doi: 10.32023/0001-5237/71.1.2

ACTA ASTRONOMICA
Vol. **71** (2021) pp. 25–53

Massive Search for Spot- and Facula-Crossing Events in 1598 Exoplanetary Transit Light Curves

R. V. Baluev[1], E. N. Sokov[2,1], I. A. Sokova[2,1], V. Sh. Shaidulin[1], A. V. Veselova[1], V. N. Aitov[3], G. Sh. Mitiani[3], A. F. Valeev[3,4], D. R. Gadelshin[3], A. G. Gutaev[3,5], G. M. Beskin[3,5], G. G. Valyavin[3,4,1], K. Antonyuk[4], K. Barkaoui[6,7], M. Gillon[6], E. Jehin[8], L. Delrez[6,8], S. Guðmundsson[9], H. A. Dale[10], E. Fernández-Lajús[11,12], R. P. Di Sisto[11,12], M. Bretton[13], A. Wunsche[13], V.-P. Hentunen[14], S. Shadick[15], Y. Jongen[16], W. Kang[17], T. Kim[17,18], E. Pakštienė[19], J. K. T. Qvam[20], C. R. Knight[21], P. Guerra[22], A. Marchini[23], F. Salvaggio[23], R. Papini[23], P. Evans[24], M. Salisbury[25], J. Garlitz[26], N. Esseiva[27], Y. Ogmen[28], P. Bosch-Cabot[29], A. Selezneva[29] and T. C. Hinse[30,31]

[1] Saint Petersburg State University, 7–9 Universitetskaya Emb., St Petersburg 199034, Russia
[2] Central Astronomical Observatory at Pulkovo of Russian Academy of Sciences, Pulkovskoje sh. 65/1, St Petersburg 196140, Russia
[3] Special Astrophysical Observatory, Russian Academy of Sciences, Nizhnii Arkhyz, 369167, Russia
[4] Crimean Astrophysical Observatory, Russian Academy of Sciences, Nauchny, 298409, Russia
[5] Kazan (Volga Region) Federal University, Kazan, 420008, Russia
[6] Astrobiology Research Unit, Université de Liège, Alleé du 6 Août 19C, B-4000 Liège, Belgium
[7] Oukaimeden Observatory, High Energy Physics and Astrophysics Laboratory, Cadi Ayyad University, Marrakech, Morocco
[8] Space Sciences, Technologies and Astrophysics Research (STAR) Institute, Université de Liège, Allée du 6 Août 19C, B-4000 Liège, Belgium
[9] Nes Observatory, Hraunholl 5, 781 Hornafjörður, Southeast Iceland
[10] Emory University Department of Physics, 400 Dowman Drive, Suite N201, Atlanta, GA 30322, USA
[11] Facultad de Ciencias Astronómicas y Geofísicas - Universidad Nacional de La Plata, Paseo del Bosque S/N - 1900 La Plata, Argentina
[12] Instituto de Astrofísica de La Plata (CCT La Plata - CONICET/UNLP), Argentina
[13] Baronnies Provençales Observatory, Hautes Alpes - Parc Naturel Régional des Baronnies Provençales, F-05150 Moydans, France

[14] Taurus Hill Observatory, Warkauden Kassiopeia ry., Härkämäentie 88, 79480 Kangaslampi, Finland
[15] Physics and Engineering Physics Department, University of Saskatchewan, 116 Science Place, Saskatoon, Saskatchewan, S7N 5E2, Canada
[16] Observatoire de Vaison la Romaine, 1075 RD 51, Le Palis, 84110 Vaison-la-Romaine, France
[17] National Youth Space Center, Goheung, Jeollanam-do, 59567, South Korea
[18] Department of Astronomy and Space Science, Chungbuk National University, Cheongju-City, 28644, South Korea
[19] Institute of Theoretical Physics and Astronomy, Vilnius University, Sauletekio al. 3, Vilnius 10257, Lithuania
[20] Horten Videregående Skole, Strandpromenaden 33, 3183 Horten, Norway
[21] Ngileah Observatory, 144 Kilkern Road, RD 1. Bulls 4894, New Zealand
[22] Observatori Astronòmic Albanyà, Camí de Bassegoda s/n, 17733 Albanyà, Spain
[23] Astronomical Observatory, DSFTA - University of Siena, Via Roma 56, 53100 – Siena, Italy
[24] El Sauce Observatory, Coquimbo Province, Chile
[25] School of Physical Sciences, The Open University, Milton Keynes, MK7 6AA, UK
[26] AAVSO, Private Observatory, Elgin, OR 97827, USA
[27] Observatory Saint Martin, code k27, Amathay Vesigneux, France
[28] Green Island Observatory, Code B34, Gecitkale, Famagusta, North Cyprus
[29] Observatori Astronòmic Albanyà, 17733 Albanyà, Girona, Spain
[30] Institute of Astronomy, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, Grudziadzka 5, 87-100 Toruń, Poland
[31] Chungnam National University, Department of Astronomy, Space Science and Geology, Daejeon, South Korea

Received March 15, 2021

ABSTRACT

We developed a dedicated statistical test for a massive detection of spot- and facula-crossing anomalies in multiple exoplanetary transit light curves, based on the frequentist p-value thresholding. This test was used to augment our algorithmic pipeline for transit light curves analysis. It was applied to 1598 amateur and professional transit observations of 26 targets being monitored in the EXPANSION project. We detected 109 statistically significant candidate events revealing a roughly $2:1$ asymmetry in favor of spots-crossings over faculae-crossings. Although some candidate anomalies likely appear non-physical and originate from systematic errors, such asymmetry between negative and positive events should indicate a physical difference between the frequency of star spots and faculae. Detected spot-crossing events also reveal positive correlation between their amplitude and width, possibly due to spot size correlation. However, the frequency of all detectable crossing events appears just about a few per cent, so they cannot explain excessive transit timing noise observed for several targets.

Key words: *Planetary systems – Techniques: photometric – starspots*

1. Introduction

There is already a long record of starspot studies, including the detection of spot-crossing events during an exoplanetary transit. Silva (2003) tested this method on HD 209458, based on its light curves analysis, and obtained parameters of its spots (or groups of spots). It was supposed that such an approach can be used to evaluate some characteristics of spots, such as size and position, and consecutive transits may provide information about spots evolution. This method was later applied to several stars with known exoplanets. For example, Silva-Valio *et al.* (2010) studied transit light curves of CoRoT-2, the observed data were fitted using starspots models with different parameters, such as spot radius, intensity and longitude.

Later on, Tregloan-Reed *et al.* (2013) developed a method for modeling the transit and spots simultaneously and introduced an IDL computer code PRISM and the optimization algorithm GEMC. Their method was applied to transit light curves of the WASP-19 system and allowed calculating the stellar rotation period and the sky-projected obliquity of the system. The model was later updated in Tregloan-Reed *et al.* (2015) and used for modeling transits in WASP-6 system. A plenty of other transit modeling routines is available, such as KSint (Montalto *et al.* 2014) or StarSim (Herrero *et al.* 2016). On the other side, Southworth *et al.* (2019) applied simply a visual detection of starspots anomalies and noticed that it was efficient enough for their goals. Močnik *et al.* (2017) revealed recurring sequences of spots in Kepler data of Qatar-2. This allowed them to accurately measure star rotation period as well as planet-star spin-orbit alignment angle.

Bradshaw and Hartigan (2014) studied the lifetimes of spots on the Sun and other stars, taking into account their magnetic stellar activity. In particular, for three main-sequence stars with planets (Kepler-17, CoRoT-2, CoRoT-6), the sizes and lifetimes of spots resembled scaled values for the Sun. The authors emphasized the importance of combined usage of the photometric data, Doppler imaging, and analysis of exoplanet transits.

Namekata *et al.* (2019) studied the evolution of starspot regions based on the analysis of local minima of light curves. The lifetimes and emergence and decay rates of the spots were estimated for more than 50 star spots on solar-type active stars in the Kepler database.

Zaleski *et al.* (2019) studied differential rotation of the young solar-type star Kepler-71. Spots and faculae were characterized using transit light curves, and these results were translated into the maps of magnetic activity. The characteristics of light curve variations were determined based on the light curve model of Silva (2003), and the authors also described (very detailedly) the construction of model light curves taking into account manifestations of stellar magnetic activity. They applied a pioneer method of using faculae to estimate the rotation period of a star, and the estimate was consistent with the value obtained from starspots.

Aronson and Piskunov (2019) presented a model-free method (the transit imaging technique) for obtaining a map of brightness variations across the disk of a star based on information from several transit light curves. They aimed to produce a large database of stellar spot coverage. A map of the star brightness distribution without taking into account spots is obtained by analyzing the median light curve for several transits, then synthetic light curves are constructed and compared with observations, on the basis of which the map is updated.

In Netto and Valio (2020) spots on the young solar-type star Kepler-63 are studied. They applied Silva (2003) method and fitted transit light curves, taking into account possible spot-crossing anomalies. Almost three hundred starspots were characterized, and it was found that some spots could have existed for at least 75 d. Yet another attempt to study the starspot evolution was made for Kepler-17 by Namekata *et al.* (2020). The authors claimed that the evolution and location of spots derived from rotational modulations are significantly different from those derived from in-transit spots. However, with an accuracy of up to an order of magnitude, their estimates for the rate of emergence and decay of spots are consistent with similar values for sunspots. The authors therefore suggested the similarity of the processes of spot formation for solar-type stars.

The issue of starspots can also be viewed from another point, namely how they may affect the best fitting exoplanetary parameters. Czesla *et al.* (2009) considered the effect of starspots and faculae on transit light curves and on the normalization of transit profiles. They redetermined the inclination of the orbit and the radius of the planet in the CoRoT-2 system, taking into account data on the spot activity. This asserts the need to take into account the effects of stellar activity when obtaining the parameters of exoplanets with an accuracy of better than a percent level.

There are multiple ways how spots can affect estimations of exoplanetary parameters. Spots behind a transiting planet lead to an underestimation of its radius, and if these spots are located near the limb they may cause inaccuracies in transit duration, hence, in orbital semi-major axis. Near-limb spots can also trigger a spurious transit timing variation (TTV). Silva-Valio *et al.* (2010) considered CoRoT-2 system and showed that spot-crossing events disturb planet parameters estimates by several percent.

Sanchis-Ojeda *et al.* (2011) considered this issue in the context of verifying the spin–orbit alignment. They used several transits of WASP-4b and analyzed them taking the effect of starspot occultations. It was claimed that such an approach gives more constraining result for the sky-projected stellar obliquity than the Rossiter–McLaughlin method.

Kipping (2012) presented a very detailed description of the model, which takes into account the differential rotation, non-linear limb darkening, the evolution of spots, and so on. Their macula code allows reducing errors in the analysis of photometric data, as well as to speed-up calculations. Among other effects, the model can take into account the so-called $T\delta V$, or the gain of the apparent transit depth.

Juvan *et al.* (2018) developed PyTranSpot routine that allows modeling transit light curves taking into account effects of stellar activity. The technique was merged with the MCMC method. The authors tested the method on the synthetic light curves, and performed the analysis of WASP-41 system.

As we can see, different researchers agree that spots or faculae appearing along the transit chord may significantly disturb exoplanetary parameters and lead to inaccurate conclusions. Therefore, such anomalies must be detected in each transit light curve and fitted. However, numerous models and codes are available that allows approximating such spot-crossing anomalies. Different methods vary from visual perception to quite complicated codes that take into account multiple effects. However, the detection of spot-crossing anomalies is a signal detection task, after all. We find that statistical issues related to spots detection and relevant significance thresholds have not been studied well enough yet. Without that, it appears difficult to estimate the reliability of numerous individual results obtained in this domain, and in particular to resolve practical contradictions about whether a given transit light curve demonstrates statistically significant spot anomalies or not (Baluev *et al.* 2020). This becomes increasingly important when we conduct massive analysis of large number of transits like in Baluev *et al.* (2019). In this work we present some mathematical results of how to perform a statistically rigorous testing of spot anomalies. We also construct the corresponding computing pipeline and apply it to our sample of ≈ 1600 transit light curves.

The paper obeys the following scheme. We discuss transit data that we used, together with a general overview of their analysis algorithm, in Section 2. We present a solution to several mathematical and algorithmic issues of spots detection in Section 3. Finally, we discuss the results of our spot-search analysis in Section 4.

2. Transit Data and Overview of the Full Analysis Pipeline

We used a moderately expanded update of the data used by Baluev *et al.* (2019). Presently we have 1598 transit light curves for 26 targets, with $> 4 \cdot 10^5$ photometric measurements in total. As before, we use transit photometry from the EXPANSION (EXoPlanetary trANsit Search with an International Observational Network) project (Sokov *et al.* 2018), which involves a network of amateur and professional observatories. We also use transit photometry available in published literature, the sources are listed in Table 1. We did not aim here to construct a comprehensive transit database, so some objects may possibly miss some known data, especially because not all of them were updated in 2020–2021.

In this work we use a reduced version of the pipeline from Baluev *et al.* (2015, 2019), as implemented in the open-source PLANETPACK software (Baluev 2013c, 2018), though augmented with our search of spot-crossings anomalies. The latter part is described below (Section 3), and now we discuss only the basic fitting pipeline.

T a b l e 1

Sources of the photometric data (not including the EXPANSION project)

Target	References	Note
CoRoT-2	Gillon *et al.* (2010)	
	TRAPPIST	from Baluev *et al.* (2019)
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
	Christiansen *et al.* (2010)	NASA EPOXI mission
HAT-P-3	Torres *et al.* (2007)	
	Chan *et al.* (2011)	
	Nascimbeni *et al.* (2011a)	
	Mancini *et al.* (2018)	
HAT-P-4	Christiansen *et al.* (2010)	NASA EPOXI mission
HAT-P-12	Hartman *et al.* (2009)	
	Lee *et al.* (2012)	
	Hinse *et al.* (2015)	
	Sada and Ramón-Fox (2016)	These data were kindly provided by the authors
	Mancini *et al.* (2018)	
	Alexoudi *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)	
	Sada and Ramón-Fox (2016)	These data were kindly provided by the authors
HAT-P-38	Sato *et al.* (2012)	
HD 189733	Bakos *et al.* (2006)	
	Winn *et al.* (2007b)	T10APT data involve erratic HJD correction (private communication), we used data kindly provided by the authors
	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.* (2018)	

T a b l e 1

Continue

Target	References	Note
Qatar-2	Bryan *et al.* (2012)	We assumed BJD TDB for the "BJD" times.
	Mancini *et al.* (2014)	
Qatar-4	Mallonn *et al.* (2019)	
TrES-1	Winn *et al.* (2007a)	
WASP-2	Southworth *et al.* (2010)	Danish telescope clock might have a shift (J. Southworth, private communication)
WASP-3	Tripathi *et al.* (2010)	
	Nascimbeni *et al.* (2013)	
	Christiansen *et al.* (2010)	NASA EPOXI mission
WASP-4	Wilson *et al.* (2008)	
	Gillon *et al.* (2009b)	
	Winn *et al.* (2009)	Superseded by Sanchis-Ojeda *et al.* (2011)
	Southworth *et al.* (2009a)	Superseded by Southworth *et al.* (2019)
	Sanchis-Ojeda *et al.* (2011)	
	Nikolov *et al.* (2012)	
	Petrucci *et al.* (2013)	These data were kindly provided by the authors
	Hoyer *et al.* (2013)	from Baluev *et al.* (2020)
	Huitson *et al.* (2017)	from Baluev *et al.* (2020)
	Southworth *et al.* (2019)	
	TRAPPIST	from Baluev *et al.* (2020)
	TESS	from Baluev *et al.* (2020)
WASP-5	Southworth *et al.* (2009b)	
	TRAPPIST	from Baluev *et al.* (2019)
WASP-6	Gillon *et al.* (2009a)	
	Tregloan-Reed *et al.* (2015)	
	TRAPPIST	from Baluev *et al.* (2019)
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.* (2018)	
WASP-35	TRAPPIST	

T a b l e 1

Concluded

Target	References	Note
WASP-50	Gillon *et al.* (2011)	
	Sada *et al.* (2012)	These data were kindly provided by the authors
	Tregloan-Reed and Southworth (2013)	Published light curves had an erratic BJD correction, we used correct ones kindly provided by J. Southworth
	Sada (2018)	These data were kindly provided by the authors
	TRAPPIST	
WASP-52	Chen *et al.* (2017)	Multi-band transmission spectroscopy, very high accuracy data
	Mancini *et al.* (2017)	
WASP-75	Gómez Maqueo Chew *et al.* (2013)	
	TRAPPIST	
WASP-84	Anderson *et al.* (2014)	
	TRAPPIST	
WASP-122	Turner *et al.* (2016)	
	TRAPPIST	
XO-2N	Fernandez *et al.* (2009)	
	Kundurthy *et al.* (2013)	
	Damasso *et al.* (2015)	
XO-5	Burke *et al.* (2008)	
	Pál *et al.* (2009)	These data were kindly provided by the authors
	Maciejewski *et al.* (2011)	Taken from G. Maciejewski personal web page
	Sada *et al.* (2012)	These data were kindly provided by the authors
	Hinse *et al.* (2015)	
	Smith (2015)	These data were kindly provided by the authors
	Kjurkchieva *et al.* (2018)	Not clear whether the data are HJD or JD, we assumed HJD UTC

We run only two fitting stages from Baluev *et al.* (2019). Stage 1 represents an initial fit used to detect photometric outliers. Now we filtered outliers a bit more aggressively than in Baluev *et al.* (2019). The threshold was chosen close to 4σ, removing about 0.05% of individual photometric measurements. This more strict filtering was chosen because a single outlier may be misinterpreted as a spot anomaly in some cases. The spot anomalies are detected on Stage 2, which was applied to data already cleaned from outliers.

Each stage involves a maximum-likelihood fit with a dedicated Gaussian Process (GP) model that remained basically the same as in Baluev *et al.* (2019). We fitted all light curves of a particular target using the same transit parameters bound between light curves (these parameters are planet radius, impact parameter, transit duration). The limb darkening was modeled using a quadratic law with fittable coefficients, but those coefficients were bound for all light curves belonging to the same or similar spectral band. For example, light curves obtained for a particular target in R_J, R_C, r, or r' filters all involved the same limb darkening coefficients.

Aside from WASP-12 and WASP-4, which both reveal a quadratic deviation of transit times (Maciejewski *et al.* 2016, Bouma *et al.* 2019, Baluev *et al.* 2020), no other target in our list demonstrated statistically significant TTV. Therefore, in this work we also fix transit times at the quadratic ephemeris (with fittable coefficients). We expect that such a restriction would make our search of spot-crossing anomalies more reliable for certain problematic light curves.

Each light curve also included a cubic polynomial to take into account possible systematic drifts. Random photometric noise was fitted using a GP model with mandatory white and optional red component. The white noise was modeled through a fittable jitter term, using the model from Baluev (2015a), which is resistant with respect to numeric peculiarities of the likelihood function. The red noise was modeled through the exponential correlation function $\exp(-|\Delta t|/\tau)$ with fittable τ. Red noise was first detected in individual light curves as described in Baluev *et al.* (2019), and only robustly fittable red noise terms were included in the model. After that, we tried to fit the red noise in all the remaining light curves under restriction that their τ is the same, and again left only those red noise terms that had a robust fit. Light curves where the red noise remained ill-fitted both in the free-τ and shared-τ treatment were left with white-only noise model (such light curves would typically imply a negative red noise, meaning blue noise that we do not consider).

3. Search of Spot-Crossing Transit Anomalies with Strict Statistical Testing

3.1. Spot Anomalies Detection: the Statistical Theory

Each spot- or facula-crossing event triggers a bell-like anomaly in the transit curve that we model by a Gaussian shape:

$$m_{\mathrm{GA}}(t,K,\mu,\sigma) = K\exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right), \tag{1}$$

where K is the amplitude of the signal, μ being its central time, and σ being characteristic width. Such Gaussian Anomaly (GA) is added to the transit model $m_{\mathrm{transit}}(t,\mathbf{p})$, where m means magnitude and $\mathbf{p}$ is the vector of fittable transit parameters. Following this convention, $K>0$ for facula-crossings and $K<0$ for spot-crossings.

Our first task is to detect all statistically significant GAs in a set of the transit light curves. This can be done *e.g.*, by numeric minimization of the χ^2 function associated to the model $m_{\rm transit} + m_{\rm GA}$. A bit more general approach taking into account poorly known noise level is to also use a parameterized noise model and to maximize the corresponding likelihood function (Baluev 2009). The latter approach can be easily extended to treat the correlated photometric noise *via* the GP model (Baluev 2011, Baluev, 2013b, Rajpaul *et al.* 2015, Foreman-Mackey *et al.* 2017, Angus *et al.* 2018).

However, when fitting nonlinear models like Eq.(1) we have to solve a computationally complicated optimization task. This task is made so heavy because the likelihood function typically has multiple peaks corresponding to different positions in the plane of nonlinear parameters (μ,σ). Each such local maximum of the likelihood corresponds to a single local solution for Eq.(1), and different such solutions appear nearly uncorrelated in terms of their best fitting parameters. From a mathematical point of view, the cause of such a behavior comes from the following correlation measure:

$$\mathrm{corr}(\mu_{\rm GA}^{(1)},\, \mu_{\rm GA}^{(2)}) = \frac{\left| \int\limits_{-\infty}^{+\infty} \mu_{\rm GA}^{(1)}(t) \mu_{\rm GA}^{(2)}(t)\, \mathrm{d}t \right|}{\sqrt{\int\limits_{-\infty}^{+\infty} \left[\mu_{\rm GA}^{(1)}(t)\right]^2 \mathrm{d}t \int\limits_{-\infty}^{+\infty} \left[\mu_{\rm GA}^{(2)}(t)\right]^2 \mathrm{d}t}} = \sqrt{\frac{2\frac{\sigma_2}{\sigma_1}}{1+\left(\frac{\sigma_2}{\sigma_1}\right)^2}}\, \mathrm{e}^{-\frac{(\mu_2-\mu_1)^2}{2(\sigma_1^2+\sigma_2^2)}}. \quad (2)$$

We can see that it decreases for large $|\mu_2-\mu_1|$ or for large $|\log(\sigma_2/\sigma_1)|$. Therefore, if either $\mu_{1,2}$ or $\sigma_{1,2}$ differ too much, the two models $\mu_{\rm GA}^{(1)}$ and $\mu_{\rm GA}^{(2)}$ can be treated as (quasi-)independent ones even though they both are expressed by formally the same function (Eq. 1). Then the entire plane (μ,σ) is split into a set of "independence cells" such that correlations between different models $\mu_{\rm GA}$ are high within a single cell, while distinct cells are only weakly correlated on average. Then total number of local maxima of the χ^2 (or likelihood) function is roughly equal to the number of such cells, and each cell would typically contain just a single maximum. Notice that the amplitude K is a linear parameter, so it cannot generate quasi-independent models: the correlation (Eq. 2) does not depend on $K_{1,2}$. Hence, for each μ and σ there is only a single best fitting value of K.

The effect of multiple likelihood peaks owed to nonlinear parameters is explained in more details in Baluev (2013a, 2015b). We cannot know in advance which peaks would appear high or low, and we do not have restrictive enough prior information about possible parameters of GAs. So we have to directly scan some reasonable domain in the (μ,σ) plane seeking the highest peak (the global maximum inside domain). In other words, we should test multiple candidate solutions (Eq. 1), starting each fit from a point inside a separate independence cell.

A quite similar phenomenon is known for periodograms, which can be viewed in a direct relationship with the least squares and maximum likelihood fitting (Lomb

1976, Scargle 1982, Baluev 2014). In this case multiple peaks appear because of the nonlinear frequency parameter f. The width of each "independence cell" in the frequency axis is about $\Delta f \sim 1/T$, where T being the time series length. Two sinusoidal variations that have $|f_2 - f_1| \gtrsim 1/T$ appear independent in terms of the correlation measure analogous to Eq.(2). This effect simply determines the periodogram resolution: we have to scan the periodogram with the step $\sim 1/T$ at largest or we may undersample (or even miss) the global maximum. So we have to perform numerous independent fits to determine just a single nonlinear parameter, the frequency.

However, the issues coming from nonlinear parameters are more important than just the increased computing load. An important caveat is that by such wide scanning we implicitly test a large number of statistically independent solutions, each corresponding to a single likelihood peak. Since all such solutions are nearly independent statistically, this leads to an increased false positives rate. This is due to the statistical effect of multiple testing: to make a mistake with, say, 1000 peaks tested at once is roughly 1000 times more probable, compared to a single test. This effect significantly increases all the detection thresholds. In the periodograms theory this is well known as the "bandwidth penalty". In our task of GA detection a similar effect should appear, even if we test just a single light curve.

The general theory of how to treat this effect for an arbitrary nonlinear signal is given in Baluev (2013a). Mathematically, that theory was considered with periodograms and periodic signals in mind, hence all formulae include a mandatory frequency parameter. But this assumption was not critical, so all formulae can be easily promoted to non-periodic GA models like Eq.(1).

In our case the null model is $m_{\text{transit}}(t)$, and the signal is expressed by $m_{\text{GA}}(t)$. We should perform two fits: for just m_{transit} and for $m_{\text{transit}} + m_{\text{GA}}$, assuming the parameters μ and σ to be constant. Given these fits, we can construct the logarithm of the likelihood ratio, ζ, which is a function of our two nonlinear unknowns μ and σ. The maximum of $\zeta(\mu,\sigma)$ shall determine (*via* its location) the best fitting values for these arguments. Notice that μ and σ are treated separately because of their nonlinearity, while the remaining parameters are either strictly linear (like K) or can be linearized approximately about the best fitting point (like $\mathbf{p}$).

Since input data involve noise, $\zeta(\mu,\sigma)$ is a random field, while its global maximum (it has to be computed numerically) is random quantity. Large $\max\zeta(\mu,\sigma)$ indicates that our light curve cannot be explained well by just $m_{\text{transit}}(t)$ and likely also involves a GA (Eq. 1), while small value means that GAs are unlikely to exist and $m_{\text{transit}}(t)$ has satisfactory accuracy. To derive the detection threshold separating these two decisions, we should statistically quantify the levels of $\max\zeta$ under the null hypothesis (no GAs).

For that, we should compute the False Alarm Probability (FAP) function, which is complementary to the distribution function of $\max\zeta$:

$$\text{FAP}(z) = \Pr\{\max\zeta > z\} = 1 - P_{\max}(z), \quad P_{\max}(z) = \Pr\{\forall\zeta < z\}. \qquad (3)$$

The computation of $\mathrm{FAP}(z)$ is one of primary results in Baluev (2013a). For the dimension two (two nonlinear parameters μ,σ) its approximation looks like

$$\mathrm{FAP}_2(z) \simeq 2A_2 e^{-z}\sqrt{z}, \quad A_2 = \frac{1}{2\pi^{\frac{3}{2}}} \int\limits_{\mathcal{D}} \sqrt{\det \mathrm{var}(\eta')}\, \mathrm{d}\mu\, \mathrm{d}\sigma, \tag{4}$$

where $\mathcal{D}$ is the domain in the (μ,σ) plane that we scan for possible GAs, and $\mathrm{var}(\eta')$ is the 2×2 variance-covariance matrix of the gradient of an auxiliary Gaussian random field η defined as $\eta(\mu,\sigma) = \pm\sqrt{2\zeta(\mu,\sigma)}$, with the sign taken from the best fitting K.

In Eq.(4) we removed an additional correction term responsible for the boundary maxima, when the global maxima is attained on the boundary of $\mathcal{D}$. This is because in our algorithm such cases are treated as unreliable (see below) and are eliminated from the investigation, and so they cannot generate false alarms.

Now, given a small detection threshold FAP^* we may claim that our light curve reveals a statistically significant GA, if $\mathrm{FAP}(\max\zeta) < \mathrm{FAP}^*$ for the particular $\max\zeta$ computed from the actual data. The best fitting GA parameters are then given by the position of $\max\zeta$. Otherwise, if $\mathrm{FAP}(\max\zeta) > \mathrm{FAP}^*$, the light curve is consistent with a clean transit.

Eq.(4) refers to a 2D domain in the (μ,σ) plane. However, 2D scan may appear computationally hard, and we may replace it by a 1D one, in which we fix σ at a reasonable prior value. Such a simplification is justified below, but here we can give a 1D version of Eq.(4) for this case:

$$\mathrm{FAP}_1(z) \simeq 2A_1 e^{-z}, \quad A_1 = \frac{1}{2\pi} \int\limits_{T_1}^{T_2} \sqrt{\mathrm{var}(\eta'_\mu)}\, \mathrm{d}\mu. \tag{5}$$

In this formula σ is assumed constant, so we integrate only over μ.

The coefficient A in Eqs.(4–5) is responsible for the penalty of multiple peaks testing. It is not obvious yet and still needs to be computed. In Section 4 of Baluev (2013a), expressions of two types were considered: precise formulae (slow) and analytic approximations (fast). The latter ones were derived assuming a periodic signal in place of our GA, so they need to be promoted to conditions of our task. The analytic approach is based on the so-called approximation of "uniform phase coverage", where various summations of periodic functions over the discrete time series are replaced by analytic integrals over a single period. This cannot be used in our task directly, because the GA signal (Eq. 1) is non-periodic, but we can apply an equivalent approximation. Namely, we can replace the necessary summations by integrals over the entire time span, assuming that observations come with a constant cadence and the time span is large. Then, using so-modified formulae of Section 4.1 of Baluev (2013a), we obtained the following:

$$\mathrm{var}(\eta') \simeq \begin{pmatrix} \frac{1}{2\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^2} \end{pmatrix}, \quad \sqrt{\det \mathrm{var}(\eta')} = \frac{1}{2\sigma^2}, \quad \sqrt{\mathrm{var}(\eta'_\mu)} = \frac{1}{\sigma\sqrt{2}}. \tag{6}$$

Let us define the parametric domain $\mathcal{D}$ as a rectangle with $\mu \in [T_1, T_2]$ (typically, the transit duration range) and $\sigma \in [\Sigma_1, \Sigma_2]$, and then

$$A_2 = \frac{T_2 - T_1}{2\pi^{\frac{3}{2}}} \left(\frac{1}{\Sigma_1} - \frac{1}{\Sigma_2} \right), \quad A_1 = \frac{T_2 - T_1}{2\pi\sigma\sqrt{2}}. \tag{7}$$

Notice that Eqs.(6-7) require $\Sigma_{1,2} \ll T_2 - T_1$. This allowed us to make several simplifications, in particular by neglecting the correlation between GA and the null model, $\mathrm{corr}(m_{\mathrm{transit}}, m_{\mathrm{GA}})$. These approximations also do not involve correlated noise models (noise is assumed to be white).

As we can see, there are many assumptions and hence multiple potential vulnerabilities with the approximation (Eq. 7), but it remains not obvious how accurate it can be until we compare it with a better assessment.

More accurate formula for A comes from the matrix decompositions of Section 4.2 of Baluev (2013a), namely we use adapted versions of Eqs.(42-44) from that paper. First, rewrite our GA as $m_{\mathrm{GA}} = Kg(t, \mu, \sigma)$ and determine the full Fisher information matrix of our compound model $m_{\mathrm{transit}}(t) + m_{\mathrm{GA}}(t)$:

$$Q = \begin{pmatrix} \left\langle \frac{\partial m_{\mathrm{transit}}}{\partial \mathbf{p}} \otimes \frac{\partial m_{\mathrm{transit}}}{\partial \mathbf{p}} \right\rangle & \left\langle g \frac{\partial m_{\mathrm{transit}}}{\partial \mathbf{p}} \right\rangle & \left\langle \frac{\partial g}{\partial \mu} \frac{\partial m_{\mathrm{transit}}}{\partial \mathbf{p}} \right\rangle & \left\langle \frac{\partial g}{\partial \sigma} \frac{\partial m_{\mathrm{transit}}}{\partial \mathbf{p}} \right\rangle \\ \left\langle \left(\frac{\partial m_{\mathrm{transit}}}{\partial \mathbf{p}} \right)^{\mathrm{T}} g \right\rangle & \langle g^2 \rangle & \left\langle \frac{\partial g}{\partial \mu} g \right\rangle & \left\langle \frac{\partial g}{\partial \sigma} g \right\rangle \\ \left\langle \left(\frac{\partial m_{\mathrm{transit}}}{\partial \mathbf{p}} \right)^{\mathrm{T}} \frac{\partial g}{\partial \mu} \right\rangle & \left\langle g \frac{\partial g}{\partial \mu} \right\rangle & \left\langle \left(\frac{\partial g}{\partial \mu} \right)^2 \right\rangle & \left\langle \frac{\partial g}{\partial \sigma} \frac{\partial g}{\partial \mu} \right\rangle \\ \left\langle \left(\frac{\partial m_{\mathrm{transit}}}{\partial \mathbf{p}} \right)^{\mathrm{T}} \frac{\partial g}{\partial \sigma} \right\rangle & \left\langle g \frac{\partial g}{\partial \sigma} \right\rangle & \left\langle \frac{\partial g}{\partial \mu} \frac{\partial g}{\partial \sigma} \right\rangle & \left\langle \left(\frac{\partial g}{\partial \sigma} \right)^2 \right\rangle \end{pmatrix}. \tag{8}$$

Here triangular brackets designate the weighted summation over the time series (substituting the best fitting $\mathbf{p}$ from the null model).

After that we should compute the Cholesky decomposition $Q = LL^{\mathrm{T}}$, where the low-triangular matrix L would look like:

$$L = \begin{pmatrix} L_{\mathrm{null,null}} & 0 & 0 & 0 \\ L_{K,\mathrm{null}} & l_{KK} & 0 & 0 \\ L_{\mu,\mathrm{null}} & l_{\mu K} & l_{\mu\mu} & 0 \\ L_{\sigma,\mathrm{null}} & l_{\sigma K} & l_{\sigma\mu} & l_{\sigma\sigma} \end{pmatrix} \tag{9}$$

Now we only need the diagonal elements of the bottom-right square block, l_{KK}, $l_{\mu\mu}$, and $l_{\sigma\sigma}$. Using them,

$$\sqrt{\det \mathrm{var}(\eta')} = \frac{l_{\mu\mu} l_{\sigma\sigma}}{l_{KK}^2}, \quad \sqrt{\mathrm{var}(\eta'_\mu)} = \frac{l_{\mu\mu}}{l_{KK}}. \tag{10}$$

These quantities can be further integrated numerically using second formula of Eqs.(4–5), and so we obtain A.

Even this way of computing the FAP is not entirely precise, because we still used several hidden simplifying assumptions: (i) the noise is still white, (ii) we

use pure least-squares fitting, *i.e.*, there is no fittable noise (noise is known), (iii) original models from Baluev (2013a) assumed strictly linear $m_{\rm transit}$, so we performed its hidden linearization with respect to $\mathbf{p}$ in the vicinity of the best fitting points. The last two issues were already discussed in Baluev (2013a) and they are likely negligible, if our models are not ill-fitted. For example, the FAP formulae simply become "more approximate" but still valid, if in place of pure least squares we apply the maximum-likelihood method with a fittable noise. This is because the corresponding Fisher information matrix has zeros in the off-diagonal blocks responsible for correlations between the noise and curve parameters (Baluev 2009). The linearization of $m_{\rm transit}$ about the best fitting null model also should not break resulting approximations, if the fit is robust. However, the first issue (correlated noise) is important because red photometric noise is quite typical.

Correlated noise models were not considered in Baluev (2013a), but the necessary formulae are not hard to obtain by a minor modification. We need to recompute the Fisher information matrix (Eq. 8) for the general likelihood function involving a GP noise model (Baluev 2013b). It appears that we simply need to replace the time-series summation operation $\langle * \rangle$ in Eq.(8) by the following bi-linear form:

$$\langle xy \rangle \longmapsto \mathbf{x}^{\rm T} V^{-1} \mathbf{y}, \tag{11}$$

where V is the covariance matrix of the noise (at the best fitting null model). Notice that this bi-linear form can be computed faster, profiting from the Cholesky decomposition of V:

$$\langle xy \rangle \longmapsto \mathbf{x}^{\rm T} V^{-1} \mathbf{y} = (L_V^{-1}\mathbf{x})^{\rm T} (L_V^{-1}\mathbf{y}), \quad V = L_V L_V^{\rm T}. \tag{12}$$

The rest of the computation remains the same.

Summarizing all the above, a GA candidate can be tested for statistical significance using Eq.(4) or Eq.(5) and: (i) a fast entirely analytic formula (Eq. 7), (ii) slow but more accurate formulae (Eqs. 8–10) that still assume only white noise, (iii) even slower version of the last set, augmented by Eq.(11) to take the red noise into account.

We find that Eq.(7) is not very accurate in practice. The approximation Eq.(6) has satisfactory accuracy only in the middle of a transit, where $m_{\rm transit}$ varies slowly. In the ingress or egress phases $m_{\rm transit}$ varies faster, so that its correlation with $m_{\rm GA}$ is not negligible. The value of A_1 computed using Eq.(10) is typically 30–50% larger than the analytic value Eq.(7). The red noise, if present, also triggers an increase of A_1, depending on the parameters. In our practical computations numeric values for A_1 appeared mostly in the range from 2 to 5. This means that we should typically have about a few or ten likelihood peaks per each light curve. This penalty is not as large as the periodogram bandwidth penalty, but still it is a big factor that cannot be neglected.

Larger A means larger detection threshold (less number of GAs pass the test). That is, using undervalued A would lead us to excessive number of false GA detections, so we did not use the fast formula Eq.(7) for actual GA testing. However,

Eq.(7) and associated expressions are useful to understand our task better. For example, since elements of the matrix Eq.(6) are basically variances of the likelihood function gradient, their inverse values estimate average width of the likelihood peaks. This width appears $\sigma\sqrt{2}$ both in μ and in σ variable. This information can be used to construct a scan grid with an optimal resolution.

3.2. Spot Anomalies Search and Verification: Practical Aspects

When we started to test the method of GA detection in practical light curve data, several additional issues appeared. We highlight three of them: (i) slow computing speed, (ii) various subtle model inaccuracies and noisy drifts have a tendency to trigger detection of highly-correlated GAs, rendering the transit model nearly degenerate, (iii) it appeared difficult to disentangle red noise from GAs.

Concerning the issue of slow computation, it cannot be avoided completely, because we have to test all probe GAs located in distinct independence cells of the (μ,σ) plane. However, the speed can be improved if we could replace the full 2D scan by a 1D one with fixed σ. Such a replace appears justified by the following explanation.

Let us first estimate typical practical range for the spot-crossing duration (the σ parameter). This range depends, primarily, on the statistical distribution of the spot impact parameter s defined as the distance between planet trajectory and spot, divided by the planet radius r. The widest GAs would appear when planet crosses a spot by its equator ($s = 0$), while "grazing" spot-crossings (s close to 1) would generate GAs with small σ. In theory, GAs may have arbitrarily small width, but too narrow spot-crossings: (i) are statistically rare, (ii) are difficult to detect due to a small amplitude. The quantity s is distributed uniformly in the $[0,1]$ range, so its median value is 1/2, while 90% of events occur for $s < 0.9$. From the other side, if the planet disk is circular (and spot itself is small) then the spot-crossing half-duration is $\tau_{\rm spot} = \sqrt{1-s^2}\tau_{\rm pl}$, where $\tau_{\rm pl}$ is time that planet takes to pass its radius. Therefore, the median half-duration of a spot-crossing is $0.87\tau_{\rm pl}$, while 90% of cases have half-duration longer than $0.44\tau_{\rm pl}$. As we can see, narrow events are statistically rare, with only $\approx 13\%$ occurrences below half of the median duration.

Additionally, physical spot-crossings should not be very wide. Small spots cannot generate events lasting longer than $2\tau_{\rm pl}$. If a spot (or spot group) is big, compared to the planet, then the event may last somewhat longer, but spot-crossings wider than *e.g.*, twice of the above value are not very likely.

Based on these considerations, we adopted the following reasonable range for a spot-crossing width: $|\log(\sigma/\sigma_{\rm c})| \le \rho$, where $\rho = \log 2 \approx 0.7$ and $\sigma_{\rm c}$ is some central value. Basically, this range is from $\sigma_{\rm c}/2$ to $2\sigma_{\rm c}$. The value of $\sigma_{\rm c}$ should correspond to the median half-duration, $0.87\tau_{\rm pl}$. Notice, however, that we cannot just equate these two quantities here, because we approximate spot-crossing by a Gaussian shape (Eq. 1), while the actual anomaly should look more box-like (if the spot is not large). The best fitting value of $\sigma_{\rm c}$ can be obtained by maximizing

the correlation (Eq. 2) between the GA and box-shaped anomaly with a given half-duration $\tau_{\rm spot}$. By performing this maximization numerically, we obtained that $\sigma \approx 0.7\tau_{\rm spot}$, and hence $\sigma_{\rm c} \approx 0.6\tau_{\rm pl}$.

Given such σ-range for spot-crossings, its log-scale width appears $2\rho \approx 1.4$. Simultaneously, we already know that likelihood peaks should have typical width of $\sigma\sqrt{2}$ along the σ axis. This corresponds to the width of $\sqrt{2}$ in $\log\sigma$, nearly the same as our σ-range. Therefore, such σ-range may embed only a single likelihood peak, and we do not need to formally scan this range. It is quite safe to simplify our task by scanning only along the μ parameter (within the transit range), fixing $\sigma = \sigma_{\rm c}$. We first perform such initial scan to detect preliminary GA candidates using our 1D criterion ($\mathrm{FAP}_1(z)$), and after that all the detected GAs are refitted using free σ. After this fit, to ensure that all GAs would be detected in the full 2D scan as well, we re-verify them based on the 2D criterion ($\mathrm{FAP}_2(z)$). Moreover, the coefficient A_2 is also computed using simplifications, in order to avoid direct 2D integration. We first compute A_1 using accurate formulae (Eqs. 8–10) and then correct this estimate by the factor $\frac{A_2}{A_1} = 2\sqrt{\frac{2}{\pi}}\sinh\rho \approx 1.2$ that follows from solely analytic formulae (Eq. 7).

The second issue appears when our algorithm tries to use a GA model to fit something not suitable. This includes attempts to fit long-term drifts by a near-degenerate superposition of Gaussians. Such trends are already modeled by cubic polynomials and using a red noise GP model. Even if these models appear partly inaccurate, it is inadequate to use GA shapes (Eq. 1) for fitting any residual longer-term variation, as this leads to degeneracy issues and may lead to an over-fit effect. To reject such cases we verify that each our GA satisfies reliability criteria: (i) after full 2D fit our GA remains within the domain $\mathcal{D}$: inside the transit range in terms of μ and inside the required σ-range, (ii) the value of σ is smaller than $0.1T$, with T being the light curve time span, (iii) all GAs detected in the same light curve must have small enough correlations (Eq. 2), namely smaller than 1/3. If some GA failed any of these reliability criteria, it was removed and the model was refit with remaining GAs which were re-verified anew. Last detected GAs were removed first. The light curve with one or more unreliable GA was no longer tested for more GAs (in practice all such light curves demonstrated weird noisy variations that were fitted as large-magnitude correlated noise). Notice that our domain tests are applied taking into account uncertainties in μ and σ, that is we keep GAs which nominal values are formally out of the domain, but the domain still intersects with the uncertainty ranges.

The third issue is to disentangle GAs and correlated noise. Red or quasi-periodic noise often demonstrate long-living variations that can be represented through a superposition of GAs. However, if fitting the red noise through a GP model typically increases uncertainties in other parameters, fitting multiple GAs often causes an overfit effect with undervalued uncertainties, owed to a mock reduction of the residuals *rms*. Therefore, it is important to avoid erratic interpretation of

correlated noise through GAs. But the opposite misinterpretation is also undesired, since our goal is to detect spot-induced GAs, after all. Moreover, cases may exist where we cannot statistically distinguish these two interpretations, "white noise + GA" or "red noise without GA". This ambiguity is difficult to resolve in any other way but through a prior prioritization of the models.

We adopt such an algorithm that resolves this ambiguity in favor of the GA model. However, if at any step the GA term appears suspicious, *e.g.*, statistically insignificant or non-trusted due to strange values of parameters, we fallback to the red noise model without this GA. Thanks to such a behavior we do not fit light curves using GAs if the GA model itself does not look well justified. The entire algorithm is as follows.

1. Perform initial 1D detection of GA candidates, filtering away unreliable ones and assuming only white noise. This would likely produce a somewhat excessive list of GAs.

2. Run the red noise detection algorithm (Baluev *et al.* 2019) as detailed above, but also taking into account all preliminary detected GAs.

3. Retest the GAs in the 2D framework, also filtering away unreliable ones (GAs parameters might change so some of them may no longer pass the reliability tests), and using the red noise GP model. Many of the GA candidates do not survive this stage.

4. Those light curves where we removed a GA should be retested for possible red noise again (it will likely appear fittable if it was not fittable before). This assumes a return to step 2.

5. Steps 2–4 are iterated in a loop until the solution is stabilized. In the end we have all red noise terms robustly fittable and all GA candidates statistically significant and passing the reliability tests.

4. Results

In 1598 transit light curves our analysis pipeline detected 109 potential GAs. All these GA candidates are shown in Fig. 1, in the form of a 2D diagram "amplitude – width". We assumed the FAP threshold of 0.0027, which means, formally, that we should have about 4 statistical false positives in total. However, one should bear in mind that this estimate refers to particular adopted models and involves various hidden assumptions about photometric noise.

All light curves with detected GAs are plotted in Figs. 2–9. In these plots we show the original light curve data with their best fitting model, their "partial" residuals (everything subtracted except candidate GA), and model of the GA (or multiple GAs, if present). We also print additional data in each plot, including the

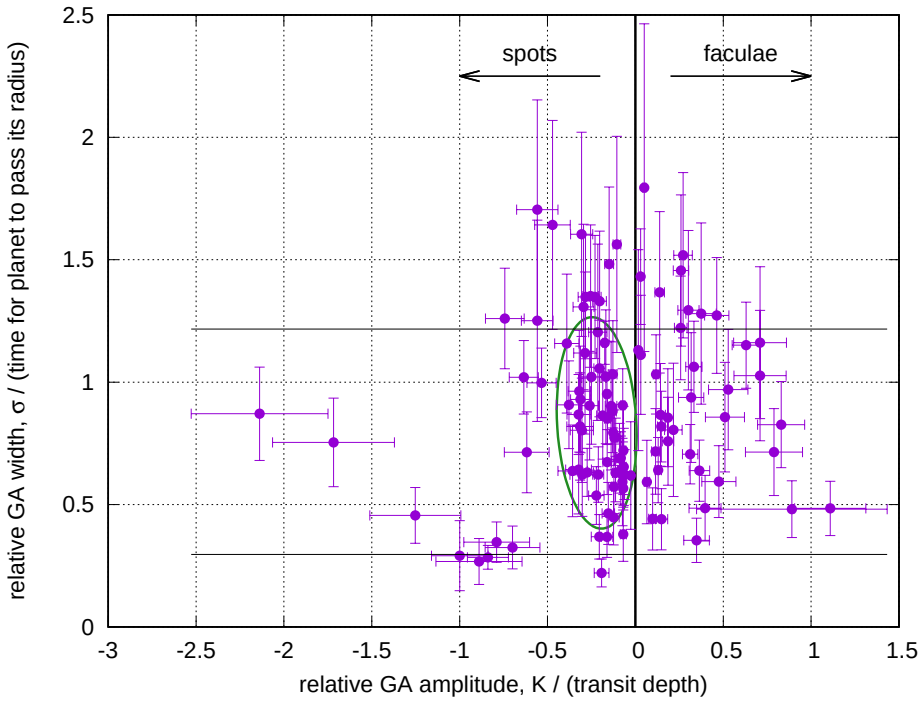

Fig. 1. All detected GAs in the amplitude–width diagram. We outline a slightly inclined concentration of negative GAs potentially reflecting physical spot-crossing events. Two horizontal lines label the range where a GA should reside (taking into account its uncertainties) to pass the reliability test.

fit *rms* and red noise parameters, if the red noise was fitted. The title is simply the file name used in the Baluev *et al.* (2019) data release. It contains the date of the observation, target name, name(s) of the observer (or first author of a paper and a year), and generic filter information.

The primary obvious property of our GA set is large asymmetry between positive and negative GAs, 38 cases against 71. The negative ones (those that may refer to spots) are clearly dominating. Such imbalance is difficult to explain by statistical errors, instrumental issues, or inaccuracies of the analysis, because then the number of positive and negative GAs should be approximately equal. Even if all 38 positive GAs were artifacts unrelated to stellar physics, the number of artifacts among negative GAs should be approximately the same, so we have no less than ≈ 33 physical spot-crossing events. Negative GAs form a clear concentration, outlined in Fig. 1. Moreover, this concentration seems slightly inclined, possibly reflecting a natural correlation with spot size (smaller spot – smaller GA width – smaller GA amplitude). We fitted the logarithmic linear regression $\log\sigma = a\log|K| + b$, restricting it to only negative GAs with $|K|$ below 1/2 of transit depth (more physically reasonable cases), and we obtained $a = 0.29 \pm 0.10$, a statistically remarkable value. Concerning positive GAs, we did not detect any clear correlation and it is not obvious in Fig. 1.

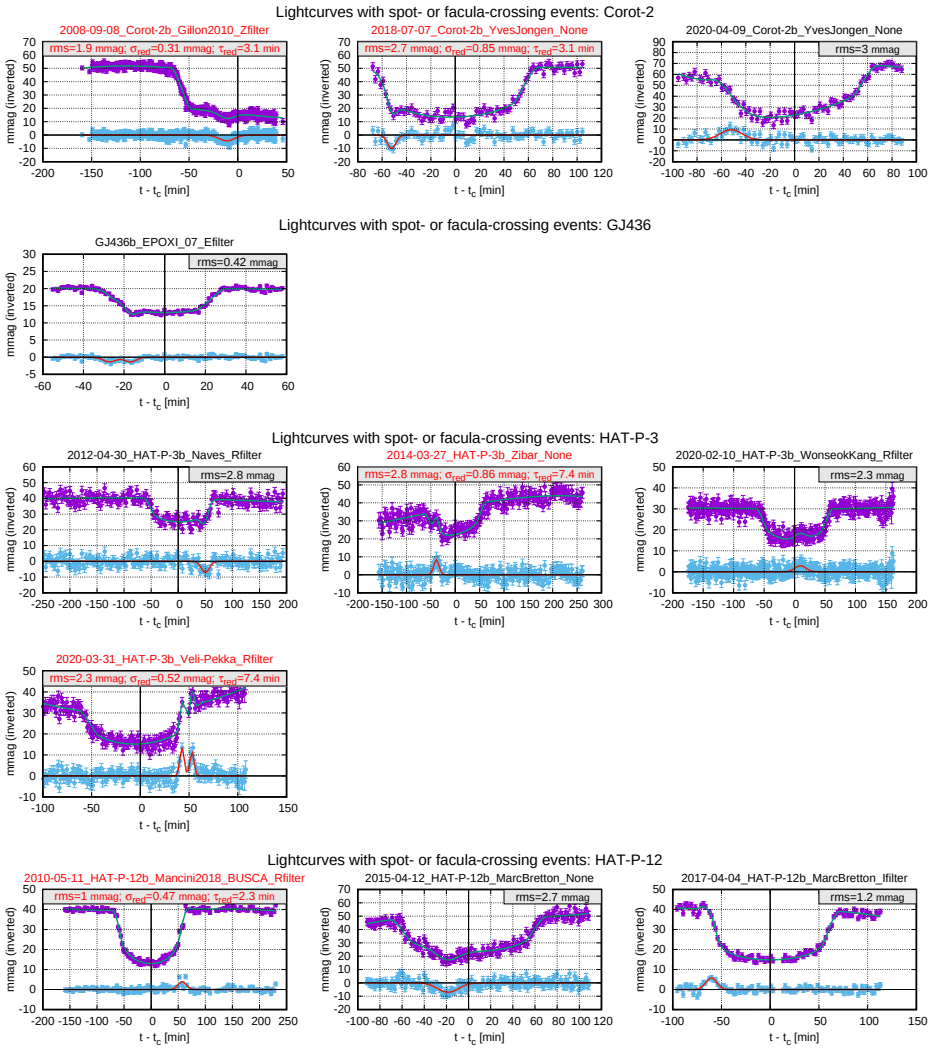

Fig. 2. GAs detected by our pipeline (Part 1: CoRoT-2, GJ 436, HAT-P-3, and HAT-P-12).

Simultaneously, the number of non-physical GAs is likely large. Some negative GAs have $|K|$ greater than transit depth. This is not physical, since a spot cannot have negative brightness. If we look into particular transit curves, far not all of them reveal a convincing GA signature. It may appear that our pipeline tried to fit some sudden noisy spikes or instrumental events that our GP model could not predict statistically. In some cases the residuals reveal hints of a non-stationary statistical behavior, *e.g.*, variable variance. Those cases are also out of our algorithm responsibility: it will approximate such a non-stationary noise by a stationary GP model with average parameters. Some GAs solely depend on just a single photometric observation, *e.g.*, a Mancini *et al.* (2014) light curve for Qatar-2, dated by

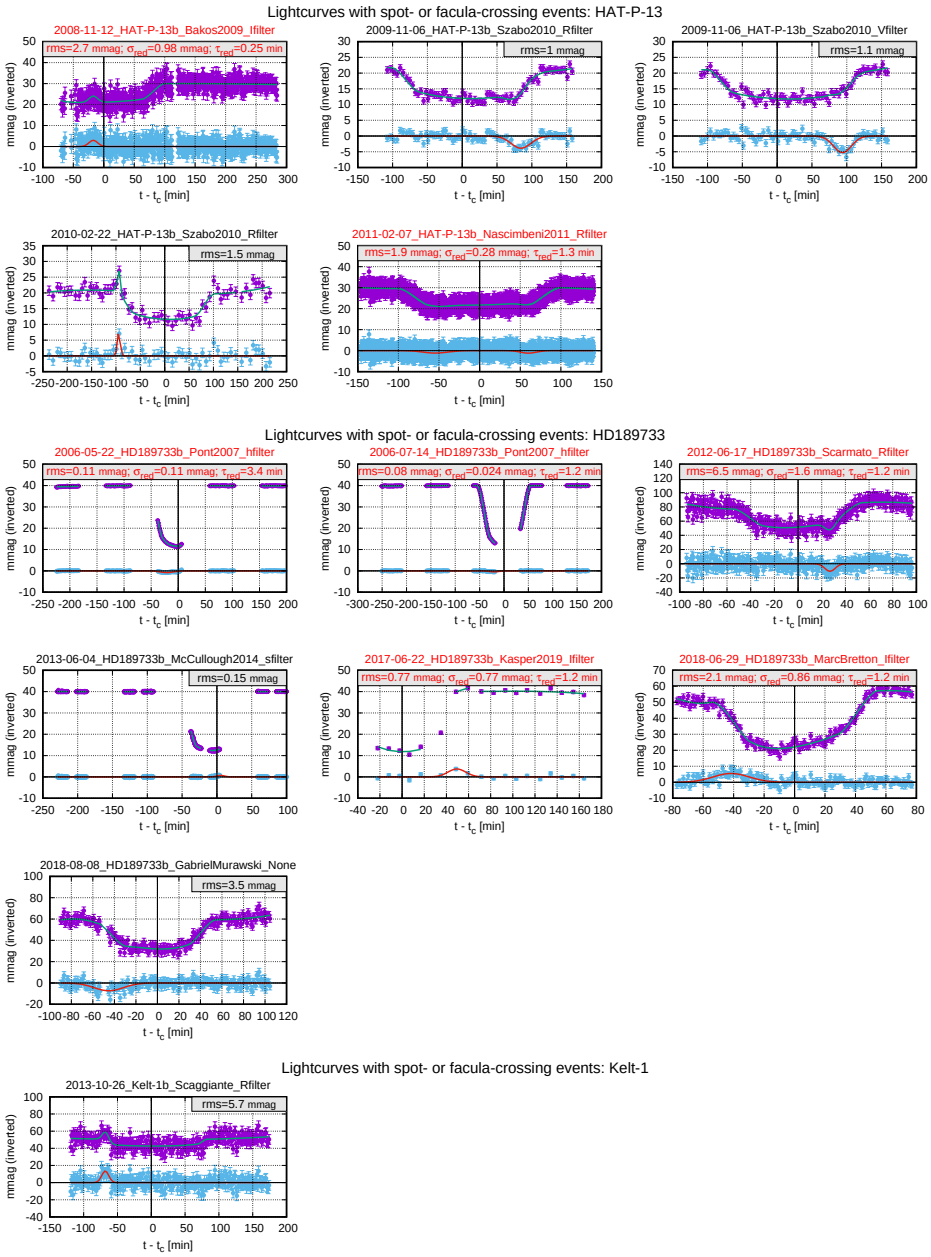

Fig. 3. GAs detected by our pipeline (Part 2: HAT-P-13, HD 189733, and Kelt-1).

2012-04-21 (Fig. 4). This point comes right after a gap in the light curve, so it seems to be a deviation caused by a cloud or an instrumental failure that was not cut away in full. As such, the GA remains unreliable even though the suspicious observation was not classified as outlier in Stage 1. Finally, in very accurate data like HST observations of HD 189733, it seems that we tried to fit either inaccura-

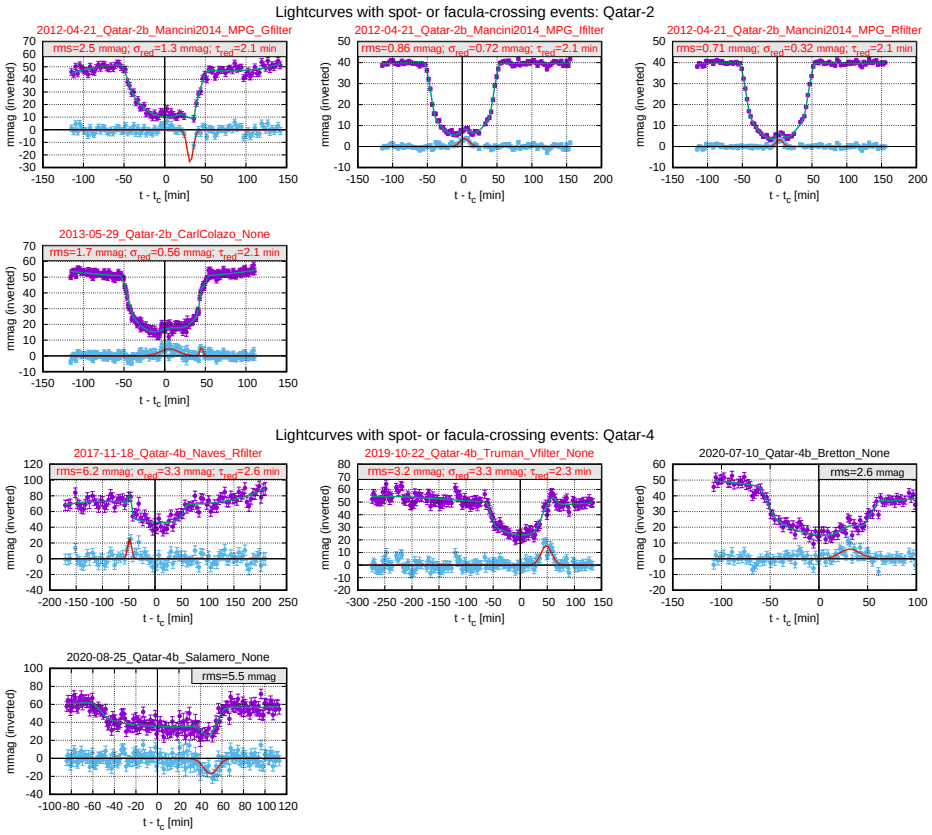

Fig. 4. GAs detected by our pipeline (Part 3: Qatar-2 and Qatar-4).

cies of the quadratic limb darkening model, or an effect of imperfect detrending, through GAs.

Many such cases have to remain inconclusive, because their classification cannot be performed based on just a single photometric curve. Although several transits in our database were observed from independent sites, this usually did not appear helpful enough, because the photometric accuracy varies for different sites.* To clearly ensure that a particular GA is a physical crossing event rather than noise artifact it is necessary to have a complex same-high-quality multi-site and multi-channel (including *e.g.*, spectral) observations of a single transit. This would be too expensive program perhaps, but in view of our results it is most important to seek such comprehensive characterization for positive GAs that indicate faculae-crossings.

*For example, the WASP-4 transit light curve on 23.08.2008 by Southworth *et al.* (2019) reveals a spot-crossing event. This transit was simultaneously observed by Hoyer *et al.* (2013) from two telescopes, but those two light curves appeared too noisy for a robust verification. Nearly the same story is about 23.09.2008 facula-crossing event detected in Southworth *et al.* (2019) data. See Fig. 6.

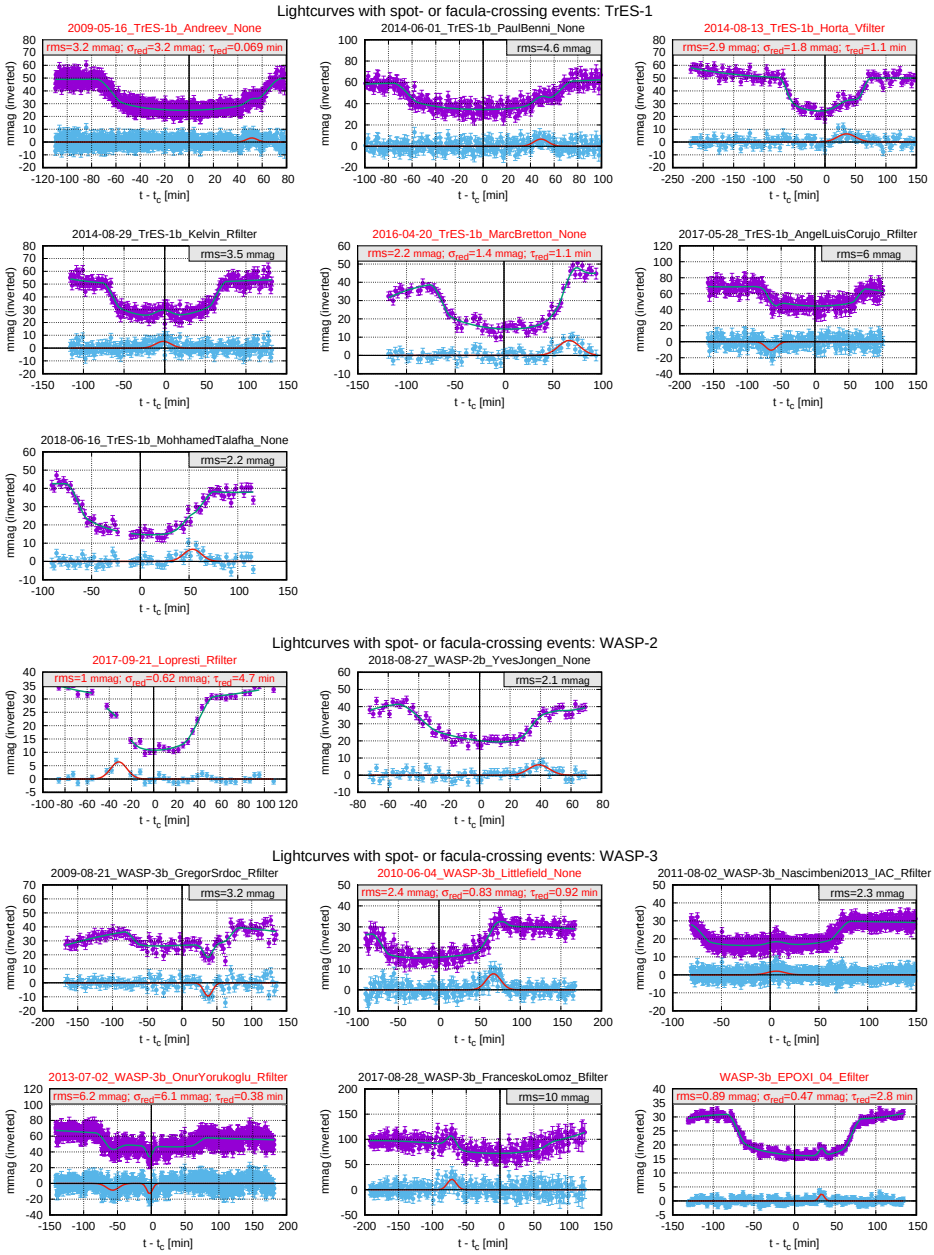

Fig. 5. GAs detected by our pipeline (Part 4: TrES-1, WASP-2, and WASP-3).

5. Discussion

We developed an algorithmic pipeline that allows us to perform massive detection of GAs that possibly refer to spot- and facula-crossing events in transit light curves. Although the algorithm was based on a statistically rigorous mathe-

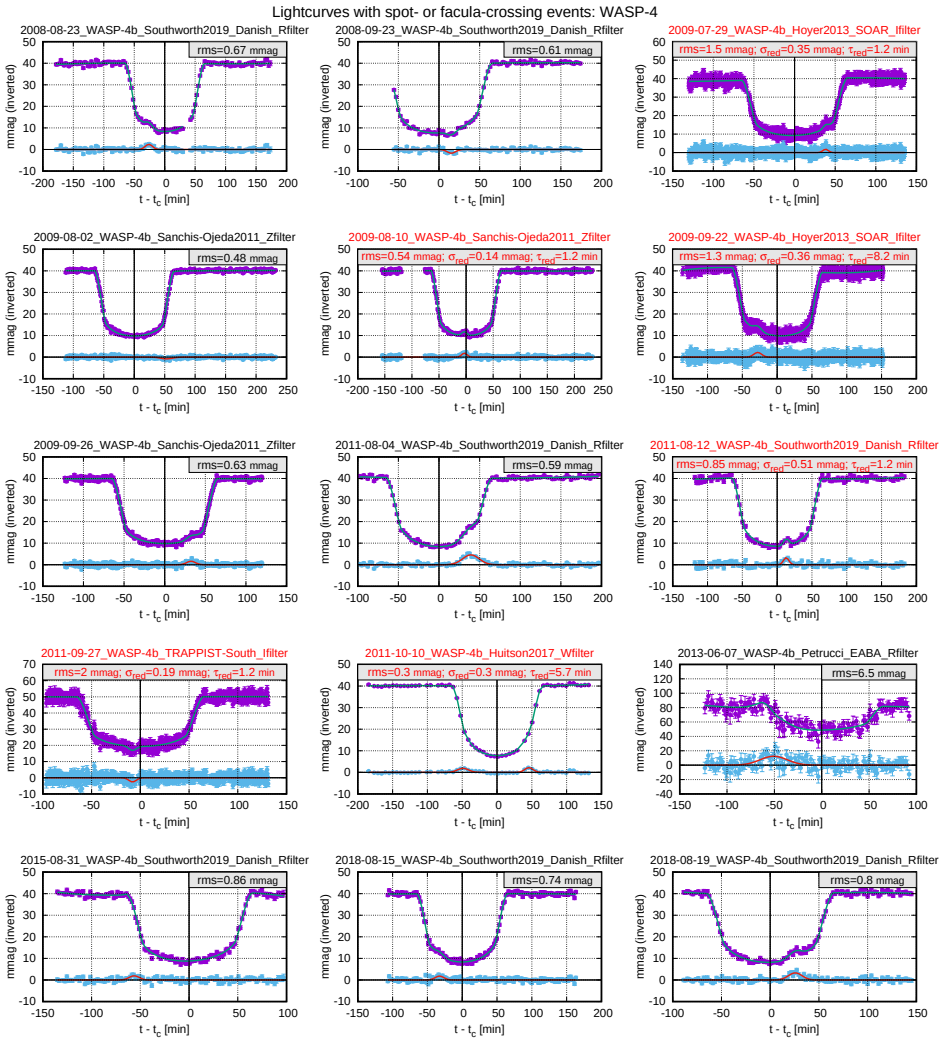

Fig. 6. GAs detected by our pipeline (Part 5: WASP-4).

matics, it relies on particular models: Gaussian model of the anomaly, quadratic limb-darkening law, exponential correlation function of the noise, and a stationary GP model. Any of these models may turn inadequate for a particular light curve, or it may involve statistically unpredictable instrumental or weather events. Because of all these factors, our algorithm collects weird systematic patterns together with physical spot- and facula-crossing. Therefore, it is necessary to develop additional post-filtering criteria rejecting unreliable GAs (*e.g.*, those that depend on just a single measurement, or which involve sudden light curve jumps, or the noise demonstrates non-stationary behavior).

However, we can already conclude that our pipeline is pretty efficient in what concerns the automated detection of spot anomalies in good data. For example, for

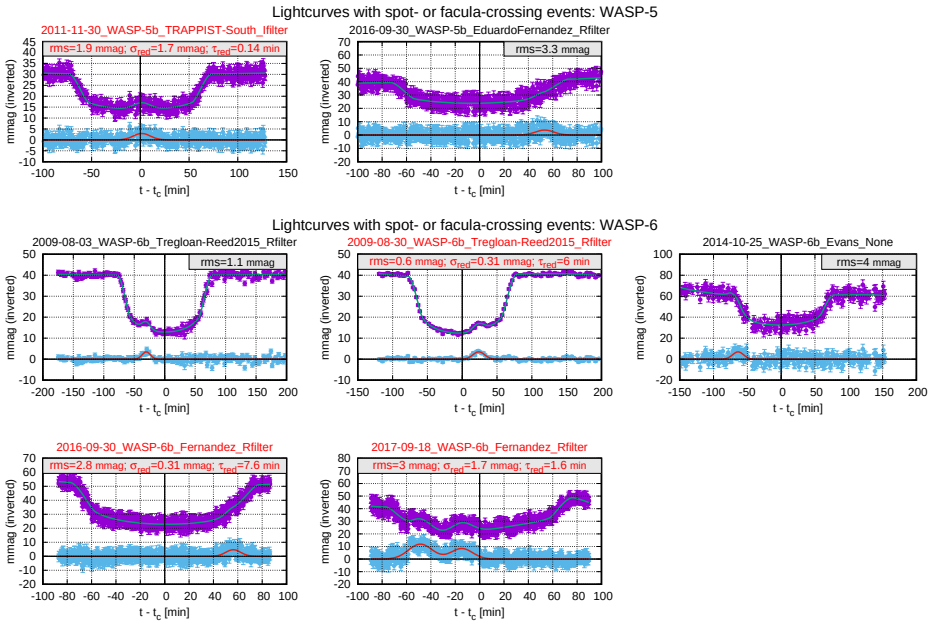

Fig. 7. GAs detected by our pipeline (Part 6: WASP-5 and WASP-6).

WASP-4 we detected 4 of 6 spot-transit events found by Southworth *et al.* (2019). One their proposed spot-crossing did not pass our reliability test (it appeared too wide) and the other one appeared statistically insignificant, but in exchange we detected two other spot-crossings and one facula-crossing in their data. We believe this agreement is good enough, and simultaneously our method has more solid math grounds in comparison with visual detection approach by Southworth *et al.* (2019).

One of our underlying goals was to improve transit timing accuracy by modeling spot anomalies. As noticed in Baluev *et al.* (2019), there is typically an excess jitter in measured transit times, and this effect clearly depends on the star. For example, HD 189733 revealed a significant TTV jitter of ≈ 1.5 min beyond uncertainties, likely indicating its larger activity in comparison with other targets. Our hypothesis was that by performing massive spot detection and modeling, this TTV jitter can be removed. However, for HD 189733 we detected approximately an average relative number of GAs, and many of them did not look convincingly robust after all. The resulting effect on TTV variance seems rather negligible, because only a few per cent of light curves revealed GAs. Simultaneously, we detected many apparently reliable spots for targets like WASP-4, WASP-52, and WASP-12, which all have paradoxically small TTV jitter. It seems that the activity-related TTV jitter comes from another physical phenomena, not directly related to detectable spots.

Yet another issue comes from ambiguous interpretation of spot anomalies. Too often the models “transit+spot” and “shifted transit” appear statistically indistinguishable. In fact, most of the moderate shifts of transit mid-times can be equally

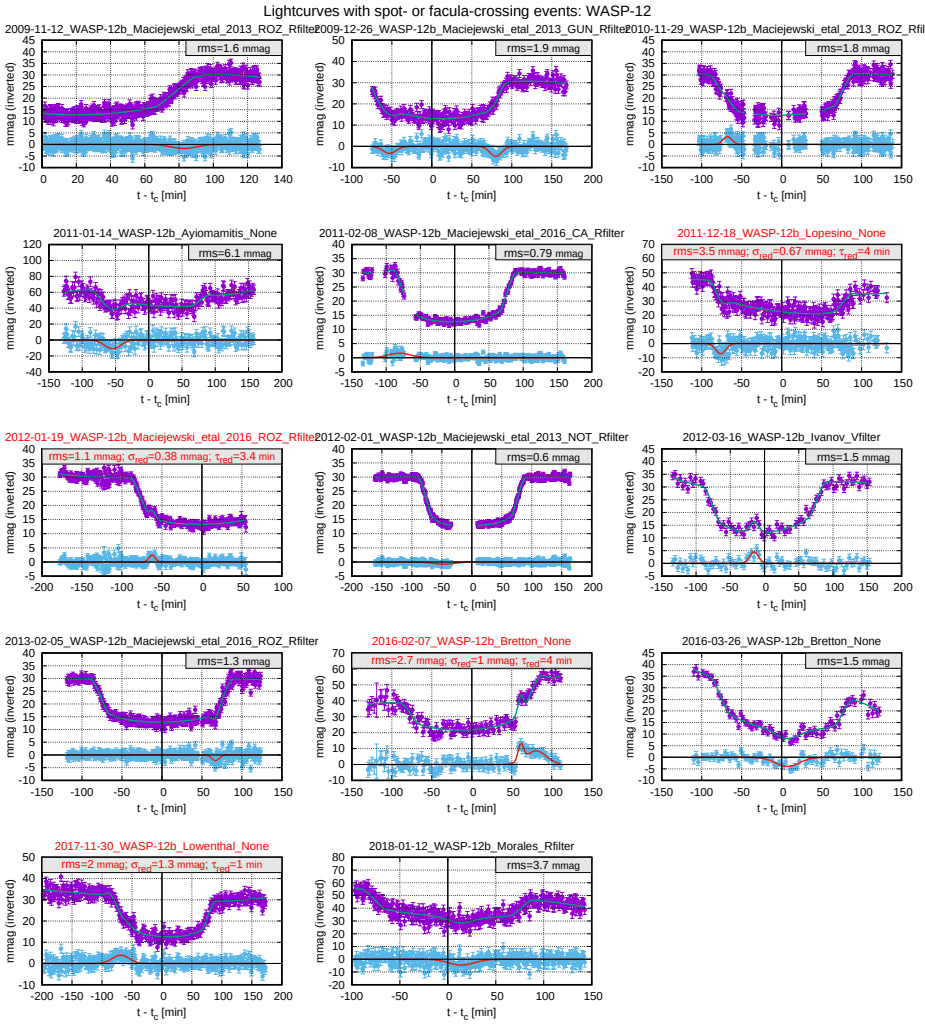

Fig. 8. GAs detected by our pipeline (Part 7: WASP-12).

explained by spot anomalies in the ingress and/or egress phases. An example is the single WASP-4 Huitson *et al.* (2017) light curve where we detected two "side" spots (Fig. 6).[†] If we did not fix its mid-time at a quadratic ephemeris, we would likely select a more simple light curve model with only one of these GAs with roughly doubled amplitude. This would result in a shifted timing, either positive or negative, depending on which GA we discard. Such cases trigger ambiguous interpretation of the data and bi-modal timing estimates. Time series of this type, with bi-modal measurements, are quite unusual and their analysis needs a better understanding.

[†]Such paired GAs may also be caused by an inaccuracy of the limb-darkening model, however it seems unlikely in this particular case, because the limb darkening was determined based on all four Huitson *et al.* (2017) light curves. So the issue is that one of these light curves, shown in the plot, has a significantly different shape than three others in average.

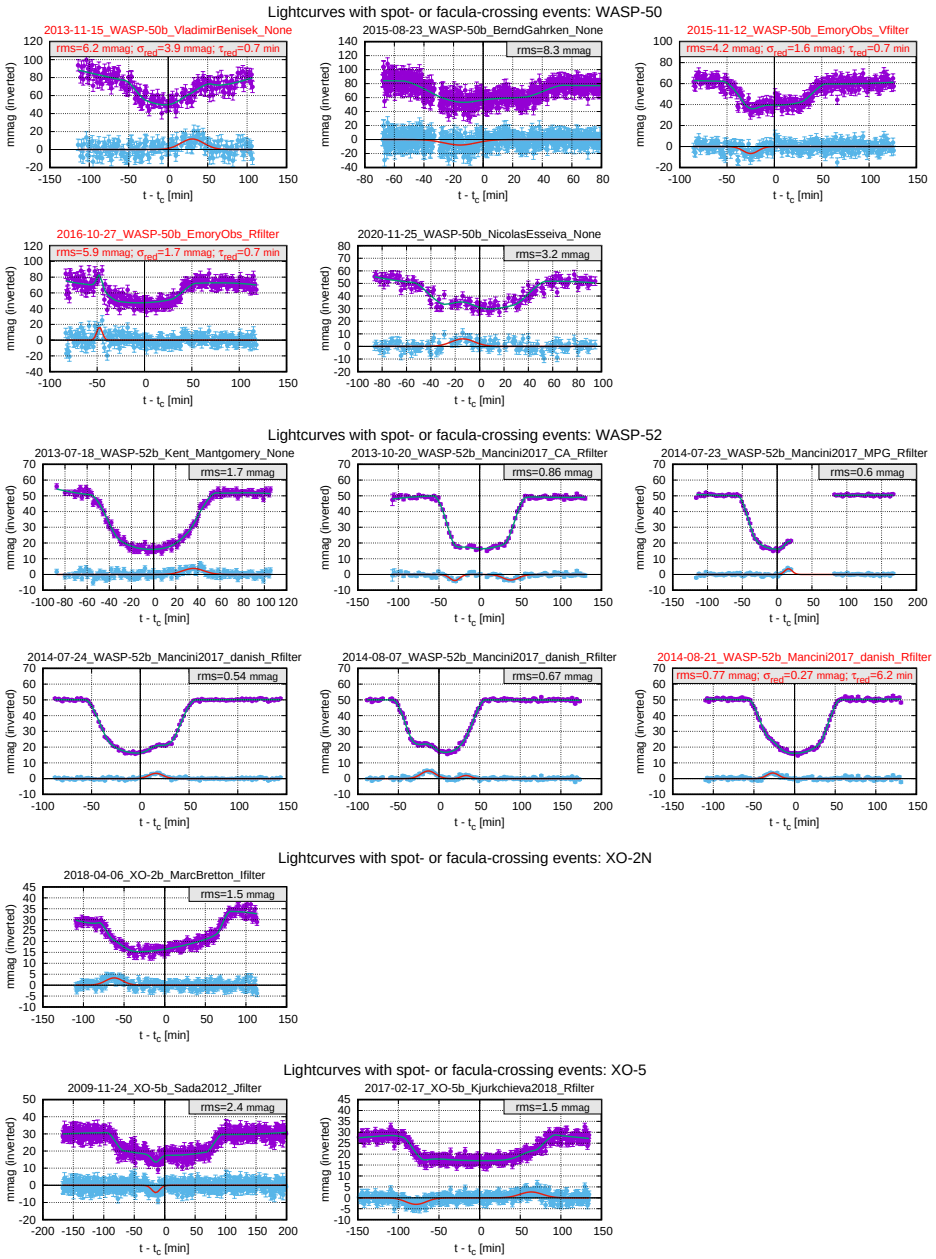

Fig. 9. GAs detected by our pipeline (Part 8: WASP-50, WASP-52, XO-2N, and XO-5).

Acknowledgements. Organization of the EXPANSION project (ENS, IAS, GGV), development of data processing algorithms (RVB, VShSh) and statistical analysis (RVB) were supported by the Russian Science Foundation, project 19-72-10023. Collecting the database of third-party exoplanetary data (RVB) and part of EXPANSION observations (VNA, GShM, AFV, DRG and GGV), in Sec-

tion 2, were supported by the Ministry of Science and Higher Education of Russian Federation, project 075-15-2020-780 (N13.1902.21.0039). AGG and GMB acknowledge "Fundamental Research" government contract of the Special Astrophysical Observatory of the Russian Academy of Sciences and the Federal program "Kazan Federal University Competitive Growth Program". EP acknowledges the Europlanet 2024 RI project funded by the European Union's Horizon 2020 Research and Innovation Programme (Grant agreement No. 871149). TCH acknowledges financial support from the National Research Foundation (NRF, No. 2019R1I1A1A01059609). TRAPPIST is funded by the Belgian Fund for Scientific Research (Fond National de la Recherche Scientifique, FNRS) under the grant FRFC 2.5.594.09.F, with the participation of the Swiss National Science Fundation (SNF). MG and EJ are F.R.S.-FNRS Senior Research Associates. TRAPPIST data sets analyzed in this study are available at the ESO Science Archive Facility at *http://archive.eso.org/*. We express gratitude to the anonymous reviewer of the manuscript for their useful comments.

REFERENCES

Alexoudi, X., Mallonn, M., von Essen, C., *et al.* 2018, *A&A*, **620**, A142.
Anderson, D.R., Collier Cameron, A., Delrez, L., *et al.* 2014, *MNRAS*, **445**, 1114.
Angus, R., Morton, T., Aigrain, S., Foreman-Mackey, D., and Rajpaul, V. 2018, *MNRAS*, **474**, 2094.
Aronson, E., and Piskunov, N. 2019, *A&A*, **630**, A122.
Bakos, G.A., Knutson, H., Pont, F., *et al.* 2006, *ApJ*, **650**, 1160.
Bakos, G.A., Howard, A.W., Noyes, R.W., *et al.* 2009, *ApJ*, **707**, 446.
Baluev, R.V. 2009, *MNRAS*, **393**, 969.
Baluev, R.V. 2011, *Celestial Mechanics and Dynamical Astronomy*, **111**, 235.
Baluev, R.V. 2013a, *MNRAS*, **431**, 1167.
Baluev, R.V. 2013b, *MNRAS*, **429**, 2052.
Baluev, R.V. 2013c, *Astronomy and Computing*, **2**, 18.
Baluev, R.V. 2014, *Astrophysics*, **57**, 434.
Baluev, R.V. 2015a, *MNRAS*, **446**, 1493.
Baluev, R.V. 2015b, *MNRAS*, **446**, 1478.
Baluev, R.V. 2018, *Astronomy and Computing*, **25**, 221.
Baluev, R.V., Sokov, E.N., Shaidulin, V.S., *et al.* 2015, *MNRAS*, **450**, 3101.
Baluev, R.V., Sokov, E.N., Jones, H., *et al.* 2019, *MNRAS*, **490**, 1294.
Baluev, R.V., Sokov, E.N., Hoyer, S., *et al.* 2020, *MNRAS*, **496**, L11.
Bean, J.L., Benedict, G.F., Charbonneau, D., *et al.* 2008, *A&A*, **486**, 1039.
Bouma, L.G., Winn, J.N., Baxter, C., *et al.* 2019, *AJ*, **157**, 217.
Bradshaw, S.J., and Hartigan, P. 2014, *ApJ*, **795**, 79.
Bryan, M.L., Alsubai, K.A., Latham, D.W., *et al.* 2012, *ApJ*, **750**, 84.
Burke, C.J., McCullough, P.R., Valenti, J.A., *et al.* 2008, *ApJ*, **686**, 1331.
Cáceres, C., Ivanov, V.D., Minniti, D., *et al.* 2009, *A&A*, **507**, 481.
Chan, T., Ingemyr, M., Winn, J.N., *et al.* 2011, *AJ*, **141**, 179.
Chen, G., Pallé, E., Nortmann, L., *et al.* 2017, *A&A*, **600**, L11.
Christiansen, J.L., Ballard, S., Deming, D., *et al.* 2010, "NASA Planetary Data System" DIF-X-HRIV-5-EPOXI-EXOPLANETS-PHOT-V1.0.
Czesla, S., Huber, K.F., Wolter, U., Schröter, S., and Schmitt, J.H.M.M. 2009, *A&A*, **505**, 1277.
Damasso, M., Biazzo, K., Bonomo, A.S., *et al.* 2015, *A&A*, **575**, A111.

Fernandez, J.M., Holman, M.J., Winn, J.N., *et al.* 2009, *AJ*, **137**, 4911.
Foreman-Mackey, D., Agol, E., Ambikasaran, S., and Angus, R. 2017, *AJ*, **154**, 220.
Fulton, B.J., Shporer, A., Winn, J.N., *et al.* 2011, *AJ*, **142**, 84.
Gillon, M., Pont, F., Demory, B.-O., *et al.* 2007, *A&A*, **472**, L13.
Gillon, M., Anderson, D.R., Triaud, A.H.M.J., *et al.* 2009a, *A&A*, **501**, 785.
Gillon, M., Smalley, B., Hebb, L., *et al.* 2009b, *A&A*, **496**, 259.
Gillon, M., Lanotte, A.A., Barman, T., *et al.* 2010, *A&A*, **511**, A3.
Gillon, M., Doyle, A.P., Lendl, M., *et al.* 2011, *A&A*, **533**, A88.
Gómez Maqueo Chew, Y., Faedi, F., Pollacco, D., *et al.* 2013, *A&A*, **559**, A36.
Hartman, J.D., Bakos, G.Á., Torres, G., *et al.* 2009, *ApJ*, **706**, 785.
Hebb, L., Collier-Cameron, A., Loeillet, B., *et al.* 2009, *ApJ*, **693**, 1920.
Herrero, E., Ribas, I., Jordi, C., *et al.* 2016, *A&A*, **586**, A131.
Hinse, T.C., Han, W., Yoon, J.-N., *et al.* 2015, *Journal of Astronomy and Space Science*, **32**, 21.
Hoyer, S., López-Morales, M., Rojo, P., *et al.* 2013, *MNRAS*, **434**, 46.
Huitson, C.M., Désert, J.-M., Bean, J.L., *et al.* 2017, *AJ*, **154**, 95.
Juvan, I.G., Lendl, M., Cubillos, P.E., *et al.* 2018, *A&A*, **610**, A15.
Kasper, D.H., Cole, J.L., Gardner, C.N., *et al.* 2019, *MNRAS*, **483**, 3781.
Kipping, D.M. 2012, *MNRAS*, **427**, 2487.
Kjurkchieva, D., Petrov, N., Ibryamov, S., Nikolov, G., and Popov, V. 2018, *Serbian Astron. J.*, **196**, 15.
Kundurthy, P., Barnes, R., Becker, A.C., *et al.* 2013, *ApJ*, **770**, 36.
Lee, J.W., Youn, J.-H., Kim, S.-L., Lee, C.-U., and Hinse, T.C. 2012, *AJ*, **143**, 95.
Lomb, N.R 1976, *Astrophysics and Space Science*, **39**, 447.
Maciejewski, G., Seeliger, M., Adam, C., Raetz, S., and Neuhäuser, R. 2011, *Acta Astron.*, **61**, 25.
Maciejewski, G., Dimitrov, D., Seeliger, M., *et al.* 2013, *A&A*, **551**, A108.
Maciejewski, G., Dimitrov, D., Fernández, M., *et al.* 2016, *A&A*, **588**, L6.
Maciejewski, G., Fernández, M., Aceituno, F., *et al.* 2018, *Acta Astron.*, **68**, 371.
Mallonn, M., von Essen, C., Herrero, E., *et al.* 2019, *A&A*, **622**, A81.
Mancini, L., Southworth, J., Ciceri, S., *et al.* 2014, *MNRAS*, **443**, 2391.
Mancini, L., Southworth, J., Raia, G., *et al.* 2017, *MNRAS*, **465**, 843.
Mancini, L., Esposito, M., Covino, E., *et al.* 2018, *A&A*, **613**, A41.
McCullough, P.R., Crouzet, N., Deming, D., and Madhusudhan, N. 2014, *ApJ*, **791**, 55.
Montalto, M., Boué, G., Oshagh, M., Boisse, I., Bruno, G., and Santos, N.C. 2014, *MNRAS*, **444**, 1721.
Močnik, T., Southworth, J., and Hellier, C. 2017, *MNRAS*, **471**, 394.
Namekata, K., Maehara, H., Notsu, Y., *et al.* 2019, *ApJ*, **871**, 187.
Namekata, K., Davenport, J.R.A., Morris, B.M., *et al.* 2020, *ApJ*, **891**, 103.
Nascimbeni, V., Piotto, G., Bedin, L.R., and Damasso, M. 2011a, *A&A*, **527**, A85.
Nascimbeni, V., Piotto, G., Bedin, L.R., *et al.* 2011b, *A&A*, **532**, A24.
Nascimbeni, V., Cunial, A., Murabito, S., *et al.* 2013, *A&A*, **549**, A30.
Netto, Y., and Valio, A. 2020, *A&A*, **635**, A78.
Nikolov, N., Henning, T., Koppenhoefer, J., *et al.* 2012, *A&A*, **539**, 159.
Pál, A., Bakos, G.Á., Fernandez, J., *et al.* 2009, *ApJ*, **700**, 783.
Petrucci, R., Jofré, E., Schwartz, M., *et al.* 2013, *ApJ*, **779**, L23.
Pont, F., Gilliland, R.L., Moutou, C., *et al.* 2007, *A&A*, **476**, 1347.
Rajpaul, V., Aigrain, S., Osborne, M.A., Reece, S., and Roberts, S. 2015, *MNRAS*, **452**, 2269.
Sada, P.V., Deming, D., Jennings, D.E., *et al.* 2012, *PASP*, **124**, 212.
Sada, P.V., and Ramón-Fox, F.G. 2016, *PASP*, **128**, 024402.
Sada, P.V. 2018, "American Astronomical Society" DPS meeting #50, 413.01.
Sanchis-Ojeda, R., Winn, J.N., Holman, M.J., *et al.* 2011, *ApJ*, **733**, 127.
Sato, B., Hartman, J.D., and Bakos, G.Á. 2012, *PASJ*, **64**, 97.
Scargle, J.D. 1982, *ApJ*, **263**, 835.

Shporer, A., Mazeh, T., Pont, F., *et al.* 2009, *ApJ*, **694**, 1559.
Silva, A.V.R. 2003, *ApJ*, **585**, L147.
Silva-Valio, A., Lanza, A.F., Alonso, R., and Barge, P. 2010, *A&A*, **510**, A25.
Siverd, R.J., Beatty, T., Pepper, J., *et al.* 2012, *ApJ*, **761**, 123.
Smith, A.M.S. 2015, *Acta Astronomica*, **65**, 117.
Sokov, E.N., Sokova, I.A., Dyachenko, V.V., *et al.* 2018, *MNRAS*, **480**, 291.
Southworth, J., Hinse, T., Burgdorf, M.J., *et al.* 2009a, *MNRAS*, **399**, 287.
Southworth, J., Hinse, T., Jorgensen, U.G., *et al.* 2009b, *MNRAS*, **396**, 1023.
Southworth, J., Mancini, L., Novati, S.C., *et al.* 2010, *MNRAS*, **408**, 1680.
Southworth, J., Bruni, I., Mancini, L., and Gregorio, J. 2012, *MNRAS*, **420**, 2580.
Southworth, J., Dominik, M., Jørgensen, U.G., *et al.* 2019, *MNRAS*, **490**, 4230.
Stevenson, K.B., Bean, J.L., Seifahrt, A., *et al.* 2014, *AJ*, **147**, 161.
Szabó, G.M., Kiss, L.L., Benkő, J.M., *et al.* 2010, *A&A*, **523**, A84.
Torres, G., Bakos, G.A., Kovács, G., *et al.* 2007, *ApJ*, **666**, L121.
Tregloan-Reed, J., and Southworth, J. 2013, *MNRAS*, **431**, 966.
Tregloan-Reed, J., Southworth, J., and Tappert, C. 2013, *MNRAS*, **428**, 3671.
Tregloan-Reed, J., Southworth, J., Burgdorf, M., *et al.* 2015, *MNRAS*, **450**, 1760.
Tripathi, A., Winn, J.N., Johnson, J.A., *et al.* 2010, *ApJ*, **715**, 421.
Turner, O.D., Anderson, D.R., Collier Cameron, A., *et al.* 2016, *PASP*, **128**, 064401.
Wilson, D.M., Gillon, M., Hellier, C., *et al.* 2008, *ApJ*, **675**, L113.
Winn, J.N., Holman, M., and Roussanova, A. 2007a, *ApJ*, **657**, 1098.
Winn, J.N., Holman, M.J., Henry, G.W., *et al.* 2007b, *AJ*, **133**, 1828.
Winn, J.N., Holman, M., Carter, J.A., *et al.* 2009, *AJ*, **137**, 3826.
Zaleski, S.M., Valio, A., Marsden, S.C., and Carter, B.D. 2019, *MNRAS*, **484**, 618.