







# Variability, periodicity, and contact binaries in *WISE*

Evan Petrosky <sup>1</sup>, <sup>1</sup>★, Hsiang-Chih Hwang <sup>1</sup>, Nadia L. Zakamska <sup>1</sup>, Vedant Chandra <sup>1</sup>  
and Matthew J. Hill <sup>2</sup>

<sup>1</sup>*Department of Physics & Astronomy, Johns Hopkins University, 3400 N Charles St, Baltimore, MD 21218, USA*

<sup>2</sup>*Knovita, 1405 Point Street, Baltimore, MD 21231, USA*

Accepted 2021 February 23. Received 2021 February 23; in original form 2020 December 8

## ABSTRACT

The time-series component of *Wide-field Infrared Survey Explorer (WISE)* is a valuable resource for the study of variable objects. We present an analysis of an all-sky sample of  $\sim 450\,000$  AllWISE+NEOWISE infrared light curves of likely variables identified in AllWISE. By computing periodograms of all these sources, we identify  $\sim 56\,000$  periodic variables. Of these,  $\sim 42\,000$  are short-period ( $P < 1$  d), near-contact, or contact eclipsing binaries, many of which are on the main sequence. We use the periodic and aperiodic variables to test computationally inexpensive methods of periodic variable classification and identification, utilizing various measures of the probability distribution function of fluxes and of time-scales of variability. The combination of variability measures from our periodogram and non-parametric analyses with infrared colours from *WISE* and absolute magnitudes, colours, and variability amplitude from *Gaia* is useful for the identification and classification of periodic variables. Furthermore, we show that the effectiveness of non-parametric methods for the identification of periodic variables is comparable to that of the periodogram but at a much lower computational cost. Future surveys can utilize these methods to accelerate more traditional time-series analyses and to identify evolving sources missed by periodogram-based selections.

**Key words:** methods: statistical – catalogues – binaries: eclipsing – binaries: general – stars: variables: general – binaries: close.

## 1 INTRODUCTION

The study of the variable sky can yield a wealth of information on a wide range of astronomical objects such as asteroids, exoplanets, stars, and active galactic nuclei. As a result of this broad importance, over the last 20 yr, there has been significant investment in high-cadence variability surveys both on the ground (OGLE – Udalski 2003; ASAS – Rucinski 2006; Catalina – Drake et al. 2009; ASAS-SN – Shappee et al. 2014; Kochanek et al. 2017; Jayasinghe et al. 2018; ZTF – Bellm et al. 2019; LSST – Ivezić et al. 2019) and in space (*Kepler* – Borucki et al. 2010; Koch et al. 2010; *TESS* – Ricker et al. 2015; *Gaia* – Gaia Collaboration 2016, 2018a).

Considerable effort goes into the sorting of data from these surveys to first differentiate between variable and non-variable objects and then to classify different types of variability. Traditionally, variable objects are classified based on the similarity of their light curves and colours to known variable prototypes (Gaia Collaboration 2019). Often, a time-series analysis tool such as a periodogram is run on objects displaying variability to differentiate between periodic and aperiodic variables. This step requires a clear understanding of the effects of the observing cadence on period recovery and a careful weighing of the pros and cons of different period-search algorithms. The results of the periodogram-based analysis are often taken, together with measures of light curve morphology and other characteristics of the source (e.g. colour) and used as inputs into a classifier (e.g. Debosscher et al. 2007, 2009; Sarro et al. 2009; Dubath

et al. 2011; Richards et al. 2011, 2012; Masci et al. 2014; Kim & Bailer-Jones 2016; Eyer et al. 2019; Jayasinghe et al. 2019a, b)

Periodograms, light-curve fitting, and machine learning classification are potent tools and will continue to be important to the astronomical community. Nevertheless, one downside of these methods is that they are often survey-dependent, difficult to physically interpret, and involve a lengthy learning curve and a lot of computational power in order to implement. To help mitigate this problem, some previous works have used non-parametric variability measures (e.g. Kinemuchi et al. 2006; Drake et al. 2013, 2014a, 2017; Palaversa et al. 2013; Rimoldini 2014; Findeisen, Cody & Hillenbrand 2015; Hillenbrand & Findeisen 2015; Torrealba et al. 2015)

In this work, we are particularly interested in eclipsing stellar binaries. Binary stars undergo interesting evolution (Stepien 1995; Fabrycky & Tremaine 2007; Duchêne & Kraus 2013; Ivanova et al. 2013; Borkovits et al. 2016; Moe & Di Stefano 2017; Hwang et al. 2020) and are relevant for a wide array of astronomical phenomena. They have been used in cosmology as distance indicators (Riess et al. 2011), in stellar astrophysics as testing grounds for precision stellar evolutionary models (Pietrinferni et al. 2004), and can even host planets (Doyle et al. 2011). Binary systems have been linked to transients such as Luminous Red Novae (Tylanda et al. 2011; Kasliwal 2012) and are thought to serve as progenitors for some of the most fascinating objects in the Universe – ultra-compact binaries of white dwarfs, neutron stars and black holes, the tantalizing source population for type Ia supernovae, kilonovae, and gravitational waves (Weisberg, Nice & Taylor 2010; Knigge, Baraffe & Patterson 2011; Maoz, Mannucci & Nelemans 2014; Postnov & Yungelson 2014; Belczynski et al. 2016; Brown et al. 2016; Abbott et al. 2017; Cowperthwaite et al. 2017; Smartt et al. 2017; Temmink et al. 2020).

\* E-mail: [epetros1@jhu.edu](mailto:epetros1@jhu.edu)

Many surveys have associated eclipsing binary (EB) catalogues (e.g. ASAS – Paczyński et al. 2006; *Kepler* – Kirk et al. 2016; Catalina – Drake et al. 2014a; OGLE – Soszyński et al. 2016; ASAS-SN – Jayasinghe et al. 2019a). Future *Gaia* data releases will include an all-sky EB catalogue and the current release identifies a variety of other types of variability (Eyer et al. 2018; Holl et al. 2018; Roelens et al. 2018; Molnár et al. 2018; Mowlavi et al. 2018; Gaia Collaboration 2019; Rimoldini et al. 2019; Clementini et al. 2019; Siopis et al. 2020).

In this work, we use non-parametric light-curve analysis techniques in conjunction with more traditional time-series analysis methods to study variability and identify a sample of eclipsing binaries in data from the *Wide-field Infrared Survey Explorer* (*WISE*; Wright et al. 2010; Mainzer et al. 2011). The all-sky and long-term coverage, decreased effects of extinction compared to optical wavelengths, and non-uniform cadence probing a wide range of variability time-scales (Hoffman et al. 2012) make *WISE* a unique probe of Galactic variability. Previously, Chen et al. (2018) used *WISE* to identify  $\sim 50\,000$  periodic variable candidates of which  $\sim 42\,000$  were binaries. Here, we expand on this work and present an analysis of  $\sim 450\,000$  light curves of variables selected using AllWISE variability metrics based on r.m.s. flux variations. Our larger period search grid allows us to detect short-period objects that were missed by Chen et al. (2018). We also employ a different period-search algorithm, use data from a more recent release, and cross-match our results with *Gaia* DR2. In the end, we identify an all-sky sample of  $\sim 56\,000$  periodic variable candidates, of which  $\sim 51\,000$  are binaries. We also present the calculation of various non-parametric variability measures for the remaining 394 000 aperiodic variables.

In Section 2, we discuss the *WISE* data and our time-series analysis. In Section 3, we explore the contents of our periodic variable selection and display the results of various cross-matches. In Section 4, we introduce our non-parametric measures and use them to classify periodic variables. In Section 5, we discuss main-sequence (MS) binaries, short-period objects, extra-galactic and young-stellar-object variability, and the application of non-parametric methods to the identification of periodic variables. We conclude in Section 6. Throughout this paper, *WISE* magnitudes are quoted on the Vega system, and the following conversions apply:  $W_{iAB} = W_{iVega} + \Delta m_i$ , with  $\Delta m_i = (2.699, 3.339, 5.174, 6.620)$ .<sup>1</sup>

## 2 WISE DATA AND METHOD

### 2.1 WISE mission

The *WISE* was launched in December 2009 and conducted observations of the entire sky in bands centred on 3.4, 4.6, 12, and 22  $\mu\text{m}$  (W1, W2, W3, and W4, respectively) until the end of September 2010 when it ran out of coolant (Wright et al. 2010; Mainzer et al. 2011). From October 2010 to February 2011 it conducted observations in the 3.4 and 4.6  $\mu\text{m}$  bands as a part of the NEOWISE post-cryogenic mission (Mainzer et al. 2011). The spacecraft was in hibernation from the end of the post-cryogenic mission until December 2013 when it was reawakened as a part of the NEOWISE Reactivation (NEOWISE-R) mission (Mainzer et al. 2014). The AllWISE data release includes data from both the original mission and the post-cryogenic mission as of 2013,<sup>2</sup> whereas the NEOWISE Reactivation

2019 Data Release includes all of the data from the time that the spacecraft was awakened from hibernation until 2019.

The *WISE* spacecraft orbits the Earth with a period of  $\sim 5700$  s ( $\sim 0.066$  d = 1.6 h)<sup>3</sup> on a polar orbit – near the dividing line between night and day – and always looks away from the Earth. Every six months, it images the same portion of the sky and obtains at least eight passes on each point of sky due to the partial overlap of the field of view on consecutive orbits. The sources outside the ecliptic are more frequently observed. The cadence of the data is such that every six months there is a collection of  $\sim 10$  data points that are each spaced one orbit apart.

The data pre-processing and extraction of the light curves are different for the AllWISE and NEOWISE releases. AllWISE stacks all the scans, identifies the objects, and then measures the per-scan magnitude of each object at a fixed position. In contrast, NEOWISE identifies objects and measures their photometry in individual scans without stacking. Mainzer et al. (2014) find systematic changes in W1 between AllWISE and NEOWISE-R to be 0.01 magnitude for sources with  $8 \text{ mag} < W1 < 14 \text{ mag}$  and on the order of 0.1 magnitude for sources with  $14 \text{ mag} < W1 < 15 \text{ mag}$ . Outside this range there are magnitude dependent systematic offsets between AllWISE and NEOWISE data. We limit our study to  $8 \text{ mag} < W1 < 15 \text{ mag}$  to ensure concordant AllWISE and NEOWISE measurements and to exclude saturated sources (Mainzer et al. 2014; Nikutta et al. 2014). We also explicitly apply a cut to ensure that the difference between the mean magnitude of AllWISE and NEOWISE be less than 0.15 mag because the time-series analysis is not reliable in the case of a large magnitude offset. A typical individual exposure for sources in the range  $8 \text{ mag} < W1 < 15 \text{ mag}$  has an uncertainty of  $\sim 0.03$  mag. A plot of the r.m.s. of single exposure flux measurements versus W1 magnitude is shown in fig. 8 of Mainzer et al. (2014). In the range  $8 \text{ mag} < W1 < 15 \text{ mag}$ , individual data point uncertainties tend to increase as a function of W1 magnitude. We incorporate these error measurements into our analysis.

### 2.2 WISE light curves

*WISE* reports a measure of the flux variation in each band based on AllWISE single-epoch photometry. Each source is assigned a single-digit `var_flg` ranging from 0 to 9 in each band, such that the probability that the object’s flux does not vary in said band is  $\propto 10^{-\text{var\_flag}}$  (Hoffman et al. 2012).

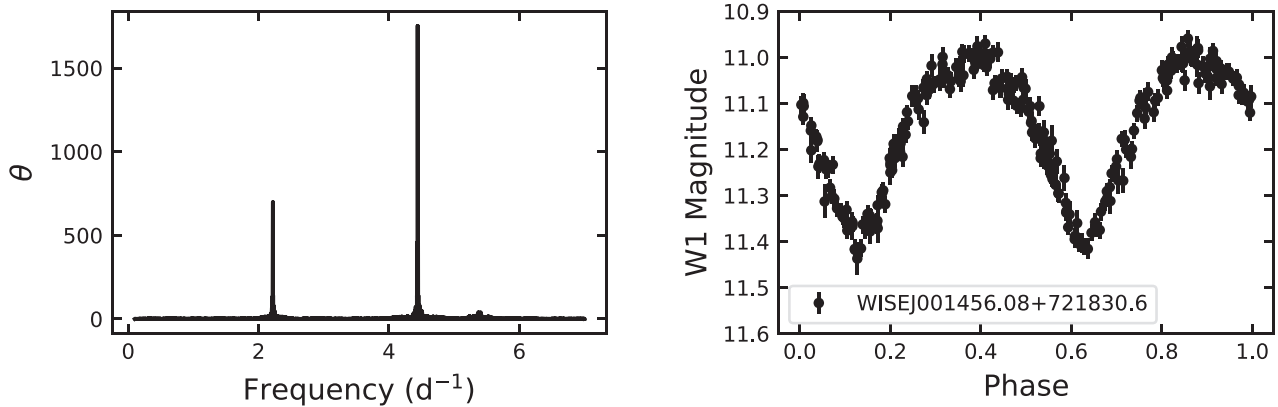
We select *WISE* variable sources having  $W1 \text{ var\_flg} \geq 6$ . We find that that proportion of variable objects found to be periodic via periodogram analysis drops steadily as `var_flg` decreases from  $\sim 23$  per cent for sources with `var_flg` == 9 to  $\sim 5$  per cent for sources with `var_flg` == 6. We elect not to extend the analysis to sources with `var_flg` == 5 because doing so would increase our storage and computational overhead by  $\sim 60$  per cent and we expect that a low proportion ( $< 5$  per cent) would be periodic. We only consider sources having `cc_flags` == 0 to ensure that there are no imaging artefacts and `ext_flg`  $\leq 1$  to ensure that it is not an extended source. After applying this selection, we download  $\approx 500\,000$  light curves from the AllWISE multi-epoch photometry table and the NEOWISE-R single exposure source table using a 1 arcsec matching distance.

Due to the  $\sim 3$  yr gap between AllWISE and NEOWISE-R, it is possible that some high-proper motion AllWISE sources are not matched to their NEOWISE-R counterpart and are omitted from our

<sup>1</sup>[https://wise2.ipac.caltech.edu/docs/release/allsky/expsup/sec4\\_4h.html](https://wise2.ipac.caltech.edu/docs/release/allsky/expsup/sec4_4h.html)

<sup>2</sup><http://wise2.ipac.caltech.edu/docs/release/allwise/>

<sup>3</sup>[http://wise2.ipac.caltech.edu/docs/release/allsky/expsup/sec1\\_1.html](http://wise2.ipac.caltech.edu/docs/release/allsky/expsup/sec1_1.html)



**Figure 1.** Left: An example periodogram made with *WISE* data. The vertical axis shows the value of the statistic,  $\theta$ , for a given frequency. The higher the value of  $\theta$ , the lower the false alarm probability. The periodogram peak occurs at a frequency of  $f_{\max} = 4.44 \text{ d}^{-1}$  ( $P_{\text{mhaov}} = 0.225 \text{ d}$ ). Right: Phase-folded light curve of the same object. This object is an eclipsing binary with approximately symmetric eclipses. This means that the MHAOV period,  $P_{\text{mhaov}}$  i.e. the period picked out by the MHAOV periodogram, is half the actual orbital period. We fold this light curve with a frequency of  $0.5 \times f_{\max} = 2.22 \text{ d}^{-1}$  which corresponds to an orbital period of  $2 \times P_{\text{mhaov}} = 0.45 \text{ d}$ .

analysis. This could cause us to miss some interesting objects such as nearby, high-proper motion brown dwarfs. To estimate the magnitude of this effect, we utilize the AllWISE/*Gaia* DR2 cross-match of Marrese et al. (2019). Out of the  $\sim 200\,000$  cross-matched objects with `var_flg` 6 through 9 that satisfy the above selection, only  $\sim 200$  have a proper motion of above  $100 \text{ mas yr}^{-1}$  so we estimate that  $\sim 0.1$  per cent of our variables suffer from a missed cross-match due to high proper motion. For an analysis of high-proper motion sources in WISE see Luhman & Sheppard (2014).

To eliminate redundancies and possible extraneous matches in the table, the `allwise_centr` listed in the NEOWISE-R single exposure table is mapped to the `source_id_mf` in the AllWISE multi-epoch photometry table.<sup>4</sup> We exclude sources whose NEOWISE-R data is mapped to more than one AllWISE source. We also exclude AllWISE sources with no corresponding NEOWISE-R data because the number of AllWISE-only observations is not sufficient for our time-series analysis.

For the AllWISE multi-epoch photometry, the following selection is applied: `saa_sep` > 5.0 deg (image outside of the South Atlantic anomaly), `moon_masked` == 0000 (frame unaffected by light scattered off the moon), and `qi_fact` > 0.9 (only the highest quality frames).<sup>5</sup> In addition, points with `null` photometric measurement uncertainty or `null` for the reduced  $\chi^2$  of the W1 profile-fit are excluded. For NEOWISE analogous cuts are applied<sup>6</sup> and we also exclude points with `null` W1 profile fit signal-to-noise ratio. The number of data points reported in our tables is the total remaining after these quality cuts are applied. Example PYTHON code for light curve download with quality flags is available through Github.<sup>7</sup>

### 2.3 Time-series analysis

The periodogram is a time-series analysis tool that allows for the identification and characterization of periodic signals. For this paper, we use the multi-harmonic analysis of variance (MHAOV)

periodogram (Schwarzenberg-Czerny 1996), which has been shown to be high-performing in comparison to other algorithms (Graham et al. 2013). We fit the data with periodic, orthogonal functions and use a statistic,  $\theta$ , which is the ratio of the squared norm of the model over the squared norm of the residuals (Schwarzenberg-Czerny 2003), to quantify the quality of the fit (Schwarzenberg-Czerny 1998). The fitting procedure is carried out for a grid of test frequencies and a periodogram shows the dependence of the statistic value on the test frequencies. Fig. 1 shows an example periodogram in frequency space with a corresponding phase-folded light curve.

For three model parameters, the MHAOV periodogram is statistically equivalent to the classic Lomb–Scargle periodogram (Lomb 1976; Ferraz-Mello 1981; Scargle 1982; Schwarzenberg-Czerny 1999). We use five model parameters to allow for better sensitivity to anharmonic oscillations. This corresponds to fitting the data in real space with a Fourier series of two harmonics (fundamental plus one additional harmonic – Schwarzenberg-Czerny 1999, 2003; Lachowicz et al. 2006; Graham et al. 2013).

We adopt a frequency grid spacing of  $0.0001 \text{ d}^{-1}$  (Graham et al. 2013). We search the frequency range  $0.1\text{--}20 \text{ d}^{-1}$ . This lower bound is selected because, due to the *WISE* observing cadence and our requirement that sources have good phase coverage, we expect low sensitivity to long-period variables. With regards to the upper bound, this extension beyond the traditional Nyquist limit of  $\sim 7.6 \text{ d}^{-1}$  that corresponds to uniform sampling at the *WISE* orbital cadence is justified when the data is not uniformly spaced (see Section 2.6 for further discussion).

### 2.4 Periodogram peak significance

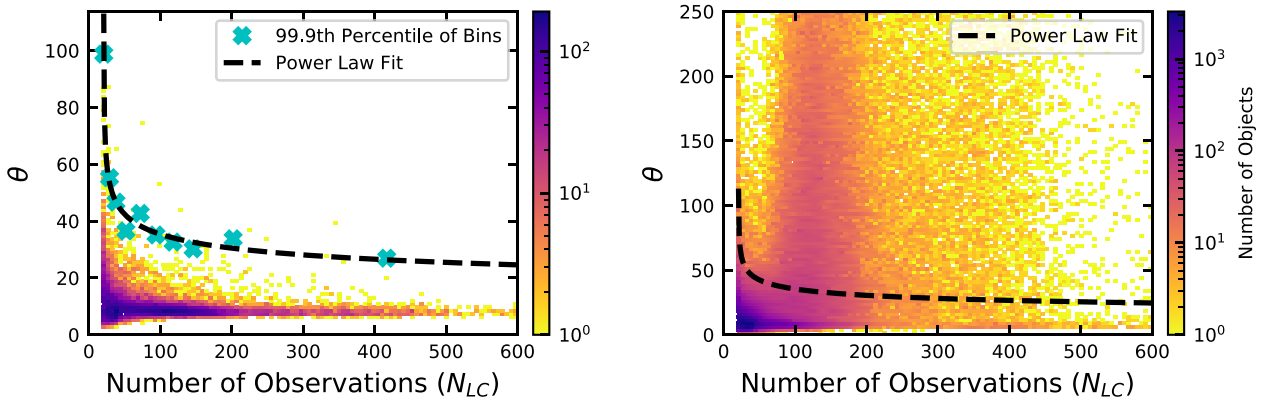
All the light curves in our sample have been determined by Hoffman et al. (2012) to exhibit variations in brightness. We seek to determine whether these variations are periodic in nature. When we run a periodogram on a source and observe a peak, we seek to reject the null hypothesis that the variation displayed by the light curve in question is pure noise and that the highest peak in the periodogram results from the chance alignment of random errors. Although some inroads have been made toward an analytical understanding of peak significance (e.g. Horne & Baliunas 1986; Koen 1990; Schwarzenberg-Czerny

<sup>4</sup>[http://wise2.ipac.caltech.edu/docs/release/neowise/expsup/sec2\\_1a.html](http://wise2.ipac.caltech.edu/docs/release/neowise/expsup/sec2_1a.html)

<sup>5</sup>[http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec3\\_1a.html](http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec3_1a.html)

<sup>6</sup>[http://wise2.ipac.caltech.edu/docs/release/neowise/expsup/sec2\\_3.html](http://wise2.ipac.caltech.edu/docs/release/neowise/expsup/sec2_3.html)

<sup>7</sup>[https://github.com/HC-Hwang/wise\\_light\\_curves](https://github.com/HC-Hwang/wise_light_curves)



**Figure 2.** Left: Distribution of maximum statistic values versus the number of observations in the light curve ( $N_{LC}$ ) for periodograms calculated on 20 000 white-noise light curves. The data are binned based on  $N_{LC}$ . The smallest bin contains about 1300 light curves and most bins have about 2 000. Cyan crosses show an estimate of the 99.9th percentile of the corresponding statistic values in each bin. The black dashed line shows the best-fitting power-law model using LMFIT (Newville et al. 2014). Right: Maximum statistic value from the periodogram versus  $N_{LC}$  for all *WISE* light curves with more than 17 data points and no peak at the *WISE* observing cadence. The black dashed line represents the power law fit from the left panel. For sources above this black line, we reject the null hypothesis that the observed light curve is consistent with white noise.

1999), the results depend in complex ways on the nature of the data and the chosen test frequencies.

We adapt the Monte Carlo method of Frescura, Engelbrecht & Frank (2007, 2008) to quantify peak significance. We generate 20 000 white-noise light curves that have means, flux deviations from mean, observing cadences, and individual data point photometric uncertainties that are representative of our actual sample. These light curves are variable at a level similar to that in the observed data, but contain no periodic signal by construction. To generate these light curves, we start with 20 000 randomly selected *WISE* light curves. For each light curve, there is an array of time measurements,  $t_i$ , an array of magnitude measurements,  $m_i$ , and an array of photometric uncertainties on each data point,  $\sigma(m_i)$  where  $i$  ranges from 1 to the number of data points in the light curve,  $N_{LC}$ . For each light curve, we calculate the mean of the magnitude distribution,  $\langle m \rangle$ , and define a measure of its width,  $w$ , as the difference between the 84.2th and 15.8th percentile (the percentiles marking, respectively, one standard deviation above and below the mean for a Gaussian distribution). Next, we create a Gaussian distribution with mean  $\langle m \rangle$  and standard deviation  $0.5 w$  and randomly draw  $N_{LC}$  values from this distribution to create a new array of Gaussian (white) noise data,  $noise_i$ . By substituting  $noise_i$  for  $m_i$  for each light curve, we create 20 000 white-noise light curves.

We run an MHAOV periodogram on each of the white-noise light curves using the same frequency grid and number of model parameters that were used on the actual data. The left-hand panel of Fig. 2 shows the resultant distribution of maximum statistic values from the periodogram ( $\theta$ ), as a function of the number of observations in the light curve ( $N_{LC}$ ). When  $N_{LC}$  is low, the light curve contains less information, it is harder to distinguish periodic and aperiodic signals, and it is expected that a higher statistic value is required to credibly reject the null hypothesis. To make the left-hand panel of Fig. 2 we bin the data based on  $N_{LC}$ . Every bin has  $\sim 2\,000$  simulated light curves save the last one on the right which has  $\sim 1300$ . In each bin, we study the resultant empirical cumulative distribution function of the calculated maximum statistic values and estimate the 99.9th percentile. Then, using LMFIT (Newville et al. 2014), we fit a power law model of the form  $\theta_{\text{cutoff}}(N_{LC}) = A(N_{LC} - N_0)^k$  to estimate the appropriate cut-off value as a function of  $N_{LC}$  to reject the null hypothesis. The best-fit values are  $A = 79 \pm 8$ ,  $N_0 = 20.7 \pm 0.2$ , and  $k$

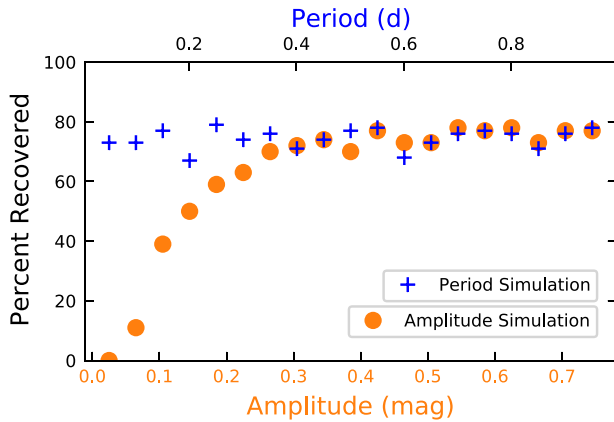
$= -0.18 \pm 0.03$ . For sources that lie above this best-fit line we reject the null hypothesis that the peak results from the chance-alignment of random errors for a white-noise light curve (see Fig. 2, right). For  $N_{LC} \leq 20$  the above fit diverges. Due to other cuts detailed below in Section 2.7, the minimum number of points that are required for periodic variable detection is 18. For sources with between 18 and 20 data points, we require a safe cut of 125 on maximum statistic. This only ends up adding one additional periodic source to the catalogue.

## 2.5 Completeness

The above cut on the maximum statistic seeks to limit the false alarm probability; however, it does not say much about the probability of periodicity itself nor the incidence of periodic signals that are passed over (VanderPlas 2018). The completeness, that is, the proportion of truly periodic signals we expect to detect, is a complex function of the nature of periodicity (Schwarzenberg-Czerny 1999), the observing cadence, and the data quality and quantity. It depends on the magnitude (worse for very faint and very bright objects), amplitude (worse for small amplitude), signal shape and phase (worse for short-duration pulses, especially if they occur in between observing epochs), period (worse for extremely short and long periods), and number of observations (worse for fewer observations).

Completeness is not the main focus of this project, and we do not attempt to characterize completeness across all of these parameters. Instead, we briefly consider just the special case of contact and near-contact binaries with sinusoidal and near-sinusoidal light curves. We explore how our ability to detect such sources varies as a function of signal amplitude and period. We randomly select 100 variable *WISE* light curves that have  $8 \text{ mag} < W1 < 15 \text{ mag}$ . We then simulate a sinusoidal signal centred at the *W1* magnitude of the light curve and sample it at the original cadence, preserving the individual data point uncertainty associated with each timestamp. In the first simulation, we fix the phase and choose an amplitude characteristic of our recovered periodic variables (0.4 mag). We pick a starting period of 0.051 d and calculate the MHAOV periodogram, phase-fold the result, and see if the source would have been classified as periodic (using the criteria detailed below in Section 2.7). We iterate over all light curves for a given period and then increase the period





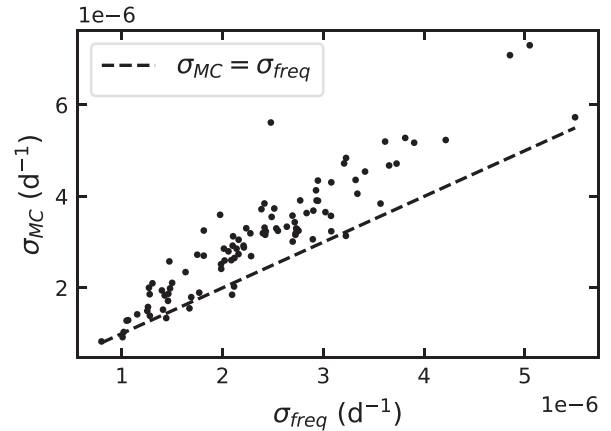
**Figure 3.** Per cent of signals correctly identified as periodic after running our method on a random sample of *WISE* light curves with a simulated sinusoidal signal. For the period simulation, the amplitude was fixed at a value representative of our close binaries. For the amplitude simulation, the period was fixed at a value representative of our close binaries.

and repeat until the period exceeds a value of 1 d. We choose periods between 0.05 and 1 d because the majority of our periodic sources are contained in this range. Next, we repeat the same procedure but this time fixing the period at a value of 0.22 d (a value typical of our close binaries) and varying the amplitude between value of 0.025 and 0.75 mag. Fig. 3 shows the results of these simulations. In summary, given our data and our method, we estimate that we can detect  $\sim 75$  per cent of near-sinusoidal signals with periods between 0.05 and 1 d and peak-to-peak amplitudes above 0.25 mag. As mentioned above, completeness is a function of signal shape, so these conclusions will not necessarily hold for strongly non-sinusoidal signals. We elect to focus solely on sinusoidal signals in this subsection because contact and near-contact EBs dominate our sample and many of their light curves are well-approximated by a sinusoid.

## 2.6 Period uncertainty

After determining that a periodogram has one or more significant peaks and represents a periodic variable candidate, we next determine whether it is indeed a truly periodic signal. If multiple periodogram peaks are present, it is important to determine which among them, if any, corresponds to the correct period. The observed light curve is a product of the continuous underlying signal and the discrete and unevenly spaced window function (i.e. observational cadence). Aliasing, or the correlation between frequencies equidistant from one-half of the inverse sampling rate (Nyquist frequency) is familiar in the case of evenly spaced sampling. Uniform observations at the *WISE* satellite period ( $\sim 95$  minutes) would correspond to a Nyquist frequency of  $\sim 7.6 \text{ d}^{-1}$ . Deviations from uniformity dampen the effects of aliasing and allow for the detection of periodic components above the traditional Nyquist limit (Eyer & Bartholdi 1999; Koen 2006).

Despite the orthogonality of the multiharmonic periodogram fit at each individual frequency, the fits on any set of frequencies are generally not independent in the case of irregular sampling (Schwarzenberg-Czerny 1998). This means that the statistic values at different frequencies can be correlated. Sometimes these correlations manifest as separate peaks. Compared to other algorithms, we find the MHAOV periodogram to be effective at damping secondary peaks across the range of probed frequencies. For the purposes of this analysis, we use the MHAOV period ( $P_{\text{mhaov}}$ ) as given by the



**Figure 4.** Frequency error derived via MCMC simulations,  $\sigma_{\text{MC}}$ , versus the error derived from periodogram methods,  $\sigma_{\text{freq}}$ , for a set of randomly selected periodic light curves. The dashed line is a plot of  $\sigma_{\text{MC}} = \sigma_{\text{freq}}$ . The two error estimates are highly correlated and the periodogram-based error estimate is systematically smaller than the MC-based error estimate.

periodogram. In the case of light curves with two identical or near-identical minima per cycle, such as an eclipsing binary composed of similar stars,  $P_{\text{mhaov}}$  corresponds to half of the orbital period.

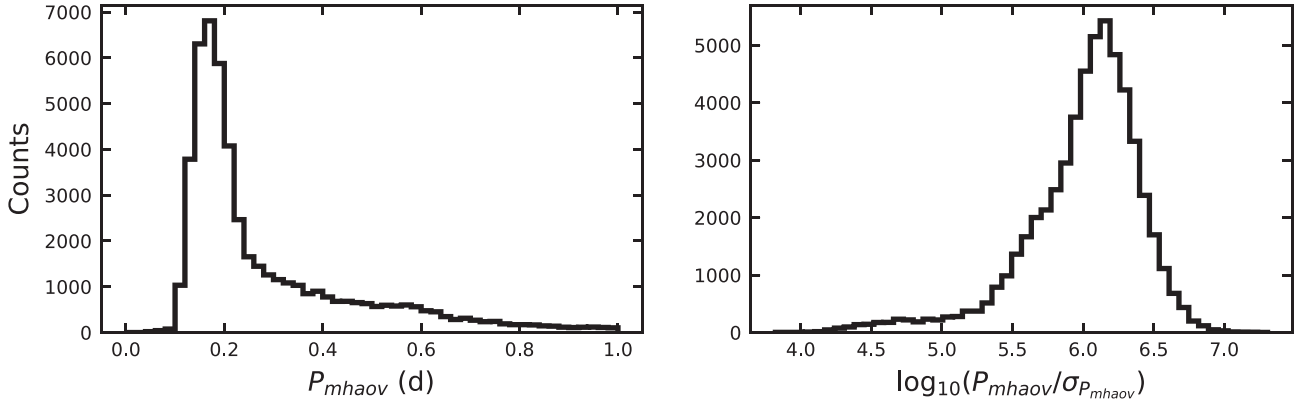
Correlations also cause the periodogram peak to have finite width and limit the precision with which the frequency can be determined from the peak. Generally, the uncertainty is a function of signal characteristics, the model used, the temporal baseline, and the quantity and quality of data (see Hartman et al. 2008; Lachowicz et al. 2009; Harding et al. 2013 for examples of different error estimation strategies).

To refine our frequency measurements and to estimate their error, we repeat the periodogram procedure with a refined frequency grid in the vicinity of the main periodogram peak. Our frequency measurement is the position of the likelihood peak on the refined frequency grid. We use the width of the peak on the fine grid and the surrounding background noise level to estimate the frequency error  $\sigma_{\text{freq}}$  (Schwarzenberg-Czerny 1991, 1995, 1996).

To check this frequency error estimation scheme, we compare with errors derived using EMCEE (Foreman-Mackey et al. 2013). To have realistic errors for realistic (non-sinusoidal) light curves, we take a random sample of 100 light curves, phase-fold them, and fit them with a Fourier series using LMFIT (Newville et al. 2014). Then, holding the Fourier coefficients fixed, we fit the light curve in the time domain with EMCEE using the phase, frequency, and constant offset from zero as parameters. The fitting is done in this way because the *WISE* cadence made simultaneously fitting the Fourier coefficients, phase, frequency, and constant offset intractable. To get the frequency distribution, we marginalize over the constant and the phase. As seen in Fig. 4, the MC-derived errors,  $\sigma_{\text{MC}}$ , are correlated with the periodogram-derived errors,  $\sigma_{\text{freq}}$ , but are systematically higher by  $\sim 25$  per cent. In what follows we use a conservative error estimate by multiplying the periodogram-derived error by a factor of 1.25. The period errors are on the order of  $\sim 10^{-6}$ – $10^{-7}$  d, but they do not capture the error of picking the wrong periodogram peak entirely and thus should be used with caution (VanderPlas 2018).

## 2.7 Selection of periodic variables

To select periodic sources, we first require that the maximum statistic of the MHAOV periodogram,  $\theta_{\text{max}}$ , exceeds the cutoff value given by



**Figure 5.** Left: Distribution of MHAOV (i.e. as calculated by the MHAOV periodogram) periods ( $P_{\text{mhaov}}$ ) for 50 493 periodic variables with MHAOV periods in the range 0.05–1 d. Right: Distribution of  $\frac{P_{\text{mhaov}}}{\sigma_{P_{\text{mhaov}}}}$  for the 56 177 identified periodic variables.

the power-law fit of Fig. 2 for the appropriate number of observations. Our cuts on false alarm probability are less strict than those of some other surveys that identify periodicity. This offers us the possibility of capturing sources undergoing interesting orbital evolution. For a higher confidence sample, a more restrictive cut on the maximum statistic value should be used. In addition to the cut on the maximum statistic value, we require phase coverage of at least 90 per cent. We estimate phase coverage by dividing the phase-folded light curve into 20 equally sized bins and then calculating the percentage of bins that contain at least 1 data point. The requirement that phase coverage is at least 90 per cent implicitly requires that the sources have at least 18 data points. Next, we exclude sources whose highest periodogram peak is at 1 or 0.5 times the orbital frequency of the *WISE* satellite ( $f_{\text{peak}} \notin [7.52, 7.68] \text{ d}^{-1}$  and  $f_{\text{peak}} \notin [14.65, 15.70] \text{ d}^{-1}$ ). Finally, we include a cut to limit phase offset due to period errors over the course of the time series. We calculate phase offset via

$$\text{Phase Offset} = \frac{\sigma_{P_{\text{mhaov}}} \times \text{Baseline}}{(P_{\text{mhaov}})^2}$$

where the baseline is the time-series duration,  $\sigma_{P_{\text{mhaov}}}$  is the period error, and  $P_{\text{mhaov}}$  is the period as given by the periodogram. The phase offset can be thought of as the ratio of the period uncertainty to the period multiplied by the complete number of cycles that the signal undergoes over the time-series duration. We require that the phase offset be less than 0.2. With the above cuts, we identify 56 177 periodic variable candidates. Fig. 5 shows the distribution of MHAOV periods and MHAOV periods over the period errors for these sources.

### 3 CATALOGUE CONTENTS

#### 3.1 Catalogue contents and *Gaia* cross-match

We cross-match our periodic variable candidates with *Gaia* DR2 (*Gaia* Collaboration 2016, 2018a; Evans et al. 2018; Arenou et al. 2018) using the pre-computed, best-neighbour *WISE/Gaia* cross-match catalogue of Marrese et al. (2019). We find 188 043 matches out of a total of 454 103 variable objects ( $\sim 41.4$  per cent) which is similar to the total percentage of AllWISE sources that have a best-neighbour cross-match (39.83 per cent – see Marrese et al. 2019). Of our 56 177 periodic variable candidates, 49 384 or  $\sim 87.9$  per cent have a best-neighbour *Gaia* cross-match. The higher match rate for

periodic variables is due to the fact that the aperiodic variables – many of them stochastically varying young stellar objects – tend to be redder, with a median (W1–W3) = 0.909 mag as compared to the periodic variables with (W1–W3) = 0.311 mag. As a further check, we cross-match our sample with the *WISE* young stellar object catalogues of Marton et al. (2016) and find 44 196 matches of which only 237 are flagged as periodic.

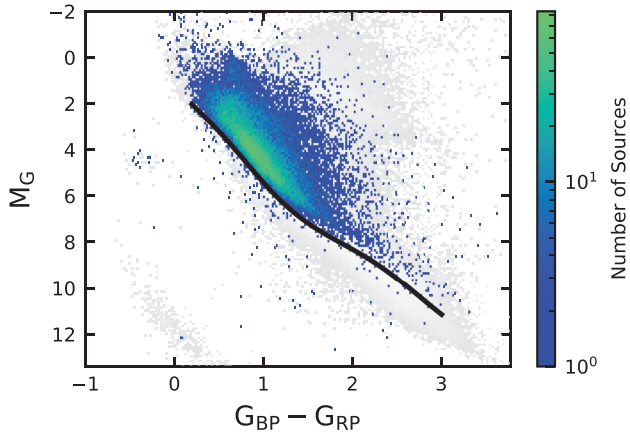
We use the *Gaia* cross-match and considerations of the limitations of *WISE* to get a sense of the catalogue contents and make the case that the sample is dominated by eclipsing binaries. We search for periods between 0.05 and 10 d, but due to the *WISE* observing cadence and our requirement of good phase coverage, the sensitivity drops for periods above 2 d. As a result, we do not expect many long-period variables in the more luminous part of the colour–magnitude diagram. The period sensitivity limit combined with the crowding in the Galactic disk cause us to detect few classical Cepheids. We expect to detect few  $\delta$  Scuti variables because many have variability amplitudes that are too low to be reliably detected (Murphy et al. 2019) and many high-amplitude  $\delta$  Scuti are faint and blue. The photometric sensitivity of *WISE* also limits our ability to detect variability on the white dwarf sequence.

For periodic objects with a *Gaia* cross-match, in order to have robust absolute magnitudes and colours, we require:

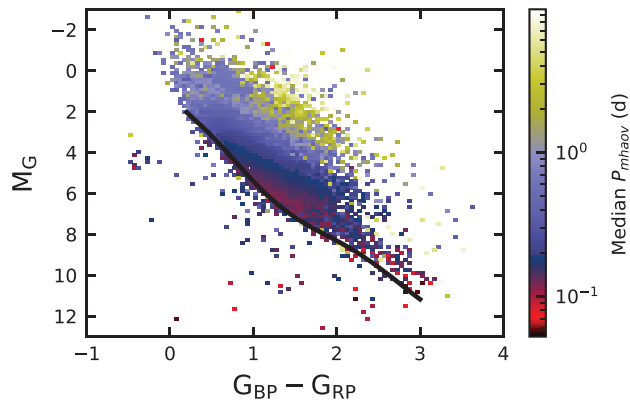
- (i) `parallax_over_error > 10`
- (ii) `phot_g_mean_flux_over_error > 50`
- (iii) `phot_bp_mean_flux_over_error > 10`
- (iv) `visibility_periods_used > 8`
- (v) `phot_rp_mean_flux_over_error > 10`

In addition, we restrict `phot_bp_rp_excess_factor` in accordance with *Gaia* Collaboration (2018b). After applying these cuts we retain 34 857 sources.

Fig. 6 shows the *Gaia* colour–absolute magnitude diagram for all the *WISE* periodic variables with a *Gaia* cross-match that satisfy the above quality cuts. The Pleiades MS fit of Hamer & Schlaufman (2019) is shown in black. Most of the *WISE* periodic variables are MS stars. On the MS, the most common form of periodic stellar variability is eclipsing binaries. These sources are located above the MS fit because there is extra flux from the presence of the second star. A few periodic sources are located below the MS. They may be white dwarf–brown dwarf binaries or cataclysmic variables. The grey



**Figure 6.** Our periodic variables cross-matched with *Gaia* and shown on the colour-absolute magnitude diagram. The black line is the Pleiades cluster spline fit from Hamer & Schlaufman (2019). The grey background sources are a sample of 500 000 *Gaia* sources chosen in accordance with Gaia Collaboration (2018b) that are shown for reference.

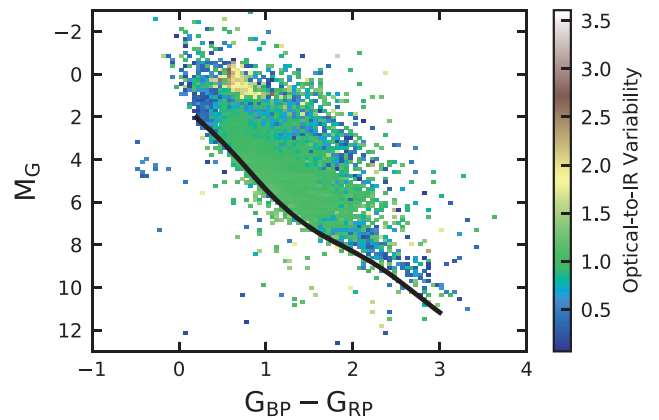


**Figure 7.** Colour-absolute magnitude diagram of our *WISE* periodic variables colour-coded by the median MHAOV period in each colour-absolute magnitude bin. The main-sequence fit of Hamer & Schlaufman (2019) is shown in black for reference.

scale background is a sample of 500 000 *Gaia* sources that meet the selection of Gaia Collaboration (2018b) that is shown for reference.

Fig. 7 again shows the *Gaia* colour-absolute magnitude for the same sources colour-coded by the median MHAOV period in each colour-absolute magnitude bin. For our sample, on average, more massive sources tend to have longer periods. For close binaries, this makes sense because larger stars have a longer limiting period before reaching contact. A group of pulsating RR Lyrae stars can be seen at  $M_G \sim 1$  and  $BP-RP \sim 0.7$  with periods between  $\sim 0.2$  and 1 d (Preston 1964; Kolenberg 2012; Das et al. 2018). These sources are in purple and are distinguished from the green band that represents the surrounding EBs in the colour-magnitude diagram.

Fig. 8 shows another view of the *Gaia* colour-absolute magnitude diagram colour-coded by the ratio of fractional variability in the optical to fractional variability in the infrared (IR). To calculate the optical fractional flux variability from *Gaia*,  $\sigma_{\text{optical}}$ , we follow the methods of Hwang et al. (2020). We compute the fractional flux variability from the *Gaia* photometric errors and then account for instrumental errors. We compute the IR fractional flux variability,  $\sigma_{\text{IR}}$ , by  $\sigma_{\text{IR}} = 0.4 \ln(10) * \text{Std}_{W1}$ , where  $\text{Std}_{W1}$  is the standard



**Figure 8.** *Gaia* colour-absolute magnitude diagram colour-coded by the median ratio of optical fractional variability from *Gaia* G band to infrared (IR) fractional variability from *WISE* W1 in each colour-absolute magnitude bin. *Gaia* fractional variability is calculated in accordance with Hwang et al. (2020). We plot only those sources that have a *Gaia* fractional variability of at least 1 per cent and have a fractional variability ratio of at least 0.05. A group of RR Lyrae can be seen in tan/brown that vary more in the optical than in the infrared. The majority of the sources are eclipsing binaries and have equivalent optical and infrared variability amplitudes (i.e. optical-to-IR ratio of  $\sim 1$ ). The black line is the Pleiades main-sequence fit of Hamer & Schlaufman (2019).

deviation of W1 in magnitudes, and the factor  $0.4 \ln(10)$  comes from the unit conversion between fluxes and magnitudes. We plot only those sources here for which the ratio of optical-to-IR fractional variability is greater than 0.05 and which have a fractional variability of at least 1 per cent in *Gaia*. Most of the sources have approximately equal optical and IR fractional variability, consistent with what is expected for eclipsing binaries. These sources are also offset to the bright side of the MS fit indicating that there is excess flux due to the presence of a second star. A group of RR Lyrae can be seen that vary more in the optical than in the infrared due to the nature of their pulsation (Das et al. 2018).

### 3.2 Comparison with Chen et al. (2018) catalogue

Chen et al. (2018) present a catalogue of 50 282 periodic variables from *WISE* identified with the Lomb–Scargle (Lomb 1976; Scargle 1982) periodogram and classified via light-curve fitting. Our work differs from that of Chen et al. (2018) in that we extend the period search grid to shorter period objects, use a different periodogram, involve a *Gaia* DR2 cross-match, use data from a more recent NEOWISE release, and apply a different selection. Chen et al. (2018) also discarded sources with varying periods from AllWISE to NEOWISE and we make no such restriction.

We cross-match our periodic variables with those of Chen et al. (2018) using a matching distance of 1 arcsec. Increasing the matching distance to 5 arcsec does not significantly alter the results. A cross-match was necessary because Chen et al. (2018) used a truncated form of the *WISE* designation that made it difficult to use this for matching purposes. Of all the sources with  $6 \leq \text{var\_flag} \leq 9$  that we analysed, 41 532 have a cross-match with the 50 282 periodic variables of Chen et al. (2018). Their remaining 8750 variables were excluded by our initial selection and were never a part of our periodogram analysis. 8703 of the variables were excluded by our cuts on `cc_flags` to avoid contamination and confusion and the remaining 47 were excluded by our cuts on `ext_flag` to remove extended objects. Of the

41 532 variables in common, we mark 36,991 as periodic. We miss some of the sources because our more stringent cuts on the quality of individual exposures reduces the number of points in the light curve and prevents us from detecting them as periodic. We explore several methods for comparing the periods from our two catalogues. First, we consider the raw period difference. For 33 870 (~92 per cent) of the cross-matched periodic sources there is a period difference  $\Delta P_{\text{half}} = |0.5P_{\text{chen18}} - P_{\text{mhaov}}| < 0.001$  d (relevant for binaries with 2 similar minima per cycle) or  $\Delta P = |P_{\text{chen18}} - P_{\text{mhaov}}| < 0.001$  d. Second, we incorporate uncertainties to see if the period differences mentioned above are significantly different than zero. We find that for more than 96 per cent of the sources, either  $\Delta P_{\text{half}}$  or  $\Delta P$  are within three standard deviations from zero. Third, we explore a more stringent selection for compatible periods by requiring:

$$\left| \frac{P_{\text{mhaov}} - P_{\text{chen18}}}{P_{\text{mhaov}}} \right| \times \text{Num.Cycles} < 0.20$$

or

$$\left| \frac{P_{\text{mhaov}} - 0.5P_{\text{chen18}}}{P_{\text{mhaov}}} \right| \times \text{Num.Cycles} < 0.20.$$

Here, Num.Cycles is the total number of periodic cycles completed over the duration of the time series,  $\text{Num.Cycles} = \text{Baseline}/P_{\text{mhaov}}$ . We find that 27 568 (~75 per cent) of the cross-matched sources satisfy this third criterion. For all three compatibility criteria, the incompatible periods generally correspond to cases in which either (a) our more stringent cuts reduce the number of data points and leads to a different period estimate, (b) the detected period is greater than 5 d, i.e. in a range poorly probed by *WISE*, or (c) we measure a period outside the range probed by Chen et al. (2018) indicating that one of the measurements is a harmonic. Of these effects, we expect (a) to be the most common. For example, for objects with  $\Delta P_{\text{half}}$  or  $\Delta P$  more than three standard deviations from zero, on average, our periodogram is computed on ~39 fewer data points (despite the fact that we include more years of NEOWISE data) due to our more stringent cuts on the quality of the individual exposures. For objects with incompatible periods as determined by the comparison method that takes into account the number of cycles, our post-quality-cut light curves had, on average, 22 fewer data points than the light curves of Chen et al. (2018). Due to the long observing baseline (~3100 d), even a slight change in the single-exposure selection criteria can lead to a period shift large enough to cause the source to be classified as incompatible under this third comparison method.

For ~89 per cent of the matches, our MHAOV periods,  $P_{\text{mhaov}}$ , are one-half the period cited by Chen et al. (2018),  $P_{\text{chen18}}$ , which is expected for a sample dominated by close eclipsing binaries where the primary and secondary eclipses are difficult to distinguish. Over 98 per cent of the sources for which we recover half periods have  $\Delta P_{\text{half}} < 0.0001$  d. Chen et al. (2018) classify 32 151 of the cross-matched sources as binaries, 710 as some type of Cepheid, and 2319 as RR Lyrae. The remainder were assigned an ambiguous classification (Cepheid/Binary, Binary/RR Lyrae, or miscellaneous) due to a lack of characteristic features in their infrared light curves.

Finally, 19 186 of our periodic variables were not in the periodic variable catalogue of Chen et al. (2018) because of our different period search method and our expanded period search grid on the short-period end.

## 4 NON-PARAMETRIC METHODS FOR LIGHT-CURVE ANALYSIS

We present a catalogue of 454 103 *WISE* variables with a variety of non-parametric measures and periodogram measurements. We explicitly flag periodic variables and eclipsing binaries. The data model is shown in Table 1. The non-parametric measures and the identification of eclipsing binaries are the subjects of this section.

### 4.1 Non-parametric measures of variability

In this section, we explore non-parametric and computationally inexpensive methods for light-curve analysis. The distribution of observed fluxes, without reference to their time dependence, carries a wealth of information. If the cadence is suitable, the distribution of the observed fluxes can be assumed to be randomly drawn from (and therefore to be representative of) the underlying flux probability density function (PDF). The various moments of the observed flux distribution, assumed to be representative of the moments of the underlying PDF, are easy to measure and computationally fast. Various measures of these moments with different weighting schemes have been successfully utilized for variable star classification in a variety of other works (e.g. Kinemuchi et al. 2006; Dubath et al. 2011; Rimoldini et al. 2012; Drake et al. 2013; Rimoldini 2014, 2019; Bass & Borne 2016).

The first and the second moment – related to the mean magnitude and the r.m.s. variations around the mean – are already directly or indirectly incorporated into our analysis. In particular the second moment of the PDF is related to the `var_flg` we use as the primary step in our selection. We also compute a non-parametric amplitude (henceforth amplitude), which is the range between the 5th and 95th percentile of the magnitude values.

The third moment (skewness) of the PDF can help distinguish between eclipsing and eruptive types of variability. For an object which stays at constant magnitude and occasionally undergoes dimmings due to occultations or eclipses, the PDF is expected to have a tail to fainter magnitudes (which correspond to the fluxes during occultations). For an object that instead undergoes occasional eruptions, we expect a tail to brighter magnitudes. In addition to the standard Fisher-Pearson skewness, an assortment of other non-parametric measures of the third moment have been introduced. One such measure is the magnitude ratio or M test of Kinemuchi et al. (2006). This is calculated as

$$M = \frac{\text{max} - \text{median}}{\text{max} - \text{min}}$$

where max, median, and min refer to the maximum, median, and minimum of the magnitude distribution respectively. Another such measure is the relative asymmetry (RelAsym), introduced by Zakamska & Greene (2014) in the context of velocity profiles but also relevant to the analysis of time-series data. This is calculated by

$$\text{RelAsym} = \frac{(P_{95} - P_{50}) - (P_{50} - P_5)}{(P_{95} - P_5)}$$

where  $P_{95}$ ,  $P_{50}$ , and  $P_5$  refer to the 95th, 50th, and 5th percentiles of the magnitude distribution respectively.

In this paper, we introduce a new non-parametric measure, fainter fraction (FF), which is related to the third moment. Specifically, for each light curve, we calculate the halfway point between the 5th and 95th percentile of magnitude and then calculate the fraction of points with magnitudes greater (fainter) than this halfway point by at least their measurement uncertainty. If FF is greater than 0.5, this means that the object spends most of its time in the faint state, becoming



**Table 1.** Column descriptions for our catalogue of *WISE* variables. After careful consideration, we elect to report the MHAOV period ( $P_{\text{mhaov}}$ ) i.e. the period as given by the periodogram. Based on our cross-match with Chen et al. (2018), we estimate that for more than  $\sim 90\%$  of the EBs,  $P_{\text{mhaov}}$  corresponds to one-half of the orbital period. However, without more detailed light-curve fitting or analysis with a different periodogram (e.g. the string-length methods of Lafler & Kinman 1965; Burke et al. 1970; Renson 1978; Dworetzky 1983) which is outside the scope of this work, there is no clear way to identify those binaries for which we detect the actual orbital period and not a harmonic. It depends on the data quality/cadence and the particular signal shape in question. When no measurement could be made for a given source and column, the value is missing. Not all non-parametric measures are calculated if the source had fewer than five data points. Each band has its own variability flag. If the variability flag for a band is ‘n’, then no flag was assigned in that band. The complete catalogue is available [online](#).

| Column      | Data Type | Units  | Description   |
|-------------|-----------|--------|---|
| wise_id     | str19     | –      | <i>WISE</i> designation   |
| ra          | float64   | deg    | Right ascension   |
| dec         | float64   | deg    | Declination   |
| sigra       | float64   | arcsec | Right ascension error   |
| sigdec      | float64   | arcsec | Declination error   |
| w1mpro      | float64   | mag    | W1 magnitude  |
| w1sigmpro   | float64   | mag    | W1 magnitude error  |
| w1snr       | float64   | mag    | W1 signal-to-noise ratio  |
| w2mpro      | float64   | mag    | W2 magnitude  |
| w2sigmpro   | float64   | mag    | W2 magnitude error  |
| w3mpro      | float64   | mag    | W3 magnitude  |
| w3sigmpro   | float64   | mag    | W3 magnitude error  |
| w4mpro      | float64   | mag    | W4 magnitude  |
| w4sigmpro   | float64   | mag    | W4 magnitude error  |
| var_flg     | bytes4    | –      | Variability flags for all four bands  |
| num_pts     | float64   | –      | Number of observations in light curve after quality cuts                        |
| median_err  | float64   | mag    | Median individual data point uncertainty for light curve                        |
| mean_mag    | float64   | mag    | Mean magnitude  |
| std_mag     | float64   | mag    | Standard deviation of magnitude   |
| median_mag  | float64   | mag    | Median magnitude  |
| amp         | float64   | mag    | Amplitude   |
| FF          | float64   | –      | Fainter Fraction  |
| rel_asym    | float64   | –      | Relative Asymmetry (Zakamska & Greene 2014)                                     |
| M           | float64   | –      | Magnitude Ratio/M-test value (Kinemuchi et al. 2006)                            |
| skew        | float64   | –      | Skewness  |
| kur         | float64   | –      | Kurtosis  |
| phase_cov   | float64   | –      | Phase coverage  |
| min_mag     | float64   | mag    | Minimum magnitude   |
| max_mag     | float64   | mag    | Maximum magnitude   |
| baseline    | float64   | d      | Time-series duration  |
| R           | float64   | –      | Ratio of variability amplitude on short time-scales to that on long time-scales |
| max_stat    | float64   | –      | Maximum statistic value   |
| cutoff_stat | float64   | –      | Cut-off maximum statistic value to reject null hypothesis                       |
| periodic    | bool      | –      | Periodic sources receive a value of True  |
| P_mhaov     | float64   | d      | MHAOV period  |
| sigP_mhaov  | float64   | d      | MHAOV period error  |
| EB          | bool      | –      | Eclipsing Binaries receive a value of True                                      |

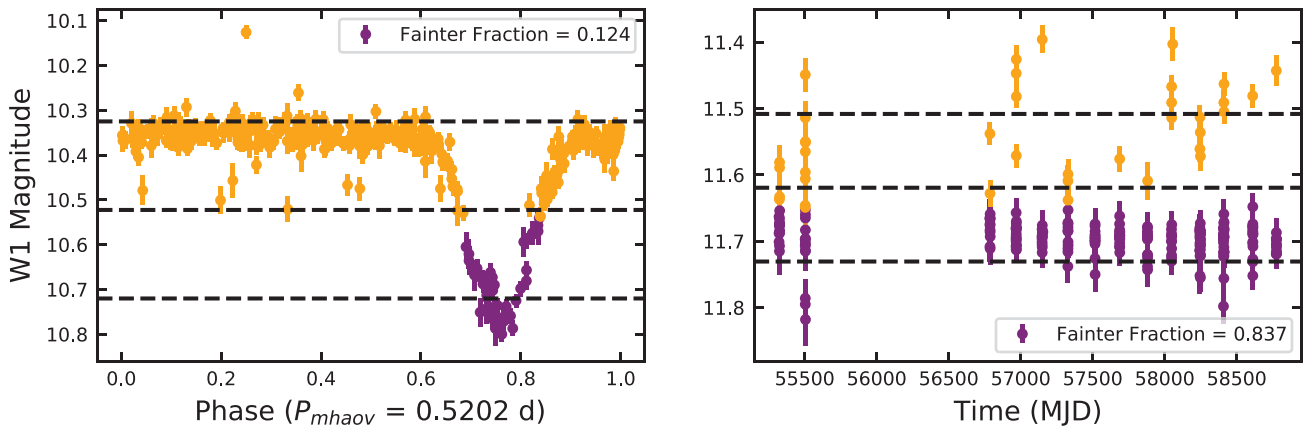
brighter less than half of the time, and it is likely to be of an eruptive type. The use of this measure is demonstrated in Fig. 9.

Each of the aforementioned measures has its own advantages and we report all of them (i.e. skewness, magnitude ratio, RelAsym, and FF) in our catalogue for all of the *WISE* variables so that users can choose which to employ. Due to the fact that we aim to apply these measures to all of the variables, periodic and not, we do not employ the moment weighting scheme proposed by Rimoldini (2014) because this would require a linear interpolation in phase (interpolation in time is not possible for the *WISE* cadence) which is not possible if the source is aperiodic. The correlations between each of these measures are shown in Table 2. All these measures are dimensionless, but the standard skewness is not bounded like  $M$ , FF, and RelAsym making it less desirable for use with a classifier.

For this work, we utilize our new measure, FF because we find that it performs well in comparison with the these other measure for our

particular use case. The magnitude ratio,  $M$ , is susceptible to effects from outliers and erroneous measurements due to its use reliance the maximum and minimum magnitude. Similar to RelAsym, FF is not as strongly affected by these outliers. Another advantage of FF is that, unlike both  $M$  and RelAsym, it is not normalized by some measure of the signal amplitude. Using FF allows us to disentangle the signal amplitude from the estimate of the third moment. This is useful, for example, in the case of eclipsing binaries with narrow eclipses as these signals should be more localized in FF space than in the space of RelAsym or  $M$ . Furthermore, in Section 4.2, we find that the separation between RR Lyrae and EBs is cleaner in the FF–amplitude space than if FF is substituted for one of these other measures.

Finally, we introduce another non-parametric measure to describe the time-scale of variability. Due to the peculiar cadence of *WISE* observations – every  $\sim$ six months there is a bundle of  $\sim 10$



**Figure 9.** Left: The light curve of an Algol-type eclipsing binary (WISEJ234343.55+812751.6) phase-folded with the MHAOV period ( $P_{mhaov}$ ) as given by the periodogram, thus showing only one eclipse instead of two. Projecting the light curve on the brightness axis gives us the flux probability density function (PDF), which can be used for non-parametric measures of the light curve. The top and bottom black, dashed lines show the 5th and 95th percentiles of the PDF. The difference between these two lines is an estimate of the signal amplitude. The middle line represents the halfway point between these two. The points in purple are the fainter fraction of the light curve. Stars with eclipses and stars with eruptions show different PDFs and can be distinguished based on these metrics. Right: Similar idea but for an eruptive variable (WISEJ005714.34-703745.5). Eruptive variables are typically aperiodic so we display the actual light curve instead of the phase-folded one.

**Table 2.** Correlation Matrix of non-parametric measures of the third moment of the flux PDF.

|          | FF        | RelAsym.  | $M$       | Skewness  |
|----------|-----------|-----------|-----------|-----------|
| FF       | 1.000000  | -0.860607 | -0.721460 | -0.571853 |
| RelAsym. | -0.860607 | 1.000000  | 0.845581  | 0.593586  |
| $M$      | -0.721460 | 0.845581  | 1.000000  | 0.783283  |
| Skewness | -0.571853 | 0.593586  | 0.783283  | 1.000000  |

observations taken  $\sim 95$  min apart – *WISE* has sensitivity to variability on a 1 d time-scale, as well as to long-term variations. We use  $R$ , the ratio of the average r.m.s. variability on day/hour-long time-scales to the r.m.s. variability of the entire light curve. To calculate the variability on short time-scales, we remove long-term variations from the light curve. We subtract off the mean from each observing bundle so that each bundle is centred at zero magnitude. Then, we find the standard deviation of the light curve composed of the zeroed bundles. Given that *WISE* has only a limited number of  $\sim$ day-long visits to the same position on the sky, this method is computationally inexpensive. For sources with equal signal-to-noise (S/N) of variability, the ratio  $R$  is expected to decline from 1 to 0 as we go from objects that vary on day-long time-scales to those with only month- or year-long variability. Sources with low S/N of variability have both their short-term and long-term variability in line with the photometric  $\sigma$ . The idea behind this non-parametric measure is similar to that behind the `MEDIAN_RANGE_HALFDAY_TO_ALL` measure used by *Gaia* (section 7.3.3 of Eyer et al. 2018), but instead of a sliding window, the 1-d intervals for computation of short-time-scale variability are naturally provided by the *WISE* observing cadence.

#### 4.2 Identification of eclipsing binaries

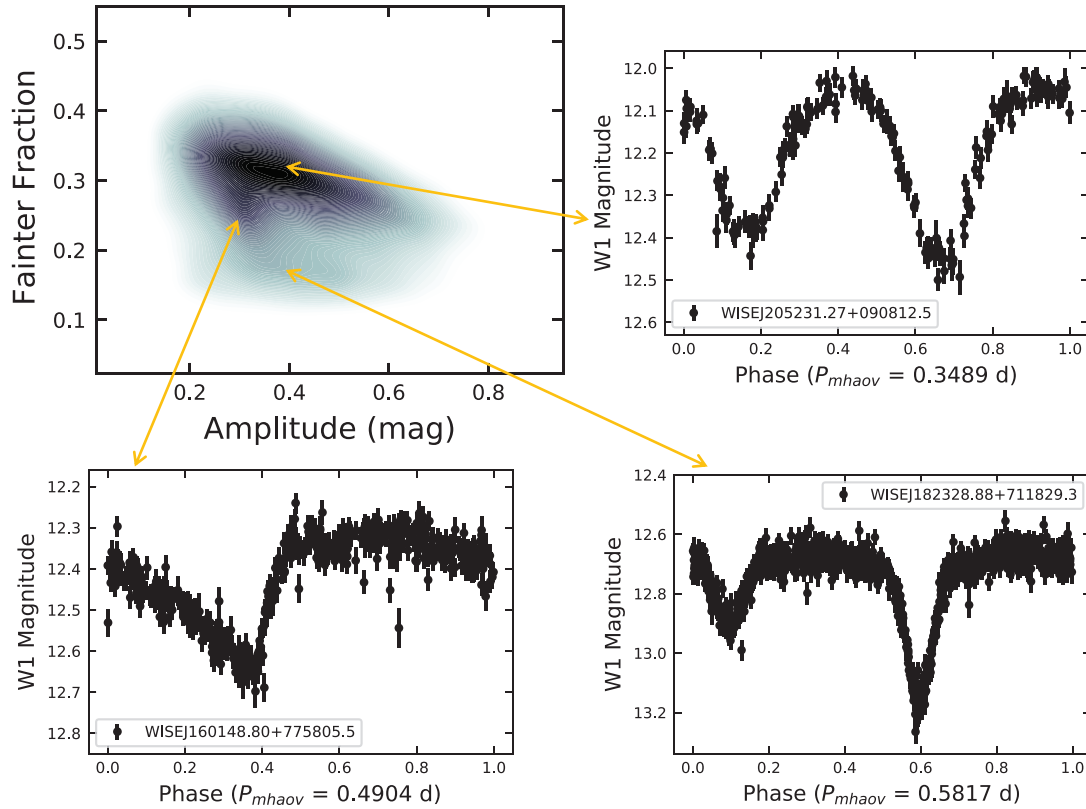
We identify a few physically motivated cuts on period and a few of our non-parametric measures that can reliably isolate eclipsing binaries from other types of variability. In Fig. 10, we show the kernel density estimate of our periodic variable candidates in the space of fainter fraction and amplitude. The sources cluster in this space and different clusters are characterized by light curves with

different morphology. Nearly all of the periodic variables have  $FF < 0.5$  indicating that they are occulting. There is an upper sequence at  $FF$  of  $\sim 0.35$  and amplitude between 0.2 and 0.7 that, based on visual inspection, seems to be dominated by near-contact and contact binaries. A lower sequence at  $FF$  of  $\sim 0.15$  and of similar amplitudes has detached binaries with more narrow eclipses. The offshoot from the upper sequence at amplitude  $\sim 0.3$  mag and  $FF \sim 0.25$  seems to contain RR Lyrae.

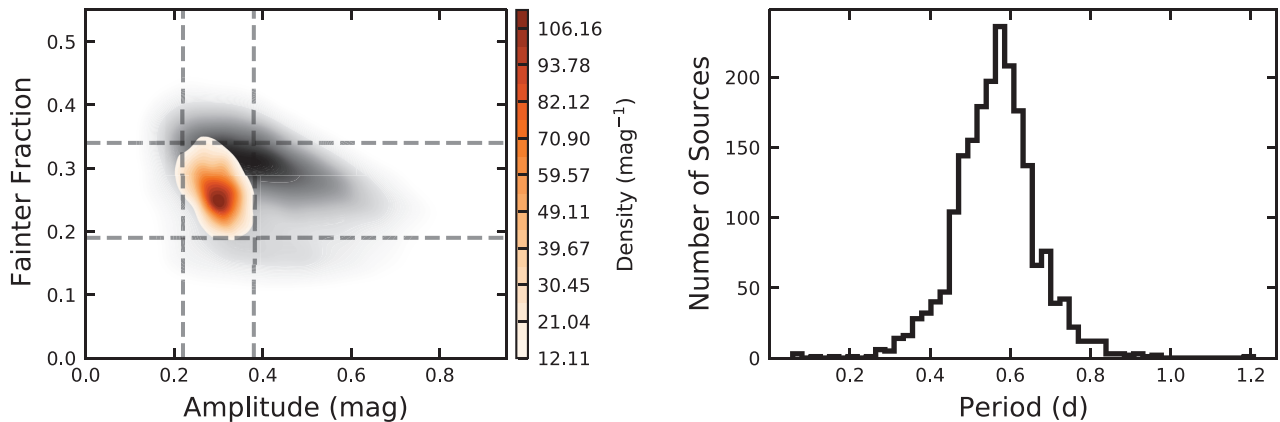
We cross-match our sample with the *Gaia* RR Lyrae catalogue of Clementini et al. (2019) and find 2027 matches. Fig. 11 shows the kernel density estimate of the cross-matched RR Lyrae in the fainter fraction–amplitude space and the period distribution of the RR Lyrae. To identify candidate RR Lyrae among *WISE* periodic variables, we apply the cut shown by the grey, dotted lines. Specifically, we require that  $0.19 < \text{fainter fraction} < 0.34$  and  $0.22 < \text{amplitude} < 0.38$ . These cuts select the entirety of the aforementioned offshoot from the upper sequence and also a portion of the upper sequence itself. To ensure that we are not excluding too many close EBs, we restrict the MHAOV periods to between 0.25 and 1 d. After applying this cut, both the upper and lower sequences disappear and the RR Lyrae offshoot becomes the dominant feature in the fainter fraction–amplitude space. We choose a lower period bound of 0.25 d because we do not expect many RR Lyrae with periods below 0.25 d and we do not want to exclude a high number of close eclipsing binaries. This figure shows that these three variables – period, amplitude, and fainter fraction – provide a means of separating RR Lyrae from EBs without resorting to the colour–magnitude diagram.

All told, this selection labels 6100 of the periodic variables as candidate RR Lyrae. Included in this are 1752 out of the 2027 ( $\sim 86$  per cent) *Gaia* RR Lyrae from the cross-matched catalogue of Clementini et al. (2019). We exclude the remaining *Gaia* RR Lyrae from the EB sample as well. In addition, we also cross-match our periodic variables with the *Gaia* DR2 single-object-study Cepheid catalogue (Clementini et al. 2019). As mentioned above, we expect few Cepheids in the sample. The cross-match reveals only 347 *Gaia* Cepheids and we remove these from the EB sample.

The above cuts remove 1801 out of 2319 ( $\sim 78$  per cent) of the RR Lyrae identified by Chen et al. (2018) while retaining



**Figure 10.** Kernel density estimate for all periodic variables in the space of two of our non-parametric measures, fainter fraction (FF) and amplitude, with example phase-folded light curves characteristic of three regions. Clockwise from top right: a contact EB found in the upper sequence phase-folded with period  $P = 2 \times P_{\text{mhaov}}$ , a detached EB found in the lower horizontal sequence phase-folded with period  $P = 2 \times P_{\text{mhaov}}$ , a candidate RR Lyrae found in the offshoot from the upper sequence. The primary mechanism driving changes in luminosity for RR Lyrae is different in the optical, where it is temperature variations, as opposed to the infrared, where it is radial pulsations (Madore et al. 2013; Braga et al. 2019). Optical and infrared light curves of RR Lyrae have different morphology and can have different skewness/FF.



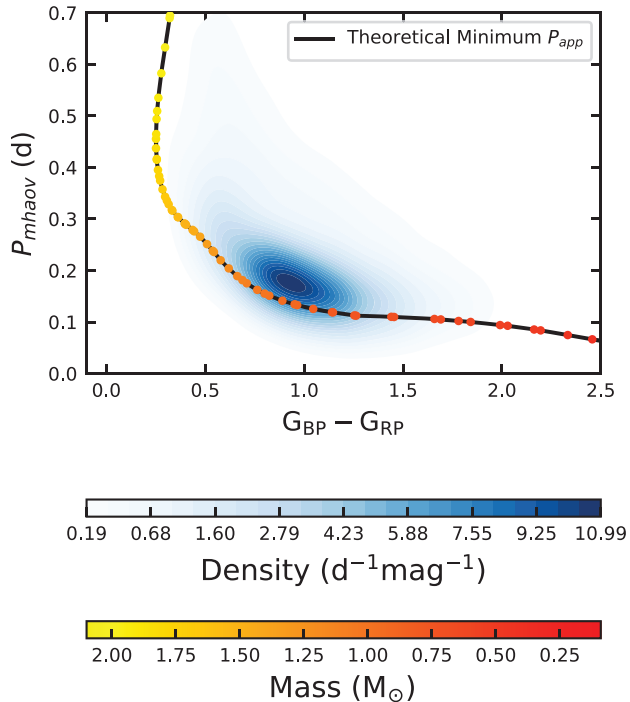
**Figure 11.** Left: Two-dimensional kernel density estimate for *Gaia* RR Lyrae (shown in red) in the space of two of our non-parametric measures, fainter fraction (FF) and amplitude. The distribution of all the periodic variables in this space is shown in grey scale. This plot shows that the RR Lyrae do indeed cluster in the area of the FF versus amplitude plot that was suggested by Fig. 10. The grey, dotted lines show our initial selection of RR Lyrae. Right: Our period measurements for these sources. As expected for RR Lyrae, the periods are nearly all between 0.2 d and 1 d.

29483 out of the 32151 sources ( $\sim 92$  per cent) classified by Chen et al. (2018) as EBs. We add these excluded sources back in to our EB sample. We remove the remaining RR Lyrae and also the 710 Cepheids classified by Chen et al. (2018). We are left with a low contamination sample of 50722 eclipsing binary candidates.

## 5 DISCUSSION

### 5.1 MS binaries

We next turn to the colour–magnitude diagram to isolate MS EBs. Of our initial sample of 50722 EB candidates, 44550 have a *Gaia* best-



**Figure 12.** Gaussian kernel density estimate in the space of MHAOV periods (which we take to be  $0.5P_{\text{orbital}}$ ) and *Gaia* BP-RP colour for MS binaries. The black line overlaid with the red-yellow points represent the theoretical minimum apparent period ( $0.5P_{\text{orbital}}$ ) for a contact, equal-mass, main-sequence binary.

neighbour cross-match (Marrese et al. 2019). Applying the quality cuts of Section 3 leaves us with 32 929 sources. To select MS binaries, we start with the Hamer & Schlaufman (2019) Pleiades spline fit shown in Fig. 6. Binaries are expected to be found above the MS for a wide range of mass ratios (Hurley & Tout 1998). We saw in Section 3 that our periodic variables, and especially those on the MS, were dominated by EBs, and the high density of sources above the MS fit can be seen in Fig. 6. We require that the source have an absolute G-band magnitude above the Pleiades spline fit and within 1.5 mag of the spline fit value for a given colour. This cut is motivated by the fact that doubling the flux corresponds to a magnitude increase of  $\sim 0.75$ . We double this value to incorporate different stellar metallicities and measurement uncertainties. This cut leaves 21 746 sources.

Fig. 12 shows the distribution of MHAOV period versus *Gaia* BP-RP colour for our MS EBs detected with *WISE*. The distribution is smooth across the entire period range and includes periods not probed by Chen et al. (2018). To make this plot, we make the approximation that  $P_{\text{mhaov}} = 0.5P_{\text{orbital}}$ , that is, the periodogram identifies half the orbital period. In our cross-match with the results of Chen et al. (2018), this is valid for  $\sim 90$  per cent of the binaries in our sample. However, many of our periodic variables that were not detected by Chen et al. (2018) have shorter periods (i.e.  $P_{\text{mhaov}} < 0.15$  d) and we expect these shorter period binaries to be more likely to have two similar minima per cycle. This could increase the percentage of sources for which we identify half the orbital period to  $>90$  per cent. The remaining sources for which we identify the actual orbital period can be found further from the black line in the upper part of Fig. 12. For binaries with two identical minima per cycle, we call half the orbital period the apparent period ( $P_{\text{app}} = 0.5P_{\text{orbital}}$ ). In blue is a kernel density estimate and the black line represents the theoretical

minimum possible apparent period for an equal-mass, contact, MS eclipsing binary of a fixed age (1 Gyr) and solar metallicity as a function of mass and colour. The calculation of this line follows Hwang et al. (2020). Briefly, from the PARSEC isochrone (Bressan et al. 2012) for an age of 1 Gyr and solar metallicity we get the stellar radius of an undistorted star (i.e. the Roche lobe volume radius,  $R_L$ ). Following Eggleton (1983), we use the relationship between  $R_L$  and the semimajor binary axis,  $a$ , to solve for  $a$  ( $R_L = 0.38a$ ). Finally, we set the masses of the two constituent stars,  $M_1$  and  $M_2$ , equal to each other and solve for the minimum possible apparent period using

$$P_{\text{app}} = 0.5P_{\text{orbital}} = \pi \left( \frac{a^3}{G(M_1 + M_2)} \right)^{1/2}$$

where  $G$  is the gravitational constant.

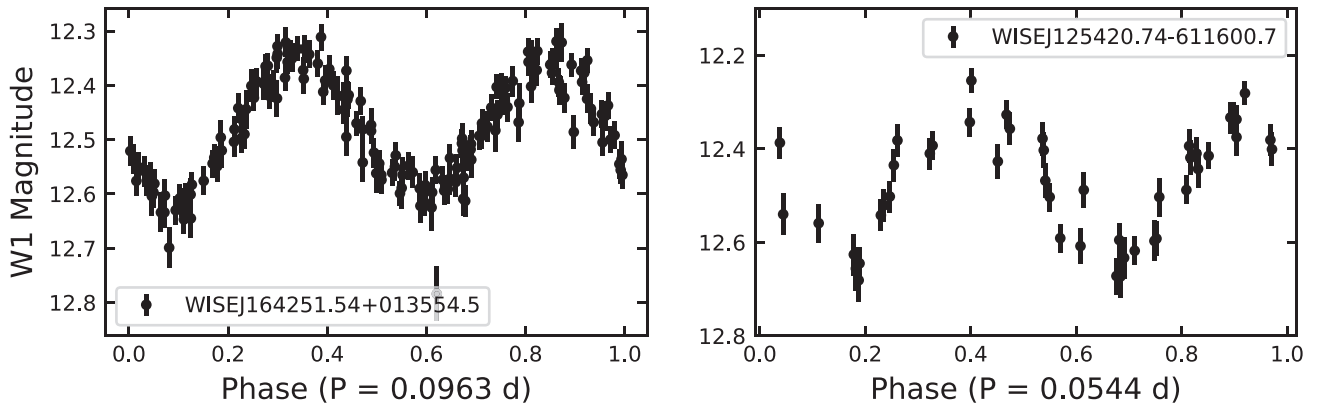
Many of our EBs are near this theoretical lower bound indicating that the majority of these systems are contact or near-contact binaries. Interestingly, some of the points are below the black, minimum-period line. These are systems for which the input assumptions break down. They can be systems in which the component stars are not of equal mass, not of identical colour, younger than 1 Gyr, or of lower-than-solar metallicity. Some RR Lyrae are located close to the MS and may be present on the blue side of the plot ( $G_{\text{BP}} - G_{\text{RP}} < 0.6$ ) accounting for some of the spillover in the bottom left. The dearth of sources hugging the line in the upper-left of the plot does not appear to be due to decreased sensitivity to those periods (see Section 2.5). It is possible that there are physical reasons for the paucity of contact binaries in the corresponding mass range – for example, if the magnetic braking mechanism is responsible for creating contact binaries (Hwang & Zakamska 2019), then it is expected to be inefficient at these masses due to lack of convection at  $M > 1.3M_{\odot}$  (Matt et al. 2011).

The peak in the period distribution of our MS EBs is located about at  $2P_{\text{mhaov}} \approx 0.34$  d. This is higher than the maximum of the period distribution of  $\sim 0.27$  d found by Rucinski (2007), but our sample is not volume-limited. Also, as mentioned above, for  $\lesssim 10$  per cent of these binaries, we have detected the actual orbital period so multiplying  $P_{\text{mhaov}}$  by two creates a tail to higher periods.

## 5.2 Shortest-period objects

There are 126 sources in our eclipsing binary candidates that have MHAOV periods less than or equal to 0.1 d. We expect higher contamination and lower period accuracy in this period range due to the limitations of the *WISE* cadence. We visually inspect the phase-folded light curves and find four are the result of bad data or an erroneous period measurement and that the remaining 122 represent real periodic signals. Fig. 13 shows some example phase-folded light curves for these objects. These sources are prime candidates for comparison with other short-period binary catalogues (e.g. Norton et al. 2011; Drake et al. 2014b) and subsequent analysis with the tools of Conroy et al. (2020). We cross-match these sources with SIMBAD with a matching distance of 1 arcsec. After removing the bad data, we find 37 matches of which 31 were classified in SIMBAD as a specific type of variability. Of these 31, we find that 4 are pulsating and the remaining 27 are all classified as some sort of binary. Three of these sources are cataclysmic variables (V\* BL Hyi; SDSS J121209.30+013627.7; RX J2218.5+1925). We find compatible periods for all of these cataclysmic variables indicating that our period measurements are accurate even in this short-period regime (Schmidt et al. 2005; Thorstensen & Halpern 2009; Avvakumova, Malkov & Kniazev 2013). In addition, five





**Figure 13.** Phase-folded light curves for a couple of the shortest-period binaries found by our search. The period,  $P$ , listed is the total period of oscillation used to fold the light curves.

sources are ellipsoidal variables indicating that the periodicity is due to the gravitational distortion of the stars (Morris 1985).

### 5.3 Young stellar objects and extra-galactic variables

We compare  $R$  values and amplitudes for blazars and young stellar objects (YSOs). We cross-match all variable sources in our sample with the blazar catalogues of D’Abrusco et al. (2019) and the YSO catalogues of Marton et al. (2016) as mentioned in subsection 3.1. We find 44,196 YSOs of which 237 were classified as periodic and 981 blazars of which 2 were classified as periodic. Seven sources classified both as blazars and YSOs are removed from subsequent analysis.

In Fig. 14 we show the results of this comparison. The distributions for  $R$  are largely similar with peaks around  $\sim 0.2$ – $0.3$  indicating that both YSOs and blazars tend to vary on longer time-scales. For amplitude, the YSOs are peaked below 2 mag while the blazar distribution is more uniform.

### 5.4 Separation of periodic and aperiodic variables with non-parametric features

Thus far, we have used our non-parametric measures for the classification of periodic variables identified with a periodogram. In this section, we discuss further applications of these measures. In particular, we show that the non-parametric features contain the requisite information to identify periodic sources.

Fig. 15 shows the distribution of periodic and aperiodic sources for a variety of the non-parametric measures. We additionally show the distribution of W1 S/N ratio, since the quality of the data can influence the determination of periodicity. For many of these measures (especially the ratio of short-term to long-term variability), there is a marked difference between the distributions of periodic and aperiodic variables.

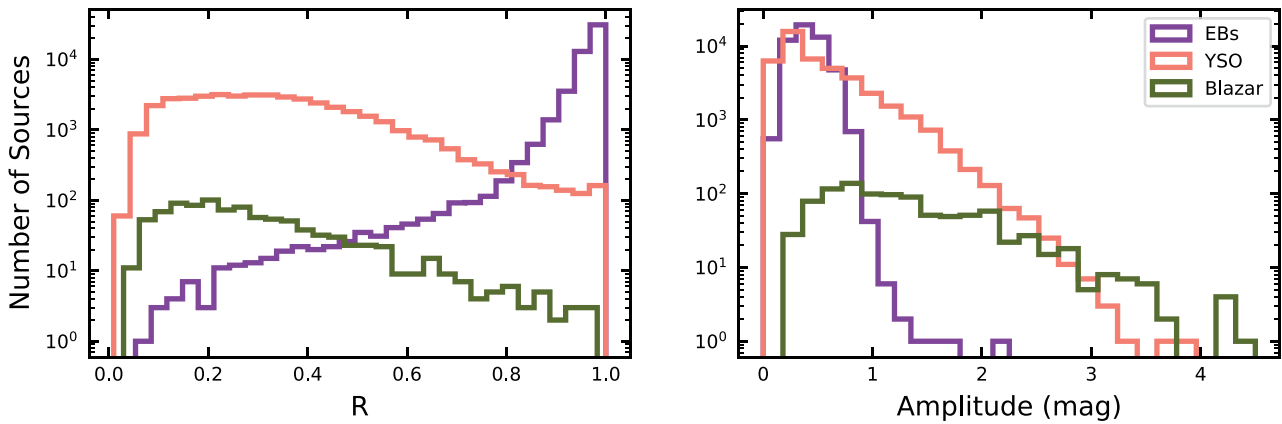
We investigate the potential of discriminating periodic variables from aperiodic variables using non-parametric features alone. Using empirical testing, we select the five most informative features – the W1 magnitude, W1 signal-to-noise ratio, estimated variability amplitude, fainter fraction, and the ratio between long-term and short-term variability. The variability amplitude and standard deviation have similar distributions (Fig. 15), but we find that the amplitude is a marginally better discriminator. We apply the quality cuts of

Section 2.7 on the W1 magnitude to exclude saturated and very faint sources and to ensure that we are not classifying trivially based on this data quality metric.

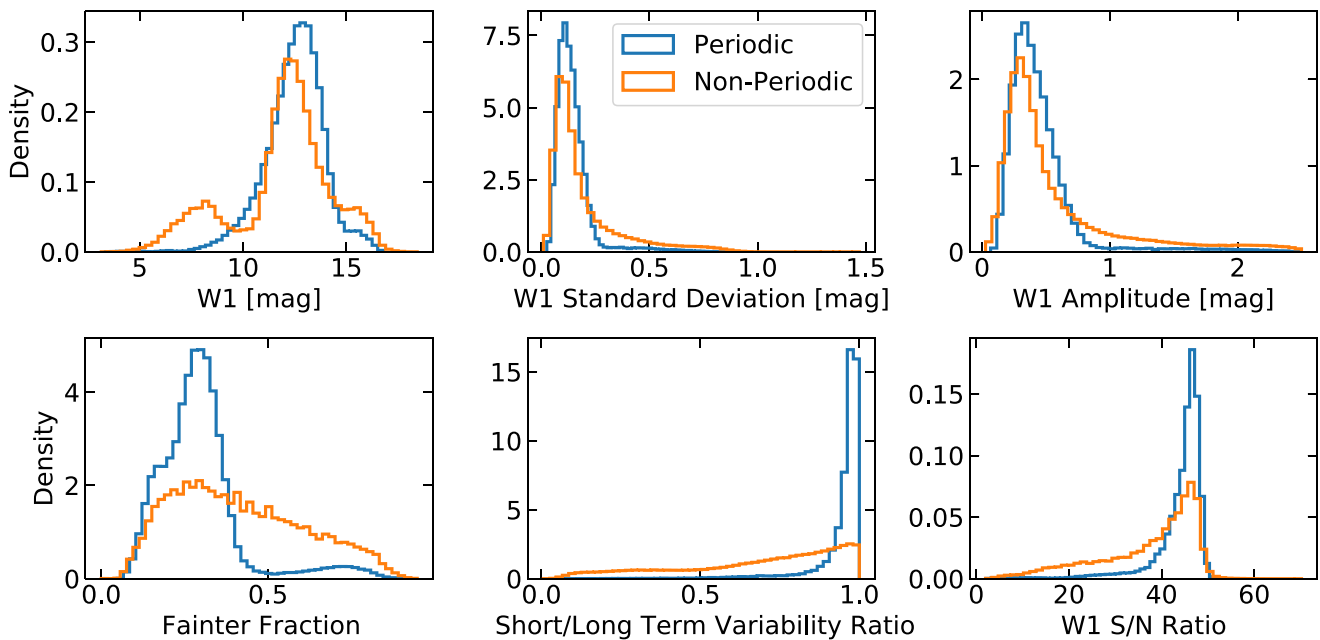
We define a two-class classification problem where the ‘positive’ class is defined as periodic variability, and the ‘negative’ class corresponds to aperiodic variability. We consider two classification models – the logistic regression and the random forest. In both cases, the general framework is the same – the model coefficients are solved for by ‘training’ on stars with known classes. The models can then perform predictions on new data, returning a probability  $\hat{p}$  that a given star exhibits periodic variability. The logistic regression model assumes a linear relation between the features and the log-odds of a star exhibiting periodic variability (Yu, Huang & Lin 2011). The random forest assumes no parametric model, instead relying on logical decision trees to map the feature space to the true and false classes (Breiman 2001).

To evaluate our methods, we randomly select 75 per cent of the stars in our sample for training, and make predictions on the remaining validation stars (25 per cent) that were left out of training. In both the train and test samples, we randomly downsample the number of aperiodic stars to make it balanced with the number of periodic stars. Repeating our experiment for different random selections (or different ratios of train/test sample size) does not significantly affect our results. We find that, on average, the classifier correctly classifies stars as aperiodic or periodic 90 per cent of the time with the logistic regression, and 94 per cent with the random forest. The random forest model outperforms the logistic regression, likely because a linear model is insufficient to describe the relationships between our features. A key metric for our particular use-case is precision – the probability that a star classified as periodic is genuinely periodic. A higher precision means a lower false positive rate, preventing wasted follow-up resources. We find that the random forest model trained on the five non-parametric features achieves an average precision of 97 per cent on our test data, an improvement over the logistic regression’s 91 per cent. The confusion matrix of the predictions on the test data set is as follows, with true aperiodic and true periodic as the columns from left to right, and classified aperiodic and classified periodic as rows from top to bottom:

$$\begin{bmatrix} 11859 & 617 \\ 856 & 12098 \end{bmatrix}$$



**Figure 14.** Left: Ratio of variation amplitudes in the short to long terms,  $R$ , for cross-matched young stellar objects and blazars. EBs are shown for comparison. Right: Amplitude for cross-matched young stellar objects and blazars.



**Figure 15.** Distribution of periodic and aperiodic variables for a variety of non-parametric measures. The peak at 8th magnitude in the W1 magnitude plot is caused by saturation, rather than by physical variability, which is why there is not a corresponding peak in the distribution of periodic variables (Nikutta et al. 2014). We remove these sources as well as very faint sources with the quality cut of Section 2.7. The fainter fraction plot shows that our sample is dominated by periodic variables of the occulting type. The histograms are normalized to have unit area, to account for the fact that our sample contains more aperiodic variables than periodic variables.

As is desirable, most objects fall on the diagonal of the confusion matrix, and the number of false positives and false negatives is similar.

An interesting application of these tools is to use the probabilities returned by the classifier to rank interesting candidates. The probabilities themselves can be incorporated into a hierarchical search that uses other prior information to inform the selection. In future large-scale surveys, such a pre-selection will be essential to efficiently allocate computational resources by running the full periodogram analysis on high-confidence periodic candidates first.

The brief demonstration of this section is intended mainly as a proof-of-concept of the information content of these features and has some important caveats. The first caveat is that one of the

features ( $S/N$  ratio) measures data quality while others deal with real information about the source from the flux PDF and time-scales of variability. The discrimination based on data quality is trivial; without good data it is difficult to flag a source as periodic. Therefore, it is important to ensure that the classifier is not biased to predict periodicity when the data quality is high. First, we quantify the relative importance of the input features. The random forest classifier has a built-in metric that ranks the importance of the input features, based on how much that feature contributes to the prediction quality (Breiman 2001). The features ordered from most to least important are the short-to-long term variability ratio, W1 signal-to-noise ratio, fainter fraction, estimated amplitude, and W1 magnitude.

Including the S/N ratio as a feature improves the classification accuracy by  $\sim 5$  per cent compared to if solely the W1 magnitude is used. We interpret this as the S/N ratio breaking a degeneracy in the variability measures between noisy sources and truly variable sources – objects with low S/N will automatically display variability due to noise, so including the S/N ratio as an explicit feature enables the classifier to incorporate this information. As another check, we run the classification on only high S/N sources, and yield nearly identical performance. Therefore, the classification appears to rely on meaningful information about the source, and does not seem to be biased by data quality.

A second caveat is that this demonstration is limited by the nature of the training sample. For example, due to the *WISE* cadence, long-period variables are almost completely absent from our training sample and so we do not expect the classifier to be adept at identifying them. Also, the use of periodogram-based classification as the ‘ground-truth’ for training biases the classifier to be more likely to detect similar types of periodicity as the periodogram, albeit at a much lower computational cost.

That said, the non-parametric approach has some important advantages over the periodogram and can even yield valuable information without reference to a periodogram-based analysis (e.g. Rimoldini et al. 2019). PDF measures like this are more robust to photometric errors than the periodogram, and are meaningful even in systems with changing periods. These systems are of great interest, but would not normally be detected by the periodogram because implicit in the periodogram approach is the assumption of a constant period.

## 6 CONCLUSIONS

In this paper, we present an analysis of  $\sim 450\,000$  *WISE* variables. The variables are identified based on their r.m.s. variability in the AllWISE survey (Hoffman et al. 2012). We combine their AllWISE and NEOWISE light curves and apply quality cuts to filter out low-quality individual exposures.

Using these combined light curves, we conduct periodogram analysis to identify  $\sim 56\,000$  periodic variables,  $\sim 19\,000$  of which were not included in the previous *WISE* periodic variable catalogue of Chen et al. (2018). We search for periodicity over a finely spaced grid in frequency between 0.1 and  $20\text{ d}^{-1}$  and use the MHAOV periodogram with five model parameters. This periodogram models the light curve as a series of periodic, orthogonal functions and uses a variance statistic to quantify the quality of the fit at each test frequency. To classify an object as periodic, we require a significant value of the statistic at a frequency non-coincident with that of the *WISE* observing cadence, a sufficiently small period uncertainty to prevent a large phase offset over the time-series duration, and good phase coverage. We cross-match with *Gaia* DR2 and use the distribution of our periodic variables in the space of colour-absolute magnitude to get a sense of our catalogue contents. We find that the sample is dominated by eclipsing binaries.

Next, we compute a variety of non-parametric variability measures for all of the  $\sim 450\,000$  *WISE* variables and show that these measures are useful both for (a) separating periodic and aperiodic variables and (b) classifying the type of periodicity. These non-parametric measures include a variety of measures of the flux distribution range and moments. We introduce a new measure of the third moment and compare it to previously existing measures. These non-parametric measures also include an estimate of the time-scale of variability ( $R$ , the ratio of variability amplitude on day time-scales to that on month/year time-scales).

In terms of classification (b), we show that these interpretable and easily implemented measures provide an effective means of isolating eclipsing binaries from other types of periodic variability. We identify an all-sky sample of  $\sim 51\,000$  eclipsing binaries in the infrared. The majority of these binaries are contact or near-contact making them prime targets for future study.

In terms of the identification of periodic variables (a), we demonstrate the high information content of the non-parametric features by using them to identify periodic variables at a much lower computational cost than the traditional, periodogram-based analysis. This type of analysis can be used to speed future studies of periodic variability and, in some cases, bypass the periodogram-based analysis entirely (e.g. Rimoldini et al. 2019). Furthermore, the non-parametric method overcomes some of the short-comings of periodograms. Importantly, because it does not implicitly assume a constant signal period as the periodogram does, it offers the possibility of identifying binaries exhibiting orbital evolution on human time-scales.

## ACKNOWLEDGEMENTS

The authors are grateful to the anonymous referee for the constructive feedback. This publication makes use of data products from the *Wide-field Infrared Survey Explorer*, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This research made use of the cross-match service provided by CDS, Strasbourg. This research has made use of the VizieR catalogue access tool, CDS, Strasbourg, France. EP was supported in part by the Provost’s Undergraduate Research Award at Johns Hopkins University. EP and HCH acknowledge support by the Space@Hopkins seed grant. The authors acknowledge support from NASA ADAP grant 80NSSC19K0581. The MHAOV periodogram and the orthogonal trigonometric function fit are implemented with the time-series analysis package of Alex Schwarzenberg-Czerny that is based on a wrapper scheme by Ewald Zietsman. This research made use of Astropy, a community-developed core PYTHON package for Astronomy (Astropy Collaboration 2013, 2018).

## DATA AVAILABILITY

The full catalogue of *WISE* variables, periodic variables, and binaries is available as an electronic supplement to this paper.<sup>8</sup> The data model is listed in Table 1.

## REFERENCES

- Abbott B. P. et al., 2017, *Phys. Rev. Lett.*, 119, 161101  
 Arenou F. et al., 2018, *A&A*, 616, A17  
 Astropy Collaboration, 2013, *A&A*, 558, A33  
 Astropy Collaboration, 2018, *AJ*, 156, 123

<sup>8</sup><https://zakamska.johnshoptkins.edu/data.htm>

- Avvakumova E. A., Malkov O. Y., Kniazev A. Y., 2013, *Astron. Nachr.*, 334, 860
- Bass G., Borne K., 2016, *MNRAS*, 459, 3721
- Belczynski K., Holz D. E., Bulik T., O'Shaughnessy R., 2016, *Nature*, 534, 512
- Bellm E. C. et al., 2019, *PASP*, 131, 018002
- Borkovits T., Hajdu T., Sztakovics J., Rappaport S., Levine A., Bíró I. B., Klagyivik P., 2016, *MNRAS*, 455, 4136
- Borucki W. J. et al., 2010, *Science*, 327, 977
- Braga V. F. et al., 2019, *A&A*, 625, A1
- Breiman L., 2001, *Mach. Learn.*, 45, 5
- Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, *MNRAS*, 427, 127
- Brown W. R., Gianninas A., Kilic M., Kenyon S. J., Allende Prieto C., 2016, *ApJ*, 818, 155
- Burke, Edward W. J., Rolland W. W., Boy W. R., 1970, *J. R. Astron. Soc. Can.*, 64, 353
- Chen X., Wang S., Deng L., de Grijs R., Yang M., 2018, *ApJS*, 237, 28
- Clementini G. et al., 2019, *A&A*, 622, A60
- Conroy K. E. et al., 2020, *ApJS*, 250, 34
- Cowperthwaite P. S. et al., 2017, *ApJ*, 848, L17
- D'Abrusco R. et al., 2019, *ApJS*, 242, 4
- Das S., Bhardwaj A., Kanbur S. M., Singh H. P., Marconi M., 2018, *MNRAS*, 481, 2000
- Debosscher J., Sarro L. M., Aerts C., Cuypers J., Vandenbussche B., Garrido R., Solano E., 2007, *A&A*, 475, 1159
- Debosscher J. et al., 2009, *A&A*, 506, 519
- Doyle L. R. et al., 2011, *Science*, 333, 1602
- Drake A. J. et al., 2009, *ApJ*, 696, 870
- Drake A. J. et al., 2013, *ApJ*, 763, 32
- Drake A. J. et al., 2014a, *ApJS*, 213, 9
- Drake A. J. et al., 2014b, *ApJ*, 790, 157
- Drake A. J. et al., 2017, *MNRAS*, 469, 3688
- Dubath P. et al., 2011, *MNRAS*, 414, 2602
- Duchêne G., Kraus A., 2013, *ARA&A*, 51, 269
- Dworetzky M. M., 1983, *MNRAS*, 203, 917
- Eggleton P. P., 1983, *ApJ*, 268, 368
- Evans D. W. et al., 2018, *A&A*, 616, A4
- Eyer L., Bartholdi P., 1999, *A&AS*, 135, 1
- Eyer L. et al., 2018, *Gaia DR2 documentation Chapter 7: Variability*
- Eyer L., Süveges M., De Ridder J., Regibo S., Mowlavi N., Holl B., Rimoldini L., Bouchy F., 2019, *PASP*, 131, 088001
- Fabrycky D., Tremaine S., 2007, *ApJ*, 669, 1298
- Ferraz-Mello S., 1981, *AJ*, 86, 619
- Findeisen K., Cody A. M., Hillenbrand L., 2015, *ApJ*, 798, 89
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Frescura F. A. M., Engelbrecht C. A., Frank B. S., 2007, preprint ([arXiv:0706.2225](https://arxiv.org/abs/0706.2225))
- Frescura F. A. M., Engelbrecht C. A., Frank B. S., 2008, *MNRAS*, 388, 1693
- Gaia Collaboration, 2016, *A&A*, 595, A1
- Gaia Collaboration, 2018a, *A&A*, 616, A1
- Gaia Collaboration, 2018b, *A&A*, 616, A10
- Gaia Collaboration, 2019, *A&A*, 623, A110
- Graham M. J., Drake A. J., Djorgovski S. G., Mahabal A. A., Donalek C., Duan V., Maker A., 2013, *MNRAS*, 434, 3423
- Hamer J. H., Schlafman K. C., 2019, *AJ*, 158, 190
- Harding L. K., Hallinan G., Boyle R. P., Golden A., Singh N., Sheehan B., Zavala R. T., Butler R. F., 2013, *ApJ*, 779, 101
- Hartman J. D., Gaudi B. S., Holman M. J., McLeod B. A., Stanek K. Z., Barranco J. A., Pinsonneault M. H., Kalirai J. S., 2008, *ApJ*, 675, 1254
- Hillenbrand L. A., Findeisen K. P., 2015, *ApJ*, 808, 68
- Hoffman D. I., Cutri R. M., Masi F. J., Fowler J. W., Marsh K. A., Jarrett T. H., 2012, *AJ*, 143, 118
- Holl B. et al., 2018, *A&A*, 618, A30
- Horne J. H., Baliunas S. L., 1986, *ApJ*, 302, 757
- Hurley J., Tout C. A., 1998, *MNRAS*, 300, 977
- Hwang H.-C., Zakamska N., 2020, *MNRAS*, 493, 2271
- Hwang H.-C., Hamer J. H., Zakamska N. L., Schlafman K. C., 2020, *MNRAS*, 497, 2250
- Ivanova N. et al., 2013, *A&A Rev.*, 21, 59
- Ivezić Ž. et al., 2019, *ApJ*, 873, 111
- Jayasinghe T. et al., 2018, *MNRAS*, 477, 3145
- Jayasinghe T. et al., 2019a, *MNRAS*, 486, 1907
- Jayasinghe T. et al., 2019b, *MNRAS*, 485, 961
- Kasliwal M. M., 2012, *Publ. Astron. Soc. Aust.*, 29, 482
- Kim D.-W., Bailer-Jones C. A. L., 2016, *A&A*, 587, A18
- Kinemuchi K., Smith H. A., Woźniak P. R., McKay T. A., ROTSE Collaboration, 2006, *AJ*, 132, 1202
- Kirk B. et al., 2016, *AJ*, 151, 68
- Knigge C., Baraffe I., Patterson J., 2011, *ApJS*, 194, 28
- Koch D. G. et al., 2010, *ApJ*, 713, L79
- Kochanek C. S. et al., 2017, *PASP*, 129, 104502
- Koen C., 1990, *ApJ*, 348, 700
- Koen C., 2006, *MNRAS*, 371, 1390
- Kolenberg K., 2012, *J. Am. Assoc. Var. Star Obs.*, 40, 481
- Lachowicz P., Zdziarski A. A., Schwarzenberg-Czerny A., Pooley G. G., Kitamoto S., 2006, *MNRAS*, 368, 1025
- Lachowicz P., Gupta A. C., Gaur H., Wiita P. J., 2009, *A&A*, 506, L17
- Lafler J., Kinman T. D., 1965, *ApJS*, 11, 216
- Lomb N. R., 1976, *Ap&SS*, 39, 447
- Luhman K. L., Sheppard S. S., 2014, *ApJ*, 787, 126
- Madore B. F. et al., 2013, *ApJ*, 776, 135
- Mainzer A. et al., 2011, *ApJ*, 731, 53
- Mainzer A. et al., 2014, *ApJ*, 792, 30
- Maoz D., Mannucci F., Nelemans G., 2014, *ARA&A*, 52, 107
- Marrese P. M., Marinoni S., Fabrizio M., Altavilla G., 2019, *A&A*, 621, A144
- Marton G., Tóth L. V., Paladini R., Kun M., Zahorec S., McGehee P., Kiss C., 2016, *MNRAS*, 458, 3479
- Masi F. J., Hoffman D. I., Grillmair C. J., Cutri R. M., 2014, *AJ*, 148, 21
- Matt S. P., Do Cao O., Brown B. P., Brun A. S., 2011, *Astron. Nachr.*, 332, 897
- Moe M., Di Stefano R., 2017, *ApJS*, 230, 15
- Molnár L., Plachy E., Juhász Á. L., Rimoldini L., 2018, *A&A*, 620, A127
- Morris S. L., 1985, *ApJ*, 295, 143
- Mowlavi N. et al., 2018, *A&A*, 618, A58
- Murphy S. J., Hey D., Van Reeth T., Bedding T. R., 2019, *MNRAS*, 485, 2380
- Newville M., Stensitzki T., Allen D. B., Ingargiola A., 2014, *Zenodo*
- Nikutta R., Hunt-Walker N., Nenkova M., Ivezić Ž., Elitzur M., 2014, *MNRAS*, 442, 3361
- Norton A. J. et al., 2011, *A&A*, 528, A90
- Paczynski B., Szczygieł D. M., Pilecki B., Pojmański G., 2006, *MNRAS*, 368, 1311
- Palaversa L. et al., 2013, *AJ*, 146, 101
- Pietrinfermi A., Cassisi S., Salaris M., Castelli F., 2004, *ApJ*, 612, 168
- Postnov K. A., Yungelson L. R., 2014, *Living Rev. Relativ.*, 17, 3
- Preston G. W., 1964, *ARA&A*, 2, 23
- Renson P., 1978, *A&A*, 63, 125
- Richards J. W. et al., 2011, *ApJ*, 733, 10
- Richards J. W., Starr D. L., Miller A. A., Bloom J. S., Butler N. R., Brink H., Crellin-Quick A., 2012, *ApJS*, 203, 32
- Ricker G. R. et al., 2015, *J. Astron. Telescopes Instrum. Syst.*, 1, 014003
- Riess A. G. et al., 2011, *ApJ*, 730, 119
- Rimoldini L., 2014, *MNRAS*, 437, 147
- Rimoldini L. et al., 2012, *MNRAS*, 427, 2917
- Rimoldini L. et al., 2019, *A&A*, 625, A97
- Roelens M. et al., 2018, *A&A*, 620, A197
- Rucinski S. M., 2006, *MNRAS*, 368, 1319
- Rucinski S. M., 2007, *MNRAS*, 382, 393
- Sarro L. M., Debosscher J., López M., Aerts C., 2009, *A&A*, 494, 739
- Scargle J. D., 1982, *ApJ*, 263, 835
- Schmidt G. D., Szkody P., Silvestri N. M., Cushing M. C., Liebert J., Smith P. S., 2005, *ApJ*, 630, L173
- Schwarzenberg-Czerny A., 1991, *MNRAS*, 253, 198



- Schwarzenberg-Czerny A., 1995, *A&AS*, 110, 405  
 Schwarzenberg-Czerny A., 1996, *ApJ*, 460, L107  
 Schwarzenberg-Czerny A., 1998, *Balt. Astron.*, 7, 43  
 Schwarzenberg-Czerny A., 1999, *ApJ*, 516, 315  
 Schwarzenberg-Czerny A., 2003, in Sterken C., ed., *ASP Conf. Ser. Vol. 292, Interplay of Periodic, Cyclic and Stochastic Variability in Selected Areas of the H-R Diagram*. Astron. Soc. Pac., San Francisco, p. 383  
 Shappee B. J. et al., 2014, *ApJ*, 788, 48  
 Siopis C., Sadowski G., Mowlavi N., Holl B., Lecoeur-Taïbi I., Eyer L., 2020, *Contrib. Astron. Obs. Skalnaté Pleso*, 50, 414  
 Smartt S. J. et al., 2017, *Nature*, 551, 75  
 Soszyński I. et al., 2016, *Acta Astron.*, 66, 405  
 Stepień K., 1995, *MNRAS*, 274, 1019  
 Temmink K. D., Toonen S., Zapartas E., Justham S., Gänsicke B. T., 2020, *A&A*, 636, A31  
 Thorstensen J. R., Halpern J. P., 2009, *Astron. Telegram*, 2177, 1  
 Torrealba G. et al., 2015, *MNRAS*, 446, 2251  
 Tylenda R. et al., 2011, *A&A*, 528, A114  
 Udalski A., 2003, *Acta Astron.*, 53, 291  
 VanderPlas J. T., 2018, *ApJS*, 236, 16  
 Weisberg J. M., Nice D. J., Taylor J. H., 2010, *ApJ*, 722, 1030  
 Wright E. L. et al., 2010, *AJ*, 140, 1868  
 Yu H.-F., Huang F.-L., Lin C.-J., 2011, *Mach. Learn.*, 85, 41  
 Zakamska N. L., Greene J. E., 2014, *MNRAS*, 442, 784

## SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

**Table 1.** Column descriptions for our catalogue of *WISE* variables.

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  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.