

# A catalogue of over 10 million variable source candidates in ZTF Data Release 1

Eran O. Ofek,<sup>1</sup> Maayane Soumagnac<sup>1,2★</sup>, Guy Nir,<sup>1</sup> Avishay Gal-Yam,<sup>1</sup> Peter Nugent,<sup>2</sup> Frank Masci<sup>3</sup> and Shri R. Kulkarni

<sup>1</sup>Department of Particle Physics and Astrophysics, Weizmann Institute of Science, 76100 Rehovot, Israel

<sup>2</sup>Computational Cosmology Center, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

<sup>3</sup>Cahill Center for Astronomy and Astrophysics, California Institute of Technology, Pasadena, CA 91125, USA

Accepted 2020 August 17. Received 2020 July 27; in original form 2020 May 31

## ABSTRACT

Variable sources probe a wide range of astrophysical phenomena. We present a catalogue of over 10 million variable source candidates found in Data Release 1 (DR1) of the Zwicky Transient Facility (ZTF). We perform a periodicity search up to a frequency of  $160 \text{ d}^{-1}$ , and we classify the light curves into erratic and smooth variables. We also present variability indicators and the results of a periodicity search, up to a frequency of  $5 \text{ d}^{-1}$ , for about 1 billion sources in the ZTF-DR1 light curve data base. We present several new short-period ( $<90 \text{ min}$ ) candidates, and about 60 new dwarf nova candidates, including two candidate eclipsing systems. Both the 10 million variables catalogue and  $\sim 1$  billion source catalogue are available online in `catsHTM` format.

**Key words:** catalogues – novae, cataclysmic variables.

## 1 INTRODUCTION

Source variability allows us to probe a wide variety of astrophysical phenomena, from binary stars and exoplanets to stellar pulsations, cataclysmic and catastrophic events, and mass accretion by massive black holes. In addition, stellar variability is of interest as it contaminates searches for transients. Therefore, comprehensive lists of variable stars are of great importance.

Several productive variable star searches have been carried out. The Catalina Real Time Survey (CRTS) published over 60 000 periodic variables (e.g. Drake et al. 2014a, b, c, 2017). Other searches include the OGLE survey (e.g. Wozniak et al. 2002), All Sky Automated Survey (e.g. Pojmanski 1997, Pojmanski 2000), Vista Vía Láctea survey (VVV; e.g. Minniti et al. 2010), SDSS stripe 82 (e.g. Sesar et al. 2007), OGLE (e.g. Udalski, Szymański & Szymański 2015; Soszyński et al. 2017, 2019), and more. Recently, Chen et al. (2020) use the ZTF-DR2 to search periodic variables up to frequency of  $40 \text{ d}^{-1}$ . They identified about 781 000 periodic variable stars and classified them into several classes.

Here, we present a search for variable star candidates in the Zwicky Transient Facility (ZTF; Bellm et al. 2019a) Data Release 1 (DR1). Variable stars are selected based on their large photometric flux root mean square (rms), compared to other stars in the field, or on some indication of periodicity in their light curves. We provide the ZTF/DR1 photometric light curve catalogue of about 1.6 billion light curves, of about a billion unique sources, in `catsHTM` format (Soumagnac & Ofek 2018). For each source, we also list some variability attributes, calculate its periodogram up to a frequency of  $5 \text{ d}^{-1}$ , and provide the periodogram highest peak significance

and frequency. Next, we present a catalogue of about 10.7 million variable star candidates. For each variable candidate we provide a list of variability indicators and a periodicity search up to a frequency of  $160 \text{ d}^{-1}$ .

The structure of this paper is as follows. In Section 2, we describe the conversion of the ZTF light curves to binary HDF5 files, and the `catsHTM` catalogue of sources and their variability properties. In Section 3, we describe the selection of variable star candidates, while in Section 4, we present some selected results. We summarize in Section 5. The data underlying this article are available in the `catsHTM` repository.

## 2 ZTF-DR1 LIGHT CURVES IN CATSHTM FORMAT

ZTF (Bellm et al. 2019a; Graham et al. 2019) utilizes the 48-in Schmidt telescope on Mount Palomar, equipped with a  $47 \text{ deg}^2$  camera. The ZTF data processing is described in Masci et al. (2019). The ZTF pipeline produces `SEXTRACTOR` (Bertin & Arnouts 1996) and `DAOPHOT` (Stetson 1987) catalogues of sources in each image. Here, we use only the light curves based on the `DAOPHOT` pipeline. The ZTF photometry is calibrated against the Pan-STARRS1 catalogue (Chambers et al. 2016). A colour term is fitted, but the magnitudes are calculated assuming the sources have a colour of  $g - r = 0 \text{ mag}$ .

As part of its DR1, the ZTF collaboration released ascii files containing 1681 215 104 light curves of about  $10^9$  unique sources detected in the  $g$  and  $r$  bands. The light curves are for non-unique sources, as they may be detected on several bands, and in several ZTF fields. Each ascii file contains all the  $g$ - and  $r$ -bands light curves of all the stars in one ZTF field (see Bellm et al. 2019a, b). These files

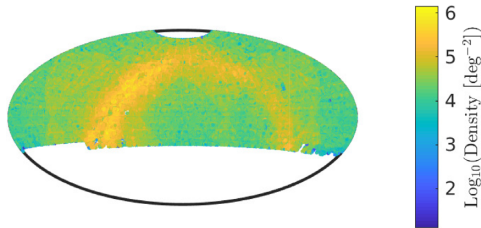
★ E-mail: maayane.soumagnac@gmail.com

**Table 1.** Columns in the ZTF/DR1 light curve data base.

Column	Description
HMJD	Heliocentric modified Julian Day.
Mag	ZTF magnitude.
MagErr	Magnitude error.
ColorCoef	Colour coefficient.
Flags	16bit flag.

**Table 2.** Columns in the `/IndAllLC` data set, in the HDF files of all the ZTF-DR1 light curves, and in the `catsHTM` catalogue of all ZTF-DR1 sources.  $\chi^2$  is the total chi-squared, calculated using the ZTF errors, where the number of degrees of freedom is  $N_{\text{ep}}$  minus one. The classical periodogram was calculated in frequency range of 0–5 d<sup>-1</sup>, with frequency steps of 1/600 d<sup>-1</sup>.

Column	Description
RA	J2000.0 right ascension (radians).
Dec	J2000.0 declination (radians).
I1	First line number of source light curve in the light curve HDF5 file.
I2	Last line number.
Nep	Number of epochs.
ID	Source ID.
FilterID	Filter ID (1 – g; 2 – r).
Field	Field ID.
RcID	CCD/quadrant ID.
MeanMag	Mean magnitude of source over all epochs.
StdMag	StD of source magnitude.
RStdMag	Robust StD of source magnitude.
MaxMag	Maximum magnitude minus mean magnitude.
MinMag	Mean magnitude minus minimum magnitude.
Chi2	$\chi^2$ , where $N_{\text{ep}}$ is the number of degrees of freedom.
MaxPower	The power (in units of sigma) of the highest peak in the periodogram.
FreqMaxPower	Frequency of maximal periodogram peak.

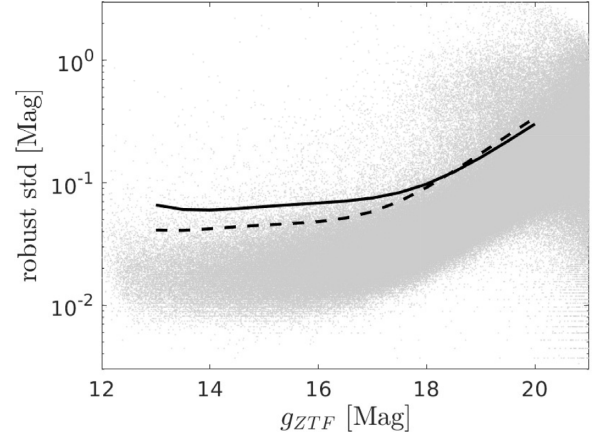
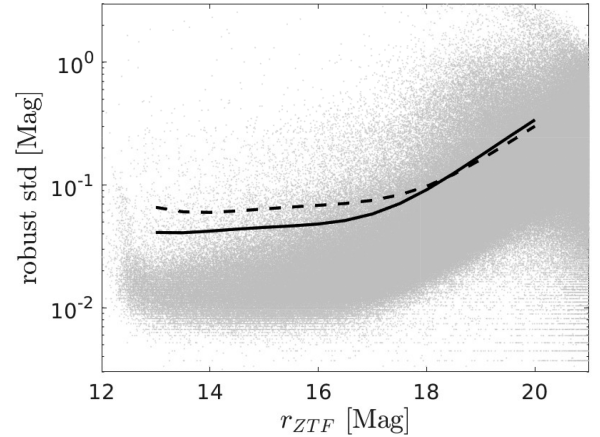
**Figure 1.** The sky surface density of sources in the ZTF-DR1 light-curve catalogue, in Aitoff projection. The source density is not uniform within a ZTF field/CCD quadrant, where the number of sources is slightly lower towards the edges of the CCD quadrants.

are available online.<sup>1</sup> We first converted these files into fast-access binary files in HDF5 format,<sup>2</sup> and have made this HDF5 catalogue public.<sup>3</sup> PYTHON and MATLAB programs to access this catalogue are available as part of the `catsHTM` toolbox. For each light curve, we also provide some basic variability indicators and list the power and frequency of the highest peak in the classical periodogram (e.g. Deeming 1975) as calculated up to a frequency of 5 d<sup>-1</sup>. The HDF5

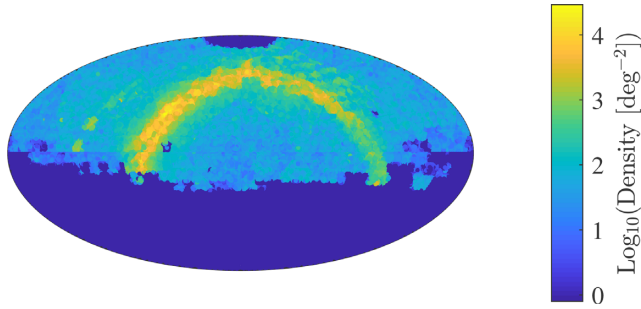
<sup>1</sup>[https://irsa.ipac.caltech.edu/data/ZTF/lc\\_dr1/](https://irsa.ipac.caltech.edu/data/ZTF/lc_dr1/)

<sup>2</sup><https://www.hdfgroup.org/solutions/hdf5/>

<sup>3</sup><https://euler1.weizmann.ac.il/catsHTM/>

**Figure 2.** The g-band robust rms of magnitude measurements of a source versus source mean magnitude for about  $3.9 \times 10^6$  random sources over the entire ZTF footprints. The solid line is the 4th degree polynomial fit plus six times the rms, calculated in 0.1 mag bins. The dashed line is like the solid line but for the *r*-band data. The lines data are given in Table 3.**Figure 3.** Like Fig. 2 but for the *r* band. The dashed line is for the *g* band.**Table 3.** The thresholds used for the selection of variable candidates. These are the  $6\sigma$  lines in the *g*- and *r*-band filters shown in Figs 2 and 3.

Mag (mag)	<i>g</i> -band threshold (mag)	<i>r</i> -band threshold (mag)
13	0.0658	0.0411
13.5	0.0605	0.0409
14	0.0598	0.0421
14.5	0.0615	0.0437
15	0.0639	0.0452
15.5	0.0662	0.0464
16	0.0683	0.0481
16.5	0.0708	0.0514
17	0.0751	0.0581
17.5	0.0831	0.0705
18	0.0976	0.0915
18.5	0.1222	0.1245
19	0.1609	0.1736
19.5	0.2189	0.2435
20	0.3016	0.3393



**Figure 4.** The sky surface density of variable candidates in the ZTF-DR1 light-curve catalogue, in Aitoff projection. There are some regions outside the Galactic plane with anomalous variable candidates surface density. This is likely due to some issues with the catalogue (e.g. underestimation of the noise in some specific fields).

file names are `ztfLCDR1.<field>.hdf5`, where `<field>` is a zero padded six digit ZTF field number. Each file contains two data sets.<sup>4</sup> The first data set, named `/AllLC`, is a matrix of all the light curves of all the stars in the field. The light curves are stored one after another. The matrix columns are listed in Table 1 (see also the ZTF-DR1 online documentation<sup>5</sup>). The second data set, named `/IndAllLC`, is a matrix in which each line corresponds to a light curve of one source/filter in the first data set. For each source/filter, we store some basic information and also calculate some light curve statistics, and the frequency and power of the highest peak in its periodogram. The matrix columns are listed in Table 2. Among the columns are the indices of the first and last line of the source's light curve in the first data set (i.e. `/AllLC`).

Next, we generated a `catsHTM`-format (Soumagnac & Ofek 2018) catalogue of all the sources/filters with their properties (as listed in Table 2). For each source/filter, we also stored the exact location of its light curve in the HDF5 light curves file. This combination of `catsHTM` and HDF5 files provides rapid search capabilities by coordinates and fast access to the entire light-curve data. Fig. 1 shows the sky source density of this catalogue.

The reformatted catalogues are available from the `catsHTM` website.<sup>6</sup> `catsHTM` access tools are available in MATLAB<sup>7</sup> and PYTHON.<sup>8</sup>

### 3 VARIABLE STAR SELECTION AND PROPERTIES

Here, we describe the variable star candidates selection, and their variability properties.

#### 3.1 Candidate selection

We first estimated the sky-averaged magnitudes rms of sources as a function of magnitude. In order to estimate this, we randomly selected 40 000 pointings across the sky, and for each pointing, we retrieved all the sources within 100 arcsec from its position. In total, about  $3.9 \times 10^6$  sources were selected within the ZTF footprints. Figs 2

and 3 show, in grey points, the robust<sup>9</sup> standard deviation (StD) of the light curves of these stars as a function of their mean magnitude in the *g* and *r* band, respectively. For the region between magnitude 13 and 20, in magnitude bins of 0.1 mag, we calculated the median and the robust StD of these points. The solid lines show the 4th degree polynomial fitted to the median-plus-six-times-the-robust-StD for the respective band, while the dashed line is the same but for the other band. The approximate 4th degree polynomials, representing these lines are given in Table 3. We selected variable-star candidates based on their high value of robust StD compared to other stars with similar magnitudes. A caveat of this approach is that we assumed the StD versus magnitude does not depend on sky position. In practice, this is likely not accurate. Verification of the ZTF photometric precision is further discussed in Masci et al. (2019).

Next, we selected variable-candidate light curves, based on two criteria: (i) Variable stars whose robust StD is larger than a band-dependent, magnitude-dependent threshold, and whose number of data points (in the selected field/band) is larger than 24. The magnitude-dependent threshold was selected to be the black solid lines in Figs 2 and 3 (see Table 3); or (ii) Stars for which the highest peak in the classical periodogram calculated up to  $5 \text{ d}^{-1}$  is above 12 (normalized to the noise) and the number of data points is larger than 24. These two criteria left us with 12 761 565 (non-unique sources) candidate light curves.

Next, we merged the sources with light curves into unique sources, if the angular distance between the sources was smaller than 1.5 arcsec. This merging process left us with 10 790 224 variable star candidates. Out of the  $\approx 10.8$  million stars, 9805 990 have one or more counterparts within 1.5 arcsec in the *GAIA*-DR2 catalogue, while 8530 535 have exactly one counterpart within 1.5 arcsec in the *GAIA*-DR2 catalogue. The stars without a counterpart are typically bad detections (e.g. ghosts, bad pixels) or transient sources. Fig. 4 shows the surface density of variable candidate on the celestial sphere. There are some regions outside the Galactic plane with anomalous variable-candidates surface density. This is likely due to some issues with the catalogue (e.g. underestimation of the noise in some specific fields).

For these  $\approx 10.8$  million stars, we calculated several additional properties, discussed in Section 3.2. This catalogue of variable star candidates along with their variability properties is available in `catsHTM` format<sup>10</sup> (Soumagnac & Ofek 2018).

#### 3.2 Properties of variable star candidates

For each variable star candidate, we calculated additional variability properties. The full list of attributes in our variable-star candidates catalogue is presented in Table 4. Here, we discuss selected properties and present their distributions.

##### 3.2.1 The five highest peaks in the periodogram

We re-calculated the classical periodogram, this time up to frequency of  $160 \text{ d}^{-1}$ . We choose to calculate the classical periodogram, and not the Lomb–Scargle periodogram (Lomb 1976; Scargle 1982), due to speed considerations. We kept in the catalogue the height and frequency of the five highest peaks in the periodogram.

<sup>4</sup>In the HDF5 terminology a data set can be regarded as a table of data.

<sup>5</sup><https://www.ztf.caltech.edu/page/dr1>

<sup>6</sup><https://euler1.weizmann.ac.il/catsHTM/>

<sup>7</sup><https://webhome.weizmann.ac.il/home/eofek/matlab/doc/catsHTM.html>

<sup>8</sup><https://github.com/maayane/catsHTM>

<sup>9</sup>Calculated from the central 50 per cent percentile of data and normalized to  $1\sigma$  assuming a Gaussian distribution.

<sup>10</sup><https://euler1.weizmann.ac.il/catsHTM/>

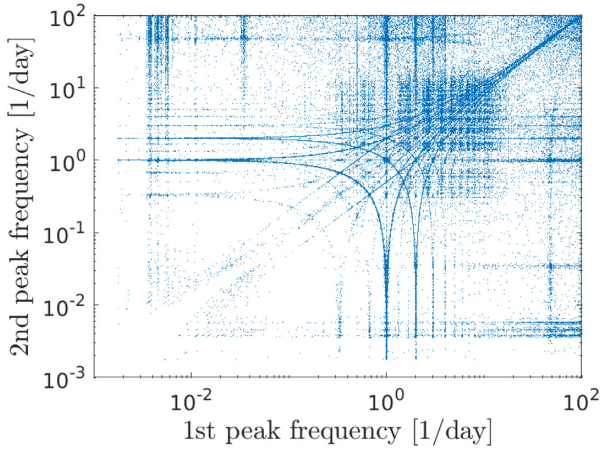
**Table 4.** Columns available in the ZTF-DR1 variable candidates catalogue.

Column	Property	Units	Explanation
1	RA	deg	J2000.0 right ascension in the ZTF catalogue (Epoch $\approx$ J2018).
2	Dec.	deg	J2000.0 declination in the ZTF catalogue (Epoch $\approx$ J2018).
3	Napp		Number of matched sources in ZTF-DR1.
4	SelBit		Flag indicating how the source was selected (e.g. STD, periodogram).
5	NcatDR1		Number of ZTF-DR1 LCs within 1.5 arcsec.
6	Nobs		Number of photometric observations in selected LCs.
7	Ntotobs		Total number of photometric observations (all fields/filters).
8	Filter		Selected filter (1 - <i>g</i> ; 2 - <i>r</i> ).
9	MeanMag	mag	Mean magnitude in selected LC.
10	MedMag	mag	Median magnitude.
11	MeanErr	mag	Mean error.
12	MedErr	mag	Median error.
13	StdMag	mag	Std of magnitudes.
14	MinMag	mag	Minimum magnitude.
15	MaxMag	mag	Maximum magnitude.
16	RstdMag	mag	Robust Std of magnitudes.
17	RangeT	d	Time range.
18	StdT	d	Std of time.
19	Chi2		$\chi^2$ of light curve, where Nobs-1 is the number of d.o.f.
20	Corr_MagColorCoef		Correlation coefficient between mag and colour term.
21	Mode1frac		Fraction of measurements in the 1st most populated magnitude bin.
22	Mode1mag	mag	Magnitude of the 1st most populated magnitude bin.
23	Mode2frac		Fraction of measurements in the 2nd most populated magnitude bin.
24	Mode2mag	mag	Magnitude of the 2nd most populated magnitude bin.
25	Mode3frac		Fraction of measurements in the 3rd most populated magnitude bin.
26	Mode3mag	mag	Magnitude of the 3rd most populated magnitude bin.
27	StdPoly10	mag	Std of residuals from fitting 10th degree polynomial.
28	MaxPSpower1		Power of 1st highest peak in periodogram.
29	MaxPSfreq1	d <sup>-1</sup>	Frequency of 1st highest peak in periodogram.
30	MaxPSpower2		Power of 2nd highest peak in periodogram.
31	MaxPSfreq2	d <sup>-1</sup>	Frequency of 2nd highest peak in periodogram.
32	MaxPSpower3		Power of 3rd highest peak in periodogram.
33	MaxPSfreq3	d <sup>-1</sup>	Frequency of 3rd highest peak in periodogram.
34	MaxPSpower4		Power of 4th highest peak in periodogram.
35	MaxPSfreq4	d <sup>-1</sup>	Frequency of 4th highest peak in periodogram.
36	MaxPSpower5		Power of 5th highest peak in periodogram.
37	MaxPSfreq5	d <sup>-1</sup>	Frequency of 5th highest peak in periodogram.
38	Ngaia		Number of counterparts in <i>GAIA</i> -DR2 within 1.5 arcsec.
39	MagG	mag	<i>GAIA</i> <i>G</i> -magnitude of nearest source.
40	MagBp	mag	<i>GAIA</i> <i>B<sub>p</sub></i> -magnitude.
41	MagRp	mag	<i>GAIA</i> <i>R<sub>p</sub></i> -magnitude.
42	Plx	mas	<i>GAIA</i> -DR2 parallax.
43	PlxErr	mas	<i>GAIA</i> -DR2 parallax error.
44	PM	mas yr <sup>-1</sup>	<i>GAIA</i> -DR2 total proper motion.
45	ExcessNoise	mas	<i>GAIA</i> -DR2 excess noise.
46	z_SDSS		SDSS redshift.
47	zErr_SDSS		SDSS redshift error.
48	class_SDSS		SDSS-spectrum object class.
49	subClass_SDSS		SDSS-spectrum object subclass.
50	objType_LAMOST		LAMOST-spectrum object type.
51	class_LAMOST		LAMOST-spectrum object class.
52	subClass_LAMOST		LAMOST-spectrum object subclass.
53	z_LAMOST		LAMOST redshift.
54	zErr_LAMOST		LAMOST redshift error.

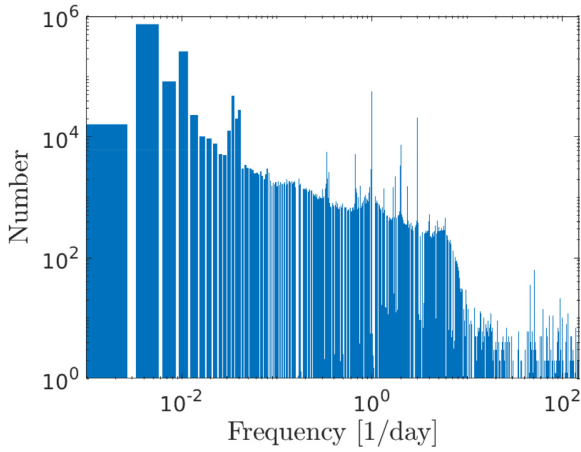
In Fig. 5, we show the first highest peak frequency versus second highest peak frequency in the periodogram for 5400 501 sources for which the highest periodogram peak is above 12. The vertical and horizontal lines that are over abundant in sources are mainly due to the window function of the data and aliases with the window

function. The lines near frequency of 0.004–0.006 d<sup>-1</sup> corresponds to several months, which is the typical yearly observation period of ZTF per source. The line near the frequency of 0.03 d<sup>-1</sup> corresponds to the lunar synodic period. Next, we can see a series of lines at about 0.33, 0.5, 1, 2, ..., and up to 13 d<sup>-1</sup>. These lines corresponds

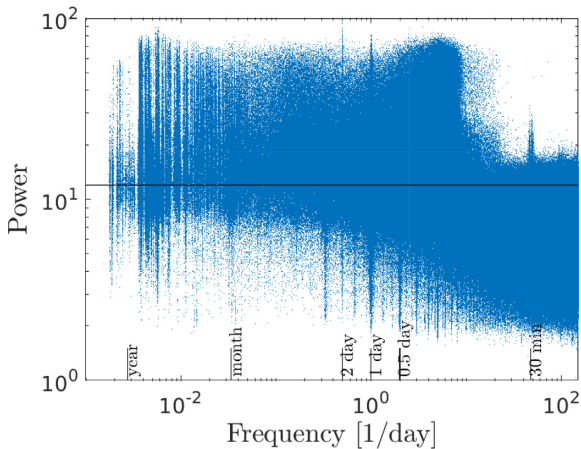




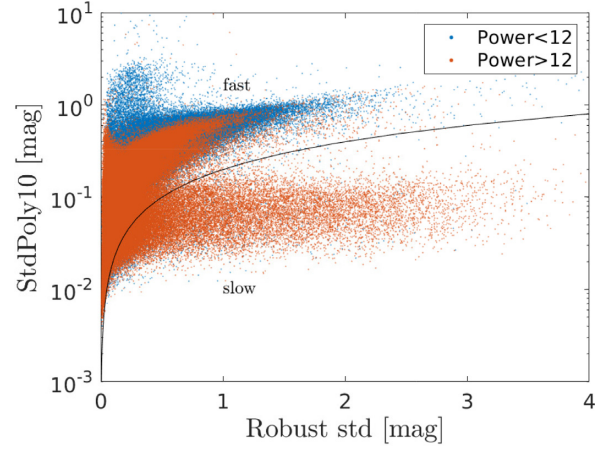
**Figure 5.** The first highest peak frequency versus second highest peak frequency in the periodogram, for sources in which the highest periodogram peak is above 12. As discussed in the text, the overabundance in specific frequencies is due to the aliasing of a true period with the window function of the data.



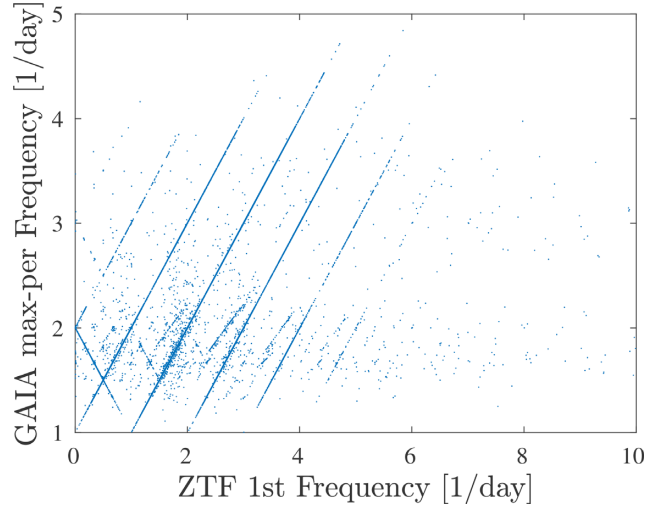
**Figure 6.** The histogram of the highest peak frequency for sources in which the power of the highest peak is larger than 12.



**Figure 7.** The frequency versus power of the highest peak in the periodogram for the sources in the variable candidates catalogue. The power is normalized by the variance, such that it roughly corresponds to the significance of the peak in units of the noise. The solid line shows the power equal 12 level.



**Figure 8.** The robust StD versus the *StdPoly10* (i.e. StD after fitting and subtracting a 10th degree polynomial from the light curve). Blue points correspond to sources whose highest peak in the periodogram is below 12, while red points denote sources whose highest peak in the periodogram is above 12.



**Figure 9.** The frequency of the highest peak in the periodogram in our catalogue versus the frequency of the maximum period among the fundamental and overtone periods in the *GAIA*-DR2 RR Lyr candidates catalogue.

to the whole multiplicities of the sidereal day period (i.e.  $0.997 \text{ d}$ ). Finally, the thick line near  $48 \text{ d}^{-1}$  corresponds to the ZTF minimal mean cadence of about 30 min. The concentration of sources near the whole multipliers of the sidereal day period is mainly due to periodic variables in which the true period (typically with long periods) generated a strong alias with the sidereal period. In fact, in such cases, two peaks are generated, one below and one above the sidereal period, and the frequencies of these two peaks are  $1.013 \pm 1/P \text{ d}^{-1}$ , where  $P$  is the true periodicity and  $1.013 \text{ d}^{-1}$  is the sidereal day frequency. This explains the bifurcation of many of the lines clearly seen in this diagram, and it allows us to estimate the true periodicity of such sources.

Fig. 6 presents the histogram of the highest peak frequency, for sources in which the power of the highest peak is larger than 12. The bins are per  $0.003 \text{ d}^{-1}$  frequency, but are plotted in logarithmic scale. The peaks discussed in the context of Fig. 5 are clearly visible. The

**Table 5.** List of selected short-period ( $<90$  min) variable candidates found in the ZTF variability catalogue. Med. Mag. is the median magnitude in the ZTF band. Period accuracy is about a few milliseconds. The comments are based on information from SIMBAD (Wenger et al. 2000).

J2000.0 RA (deg)	J2000.0 Dec. (deg)	Filter ( )	Med. Mag. (mag)	Period (min)	Comments ( )
290.81069	3.897 78	<i>g</i>	18.36	72.445 24	
288.28656	12.080 99	<i>g</i>	18.36	36.018 94	
288.67014	19.640 48	<i>g</i>	17.87	23.647 39	RGB star (SIMBAD)
322.73628	44.346 25	<i>g</i>	15.40	19.669 84	MGAB-V249; Kupfer et al. (2020b)
313.81656	46.851 80	<i>g</i>	17.75	28.737 26	Nearby X-ray source; Kupfer et al. (2020a)
320.62207	57.324 72	<i>g</i>	15.94	70.864 44	
326.03118	58.304 84	<i>r</i>	17.47	75.751 60	
5.74011	61.685 44	<i>r</i>	16.49	9.385 29	V1033 Cas (Nova)

jump in the abundance (and also the power, see Fig. 6) of sources around frequency of  $10 \text{ d}^{-1}$  is likely due to loss of efficiency above this frequency (due to the ZTF cadence), and the fact we are using the classical periodogram, which, unlike the Lomb–Scargle periodogram (Lomb 1976), does not normalize the power to the frequency response of the periodogram to a sine wave at that frequency.

Fig. 7 shows the periodogram highest peak versus power. The solid black line marks our periodicity threshold (12), while prominent frequencies are marked at the bottom.

### 3.2.2 Standard deviation after a high-order polynomial fit

We subtracted from the time of each light curve the mean time and divided it by the time range. Each time-normalized light curve was fitted with a 10th degree polynomial. Next, we calculated the standard deviation of the polynomial-subtracted light curve (i.e. *StdPoly10*). The motivation for obtaining this property is that variable stars with smooth and slowly evolving light curves (e.g. Mira stars) can be easily identified by their low *StdPoly10*, while eruptive variables [e.g. dwarf novae (DN)] tend to have a high standard deviation even after the removal of such a polynomial. The response of this estimator to eruptive variables is not uniform. The reason is that different sources were observed with a different time range and sampling.

Fig. 8 shows the robust StD versus the StD after fitting and removing the 10th degree polynomial from the light curve. The solid black line corresponds to  $y = 0.2x$ , where  $x$  and  $y$  are the plot axes. Blue points correspond to sources in which the highest peak in the periodogram is below 12, while red points denote sources in which the highest peak in the periodogram is above 12. This plot distinguishes between fast variables (mainly above the black line), and slow variables (mainly below the black line). This plot shows that slow variables are mostly periodic. Furthermore, periodic variables that are classified as fast tend to have low amplitudes compared to the population of fast variables.

### 3.2.3 Peaks in the magnitude histogram

For each light curve, we calculated a histogram of the magnitude measurements with 0.25 mag bins. We selected the three bins with the highest number of magnitude measurements. For each one of these three bins, we stored the magnitude at the middle of the bin (i.e. *mode1mag*, *mode2mag*, *mode3mag*) and the fraction of measurements in the bin compared to the total number of measurements (i.e. *mode1frac*, *mode2frac*, *mode3frac*). These indicators

maybe able to assist in identifying variable stars that mainly move from one state to another; examples include eclipsing binaries and DN. It can also be used to identify stars that are likely flagged as variable candidates due to outliers or due to the mistaken matching of multiple sources as a single star.

## 3.3 Comparison with other catalogues

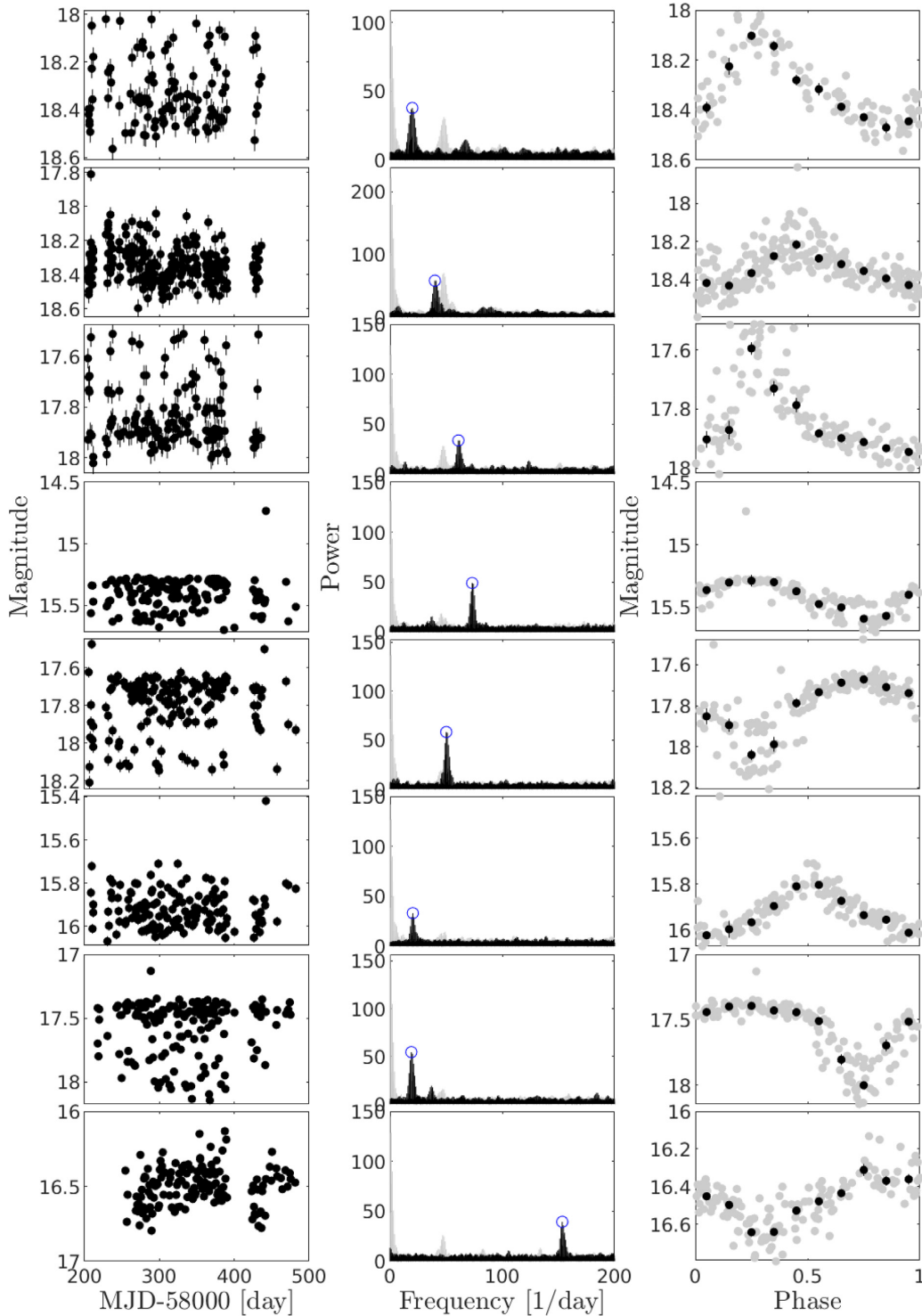
We attempted comparing the content of our catalogue to other variable source catalogues. We chose to compare our catalogue to catalogues which have photometric precision superior to that of the ZTF photometric calibration. Since the largest available homogeneous variable star candidate catalogue includes mainly periodic variables, we compared our catalogue with the *GAIA*-DR2 RR Lyrae variable candidates list (Gaia Collaboration 2018). Molnár et al. (2018) estimated that this catalogue is 70 per cent–78 per cent complete, and has purity of 92 per cent to 98 per cent. The *GAIA*-DR2 RR Lyr catalogue contains about 56 000 RR Lyr candidates brighter than  $G$  mag of 20, and above Declination of  $-30$  deg. In our catalogue, we found matches for 27 000 sources in the *GAIA*-DR2 RR Lyr catalogue. In Fig. 9, we show the frequency of the highest peak in the periodogram in our catalogue versus the frequency of the maximum period among the fundamental and overtone periods in the *GAIA*-DR2 catalogue. For a large fraction of the sources, the sources sit on straight lines in this diagram. This demonstrate (see also Fig. 5) that around periods of 1 d, our period estimates are affected by the sidereal day aliasing. As demonstrated in Fig. 5, in order to estimate the correct periodicity of the sources in our catalogue one needs to use the frequencies of the highest peaks in the periodogram.

## 4 SELECTED RESULTS

Here, we present some selected results related to variable stars. In Sections 4.1 and 4.2, we present a preliminary search for short-period variables, and DN, respectively. Variability across the HR diagram is presented in section 4.3.

### 4.1 short period variability

Given the typical minimum cadence of ZTF (about 30 min), our period search is not very effective above frequencies of  $12 \text{ d}^{-1}$  (i.e. Nyquist frequency for 30-min sampling). Since the ZTF cadence is not strict, but has some randomness, it is possible to identify periods shorter than about 1 h. However, any such search for short-period variability will have low efficiency and a high false-alarm rate.



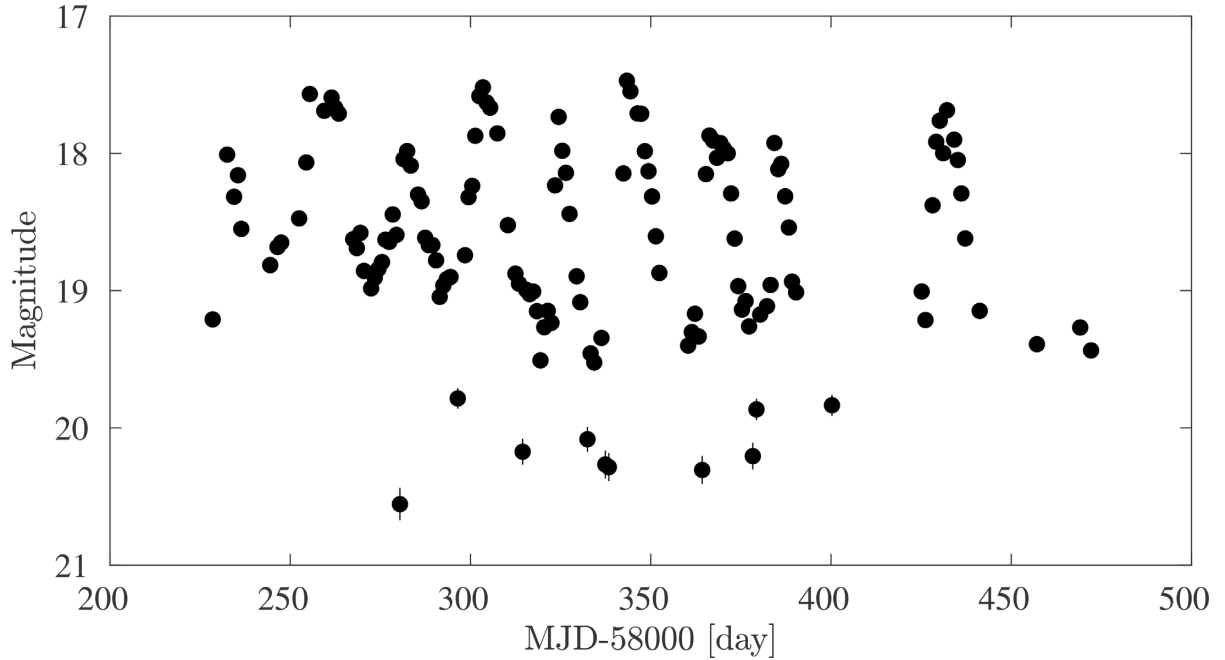
**Figure 10.** The light curves of eight short-period ( $< 90$  min) variable candidates – row per source. The left column shows the light curves. The middle column presents the power spectra (black), the window functions (grey), and the strongest power-spectrum peaks (blue circles). The right column shows the light curves folded into the period (strongest peak in the periodogram). Grey points are the measurements while black dots are median magnitudes in bins of 0.1 of the period. Candidate coordinates and periods are listed in Table 5 (in the same order).

In order to test the possibility of identifying short-period variable candidates, we selected sources using the following criteria: (i) The power of the highest peak in the periodogram  $> 25$ . (ii) The frequency of the first and second highest peaks in the periodogram are  $> 18 \text{ d}^{-1}$ . (iii) Robust StD  $> 0.02$  mag. (iv) StdPoly10  $> 1.1$ . (v) Magnitude

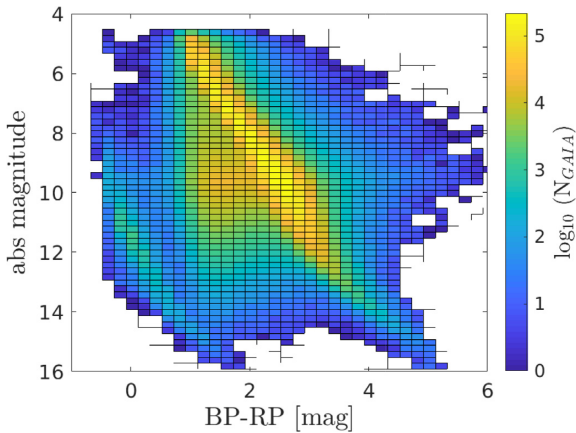
$G < 18$ . This resulted in 63 candidates. We inspected all these candidates by eye and found eight promising short-period candidates. The properties of these candidates are listed in Table 5, while their light curves, periodograms, and folded light curves, are presented in Fig. 10.

**Table 6.** List of selected DN candidates found in the ZTF variability catalogue. First five lines are shown. The full table is available in the electronic version of the paper.

J2000.0 RA (deg)	J2000.0 Dec. (deg)	Filter ( <i>i</i> )	Min. Mag. (mag)	Max. Mag. (mag)	Med. Mag. (mag)	Comments ( <i>i</i> )
313.82849	−16.44455	<i>r</i>	16.65	20.19	18.74	SDSS J205518.83−162640.4
109.17199	−6.94683	<i>g</i>	13.98	19.75	18.95	FQ Mon (Mira cand), but looks like a DN
99.22749	0.03812	<i>g</i>	12.37	17.29	16.65	CW Mon (DN)
280.32467	4.15413	<i>r</i>	17.81	20.90	20.36	
93.38401	6.95214	<i>g</i>	17.39	21.46	18.37	MGAB-V343



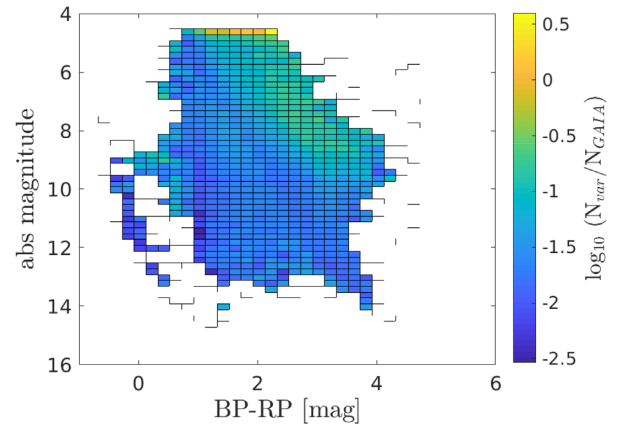
**Figure 11.** The light curve of the best eclipsing dwarf nova candidate, ZTF J222959.198+521507.78. A possible period for the eclipses is 5.9186 h. However, this is not a unique solution and further observations are required.



**Figure 12.** The density of *GAIA* sources across the HR diagram (see text for details). The colour coding is the  $\log_{10}$  of number of sources in  $0.2 \times 0.2$  mag-bins. The plot is based on about 9.8 million sources.

#### 4.2 Dwarf novae candidates

We conducted a partial search for dwarf nova eruptions. We used the following selection criteria: (i) The power of the highest peak in the periodogram  $>8$ . (ii) The frequency of the highest peak



**Figure 13.** The density of variable sources divided by the density of *GAIA* sources across the HR diagram (see text for details). The colour coding is the  $\log_{10}$  of the number ratio in  $0.2 \times 0.2$  mag-bins. The plot is based on about 0.27 million sources.

in the periodogram is between  $0.01$  and  $0.1 \text{ d}^{-1}$ . (iii) Robust  $\text{Std} > 0.2 \text{ mag}$ . (iv)  $\text{StdPoly10} > 0.3 \text{ mag}$ . (v) Magnitude range  $> 2.5 \text{ mag}$ . This resulted in 452 candidates. We inspected the candidates light curves by eye and selected 76 DN candidates. The



selection was based on the similarity of the light curves to other known classical DN. About 60 of the candidates were previously unknown. The candidates are listed in Table 6. Two candidates show some evidence of eclipses. The light curve of the best eclipsing dwarf nova candidate, ZTF J222959.198+521507.78, is shown in Fig. 11.

### 4.3 Variability across the HR diagram

We calculate the density of variable stars across the HR diagram. We selected *GAIA* sources with parallax above 2 mas, *G* magnitude between 13 and 18, and calculated their density in the colour magnitude diagram with  $0.2 \times 0.2$  mag-bins. Fig. 12 presents the density of these *GAIA* sources across the HR diagram. Next, we selected ZTF variable candidates with the same properties (i.e. brighter than *G* magnitude 18 and parallax above 2 mas). Fig. 13 shows the density ratio, across the HR diagram, between variable sources and the *GAIA* sources.

These plots suggest that variable star candidates in our catalogue are found preferentially outside the main sequence. This also serve as a validation for our catalogue. For example, in case our catalogue would contain a large number of false candidates due to poor photometric precision of the ZTF catalogue, then we would expect these sources to track the *GAIA* source density.

## 5 SUMMARY

We present a catalogue of over 10 million variable star candidates identified in the ZTF-DR1 light curves data base. We provide the entire ZTF-DR1 light curves data base in HDF5 format and an index catalogue in `catsHTML` format. We also generated a catalogue of variable candidates in `catsHTML` format. We provide some variability indicators for the variable star candidates, including a division into erratic (fast) and smooth (slow) variable sources.

We conducted a preliminary search for variable sources with a periodicity shorter than 90 min. We identified eight candidates, one of which is a known AM CVn star. We also searched for outbursting DN, and found about 60 new DN candidates. This catalogue may also be useful for transient searches where a catalogue of stellar variability helps in removing false alarms.

## ACKNOWLEDGEMENTS

EOO is grateful for the support by grants from the Israel Science Foundation, Minerva, Israeli Ministry of Technology and Science, the US-Israel Binational Science Foundation, Weizmann UK, Weizmann-Yale, and the Weizmann-Caltech grants.

## REFERENCES

- Bellm E. C. et al., 2019a, *PASP*, 131, 018002  
 Bellm E. C. et al., 2019b, *PASP*, 131, 068003  
 Bertin E., Arnouts S., 1996, *A&AS*, 117, 393  
 Chambers K. C. et al., 2016, preprint (arXiv:1612.05560)  
 Chen X., Wang S., Deng L., de Grijs R., Yang M., Tian H., 2020, preprint (arXiv:2005.08662)  
 Deeming T. J., 1975, *Ap&SS*, 36, 137  
 Drake A. J. et al., 2014a, *ApJS*, 213, 9  
 Drake A. J. et al., 2014b, *MNRAS*, 441, 1186  
 Drake A. J. et al., 2014c, *ApJ*, 790, 157  
 Drake A. J. et al., 2017, *MNRAS*, 469, 3688  
 Gaia Collaboration, 2018, *A&A*, 616, A1  
 Graham M. J. et al., 2019, *PASP*, 131, 078001  
 Kupfer T. et al., 2020a, *ApJ*, 898, 25  
 Kupfer T. et al., 2020b, *ApJ*, 891, 45  
 Lomb N. R., 1976, *Ap&SS*, 39, 447  
 Masci F. J. et al., 2019, *PASP*, 131, 018003  
 Minniti D. et al., 2010, *New Astron.*, 15, 433  
 Molnár L., Plachy E., Juhász Á. L., Rimoldini L., 2018, *A&A*, 620, A127  
 Pojmanski G., 1997, *Acta Astron.*, 47, 467  
 Pojmanski G., 2000, *Acta Astron.*, 50, 177  
 Scargle J. D., 1982, *ApJ*, 263, 835  
 Sesar B. et al., 2007, *AJ*, 134, 2236  
 Soszyński I. et al., 2017, *Acta Astron.*, 67, 103  
 Soszyński I. et al., 2019, *Acta Astron.*, 69, 321  
 Soumagnac M. T., Ofek E. O., 2018, *PASP*, 130, 075002  
 Stetson P. B., 1987, *PASP*, 99, 191  
 Udalski A., Szymański M. K., Szymański G., 2015, *Acta Astron.*, 65, 1  
 Wenger M. et al., 2000, *A&AS*, 143, 9  
 Wozniak P. R., Udalski A., Szymański M., Kubiak M., Pietrzyński G., Soszynski I., Zebrun K., 2002, *Acta Astron.*, 52, 129

## SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

**Table 6.** List of selected DN candidates found in the ZTF variability catalogue.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.