that the TTV uncertainties are scaled by a fittable constant factor in the way preserving their relative weights. The scale factor is 1.35 or 1.33, the values of $\sqrt{\chi^2_{\rm TTV}}$ from Table 2.

We also considered the so-called regularized noise model from (Baluev 2015), which in our conditions is almost equivalent to the 'additive' model. In this model, the noise is represented as a quadrature sum of the derived TTV uncertainty and of a 'jitter'. That 'jitter' describes some additional source of noise (e.g. the effect of random spot-transit events disturbing the derived timings). With this model we obtain $T_{\rm d} = 3.55 \pm 0.31$ Myr (3.60 ± 0.31 Myr from only HQ TTVs), practically the same values. The best fitting TTV jitter for our data is estimated to be 20.8 ± 2.5 sec (18.9 ± 2.5 sec for the HQ subsample).

Therefore, this result is practically model-invariant and thus very trustable. As such, the tidal quality factor remains at $Q_* \sim 2 \cdot 10^5$, the value from (Patra et al. 2017). The TTV residuals and the corresponding trend best fitting models are plotted in Fig. 2.

Notice that we did not include secondary eclipses in our analysis, and did not use some transit timings published without lightcurves that we could reprocess in a homogeneous manner.

From only the transit timing data, we did not obtain any qualitatively new result for WASP-12. Initially we intended to use it mainly as a good benchmark to test our analysis pipeline: the fact that we basically reproduced the literature results at the same accuracy level ensured us that we are quite competitive in the detection of TTVs of such type. But radial velocity data for this star brought a significant additional information about the nature of this TTV trend. This is discussed in more details in Section 6.

In fact, in this task even more complicated models of the TTV noise might appear reasonable, treating differently the high-quality and low-quality timing data. For example, such models may involve multiple fittable parameter determining how the actual error variance depends on the derived formal TTV uncertainty. Contrary to the Doppler radial-velocity time series, adding more degrees of freedom to the TTV noise model may appear practically feasible because the trend model like (5) is quite simple and has only three

Table 2. Fitted parameters of exoplanetary transit curves after Stages 4 and 5 processing.

transiter host	total number of transits N	Assuming fittable transit times: number of red-noised lightcurves[3]	total quality[1] $\mathcal{Q}'_{\rm sum}$	radii ratio $r = R_{\rm pl}/R_\star$	half-duration $t_d/2$ [days]	impact par. b	$\sqrt{\chi^2_{\rm TTV}}$ [2]	mid-times correl. MAD	MAX	Fixing timings at a quadratic model[2]: ref. mid-time[5] T_0 [BJD$_{\rm TDB}-$ -2450000]	orbital period[5] P [days]	TTV trend[5] dP/dn [10^{-10}day]
Corot-2	38	13 + 12	893	0.16524(93)	0.04726(20)	0.158(70)	1.37	0.00034	0.006	7622.43669(11)	1.74299767(36)	2.6(5.1)
GJ436[4]	47	8 + 7	857	0.0847(10)	0.02108(21)	0.8612(66)	1.10	0.0022	0.66	4439.41624(10)	2.64389938(67)	−28(12)
HAT-P-13	51	19 + 12	1470	0.08826(78)	0.06919(38)	0.7500(77)	1.55	0.0011	0.036	5476.91220(19)	2.9162394(14)	92(33)
HAT-P-3	66	16 + 16	1400	0.11091(48)	0.04335(17)	0.615(12)	1.48	0.00015	0.011	7237.38678(10)	2.89973797(38)	8.0(9.6)
HD189733	106	27 + 32	12800	0.15703(32)	0.037514(48)	0.6646(20)	2.10	0.00029	0.17	3968.837026(20)	2.21857545(15)	−5.3(1.9)
Kelt-1	34	12 + 14	938	0.07584(82)	0.05682(28)	0.05(64)	1.91	0.00092	0.023	8026.51487(14)	1.21749220(71)	−20.6(8.6)
Qatar-2	59	12 + 20	1750	0.16165(91)	0.03812(11)	0.129(63)	1.26	0.00021	0.014	6045.458848(35)	1.33711643(26)	4.8(3.6)
TrES-1	56	13 + 23	1370	0.13799(83)	0.05233(17)	0.238(45)	1.17	0.00029	0.0067	4350.354597(84)	3.03006957(34)	1.0(7.1)
WASP-2	68	20 + 19	1530	0.13315(50)	0.03727(15)	0.7382(50)	1.47	0.00024	0.0042	5513.13577(12)	2.15222222(39)	3.0(8.0)
WASP-3	69	17 + 20	1710	0.10637(58)	0.05700(18)	0.492(17)	1.53	0.00025	0.068	5325.825419(88)	1.84683507(26)	5.4(4.0)
WASP-4	66	22 + 22	4250	0.15488(32)	0.044907(53)	0.130(29)	1.31	0.00019	0.0094	5045.738470(22)	1.338231531(83)	−0.98(94)
WASP-5	17	9 + 3	1220	0.11459(79)	0.05030(22)	0.453(25)	1.62	0.002	0.022	5896.57891(15)	1.62843035(71)	−10(17)
WASP-6	18	8 + 3	1460	0.14310(88)	0.05368(19)	0.222(50)	1.68	0.006	0.68	5379.546164(80)	3.36100260(57)	−28(22)
WASP-12	230	84 + 72	9070	0.11840(17)	0.062362(49)	0.4312(48)	1.35	0.00013	0.018	5994.401004(25)	1.091420405(51)	−9.51(64)
WASP-52	72	25 + 20	2010	0.16538(51)	0.03875(11)	0.5985(62)	1.56	0.00063	0.1	6904.792855(60)	1.74978179(37)	−19(11)
XO-2N	73	29 + 20	3930	0.10348(38)	0.055937(92)	0.194(40)	1.44	0.0004	0.054	5167.935634(40)	2.61585965(16)	−4.1(4.1)
XO-5	28	9 + 4	678	0.1026(15)	0.06354(58)	0.537(40)	2.12	0.00075	0.01	7760.49334(36)	4.1877641(22)	258(67)
Corot-2	25	10 + 9	856	0.16557(99)	0.04728(22)	0.190(63)	1.55	0.00061	0.0074	7629.40868(12)	1.74299768(47)	3.2(6.5)
GJ436[4]	11	2 + 0	757	0.0829(15)	0.02107(26)	0.8664(78)	1.32	0.041	0.66	4280.78227(12)	2.6439052(85)	−160(190)
HAT-P-13	38	16 + 10	1420	0.08757(81)	0.06895(39)	0.7456(82)	1.54	0.0015	0.037	5511.90699(19)	2.9162394(14)	96(34)
HAT-P-3	34	7 + 7	1340	0.11075(50)	0.04337(18)	0.622(12)	1.54	0.00034	0.012	7237.38674(10)	2.89973825(42)	14(11)
HD189733	75	24 + 19	12800	0.15696(33)	0.037525(48)	0.6645(20)	2.39	0.00046	0.17	3968.837031(20)	2.21857545(16)	−5.6(2.0)
Kelt-1	18	6 + 5	887	0.07523(89)	0.05694(31)	0.13(26)	1.57	0.0018	0.019	8026.51487(15)	1.21749273(88)	−14(11)
Qatar-2	25	8 + 10	1720	0.16159(93)	0.03810(11)	0.125(66)	1.40	0.0005	0.014	6034.761900(36)	1.33711631(33)	4.0(5.6)
TrES-1	44	14 + 18	1340	0.13772(85)	0.05228(17)	0.235(47)	1.13	0.00036	0.0071	4347.324507(86)	3.03006949(35)	4.2(7.4)
WASP-2	32	12 + 10	1450	0.13328(52)	0.03730(16)	0.7378(54)	1.64	0.00037	0.0059	5405.52461(14)	2.15222181(58)	10(10)
WASP-3	45	10 + 16	1660	0.10630(60)	0.05704(19)	0.495(18)	1.52	0.00046	0.071	5325.825409(91)	1.84683487(28)	8.2(4.2)
WASP-4	50	17 + 19	4240	0.15488(32)	0.044915(53)	0.134(28)	1.22	0.00025	0.0093	5045.738469(23)	1.338231514(85)	−1.0(1.0)
WASP-5	15	8 + 3	1220	0.11477(79)	0.05037(23)	0.459(25)	1.71	0.0026	0.022	5896.57894(15)	1.62843056(75)	−16(19)
WASP-6	17	8 + 3	1460	0.14308(88)	0.05368(19)	0.220(51)	1.72	0.0067	0.68	5379.546159(80)	3.36100249(57)	−30(22)
WASP-12	203	78 + 58	9090	0.11839(17)	0.062367(49)	0.4316(48)	1.33	0.00015	0.018	6003.132366(25)	1.091420385(52)	−9.40(65)
WASP-52	44	19 + 10	1970	0.16530(53)	0.03877(12)	0.5965(65)	1.59	0.0013	0.11	6904.792887(62)	1.74978159(40)	−14(12)
XO-2N	54	24 + 14	3920	0.10358(38)	0.055930(92)	0.207(37)	1.56	0.00069	0.054	5212.405250(40)	2.61585963(16)	−2.5(4.3)
XO-5	12	3 + 0	595	0.1024(20)	0.06400(76)	0.560(48)	2.70	0.0015	0.0094	7802.37086(43)	4.1877642(29)	272(99)

The fitting uncertainties are given in parenthesis after each estimation, in the units of the last two figures. Most of the columns have the same meaning as in Table 4 from (Baluev et al. 2015)

[1] Defined as $\mathcal{Q}'_{\rm sum} = \sqrt{\sum_{i=1}^{N} Q_i'^2}$.

[2] The quadratic TTV ephemeris and the value of $\chi^2_{\rm TTV}$ do not include the Southworth et al. (2009a,b, 2010) DFOSC data, because they may be affected by clock errors.

[3] Number of lightcurves fitted with individual red noise term + number of lightcurves fitted with shared τ.

[4] Orbital eccentricity of ~ 0.15 is not taken into account, see Sect. 6 for a self-consistent fit.

[5] The realistic uncertainty also depends on the observed TTV scatter $\chi^2_{\rm TTV}$, likely inspired by the star activity, see text.

free parameters. However such an investigation is out of the scope of the present paper.

5.5 WASP-4: yet another TTV trend?

We suspected the nonlinear trend in WASP-4, analogous to WASP-12 one, right after the new EXPANSION lightcurves from 2017 observing season were processed. The magnitide of the trend corresponded to $T_{\rm d} \sim 10$ Myr (surprisingly close to what was recently claimed by Bouma et al. 2019). However, that time the trend interpretation depended on just a few data points from 2017. To confirm or retract the trend hypothesis we initiated in 2018 a prioritized observing campaign of WASP-4 within the EXPANSION project. By the end of 2018 we acquired 10 new transit lightcurves.

Table 3 shows the observation log, including the EXPANSION data, as well as a few older lightcurves found in the ETD and AXA databases. This table does not include data taken from the literature (29 lightcurves) and 6 archive lightcurves from the TRAPPIST-South telescope. The total number of WASP-4 lightcurves reprocessed in this work was 66 (plus one outlier not included in the final analysis). The trend information mainly comes from 14 observations made in 2017-2018, among them 10 were taken by P. Evans with a 36 cm Planewave CDK telescope equipped by a SBIG STT 1603-3 CCD and hosted at El Sauce Observatory, Chilie. This is a top-class amateur type equipment, and thanks to good astroclimate of the site the transit lightcurves were of a noble quality, and the corresponding TTV measurements appear in turn quite competitive with even TESS ones (which were released later).

The new EXPANSION data did not confirm the trend: our updated TTV time series became consistent with strictly linear ephemeris, so we decided that our trend hypothesis was wrong. Surprisingly, very recently Bouma et al. (2019) reported a detection of this trend based on the new TESS transit data, obtained practically simultaneously with our observations in the EXPANSION network. Moreover, the reported trend magnitude appeared almost the same as we suspected initially (before 2018).

To shed more light on this apparent controversy, we then performed additional analysis, including the TTV data published in the literature without lightcurves and the new TESS timings. This includes very accurate transit times derived from the transmission spectroscopy by Huitson et al. (2017), transit times by Hoyer et al. (2013), as well as transit times by (Wilson et al. 2008) and two early WASP timings given in (Gillon et al. 2009a). We did not use the HST spectral observations by Ranjan et al. (2014). These data might appear inaccurate because the spectra were partly overexposed and hence the flux measurements are likely not very reliable.

The full TTV time series is shown in Fig. 3. Now, with the new TESS transit times added, the quadratic term of the trend indeed appears significant, according to our analysis. However, we obtain a smaller magnitude and significance than Bouma et al. (2019) reported. The trend is still not detectable with the use of only the homogeneously derived por-

Table 3. Observation log for 31 of the WASP-4 transit lightcurves.

Obs. Date	Aperture [m]	Filter	Cadence [min]	Airmass	Observer
2008-09-23		Rc	0.46	1.26 ⟼ 1.03	Fernando Tifner (AXA)
2009-09-22		Rc	1.18	1.27 ⟼ 1.02	Fernando Tifner (AXA)
2010-07-08		Rc	1.02	1.11 ⟼ 1.19	Thomas Sauer (ETD)
2010-08-17	2.15	Ic	1.13	1.07 ⟼ 1.34	Eduardo Fernandez-Lajus, Romina P. Di Sisto
2010-10-04		Clear	0.46	1.02 ⟼ 1.15	Gavin Milne (ETD)
2010-11-01		Rc	0.41	1.03 ⟼ 1.38	TG Tan (ETD)
2010-11-05		Clear	0.52	1.13 ⟼ 1.85	Ivan Curtis (ETD)
2010-12-20	0.60	'I+z'	0.53	1.14 ⟼ 2.02	TRAPPIST
2011-09-15	0.60	'I+z'	0.33	1.62 ⟼ 1.04	TRAPPIST
2011-09-27	0.60	Ic	0.33	1.23 ⟼ 1.05	TRAPPIST
2011-10-21	0.60	Ic	0.33	1.04 ⟼ 1.61	TRAPPIST
2011-12-19	0.60	'I+z'	0.33	1.14 ⟼ 2.07	TRAPPIST
2012-06-07	0.60	Ic	0.35	1.41 ⟼ 1.03	TRAPPIST
2012-09-11	0.25	Clear	0.82	1.59 ⟼ 1.14	Phil Evans
2013-09-21		Clear	0.79	1.19 ⟼ 1.07	Colazo, C. Schneiter, E. M.
2013-10-07		Rc	0.30	1.41 ⟼ 1.05	Erin Miller (ETD)
2013-12-13		Rc	0.31	1.60 ⟼ 1.10	Erin Miller (ETD)
2014-08-03	2.15	Clear	3.17	1.10 ⟼ 1.04	Eduardo Fernandez-Lajus, Romina P. Di Sisto
2014-08-16		Clear	0.64	1.05 ⟼ 1.60	Martin Masek (ETD)
2014-08-20	2.15	Ic	6.04	1.01 ⟼ 1.27	Eduardo Fernandez-Lajus, Romina P. Di Sisto
2014-08-20	1.54	Clear	3.26	1.01 ⟼ 1.42	Carlos Colazo, Carolina Villarreal
2014-10-25	1.54	Rc	0.75	1.30 ⟼ 2.91	Cecilia Quinones
2015-08-15	2.15	Ic	3.05	1.02 ⟼ 1.22	Eduardo Fernandez-Lajus, Romina P. Di Sisto
2017-07-26	0.36	Rc	1.51	1.42 ⟼ 1.04	Phil Evans
2017-09-07	0.36	Rc	1.19	2.44 ⟼ 1.02	Phil Evans
2017-09-23	0.36	Rc	2.16	1.38 ⟼ 1.02	Phil Evans
2017-09-24	0.46	Clear	1.02	1.19 ⟼ 1.15	H. Durantini Luca, P. Baez, C. Colazo
2018-05-23	0.36	Rc	2.07	2.74 ⟼ 1.09	Phil Evans
2018-06-20	0.36	Rc	2.03	1.05 ⟼ 1.02	Phil Evans
2018-07-25	0.36	Rc	2.26	1.75 ⟼ 1.02	Phil Evans
2018-08-10	0.36	Rc	2.06	1.08 ⟼ 1.31	Phil Evans
2018-08-12	0.30	Rc	1.55	1.10 ⟼ 1.30	Carl R. Knight
2018-08-14	0.36	Rc	2.22	1.03 ⟼ 1.32	Phil Evans
2018-08-15	2.15	Ic	3.05	1.05 ⟼ 1.28	Eduardo Fernandez-Lajus, Romina P. Di Sisto
2018-08-22	0.36	Rc	2.12	1.02 ⟼ 1.51	Phil Evans
2018-08-26	0.36	Rc	2.06	1.02 ⟼ 1.58	Phil Evans
2018-10-14	0.30	Rc	1.61	1.10 ⟼ 1.29	Carl R. Knight

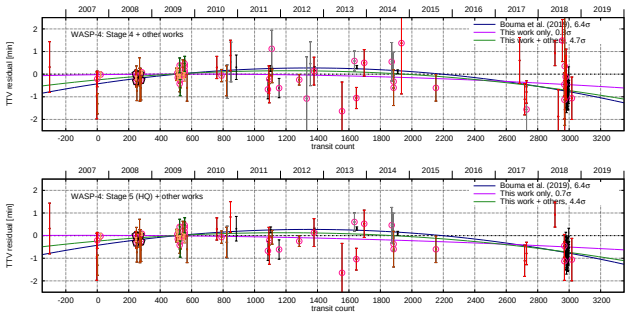


Figure 3. Transit times of WASP-4, including the homogeneous sample from this work (without DFOSC data possibly affected by clock errors), and the timing data published in literature without lightcurves. Top panel is for all the TTV data (Stage 4), bottom panel is for only HQ ones (Stage 5). Several models of the quadratic TTV trend are also plotted with the trend significance labelled in the legend (for the regularized noise model and simply merging heterogeneous TTV data).

tion of TTV data from this work. That is, the information about the trend comes mainly from the third-party observations rather than from our data release. By inspecting Fig. 3 we may suspect that the trend depends primarily on just the four high-accuracy timings provided by Huitson et al. (2017). The TESS timing does not in fact contradict anything and visually they are in a satisfactory agreement with what was obtained in the EXPANSION project in 2018.

However, justifying the trend detection based on just four data points, even accurate ones, might be quite dangerous. Digging into the details of the (Huitson et al. 2017) TTV data, they were based on just the *linear* limb-darkening model with a single coefficient. Although the authors ensured that based on some preliminary analysis their results (including fit uncertainties) did not change significantly for linear and for more complicated limb-darkening models, we remain to be concerned about this. Also, we could not find a clear confirmation in the text that the red noise was taken into account when fitting the lightcurves. Although it is mentioned that some 'systematics' are fitted, from the description given in the text this 'systematics' appears to be a deterministic parametric function rather than an autocorrelated random process.

In view of this we also notice that when analysing with our pipeline the similar transmission spectroscopy lightcurves for WASP-12 (Stevenson et al. 2014), we robustly detected a significant red noise. Inclusion of this red noise in the lightcurve model roughly doubled the derived transit timing uncertainties from ~ 3 sec to ~ 5 sec. Significant red noise was also detected in the WASP-52 transmission spectroscopy lightcurve from (Chen et al. 2017), though not found in the HD 189733 data by Kasper et al. (2019). The latter, however, revealed an anomalous transit time deviating by ~ 2 min from the quadratic TTV ephemeris (compared to an intrinsic statistical uncertainty of just about 12 sec).

Therefore, regardless of the undoubtfully outstanding quality of the lightcurves acquired and analysed by Huitson et al. (2017) within the context of this programme their TTV uncertainties might be understated. A public release of the Huitson et al. (2017) lightcurves is not available, so we did not reanalyse them in our pipeline. We therefore decided to investigate this issue using a different approach.

In practice, formally declared TTV uncertainties never appear entirely accurate: the actual scatter of TTV residuals may be systemtically different (usually larger). This may appear because of some subtle astrophysical effects (e.g. spot-transit events), additional instrumental errors, or shortcomings of the adopted models or analysis algorithm. However, different teams may process data quite differently, and hence each team might have its own bias in the reported TTV uncertainties. Therefore, different portions of such a heterogeneous TTV compilation may need to be weighted differently to balance this effect. However, those weights are not known to us a priori, so they need to be estimated from the TTV data 'on-the-fly', e.g. based on the actually observed scatter of the TTV residuals in each homogeneous portion.

We therefore separated all our TTV data into the following four more or less homogeneous classes: (i) the 'main' subset including transit timings derived in this work and three old timings given in (Gillon et al. 2009a) without public lightcurve data; (ii) the rich TTV subset by Hoyer et al. (2013); (iii) the four high-accuracy timings by Huitson et al. (2017); and (iv) the TESS timings from (Bouma et al. 2019). All these data sets should be treated separately in the sense that each should have an independently fittable noise parameter.

The noise in each of these data blocks was modelled by

one of two noise models discussed in (Baluev 2015), namely by (i) the multiplicative model, or (ii) the so-called regularized model. These 'noise models' represent a parametrized model for the variance of each TTV measurement, in which a single free parameter basically regulates the weight of the corresponding TTV data set as a whole. See the cited paper for further mathematical details and motivations.

Since this approach involves a separate and largely independent treatment of each TTV data set, we call this as 'separated' model of the TTV data. Such a model with parametrized noise can be fitted by using the maximum-likelihood method, as discussed in (Baluev 2009). In such a way the relative weighting of different TTV subsets is determined adaptively and basically tied to the corresponding TTV residuals RMS.

For a comparison, we also performed an analysis of the TTV data plainly merged into a single time series without any relative weighting. This analysis was also performed for two noise models: for the multiplicative and the regularized one.

The TTV trend itself was always modelled by the quadratic function (5) with three free coefficients.

As we expected, it appeared that the magnitude of the quadratic term and especially its derived uncertainty is sensitive to the choice of the noise model. In the case of a 'separated' model the trend uncertainty gets increased. This means that by allowing some TTV data to be actually less accurate than stated, the significance of the trend may reduce. For example, it may reduce if the four Huitson et al. (2017) transit times are less accurate than formally stated. And because of a small number of these data (just the four), their RMS does not accurately constrain the noise level, so this level can be varied relatively freely.

In Fig. 4, we demonstrate this effect in the shape of the likelihood function $\mathcal{L}$. For this goal we consider the log-likelihood-ratio statistic Z determined in accordance with (Baluev 2009). We compute (i) the global maximum of the likelihood function $\mathcal{L}_{\max}$ with respect to all the noise parameters and all three TTV trend coefficients, and (ii) the value $\mathcal{L}'_{\max}(q)$ maximized with respect to all parameters except for the quadratic coefficient $dP/dn = P^2 T_{\rm d}^{-1} = qP^2$, where $q = T_{\rm d}^{-1}$ is treated as independent variable. The quantity $Z(q) = \log[\mathcal{L}_{\max}/\mathcal{L}'_{\max}(q)]$ therefore indicates whether the given q is statistically consistent with the best-fitting value $\hat{q}$ which corresponds to the global maximum of $\mathcal{L}$. We always have $Z(q) \geq 0$, and the larger is Z, the more significant is deviation of q from $\hat{q}$ and the less consistent with the data this q is. If our models are linearizable than $Z(q)$ should have an almost parabolic shape with a single minimum at $\hat{q}$.

We use two approaches to calibrate the levels of $Z(q)$, both rely on the assumption that the model is linearizable and hence $Z(q)$ is quadratic (while the likelihood ratio $\exp(-Z(q))$ is Gaussian). The first approach is the frequentist χ^2 test, and the second one is the Bayesian Information Criterion (BIC).

In this frequentist treatment, the significance level of a given Z is approximately the $\chi^2_1(2Z)$, or the χ^2-distribution with $d = 1$ degree of reedom (one degree because we have just one free parameter left in $Z(q)$). This would simply mean that the significance level for a given q would correspond to $\sqrt{2Z}$ in the n-sigma notation, or vice versa, any n-sigma significance level would correspond to the threshold level $Z(q) = n^2/2$.

The BIC is defined as $\mathrm{BIC} = 2\log\mathcal{L} - k\log N$, where k is the total number of free parameters in the model, and N is the number of observations (number of transit times). To compare different models with k_1 and k_2 parameters we use the difference $\Delta\mathrm{BIC} = 2Z - d\log N$ with $d = k_1 - k_2 = 1$ in our case. Hence, the significance threshold for $Z(q)$ becomes $Z = (\Delta\mathrm{BIC} + \log N)/2$. Here ΔBIC is deemed to be an input parameter determining the requested significance level (typical practical values are 2, 4, 6, 10).

We especially notice that the value $Z(0)$ indicates the significance of the nonlinear trend itself (i.e. how much $q = 0$ is consistent with the TTV data, within the adopted noise model).

In Fig. 4, we plot this statistic $Z(q)$ for three data compilations, including (i) only the homogeneous data from this work, (ii) all TTV data, (iii) all TTV data excluding (Huitson et al. 2017), and for all our noise models, including (i) the plain merging of heterogeneous datasets and (ii) adaptive merging of heterogeneous datasets with individually fittable noise parameters. For each of these model layouts we adopt either a multiplicative or regularized noise model, defined in (Baluev 2015).

As we can see, the shape of the likelihood function may change a lot depending on the model and TTV data involved. We can draw the following conclusions from these graphs:

(i) The homogeneously derived TTV data from this work does not support the existence of any quadratic trend. The data are consistent with linear a ephemeris below 1-sigma level.

(ii) Simultaneously, the value of $T_{\rm d} = 9.2$ Myr from (Bouma et al. 2019) seems to be poorly consistent with our homogeneous TTV subsample, at the level above 3-sigma in terms of the χ^2 test or with $\Delta\mathrm{BIC} \sim 4-6$ (depending on the model). We believe this may appear, at least in some part, because Bouma et al. (2019) did not take into account the heterogeneous nature of the TTV data they used, plainly merging them into a single time series.

(iii) Joining our data with the remaining third-party TTV measurements allows us to refine the localization of the parameter q greatly and even suggests that this q can be significantly non-zero. However, the significance of this conclusion, as well as possible confidence ranges for q appear very model dependent. If we plainly merge all the TTV data, we obtain that q is inconsistent with zero at the high $\sim 5-7$-sigma level. But using our adaptive separated noise model, this significance drops to merely $3-4$-sigma.

(iv) The shape of the likelihood function becomes significantly non-parabolic in the case of our adaptive separated noise model. This indicates that this model may be too much non-linear and therefore our significance estimates may appear inaccurate. It may even appear that the significance of the trend is reduced even further below the $3-4$-sigma level mentioned above.

(v) The most trustable and model-stable behaviour appears when we remove the TTV data by Huitson et al. (2017) from the analysis. In this case $Z(q)$ behaves nicely as a parabolic function, indicating an almost-linear model and nearly Gaussian likelihood. In this case, the quadratic trend

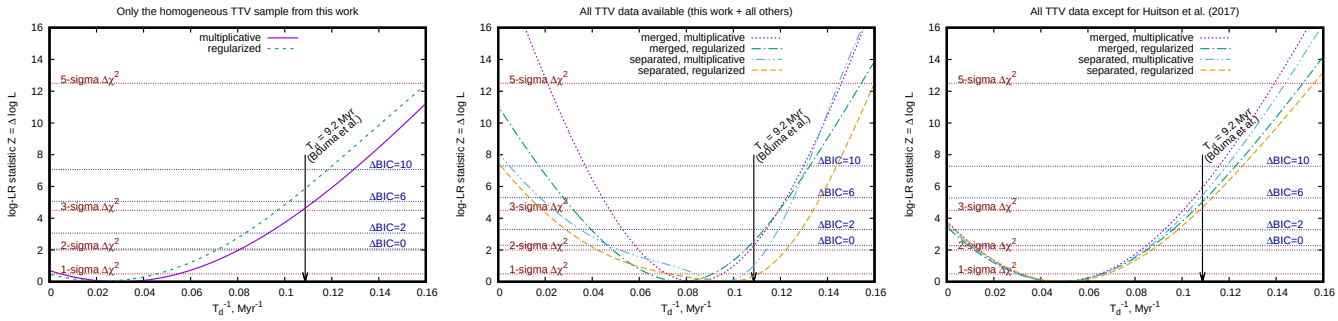


Figure 4. Logarithm of the likelihood ratio statistic for WASP-4, $Z(q)$, as a function of $q = T_{\rm d}^{-1}$. Three graphs correspond to different compilations of TTV data (homogeneous from this work / all / all but Huitson et al. (2017)). The curves within each graph correspond to different models of the TTV noise (merged/separated and multiplicative/regularized). In each graph a set of the significance threshold levels is also shown, corresponding to the frequentist χ^2 test or to the Bayesian information criterion (BIC). See text for more details and a discussion.

has the significance 2.8 − 3-sigma or ΔBIC $\sim$ 4, which is very remarkable but still needs further confirmation by more observations. The magnitude of the best fitting trend then becomes $T_{\rm d} \sim 20$ Myr with large uncertainty. Curiously, the value of $T_{\rm d} = 9.2$ Myr given by Bouma et al. (2019) appears in this case even less likely than the no-trend model ($T_{\rm d} = \infty$).

(vi) In any case, the magnitude of the trend is very uncertain, and its putative confidence ranges appear very asymmetric and non-Gaussian in the separated noise model. The value $T_{\rm d} = 9.2$ Myr given in (Bouma et al. 2019) looks more like a lower limit on $T_{\rm d}$, while the actual value may reach even ~ 100 Myr, given the large uncertainty of this parameter.

Therefore our general conclusion on WASP-4 is that its putative TTV trend magnitude and the detection significance are severely model-dependent. They solely depend on how do we treat the heterogeneous nature of the TTV data, in particular on the (Huitson et al. 2017) data. Moreover, as recognized by Bouma et al. (2019), $T_{\rm d}$ as small as 9.2 Myr is inconsistent with theoretical predictions of the tidal quality parameter. Given our discussion, we believe that it is too early to definitely claim the detection of this trend until more homogeneous TTV data are collected. At least, it is too early to claim that this object breaks any theory predictions. However, WASP-4 remains a very interesting target that may indeed hide serendipitous discoveries. In 2019, we will continue to ascquire more data in the EXPANSION project for WASP-4.

6 SELF-CONSISTENT ANALYSIS OF RADIAL VELOCITY AND TRANSIT DATA

This analysis here is relatively preliminary and rather simplistic. A more careful analysis of our RV data, including e.g. the comprehensive search of residual RV variations in our targets, is sheduled for a future work, while our goals here are:

(i) Derive a more complete set of parameters in a self-consistent model, in particular planetary masses and physical radii (rather than merely the planet/star radii ratio).

(ii) Derive a more realistic fit of GJ 436 b, taking into account its significant orbital eccentricity.

(iii) For planets with a (possible) TTV trend, WASP-12 and possibly WASP-4, test whether the trend could appear through the effect of variable light arrival delay, caused by the gravity of a distant unseen companion. In such a case the quadratic TTV trend would simultaneously appear in radial velocities as a linear trend (a radial acceleration).

(iv) Derive the rotation parameters of the stars via the Rossiter-McLaughlin (hereafter RM) effect, and test how much it is sensitive to the correction coeffiecients suggested in (Baluev & Shaidulin 2015).

Notice that even the combination of transit and radial velocity data does not allow to determine the star mass from a self-consistent fit. The informations about the star mass has to be derived from astrophysical models of stellar spectra, e.g. based on stellar evolutionary tracks. Such models in fact provide certain constraint on the star mass $M_\star$ and radius $R_\star$ that can be used to provide an entirely self-consistent fit. However, in this work we were more interested to estimate the uncertainties inferred by the transit and radial velocity data, so we still prefer not to mix them with the uncertainties of astrophysical models that may also contain an additional systematic error.

Therefore, we adopted the 'reference' values of $M_\star$ for our ten targets, as given in Table 4. We did not take into account the uncertainties of $M_\star$ when computing our fits. In case if the adopted $M_\star$ is wrong, the fit can be easily rebased to another $M'_\star$ based on the following simple laws:

$$\begin{aligned} R'_\star &= R_\star \left(M'_\star/M_\star\right)^{\frac{1}{3}}, \\ r'_{\rm pl} &= r_{\rm pl} \left(M'_\star/M_\star\right)^{\frac{1}{3}}, \\ a' &= a \left(M'_\star/M_\star\right)^{\frac{2}{3}}, \\ m'_{\rm pl} \sin i' &= m_{\rm pl} \sin i \left(M'_\star/M_\star\right)^{\frac{1}{3}}, \\ \cos i' &= \cos i \left(M'_\star/M_\star\right)^{-\frac{1}{3}}. \end{aligned} \qquad (7)$$

The first formula of this list comes from the known property that a transit lightcurve fit actually constrains the star density $\rho_\star \propto M_\star R_\star^{-3}$, rather than $R_\star$ or $M_\star$ separately (Mandel & Agol 2002). The second one appears because the transit data actually constrain the ratio $r_{\rm pl}/R_\star$, so the scale

Table 4. Star masses adopted in the joint transit+RV fits.

host star	$M_\star[M_\odot]$	reference
Corot-2	0.97(6)	Alonso et al. (2008)
GJ436	$0.452\left(^{+14}_{-12}\right)$	Torres et al. (2008)
HAT-P-13	$1.22\left(^{+5}_{-10}\right)$	Bakos et al. (2009)
HD189733	0.806(48)	Torres et al. (2008)
TrES-1	$0.878\left(^{+38}_{-40}\right)$	Torres et al. (2008)
WASP-2	$0.84\left(^{+11}_{-12}\right)$	Triaud et al. (2010)
WASP-3	$1.24\left(^{+6}_{-11}\right)$	Pollacco et al. (2008)
WASP-4	$0.930\left(^{+54}_{-53}\right)$	Triaud et al. (2010)
WASP-5	$1.000\left(^{+63}_{-64}\right)$	Triaud et al. (2010)
WASP-6	$0.88\left(^{+5}_{-8}\right)$	Gillon et al. (2009b)
WASP-12	$1.434\left(^{+110}_{-90}\right)$	Collins et al. (2017)
XO-2N	0.96(5)	Damasso et al. (2015)
XO-5	0.88(3)	Pál et al. (2008)

law of $r_{\rm pl}$ is the same as for $R_\star$. The third and the fourth formulae follow from the basic properties of the Doppler method and can be found e.g. in (Baluev 2013). The last one for $\cos i$ follows because i is constrained by only the transit data, via the impact parameter $b = a\cos i/R_\star$, so the scale law for $\cos i$ corresponds to $R_\star/a$. The last two formulae can be combined together to obtain

$$m'_{\rm pl} = m_{\rm pl}\frac{(M'_\star/M_\star)^{\frac{1}{3}}}{\sqrt{1+\left(1-(M'_\star/M_\star)^{-\frac{2}{3}}\right)\operatorname{cotan}^2 i}}. \qquad (8)$$

Since $\operatorname{cotan} i$ is below 0.15 for all our targets, the term beneath the square root represents only a negligible correction.

In fact, some scale formulae above are not entirely accurate, neglecting certain second-order corrections, but since the values of $M_\star$ for all our targets are now restricted to quite narrow ranges, the practical accuracy of (7,8) is satisfactory.

To perform the self-consistent analysis we used for the transit data basically the same model as above, except that the planet motion was assumed Keplerian rather than circular. The radial velocity for each target was modelled by the Keplerian curve plus a linear trend (to account e.g. for possible long-period companions in the system). We also include a quadratic term in the planetary longitude, to take into account possible TTV trends, see (Baluev 2018).

For many of our targets the RV data contained substantial in-transit runs. They were obviously aimed to detect the RM effect. For these targets we therefore included in our compound RV model the RM effect based on the approach by Baluev & Shaidulin (2015). To accurately approximate this effect, we must know some effective values of the limb-darkening coefficients $A_{\rm RV}$ and $B_{\rm RV}$, corresponding to the spectral range used to extract the Doppler information. Also, we need to specify two correction coefficients ν and μ that depend on the average characteristics of spectral lines and on the method used to derive the radial velocity from the spectrum. But unfortunately, these four quantities are very difficult to derive reliably from the spectra. Instead, it is reasonable to treat them as fittable parameters of the RV model. However, in such a case the model would appear nearly degenerate, because as discussed in (Baluev & Shaidulin 2015), the parameters ν and μ are strongly correlated with A and B: by varying any pair of coefficients we may achieve an accurate fit with almost the same RV curve.

We therefore adopted the following hybrid approach. First, we assumed that $A_{\rm RV}$ and $B_{\rm RV}$ are equal to the corresponding values of the photometric data obtained with a clear aperture. After the values $A_{\rm RV}$ and $B_{\rm RV}$ were constrained in such a way, we considered ν and μ separately for different instruments (like HARPS, HARPS-N, etc.) This should take into account possible corrections of the RM effect, jointly with possible inaccuracies of the limb-darkening coefficients, on the per-instrument basis.

When computing the fits, we of course treated separately the RV data obtained at different instruments. Moreover, if the RV data belonging to the same instrument contained in-transit pieces, we separated these portions of the data from each other and treated them as individual RV data sets. This makes the model more adequate, because e.g. the scatter of the RV residuals within each such short run covering just a few hours is significantly smaller than for the entire dataset covering years. This is basically the effect of the red noise in the RV data. Also, the red noise may result in a small individual offset of each such in-transit run.

We also found several compact series of out-of-transit runs covering just a single night. Such portions of the RV data were treated separately too, with an individual offset and individual noise parameters.

After performing a preliminary fit of the joint transit+RV model described above, we runned the red-noise detection algorithm described above for transit data, but now we extended it to all the RV data sets as well. This yielded our the final self-consistent fits per each of ten targets.

For Corot-2, WASP-4, and WASP-12 we computed an additional alternative fit without splitting the RV data belonging to the same instrument. In this case, the RV offsets between different compact runs were taken into account implicitly, via a single red noise model of the merged data set.

All the analysis was performed with the PlanetPack software of version 3 (Baluev 2018).

We separate our results in two parts: Table 5 gives some most important non-planetary parameters of our fits, and Table 6 contains only planetary parameters. The tables are presented here in a reduced form; their full versions that include e.g. RM correction coefficients can be found in the online-only supplement. We do not give the full set of our fit parameters, since they would be to long for a print presentation. In particular, we omit all the white and red noise parameters, and individual limb-darkening parameters for each lightcurve, since there are multiple (up to 14) RV datasets and tens or even hundreds lightcurves for each target.

First of all, we notice that the RM correction coefficients ν and μ are usually consistent with zero, given their uncertainties. We found only the following targets convincingly demonstrating significant nonzero values: Corot-2 (μ for the HARPS RV data), HD 189733 (ν for Keck/HIRES and SOPHIE, and μ for HARPS, HARPSN, and Keck/HIRES), WASP-5 (μ for HARPS), and possibly GJ 436 (ν for GJ 436 HARPS, HARPSN, Keck/HIRES). In theory, ν should usually be zero, since most our RV data were derived with

Table 5. Self-consistent fits of transit and radial velocity data: stellar parameters and RV trends.

host star	rad. accel. c_1 [m/s/yr]	star radius $R_\star$ [$R_\odot$]	rotation vel. $v \sin i$ [m/s]	spin-orbit ang. λ [°]
Corot-2	−74(11)	0.925(19)	8930(400)	3.1(6.6)
Corot-2'	−0.5(3.1)	0.885(29)	8570(430)	3.0(8.2)
GJ436	0.02(11)	0.4266(89)	1800(1000)	354.5(4.9)
HAT-P-13	17.2(1.0)	1.691(25)	2300(1100)	359.5(5.5)
HAT-P-13'	17.56(85)	1.692(25)	2600(1300)	0.0(5.4)
HD189733	−1.07(75)	0.7583(31)	2280(110)	0.08(36)
TrES-1	−1.9(1.4)	0.819(13)	—	—
WASP-2	−3.9(1.8)	0.8200(82)	960(850)	345(18)
WASP-3	−7.9(2.7)	1.337(28)	—	—
WASP-4	−0.7(1.1)	0.9029(43)	1890(220)	344(16)
WASP-4'	−0.8(1.2)	0.9045(46)	1890(220)	343(16)
WASP-5	−0.03(78)	1.127(16)	2470(400)	0.2(6.4)
WASP-6	−420(260)	0.829(23)	1620(130)	352(11)
WASP-12	−5.4(2.0)	1.657(12)	600(1200)	9(48)
WASP-12'	−7.5(2.2)	1.660(16)	1300(1500)	75(30)
XO-2N	−1.2(1.2)	0.986(14)	—	—
XO-5	1.3(1.2)	1.068(51)	—	—

The fitting uncertainties are given in parenthesis after each estimation, in the units of the last few figures. The star mass from Table 4 was assumed constant here and its uncertainties were not included in the fit. Using the full set of transits for GJ436, while only the HQ (Stage 5) transits for other targets. A stroke stands for an alternative fit computed without splitting the RV data belonging to the same instrument (see text).

TERRA, which is a kind of a spectrum modelling method (Anglada-Escudé & Butler 2012). A nonzero ν may appear only if the RV data were derived by the cross-correlation technique, and simultaneously the spectral lines have some asymmetry on average (Baluev & Shaidulin 2015). Since this is not the case, a few significantly nonzero ν estimates may indicate that the adopted limb-darkening coefficients $A_{\rm RV}$ and $B_{\rm RV}$ are inaccurate for the relevant RV dataset. In such a case, the value of $v \sin i$ and λ may involve an additional bias which is difficult to estimate without a better guess for $A_{\rm RV}$ and $B_{\rm RV}$.

Even though we did not detect here very many occurences of significantly nonzero RM correction coefficients, it is still important to preserve them in the RV model as free parameters, in order to have a more realistic (increased) uncertainty in $v \sin i$ and λ, as well as in other RV-derived parameters, like e.g. the RV trends.

Now let us consider the putative RV trends (radial accelerations). They are modelled by a linear function

$$\mathrm{RV} = c_0' + c_1(t - t_0), \tag{9}$$

where c_0' is a dataset-specific RV offset, and c_1 is the trend coefficient (common for all datasets).

If not an instrumental effect or some long-term astrophysical variation, such a trend can be explained through a graviational effect from a distant unseen companion. But such a companion would also affect the observed transit times via the variable light arrival delay. Basically, the light arrival delay is determined by the position of the tight star-planet system along the line of sight, which can be obtained by a plain integration of the radial velocity. Therefore, the transit time delay (TTD) effect corresponding to (9) would be a quadratic trend:

$$\begin{aligned} \mathrm{TTD} &= \mathrm{const} + \frac{c_0}{\mathcal{C}}(t - t_0) + \frac{c_1}{\mathcal{C}}\frac{(t-t_0)^2}{2} \\ &= \mathrm{const} + \frac{P c_0}{\mathcal{C}}(n - n_0) + \frac{P^2 c_1}{\mathcal{C}}\frac{(n-n_0)^2}{2}, \end{aligned} \tag{10}$$

where $\mathcal{C}$ is the speed of light. This formula becomes mathematically identical to the quadratic TTV trend (5). But now the second term in (10) with c_0 (the absolute RV) represent a minor Doppler-like correction of the orbital period P, while c_1 is basically the same TTV effect as observed for WASP-12. We simply have $T_{\rm d} = \mathcal{C}/c_1$. Taking into account the adopted measurement units,

$$T_{\rm d}[\mathrm{Myr}] \simeq -\frac{300}{c_1[\mathrm{m\,s^{-1}\,yr^{-1}}]}. \tag{11}$$

The minus sign appears because positive RV trend means increasingly late light arrival (apparent delay of the transits, as if the planet was spiraling out).

We found the statistically significant RV trend in the Corot-2, about -74 ± 11 m/s per year. However, this trend can also be a hidden red-noise effect, because it disappears in the alternative fit without splitting the HARPS data per subsets. One of the HARPS subsets appeared to have just seven observations, so the "splitted" fit might appear rather unreliable statistically. Finally, we found several evident RV outliers for Corot-2, possibly indicating some hidden anomalies in its spectrum that inferred potential inaccuracies of the derived Doppler information. Therefore, we conclude that the suspected radial acceleration in Corot-2 remains controversial and needs further confirmation.

For Corot-2, the formula above implies $T_{\rm d} = (4.0 \pm 0.6)$ Myr, or $q = T_{\rm d}^{-1} \sim (0.25 \pm 0.04)$ Myr^{-1}. The current TTV data are unable to reliably detect such a trend. We obtain $q \sim -0.1 \pm 0.1$ from the Corot-2 transit times. The difference from the RV estimate is inconclusive given the high level of model dependency observed for this target. Nonetheless, Corot-2 remains to be an interesting target for further monitoring, because it may provide the first detection of the unseen object simultaneously by the RV Doppler and TTV Roemer effects.

Contrary to the Corot-2, in WASP-12 we detect a more model-stable radial acceleration about $-(5-8)$ m/s/yr. The trend information mainly comes from the rich set of SOPHIE data available for this target (> 100 points). The RV trend looks rather reliable given its uncertainty. This value corresponds to $T_{\rm d} \sim 40 - 60$ Myr, so the "observed" $T_{\rm d}$ need to be cleaned from the nontidal portion. This is performed in Sect. 7 below.

For WASP-4, the RV trend estimation is consistent with zero, but given the uncertainty the value of c_1 may be as small as -4 m/s/yr (the three-sigma limit). This infers a TTV trend with $T_{\rm d} \sim 70$ Myr, which is consistent with the observed transit times within two sigma (see third panel in Fig. 4). Therefore, we cannot conclusively rule out or confirm the tidal nature of the putative WASP-4 TTV trend, even if this trend exists. More Doppler observations are needed to rule out the Roemer effect interpretation, or to determine in which fraction this TTV trend can be explained so. Using the same method as for WASP-12 (see Sect. 7), we obtain for WASP-4 an estimate of the tidal portion of the TTV trend of $q^{\rm tidal} = (0.019 \pm 0.025)$ Myr^{-1}, again a statistically insignificant value (formally, this corresponds to $T_{\rm d}^{\rm tidal} = (54 \pm 73)$ Myr).

Yet another possible candidate with radial acceleration is WASP-3 (-8 m/s/yr), but this is based on just 13 RV measurements. Also, HAT-P-13 has an RV trend about 18 m/s/yr, established long ago (Winn et al. 2010). Both the cases cannot be currently verified by TTV data, because of the latters' insufficient amount and accuracy.

Concerning the search of possible periods in the RV residuals, we leave this for a future work. But here we confirm that Corot-2, HAT-P-13, WASP-4, and WASP-12 re-

Table 6. Self-consistent fits of transit and radial velocity data: planetary parameters.

host star	planet mass $m_{\rm pl}$ [$M_{\rm Jup}$]	planet radius $r_{\rm pl}$ [$R_{\rm Jup}$]	orbital period[1] P [d]	TTV trend[1,2,3] $T_{\rm d}$ [Myr]	mean longitude[1] l [°]	inclination i [°]	eccentricity e	pericenter arg. ω [°]
Corot-2	2.744(51)	1.523(36)	1.74299686(44)	−11.2(9.4)	359.4(1.6)	88.15(56)	0.042(18)	60(18)
Corot-2'	2.85(10)	1.455(51)	1.74299682(45)	−10.4(8.2)	0.7(2.4)	88.32(52)	0.013(26)	320(130)
GJ436	0.06896(56)	0.3581(97)	2.64389856(36)	7.1(3.2)	335.82(44)	86.83(10)	0.1666(57)	324.9(2.4)
HAT-P-13 b	0.8528(59)	1.465(30)	2.9162384(17)	−2.42(86)	158.64(41)	82.12(22)	0.0126(46)	219(29)
c	14.17(28)	—	446.32(24)	—	63.99(18)	—	0.6621(58)	175.28(37)
HAT-P-13' b	0.8532(53)	1.466(30)	2.9162383(17)	−2.41(85)	158.58(35)	82.10(22)	0.0117(41)	218(28)
c	14.16(25)	—	446.29(21)	—	63.99(16)	—	0.6614(50)	175.28(32)
HD189733	1.1542(74)	1.1840(52)	2.218575123(57)	28(12)	20.45(38)	85.712(36)	0.0028(38)	62(64)
TrES-1	0.6967(82)	1.122(22)	3.03006960(18)	−5.7(9.5)	298.67(49)	88.69(28)	0.003(12)	263(86)
WASP-2	0.8711(73)	1.087(14)	2.15222163(67)	−12(11)	214.37(25)	84.82(10)	0.0134(56)	253(11)
WASP-3	1.982(49)	1.419(32)	1.84683480(30)	−11.1(5.5)	274.88(83)	84.24(32)	0.010(15)	41(74)
WASP-4	1.1949(65)	1.3915(82)	1.338231501(75)	47(45)	235.80(26)	88.63(30)	0.0068(35)	258(20)
WASP-4'	1.1976(68)	1.3940(86)	1.338231501(75)	47(45)	235.87(26)	88.63(30)	0.0053(38)	247(28)
WASP-5	1.5351(80)	1.294(25)	1.6284311(14)	5.2(6.3)	343.01(32)	84.57(33)	0.0086(46)	66(22)
WASP-6	0.458(20)	1.175(35)	3.36100264(65)	14(14)	32.6(2.4)	89.00(36)	0.036(24)	116(36)
WASP-12	1.422(14)	1.953(15)	1.091421080(96)	3.46(24)	37.00(51)	81.86(16)	0.0259(74)	250(11)
WASP-12'	1.413(15)	1.956(20)	1.091421078(96)	3.47(24)	36.82(49)	81.96(18)	0.024(11)	252(11)
XO-2N	0.5924(68)	1.017(16)	2.61585963(16)	80(140)	198.72(48)	88.33(25)	0.008(13)	91(29)
XO-5	1.050(15)	1.061(67)	4.1877477(38)	−1.82(68)	83.5(1.2)	86.82(51)	0.009(12)	200(94)

Same comments as in Table 5 also apply here.
[1]These parameters refer to $T_0 = 2455197.5$ (1 Jan, 2015) in the BJD TDB system.
[2]Except for WASP-12 case, the uncertainties for $T_{\rm d}$ are rather formal here, because this parameter becomes very nonlinear and nongaussian whenever it is comparable to the uncertainty. The linear parameter is $q = 1/T_{\rm d} = -\dot{P}/P$ with the uncertainty $\sigma_q = \sigma_{T_{\rm d}}/T_{\rm d}^2$.
[3]The realistic uncertainty in $T_{\rm d}$ also depends on the observed TTV scatter $\chi^2_{\rm TTV}$ from Table 2, see Section 5.3.

veal no hints of such signals, so the corresponding RV trend estimations should not be affected by possible periodic signals. Also, no residual variations were detected, including trends, in the updated RV data for GJ 436 (though a spurious RV trend might be suspected from the older data release by Baluev et al. 2015).

7 TRULY TIDAL PORTION IN THE WASP-12 TTV TREND

According to (Bechter et al. 2014), WASP-12 is a member of a triple star system. It has a stellar companion which appears as binary itself. The total mass of these two components B and C is $0.75 M_\odot$, and they may induce the radial acceleration of up to 0.33 m/s/yr on the primary star (Bailey & Goodman 2019).

Therefore, the radial acceleration of −7.5 m/s/yr would definitely belong to some other unseen companions, possibly other distant planets, or brown dwarfs, or even to an unresolved cool star.

Whatever object or multiple objects induced this radial acceleration, they should also induce a quadratic TTV trend according to the formula (10). Notice that even if the RV trend was not significant at all, we should include it in the model in order to have a more realistic uncertainty in the tidal part of $T_{\rm d}$. Our RV trend estimation implies the same sign as the observed cumulative value of dP/dn, so the tidal portion is smaller than the total observed TTV. To determine how much it is smaller, and how uncertain this value is, we perform a brief additional calculation.

First of all, we notice that information about the RV and TTV trends comes from qualitatively different observations: either solely from RV data, or solely from transit timings. Therefore, we may expect that these two quantities are practically uncorrelated (even though some negligible correlation may appear because e.g. the planetary orbital period P is shared between the RV and transit data). Thanks to this property, the corresponding trend coefficients may be treated independently from each other.

Concerning the TTV trend, its magnitude is defined by $q = T_{\rm d}^{-1}$, where $T_{\rm d}$ from Table 6 is (3.47±0.24) Myr. Here we notice again that this uncertainty is likely underestimated by the factor of $\sqrt{\chi^2_{\rm TTV}}$ from Table 2, or 1.33. Therefore, a more realistic estimate is $T_{\rm d} = (3.47 \pm 0.32)$ Myr.

Concerning the radial acceleration c_1, it is estimated in Table 5. Based on the Corot-2 example, we believe that the alternative fit WASP-12' is more realistic in what concerns the RV trend, because in the basic fit (for splitted RV data) the estimated radial acceleration may also include a local red noise effect appearing within individual short in-transit runs. We do not expect that the estimated c_1 value may have any significant escaped random uncertainty, since the stellar activity was already modelled through the white and red RV jitter. Based on the formula (11), from $c_1 = (-7.5 \pm 2.2)$ m/s/yr we obtain the apparent orbital decay time $T_{\rm d}^{\rm RA} = (40 \pm 12)$ Myr.

The tidal part of the TTV trend is therefore defined as the following difference:

$$\frac{1}{T_{\rm d}^{\rm tidal}} = \frac{1}{T_{\rm d}} - \frac{1}{T_{\rm d}^{\rm RA}}. \qquad (12)$$

Since $T_{\rm d}$ and $T_{\rm d}^{\rm RA}$ are practically uncorrelated, this formula yields an estimation $T_{\rm d}^{\rm tidal} = (3.80 \pm 0.40)$ Myr. Therefore, the truly tidal orbital decay time is by ~ 10 per cent larger than the observed (apparent) $T_{\rm d}$, and also it has a larger uncertainty. Although the bias appeared merely about 1σ, the increase of the uncertainty is noticable.

In this work we did not consider the alternative explanation of the WASP-12 TTV via the apsidal precession (Patra et al. 2017), although it is obvious that this model should also be corrected for the nontidal part of the TTV.

8 CONCLUSIONS AND DISCUSSION

The main conclusion of our work is that TTV data present in published literature are significantly inhomogeneous. Therefore, they cannot be plainly merged with each other before a processing. Substantial efforts must be made to increase the degree of homogeneity of the TTV data by reprocessing the archival transit lightcurves, or by employing sophisticated

noise models separating different TTV subsets from each other. Incareful analysis of heterogeneous TTV data may prove dangerous, potentially leading to spurious analysis artifacts. In view of this it appears necessary to always verify our conclusions against different data models, in order to gain an impression of how much they are model-dependent.

In what concerns the particular cases of WASP-12 and WASP-4, we notice that the Roemer effect induced by additional companions may cause biased interpretation of the results. In particular, it may bias the apparent value of the TTV trend, even mimicing to tidal phenomena. The Doppler observations a crucial in verifying such interpretations. In particular, they may help to assess a more realistic uncertainty of the truly tidal part of a TTV trend, based on the Doppler-derived estimation of the RV trend. Notice that even if the latter estimation is consistent with zero it is important to keep it as a free parameter of the fit, because its uncertainty increases the uncertainty in the tidal portion of a TTV.

When performing the lightcurve analysis, the primary nuisance effect that currently remains rather poorly modelled is the spotting activity that implies anomalies in the transit curve inducing additional TTV noise. This additional noise is problematic because currently it is difficult to predict or assess, except through the post-hoc estimation of the TTV scattering. Various self-consistent fits that avoid using TTVs as an intermediate data do not include, as a rule, the spot-transit effect in their uncertainties. To further improve their quality, we need to spend more efforts to reducing the spot-transit event, e.g. perform some kind of their automated detection followed by a dedicated fitting or removing the associated piece of a lightcurve. In our work the red noise model did not appear effective in removing the effect of spot-transit anomalies, since from the statistical point of view such anomalies may behave closer to a blue noise rather than red.[5]

Concerning the coefficients of the limb-darkening models by Claret & Bloemen (2011), based on our empiric verification we conclude that their FCM versions are rather accurate, though there is a detectable remaining bias in the coefficient B of about 0.1. The LF version of these coefficients is rather poor for practical use, as well as the older coefficients from (Claret 2000, 2004).

ACKNOWLEDGEMENTS

The observational campaign within the EXPANSION project (E.N. Sokov, I.A. Sokova), the programming and data analysis work (R.V. Baluev), and collection of literature data (V.Sh. Shaidulin) were supported by the Russian Foundation for Basic Research grant 17-02-00542 A. R.V. Baluev also acknowledges the support by the Russian Academy of Sciences programme No. 12, *Questions of the Universe origins and evolution*, for organizing the computing simulations.

Students of Horten Upper Secondary School (Horten Videregående Skole) H. Fjeldstad, S. Herstad, D.N. Lønvik, I. Oknes, and W. Zhao are acknowledged for taking part in observing WASP-12 b on the Nordic Optical Telescope.

E. Pakstiene acknowledges support from the Research Council of Lithuania (LMT) through grant 9.3.3-LMT-K-712-01-0103.

Administration of Ulugh Beg Astronomical Institute (Tashkent, Uzbekistan), M.A. Ibrahimov, R.G. Karimov, and I.M. Asfandiyarov acknowledge financial support and exchange visitor support from Max Planck Institute for Astronomy (Heidelberg, Germany).

[5] Notice that blue noise, just like white noise, has an infinite total power and is therefore a mathematical abstraction. Here under a blue noise we in fact mean some *banded* noise that possibly can be approximated by a blue-noise model.

REFERENCES

Agol E., Fabrycky D. C., 2017, in Deeg H. J., Belmonte J. A., eds, , Handbook of Exoplanets. Springer, Cham
Albrecht S., et al., 2012, ApJ, 757, 18
Alonso R., et al., 2008, A&A, 482, L21
Anglada-Escudé G., Butler R. P., 2012, ApJS, 200, 15
Bailey A., Goodman J., 2019, MNRAS, 482, 1872
Bakos G. A., et al., 2006, ApJ, 650, 1160
Bakos G. A., et al., 2009, ApJ, 707, 446
Baluev R. V., 2008, MNRAS, 385, 1279
Baluev R. V., 2009, MNRAS, 393, 969
Baluev R. V., 2013, Astron. & Comp., 2, 18
Baluev R. V., 2015, MNRAS, 446, 1493
Baluev R. V., 2018, Astron. & Comp., 25, 221
Baluev R. V., Shaidulin V. S., 2015, MNRAS, 454, 4379
Baluev R. V., et al., 2015, MNRAS, 450, 3101
Bean J. L., et al., 2008, A&A, 486, 1039
Bechter E. B., et al., 2014, ApJ, 788, 2
Bouma L. G., et al., 2019, AJ, 157, 217
Brown A. G. A., Vallenari A., Prusti T., de Bruijne J. H. J., Babusiaux C., Bailer-Jones C. A. L., et al., 2018, A&A, 616, A1
Bryan M. L., et al., 2012, ApJ, 750, 84
Cáceres C., Ivanov V. D., Minniti D., Naef D., Melo C., Mason E., Selman F., Pietrzynski G., 2009, A&A, 507, 481
Chan T., Ingemyr M., Winn J. N., Holman M. J., Sanchis-Ojeda R., Esquerdo G., Everett M., 2011, AJ, 141, 179
Chen G., Pallé E., Nortmann L., Murgas F., Parviainen H., Nowak G., 2017, A&A, 600, L11
Claret A., 2000, A&A, 363, 1081
Claret A., 2004, A&A, 428, 1001
Claret A., Bloemen S., 2011, A&A, 529, A75
Collins K. A., Kielkopf J. F., Stassun K. G., 2017, AJ, 153, 78
Damasso M., et al., 2015, A&A, 575, A111
Eastman J., Siverd R., Gaudi B. S., 2010, PASP, 122, 935
Fernandez J. M., Holman M. J., Winn J. N., Torres G., Shporer A., Mazeh T., Esquerdo G. A., Everett M. E., 2009, AJ, 137, 4911
Foreman-Mackey D., Agol E., Ambikasaran S., Angus R., 2017, AJ, 154, 220
Fulton B. J., Shporer A., Winn J. N., Holman M. J., Pál A., Gazak J. Z., 2011, AJ, 142, 84
Gillon M., et al., 2007, A&A, 472, L13
Gillon M., et al., 2009a, A&A, 496, 259
Gillon M., et al., 2009b, A&A, 501, 785
Gillon M., et al., 2010, A&A, 511, A3
Hebb L., et al., 2009, ApJ, 693, 1920
Heyrovský D., 2007, ApJ, 656, 483
Hoyer S., et al., 2013, MNRAS, 434, 46
Huitson C. M., Désert J.-M., Bean J. L., Fortney J. J., Stevenson K. B., Bergmann M., 2017, AJ, 154, 95
Kasper D. H., et al., 2019, MNRAS, 483, 3781
Kipping D. M., 2013, MNRAS, 435, 2152

Knutson H. A., et al., 2014, ApJ, 785, 126
Kundurthy P., Barnes R., Becker A. C., Agol E., Williams B. F., Gorelick N., Rose A., 2013, ApJ, 770, 36
Maciejewski G., et al., 2010, MNRAS, 407, 2625
Maciejewski G., et al., 2013, A&A, 551, A108
Maciejewski G., et al., 2016, A&A, 588, L6
Maciejewski G., Stangret M., Ohlert J., Basaran C. S., Maciejczak J., Puciata-Mroczynska M., Boulanger E., 2018a, Information Bulletin on Variable Stars, 63, 6243
Maciejewski G., et al., 2018b, Acta Astronomica, 68, 371
Mancini L., et al., 2014, MNRAS, 443, 2391
Mancini L., et al., 2017, MNRAS, 465, 843
Mancini L., et al., 2018, A&A, 613, A41
Mandel K., Agol E., 2002, ApJ, 580, L171
McCullough P. R., Crouzet N., Deming D., Madhusudhan N., 2014, ApJ, 791, 55
Montalto M., et al., 2012, MNRAS, 427, 2757
Nascimbeni V., Piotto G., Bedin L. R., Damasso M., 2011a, A&A, 527, A85
Nascimbeni V., Piotto G., Bedin L. R., Damasso M., Malavolta L., Borsato L., 2011b, A&A, 532, A24
Nascimbeni V., et al., 2013, A&A, 549, A30
Nikolov N., Henning T., Koppenhoefer J., Lendl M., Masiejewski G., Greiner J., 2012, A&A, 539, 159
Pál A., et al., 2008, ApJ, 700, 783
Pál A., Sárneczky K., Szabó G. M., Szing A., Kiss L. L., Mezö G., Regály Z., 2011, MNRAS, 413, L43
Patra K. C., Winn J. N., Holman M. J., Yu L., Deming D., Dai F., 2017, AJ, 154, 4
Petrucci R., Jofré E., Schwartz M., Cúneo V., Martínez C., Gómez M., Buccino A. P., Mauas P. J. D., 2013, ApJ, 779, L23
Pollacco D., et al., 2008, MNRAS, 385, 1576
Pont F., Zucker S., Queloz D., 2006, MNRAS, 373, 231
Pont F., et al., 2007, A&A, 476, 1347
Queloz D., Shao M., Mayor M., 2001, in American Astronomical Society Meeting Abstracts. p. 1356
Ranjan S., Charbonneau D., Désert J.-M., Madhusudhan N., Deming D., Wilkins A., Mandell A. M., 2014, AJ, 147, 161
Sanchis-Ojeda R., Winn J. N., Holman M. J., Carter J. A., Osip D. J., Fuentez C. I., 2011, ApJ, 733, 127
Santos N. C., et al., 2013, A&A, 556, A150
Shporer A., Mazeh T., Pont F., Winn J. N., Holman M. J., Latham D. W., Esquerdo G. A., 2009, ApJ, 694, 1559
Siverd R. J., et al., 2012, ApJ, 761, 123
Sokov E. N., et al., 2018, MNRAS, 480, 291
Southworth J., 2015, Astrophysics Source Code Library, ascl:1511.016
Southworth J., et al., 2009a, MNRAS, 396, 1023
Southworth J., et al., 2009b, MNRAS, 399, 287
Southworth J., et al., 2010, MNRAS, 408, 1680
Southworth J., Bruni I., Mancini L., Gregorio J., 2012, MNRAS, 420, 2580
Stevenson K. B., Bean J. L., Seifahrt A., Désert J.-M., Madhusudhan N., Bergmann M., Kreidberg L., Homeier D., 2014, AJ, 147, 161
Szabó G. M., et al., 2010, A&A, 523, A84
Torres G., 2007, ApJ, 671, L65
Torres G., Winn J. N., Holman M. J., 2008, ApJ, 677, 1324
Tregloan-Reed J., et al., 2015, MNRAS, 450, 1760
Triaud A. H. M. J., et al., 2010, A&A, 524, A25
Tripathi A., et al., 2010, ApJ, 715, 421
Wilson D. M., et al., 2008, ApJ, 675, L113
Winn J. N., et al., 2007a, AJ, 133, 1828
Winn J. N., Holman M., Roussanova A., 2007b, ApJ, 657, 1098
Winn J. N., Holman M., Carter J. A., Torres G., Osip D. J., Beatty T., 2009, AJ, 137, 3826
Winn J. N., et al., 2010, ApJ, 718, 575
Wright J. T., 2005, PASP, 117, 657

This paper has been typeset from a TEX/LATEX file prepared by the author.