Variables, periodic variables and contact binaries in WISE

Evan Petrosky,1 Hsiang-Chih Hwang,1 Nadia L. Zakamska,1 Vedant Chandra,1 Matthew J. Hill,2

1Department of Physics & Astronomy, Johns Hopkins University, 3400 N Charles St, Baltimore, MD 21218, USA
2Knovita, 1405 Point Street, Baltimore, MD 21231, USA

(Dated: December 10, 2020)

ABSTRACT

The time-series component of WISE is a valuable resource for the study of stellar variability. We present an analysis of an all-sky sample of ∼450,000 AllWISE+NEOWISE infrared light curves of likely variables identified in AllWISE. By computing periodograms of all these sources, we identify ∼63,500 periodic variables. Of these, ∼44,000 are short-period (P < 1 day), 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 timescales of variability. Combining variability measures from our periodogram and non-parametric analyses with WISE infrared colors and absolute magnitudes, colors and root-mean-square variability from Gaia yields a powerful engine 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.

Keywords: binaries: close – binaries: eclipsing – binaries: general – catalogues – methods: statistical – stars: variables: general

1. INTRODUCTION

Binary systems 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 proving 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 (Tylenda et al. 2011; Kasliwal 2012) and they 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 et al. 2010; Knigge et al. 2011; Maoz et al. 2014; Postnov & Yungelson 2014; Brown et al. 2016; Belczynski et al. 2016; Smartt et al. 2017; Cowperthwaite et al. 2017; Abbott et al. 2017; Temmink et al. 2020).

Over the last 20 years, the study of binaries has been aided by an abundance of 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) and in space (Kepler – Borucki et al. 2010; Koch et al. 2010; TESS – Ricker et al. 2015). Future surveys like LSST (Ivezic et al. 2019) will further add to this plethora of data. Many surveys have associated eclipsing binary (EB) catalogs (e.g. ASAS – Paczynski 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 catalog and the current release identifies a variety of other types of variability (Holl et al. 2018; Gaia Collaboration et al. 2019; Rimoldini et al. 2019; Clementini et al. 2019; Siopis et al. 2020).

The growth in variability datasets has been accompanied by a rise in complexity of methods for variable classification. Traditionally, variable objects are classified based on the similarity of their light curves and colors to known variable prototypes (Gaia Collaboration et al. 2019). Often, a time-series analysis tool such as a periodogram is run to differentiate between peri-
odic and aperiodic variables. This step requires a clear understanding of the effects of 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., color) and used as inputs into a classifier (e.g., Debosscher et al. 2007; Richards et al. 2011; Dubath et al. 2011; Richards et al. 2012; Masci et al. 2014; Jayasinghe et al. 2019a,b; Eyer et al. 2019). There are publicly available light curve classifying engines that rely on machine learning techniques (e.g., Kim & Bailer-Jones 2016). 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 reduce computational overhead, some previous works have used non-parametric variability measures (Kinemuchi et al. 2006; Palaversa et al. 2013; Drake et al. 2013, 2014a, 2017; Torrealba et al. 2015; Hillenbrand & Findeisen 2015; Findeisen et al. 2015). We build upon this work and develop new, non-parametric light curve analysis techniques and apply these techniques on 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 timescales (Hoffman et al. 2012) make WISE a unique probe of Galactic variability. Previously, Chen et al. (2018) used WISE to select ∼50,000 variable candidates of which ∼42,000 were binaries. Here, we expand on this work and present an analysis of ∼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 produce an all-sky sample of ∼63,500 periodic variable candidates, of which ∼56,000 are binaries.

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: $W_{\text{AB}} = W_{\text{Vega}} + \Delta m_i$, with $\Delta m_i = (2.699, 3.339, 5.174, 6.620)^3$.

## 2. WISE DATA AND METHOD

### 2.1. WISE Mission

The Wide-field Infrared Survey Explorer (WISE) was launched in December 2009 and conducted observations of the entire sky in bands centered on 3.4, 4.6, 12 and 22 μm (W1, W2, W3, and W4 respectively) until 2010 when it ran out of coolant (Wright et al. 2010). From 2010 to 2011 it conducted observations in the 3.4 and 4.6 μm 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 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 2013², 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 ∼5700 seconds (∼0.066 days=1.6 hours)³ 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 ∼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 due to differences in the mission during the two phases. 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)

---

1. [https://wise2.ipac.caltech.edu/docs/release/allsky/expsup/sec4_4h.html](https://wise2.ipac.caltech.edu/docs/release/allsky/expsup/sec4_4h.html)
2. [http://wise2.ipac.caltech.edu/docs/release/allwise/](http://wise2.ipac.caltech.edu/docs/release/allwise/)
3. [http://wise2.ipac.caltech.edu/docs/release/allsky/expsup/sec1_1.html](http://wise2.ipac.caltech.edu/docs/release/allsky/expsup/sec1_1.html)
find systematic changes in W1 between AllWISE and NEOWISE-R to be 0.01 magnitude for sources with $8<W1<14$ magnitude and on the order of 0.1 magnitude for sources with $14<W1<15$ magnitude. Outside this range there are magnitude dependent systematic offsets between AllWISE and NEOWISE data. We limit our sources to $8<W1<15$ mag to ensure concordant AllWISE and NEOWISE measurements and to exclude saturated sources (Nikutta et al. 2014; Mainzer 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 because the time-series analysis is not reliable in the case of a large magnitude offset. A typical flux measurement in the range $8<W1<15$ mag has an error of $\sim 0.025$ mag. We incorporate these flux 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 $\text{var}_{\text{fig}}$ ranging from 0 to 9 in each band, such that the probability that the object’s flux does not vary in said band is $\propto 10^{-\text{var}_{\text{fig}}}$ (Hoffman et al. 2012).

We select WISE variable sources having $\text{W1}\text{ var}_{\text{fig}} \geq 6$. We only consider sources having $\text{cc_flags} == 0$ to ensure that there are no imaging artifacts and $\text{ext}_{\text{flag}} \leq 1$ to ensure that it is not an extended source. After applying these quality cuts, we download $\approx 500,000$ light curves from the AllWISE multi-epoch photometry table and the NEOWISE-R single exposure source table using a $1''$ matching distance.

To eliminate redundancies and possible extraneous matches in the table, the $\text{allwise\_cntr}$ listed in the NEOWISE-R single exposure table is mapped to the $\text{source\_id\_mf}$ in the AllWISE multi-epoch photometry table. 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 quality cuts were applied: $\text{saas\_sep}>5.0$ deg (image outside of the South Atlantic anomaly), $\text{moon\_masked} == 0000$ (frame unaffected by light scattered off the moon), and $\text{qi\_fact} > 0.9$ (only the highest quality frames). 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 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 lightcurve download with quality flags is available through Github.

2.3. Time Series Analysis

The periodogram is a time-series analysis tool that allows for the location 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. Figure 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; Scargle 1982; Ferraz-Mello 1981; Schwarzenberg-Czerny 1999). We use 5 model parameters to allow for better sensitivity to anharmonic oscillations. This corresponds to fitting the data in real space with with a Fourier series of 2 harmonics (Schwarzenberg-Czerny 2003, 1999; Graham et al. 2013; Lachowicz et al. 2006).

We adopt a frequency grid spacing of 0.0001 days$^{-1}$ (Graham et al. 2013). We search the frequency range 0.1 to 20 days$^{-1}$. This extension beyond the traditional Nyquist limit of $\sim 7.6$ days$^{-1}$ that corresponds to uniform sampling at the WISE orbital cadence is justified when the data is not uniformly spaced (see subsection 2.6 for further discussion).

2.4. Periodogram Peak Significance

Upon observing a periodogram peak, we seek to reject the null hypothesis that the light curve in question consists of pure noise and that the highest peak in the periodogram results from the chance alignment of random

---

4 http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec2_1a.html
5 http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec3_1a.html
6 http://wise2.ipac.caltech.edu/docs/release/neowise/expsup/sec2_3.html
7 https://github.com/HC-Hwang/wise_light_curves
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 higher the likelihood of periodicity. The periodogram peak occurs at a frequency of $f_{\text{max}} = 4.44$ days$^{-1}$ ($P_{\text{app}} = 0.225$ days). **Right:** Phase-folded light curve of the same object. This object is an eclipsing binary with approximately symmetric eclipses. This means that the apparent period, $P_{\text{app}}$ i.e. the period picked out by the periodogram, is half the actual orbital period. We fold this light curve with a frequency of $0.5 \times f_{\text{max}} = 2.22$ days$^{-1}$ which corresponds to a period of $2 \times P_{\text{app}} = 0.45$ days.

Figure 2. **Left:** Distribution of maximum statistic values versus the number of observations in the light curve ($N_{\text{LC}}$) for periodograms calculated on 12,000 white-noise light curves. The data are binned based on $N_{\text{LC}}$. The smallest bin contains about 1,000 light curves and most bins have about 1,500. Cyan crosses show an estimate of the 99.9$^{\text{th}}$ percentile of the corresponding statistic values in each bin. The black dashed line shows the best-fit power law model using LMFIT (Newville et al. 2014). **Right:** Maximum statistic value 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.

Although some inroads have been made toward an analytical understanding of peak significance (e.g. Horne & Baliunas 1986; Koen 1990; Schwarzenberg-Czerny 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 et al. (2007, 2008) to quantify peak significance. We generate 12,000 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. To do this, we start with 12,000 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, $N_{\text{LC}}$. 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 84.2$^{\text{th}}$ and 15.8$^{\text{th}}$ percentile. Next, we create a Gaussian distribution with mean $\langle m \rangle$ and standard deviation $0.5w$ and randomly draw $N_{\text{LC}}$ values from this distri-
bution to create a new array of Gaussian (white) noise data, \( \text{noise}_i \). By substituting \( \text{noise}_i \) for \( m_i \) for each light curve, we create 12,000 white-noise light curves.

We run a 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. Figure 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 \( N_{\text{LC}} \). When \( N_{\text{LC}} \) 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 Figure 2, we bin the data based on \( N_{\text{LC}} \). Every bin has \(~1,500\) simulated light curves save the last one on the right which has \(~1,000\). In each bin, we study the resultant empirical cumulative distribution function of the calculated maximum statistic values and estimate the 99.9\(^{th}\) percentile. Then we fit a power law model of the form \( \theta_{\text{cutoff}}(N_{\text{LC}}) = A(N_{\text{LC}} - N_0)^k \) to estimate the appropriate cut-off value as a function of \( N_{\text{LC}} \) to reject the null hypothesis. The best-fit values are \( A = 154 \pm 57 \), \( N_0 = 14.9 \pm 4.7 \), and \( k = -0.344 \pm 0.082 \). 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. The bin with the highest number of light-curve data points (i.e. \( N_{\text{LC}} \gtrsim 300 \)) had the fewest number of sources spread over the largest range. To account for the less-precise estimate of the cumulative distribution function in this region and for the fact that the power law fit approaches zero as \( N_{\text{LC}} \) increases, we adopt a minimum statistic value of \( \theta = 10 \) to reject the null hypothesis.

### 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 truly periodic signals that are passed over (VanderPlas 2018). The completeness, i.e. the proportion of truly periodic signals we expect to detect given our method and data, is a complex function of the nature of periodicity (Schwarzenberg-Czerny 1999). 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 use simulations to gauge our ability to detect contact binaries with near-sinusoidal light curves as a function of signal amplitude and period. We randomly select 100 light curves that have W1 magnitudes between 10 and 14, the range best probed by WISE. We then simulate a sinusoidal signal centered 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 days and calculate the MHAOV periodogram, phase-fold the result and see if the source would have been classified as periodic. We iterate over all light curves for a given period and then increase the period and repeat until the period exceeds a value of 2 days. We choose periods between 0.05 and 2 days because the majority of our periodic sources are contained in this range. In this interval, aside from the periods that coincide with the WISE observing cadence, the recovery rate is nearly constant at \(~77\%\) of the simulated periodic variables. Next, we repeat the same procedure but this time fixing the period at a value of 0.22 days (a value typical of our close binaries) and varying the amplitude between value of 0.05 magnitude and 2 magnitude. For amplitudes of 0.2 mag and above, the distribution is constant with a median recovery rate of 82%. Below 0.2 mag, the recovery rate drops to 14\%\) at a magnitude of 0.05 mag. In summary, we estimate that we can detect \(~80\%\) of near-sinusoidal signals with periods between 0.05 and 2 days and peak-to-peak amplitudes above 0.2 magnitude.

### 2.6. Period Uncertainty

After rejecting the null hypothesis that the object is aperiodic, the next task is to select 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 the inverse sampling rate (Nyquist frequency) is familiar in the case of evenly-spaced sampling. Uniform observations at the WISE satellite period (\(~95\) minutes) would correspond to a Nyquist frequency of \(~7.6\) days\(^{-1}\). 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 multi-harmonic 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 1999).
The dashed line is a plot of the MC frequency error derived via MCMC simulations, as an eclipsing binary composed of similar stars, periodogram. In the case of symmetric light curves, such as 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).

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 apparent period ($P_{app}$) as given by the periodogram. In the case of symmetric light curves, such as an eclipsing binary composed of similar stars, $P_{app}$ corresponds to half of the orbital period.

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 $\sigma_{freq}$ (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 Figure 3, the MC-derived errors, $\sigma_{MC}$, are correlated with the periodogram-derived errors, $\sigma_{freq}$, but are systematically higher by ~ 25%. 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 $\sim 10^{-6}$ – $10^{-7}$ days, 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, $\theta_{max}$, exceeds the cutoff value given by the power-law fit of Figure 2 for the appropriate number of points. 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%. 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% implicitly requires that the sources have at least 18 data points. Finally, we exclude sources whose highest periodogram peak is at 1 or 0.5 times the orbital frequency of the WISE satellite ($f_{peak} \notin [7.52, 7.68]$ days$^{-1}$ and $f_{peak} \notin [14.65, 15.7]$ days$^{-1}$). With the above cuts, we find 63,526 periodic variable candidates. Figure 4 shows the distribution of apparent periods and apparent periods over the period errors for these sources.

### 3. CATALOG CONTENTS

#### 3.1. Catalog Contents and Gaia Cross-Match

We cross-match our periodic variable candidates with Gaia DR2 (Gaia Collaboration et al. 2016, 2018a; Evans et al. 2018; Arenou et al. 2018) using the pre-computed, pre-calibrated Gaia catalog of photometric properties (Arenou et al. 2018).
best-neighbor WISE/Gaia cross-match catalog of Marrese et al. (2019). We find 188,043 matches out of a total of 454,103 variable objects (∼ 41.4%) which is similar to the total percentage of AllWISE sources that have a best-neighbor cross-match (39.83% - see Marrese et al. 2019). Of our 63526 periodic variable candidates, 54,756 or ∼ 86.1% have a best-neighbor 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⟩ = 0.916 mag as compared to the periodic variables with ⟨W1-W3⟩ = 0.334 mag. As a further check, we cross-match our sample with the WISE young stellar object catalogs of Marton et al. (2016) and find 44,196 matches of which only 289 are flagged as periodic.

We use the Gaia cross-match and considerations of the limitations of WISE to get a sense of the catalog contents and make the case that the sample is dominated by eclipsing binaries. We search for periods between 0.05 and 10 days, but due to the WISE observing cadence, the sensitivity drops for periods above 2 days. As a result, we do not expect many long-period variables in the more luminous part of the color-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 δ Scuti variables because their variability amplitudes are too low to be reliably detected (Murphy et al. 2019). 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 colors, we require:

1. parallax\_over\_error > 10
2. phot\_g\_mean\_flux\_over\_error > 50
3. phot\_bp\_mean\_flux\_over\_error > 10
4. visibility\_periods\_used > 8
5. phot\_rp\_mean\_flux\_over\_error > 10

In addition, we restrict phot\_bp\_rp\_excess\_factor in accordance with Gaia Collaboration et al. (2018b). After applying these cuts we retain 37,962 sources.

Figure 5 shows the Gaia color-absolute magnitude diagram of the WISE periodic variables that satisfy the above quality cuts. They are color-coded by the median apparent period in each color-magnitude bin. Most of the WISE periodic variables are main-sequence stars. On the main sequence, the most common form of periodic stellar variability is eclipsing binaries. 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 MG ∼ 1 and BP-RP ∼ 0.7 with periods between ∼0.2 days and 1 day (Preston 1964; Kolenberg 2012; Das et al. 2018), significantly different from the surrounding EBs in the color-magnitude diagram. A few periodic sources are located below the main sequence. They may be white dwarf-brown dwarf binaries or cataclysmic variables.

Figure 6 shows another view of the Gaia color-absolute magnitude diagram color-coded by the ratio of optical variability to infrared (IR) variability. To calculate the optical flux variability from Gaia, we follow the methods of Hwang et al. (2020) by computing the fractional variability from the Gaia photometric errors and then accounting for instrumental errors. We plot only those sources here that display appreciable variability

Figure 4. Left: Distribution of apparent (i.e. as calculated by the periodogram) periods ($P_{app}$) for 54,662 periodic variables with apparent periods in the range 0.05 to 1 d. Right: Distribution of $\log_{10}(P_{app}/\sigma_{P_{app}})$ for the 63,526 identified periodic variables.
3.2. Comparison with Chen et al. (2018) Catalog

Chen et al. (2018) present a catalog 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 different quality cuts to the initial 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 arcsecond. Increasing the matching distance to 5 arcseconds does not significantly alter the results. Of all the sources with $6 \leq \text{var\_flag} \leq 9$ that we analyzed, 41,532 have a cross-match with the 50,282 periodic variables of Chen et al. (2018). Their remaining 8,750 variables were excluded by our initial quality cuts and were never a part of our periodogram analysis. 8,703 of the variables were excluded by our cuts on $\text{cc\_flags}$ to avoid contamination and confusion and the remaining 47 were excluded by our cuts on $\text{ext\_flg}$ to remove extended objects. Of the 41,532 variables in common, we mark 38,681 as periodic. We miss some of the sources due to our more stringent cuts on the quality of individual exposures, reducing the number of points in the light curve and preventing us from detecting them as periodic. We find compatible periods for 38,297 (99%) of the cross-matched sources. Sources are said to have compatible periods if they satisfy:

$$| \frac{P_{\text{app}} - P_{\text{chen18}}}{P_{\text{app}}} | < 0.10$$

or

$$| \frac{P_{\text{app}} - 0.5P_{\text{chen18}}}{P_{\text{app}}} | < 0.10$$

The 384 outliers with non-compatible periods correspond to cases where either (a) our more stringent cuts significantly reduce the number of data points and leads to a different period estimate, (b) the detected period is greater than 5 days 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.

For $\sim$88% of the matches, our apparent periods, $P_{\text{app}}$, 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 indistinguishable. We define the difference between the half periods $\Delta P = 0.5P_{\text{chen18}} - P_{\text{app}}$. Over

Figure 5. Color-absolute magnitude diagram. Gray scale bins represent a sample of 500,000 Gaia sources that meet the selection of Gaia Collaboration et al. (2018b). Overlaid in color are our WISE periodic variables color-coded by the median apparent period in each color-magnitude bin.

Figure 6. Gaia color-magnitude diagram color-coded by the ratio of optical variability from Gaia G band to infrared (IR) variability from WISE W1. A group of RR Lyrae can be seen in dark blue 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. The black line is the Pleiades main-sequence fit.
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 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 Figure 7. We also report the standard skewness in the catalog.

Finally, we introduce another non-parametric measure to describe the timescale of variability. Due to the peculiar cadence of WISE observations – multiple observations within a few days’ time, but then not again for about six months – WISE has sensitivity to variability on a 1 day time scale, as well as to long-term variations. We use *R*, the ratio of the average r.m.s. variability within any day-long stretch to the r.m.s. variability of the entire light curve. Given that WISE has only a limited number of ~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 timescales 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 *σ*.

### 4. 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 Figure 8 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 ~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 ~0.15 and of similar amplitudes has detached binaries with more narrow eclipses. The offshoot from the upper sequence at *amplitude* ~0.3 mag and *FF* ~0.25 seems to contain RR Lyrae.

We cross-match our sample with the Gaia RR Lyrae catalog of Clementini et al. (2019) and find 2172 matches. Figure 9 shows the kernel density estimate
Figure 7. Left: The light curve of an Algol-type eclipsing binary (WISEJ234343.55+812751.6) phase-folded with the apparent period \( P_{\text{app}} \) 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 \textit{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.

Figure 8. Kernel density estimate for all periodic variables in the space of two of our non-parametric measures, \textit{fainter fraction} (\( FF \)) and \textit{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 \( P = 2 \times P_{\text{app}} \), a detached EB found in the lower horizontal sequence phase-folded with period \( P = 2 \times P_{\text{app}} \), a candidate RR Lyrae found in the offshoot from the upper sequence.
| Column         | Data Type | Units | Description                                           |
|---------------|-----------|-------|-------------------------------------------------------|
| wise_id       | str19     | –     | WISE 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 |
| mean_mag      | float64   | mag   | Mean magnitude                                         |
| std_mag       | float64   | mag   | Standard deviation of magnitude                       |
| amp           | float64   | mag   | Amplitude                                             |
| FF            | float64   | –     | Fainter fraction                                      |
| phase_cov     | float64   | –     | Phase coverage                                        |
| R             | float64   | –     | Ratio of short- to long-term variability              |
| max_stat      | float64   | –     | Maximum statistic value                                |
| skew          | float64   | –     | Skewness                                              |
| kur           | float64   | –     | Kurtosis                                              |
| cutoff_stat   | float64   | –     | Cutoff maximum statistic value to reject null hypothesis |
| periodic      | bool      | –     | Periodic sources receive a value of True              |
| Papp          | float64   | d     | Apparent period                                       |
| sigP          | float64   | d     | Apparent period error                                 |
| EB            | bool      | –     | Eclipsing Binaries receive a value of True            |

**Table 1.** Column descriptions for our catalog of WISE variables. A default value of -1.0 is assigned when no measurement could be made for a given source and column. 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 catalog is available online.
of the cross-matched RR Lyrae in the \textit{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 gray, 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 apparent periods to between 0.25 and 1 days. After applying this cut, both the upper and lower sequences disappear and the RR Lyrae offshoot becomes the dominant feature in the \textit{fainter fraction-amplitude} space. We choose a lower period bound of 0.25 days because we do not expect many RR Lyrae with periods below 0.25 days$^{-1}$ and we do not want to exclude a high number of close eclipsing binaries. This figure shows that these three variables - period, \textit{amplitude}, and \textit{fainter fraction} - provide a means of separating RR Lyrae from EBs without resorting to the color-magnitude diagram.

All told, this selection labels 6,632 of the periodic variables as candidate RR Lyrae. Included in this are 1,868 out of the 2,172 ($\sim 86\%$) Gaia RR Lyrae from the cross-matched catalog 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 catalog (Clementini et al. 2019). As mentioned above, we expect few Cepheids in the sample. The cross-match reveals only 386 Gaia Cepheids and we remove these from the EB sample.

The above cuts remove 1,906 out of 2,464 ($\sim 77\%$) of the RR Lyrae identified by Chen et al. (2018) while retaining 30,656 out of the 33,426 sources ($\sim 92\%$) 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 780 Cepheids classified by Chen et al. (2018). We are left with a low contamination sample of 56,323 eclipsing binary candidates.

5. DISCUSSION

5.1. Main-Sequence Binaries

We next turn to the color-magnitude diagram to isolate main-sequence EBs. Of our initial sample of 56,323 EB candidates, 48,902 have a Gaia best-neighbor cross-match. Applying the quality cuts of Section 3 leaves us with 35,684 sources. To select main-sequence binaries, we start with the Hamer & Schlaufman (2019) Pleiades spline fit shown in Figure 6. Binaries are expected to be found above the main sequence for a wide range of mass ratios (Hurley & Tout 1998). We require that the source have an absolute G-band magnitude above the Pleiades spline fit and within 1.5 magnitudes of the spline fit value for a given color. This leaves 23,079 sources.

Figure 10 shows the distribution of apparent period ($P_{app} = 0.5P_{orbital}$) versus Gaia BP-RP color for our main-sequence EBs detected with WISE. The distribution is smooth across the entire period range and includes periods not probed by Chen et al. (2018). In blue is a kernel density estimate and the black line represents the theoretical minimum possible apparent period for an equal-mass, contact, main sequence eclipsing binary of a fixed age (1 Gyr) and solar metallicity as a function of mass and color. 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 undis
torted star (i.e. the Roche lobe volume radius, \( R_L \)). Following Eggleton (1983), we use the relationship between \( R_L \) and the semi-major 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

\[
P_{app} = 0.5P_{\text{orbital}} = \frac{a^3}{G(M_1 + M_2)}^{\frac{1}{2}}
\]

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 color, younger than 1 Gyr, or of lower-than-solar metallicity. Some RR Lyrae are located close to the main sequence and may be present on the blue side of the plot \((G_{BP} - G_{RP} < 0.6)\) accounting for the spillover in the bottom left. The dearth of sources hugging the line in the upper-left of the plot does not appear be due to decreased sensitivity to those periods (see subsection 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 main-
sequence EBs is located about at \( 2P_{app} \approx 0.34 \) days. This is higher than the maximum of the period distribution of \( \sim 0.27 \) days found by Rucinski (2007), but this sample is not volume-limited.

5.2 Shortest-Period Objects

There are 146 sources in our eclipsing binary candidates that have apparent periods below 0.1 days. 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 nine are the result of bad data or an erroneous period measurement and that the remaining 137 represent real periodic signals. Figure 11 shows some example phase-folded light curves for these objects. These sources are prime candidates for comparison with other short-period binary catalogs (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 arcsecond. After removing the bad data, we find 42 matches of which 37 were classified in SIMBAD as a specific type of variability. Of these 37, we find that four are pulsating, one is a long-period variable (that also exhibits short-period variability), and one is a Carbon star. The remaining 31 are all classified as some sort of binary. Five of these sources are cataclysmic variables (V* BL Hya; SDSS J121209.30+013627.7; V* V347 Pav; RX J2218.5+1925; V* LX Ser). We find compatible periods for all of these cataclysmic variables indicating that our period measurements are accurate even in this short-period regime (Ramsay et al. 2004; Schmidt et al. 2005; Thorstensen & Halpern 2009; Avvakumova et al. 2013; Drake et al. 2014a). In addition, five 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 vari-
able sources in our sample with the blazar catalogs of D’Abrusco et al. (2019) and the YSO catalogs of Marton et al. (2016) as mentioned in subsection 3.1. We find 44,196 YSOs of which 289 were classified as periodic and 981 blazars of which 16 were classified as periodic. Seven sources classified both as blazars and YSOs are removed from subsequent analysis.

In Figure 12 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 timescales. For amplitude, the YSOs are peaked below 2 magnitude 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.

Figure 13 shows the distribution of periodic and aperiodic sources for a variety of the non-parametric measures. We additionally show the distribution of W1 signal-to-noise (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 (Figure 13), 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 et al. 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 90% of the stars in our sample and train the models on them – we then make predictions on the remaining validation stars (10%) that were left out of training. Repeating our experiment for different random selections (or even different ratios of train/test sample size) does not affect our results. We find that we can correctly identify stars as periodic using only the non-parametric features with 95% accuracy. 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...
Figure 12. Left: Ratio of short- to long-term variation, $R$, for cross-matched young stellar objects and blazars. EBs are shown for comparison. Right: Amplitude for cross-matched young stellar objects and blazars.

Figure 13. 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.

A lower false positive rate, preventing wasted follow-up resources. Additionally, the precision better reflects the performance of the model in this application due to the imbalance between classes – there are six times more aperiodic variables than periodic variables. We find that the random forest model trained on the five non-parametric features achieves an average precision of 90% on our test data, significantly better than the logistic regression’s 70%. The confusion matrix of the classifier is balanced, i.e. the number of false positives and false negatives is similar. This indicates that the classifier is not biased to arbitrarily favor one class over the other.

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 timescales of variability. The discrimination based on data quality is trivial; without good data it is impossible 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 verify that the S/N ratio is not the most important feature for the classification: by testing different combinations of features we find that several of other measures are more informative, and the majority of the discriminating power comes from the physical metrics and not just the data quality. Including the S/N ratio does improve the classification accuracy by \( \sim 5\% \). 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.

The 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. 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 \(~450,000\) 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 lightcurves and conduct periodogram analysis to identify \(~63,500\) periodic variable candidates. To this end, we search for periodicity over a finely spaced grid in frequency between 0.1 and 20 days\(^{-1}\). For light curves that meet our quality cuts, we model the light curve as a series of periodic, orthogonal functions with five parameters and compute a variance statistic to quantify the quality of the fit. Good phase coverage and a significant value of the statistic at a frequency non-coincident with that of the WISE observing cadence are required to classify an object as periodic.

With our periodic variables in hand, we cross-match with Gaia DR2. We use the distribution of our periodic variables in the space of color-absolute magnitude to get a sense of our catalog contents. We also cross-match with a previous catalog of WISE periodic variables. We find that the sample is dominated by eclipsing binaries.

Next, we turn to less computationally expensive measures of variability. Collapsing the light curve into a flux distribution and measuring non-parametric estimates of its range and second and third moments yields a wealth of information. Together with our non-parametric measure of the timescale of variability \(R\), the ratio of variability on day timescales to that on month/year timescales), these measures can be used to cheaply and effectively (a) identify periodic variables and (b) classify the type of periodicity.

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 \(~56,000\) eclipsing binaries in the infrared. Binary formation and evolution is interesting in its own right (Stepien 1995; Fabrycky & Tremaine 2007; Ivanova et al. 2013; Duchêne & Kraus 2013; Borkovits et al. 2016; Moe & Di Stefano 2017; Hwang et al. 2020) and binary systems have been linked to a wide range of astronomical phenomena (e.g. Kasliwal 2012; Belczynski et al. 2016). The majority of these binaries are contact or near-contact making them prime targets for future study.

In terms of 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 could be used to speed future studies
of periodic variability and, in some cases, bypass the periodogram-based analysis entirely. Furthermore, the non-parametric method overcomes some of the shortcomings 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 timescales.

Data Availability: The full catalog of WISE variables, periodic variables, and binaries is available as an electronic supplement to this paper. The data model is listed in Table 1.

ACKNOWLEDGMENTS

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 and by NASA ADAP 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 et al. 2013, 2018).

REFERENCES

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, PhRvL, 119, 161101, doi: 10.1103/PhysRevLett.119.161101
Arenou, F., Luri, X., Babusiaux, C., et al. 2018, A&A, 616, A17, doi: 10.1051/0004-6361/201833234
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f
Avvakumova, E. A., Malkov, O. Y., & Kniazev, A. Y. 2013, Astronomische Nachrichten, 334, 860, doi: 10.1002/asna.201311942
Belczynski, K., Holz, D. E., Bulik, T., & O’Shaughnessy, R. 2016, Nature, 534, 512, doi: 10.1038/nature18322
Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002, doi: 10.1088/1538-3873/aacbe
Borkovits, T., Hajdu, T., Sztakovics, J., et al. 2016, MNRAS, 455, 4136, doi: 10.1093/mnras/stv2530
Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977, doi: 10.1126/science.1185402
Breiman, L. 2001, Machine Learning, 45, 5, doi: 10.1023/A:1010933404324
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127, doi: 10.1111/j.1365-2966.2012.21948.x
Brown, W. R., Gianninas, A., Klicic, M., Kenyon, S. J., & Allende Prieto, C. 2016, ApJ, 818, 155, doi: 10.3847/0004-637X/818/2/155
Chen, X., Wang, S., Deng, L., de Grijs, R., & Yang, M. 2018, ApJS, 237, 28, doi: 10.3847/1538-4365/aad32b
Clementini, G., Ripepi, V., Molinaro, R., et al. 2019, A&A, 622, A60, doi: 10.1051/0004-6361/201833374

https://zakamska.johnshopkins.edu/data.htm
Stepien, K. 1995, MNRAS, 274, 1019, doi: 10.1093/mnras/274.4.1019
Temmink, K. D., Toonen, S., Zapartas, E., Justham, S., & Gänsicke, B. T. 2020, A&A, 636, A31, doi: 10.1051/0004-6361/201936889
Thorstensen, J. R., & Halpern, J. P. 2009, The Astronomer’s Telegram, 2177, 1
Torrealba, G., Catelan, M., Drake, A. J., et al. 2015, MNRAS, 446, 2251, doi: 10.1093/mnras/stu2274
Tylenda, R., Hajduk, M., Kamiński, T., et al. 2011, A&A, 528, A114, doi: 10.1051/0004-6361/201016221
Udalski, A. 2003, AcA, 53, 291. https://arxiv.org/abs/astro-ph/0401123
VanderPlas, J. T. 2018, ApJS, 236, 16, doi: 10.3847/1538-4365/aab766
Weisberg, J. M., Nice, D. J., & Taylor, J. H. 2010, ApJ, 722, 1030, doi: 10.1088/0004-637X/722/2/1030
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868, doi: 10.1088/0004-6256/140/6/1868
Yu, H.-F., Huang, F.-L., & Lin, C.-J. 2011, Machine Learning, 85, 41, doi: 10.1007/s10994-010-5221-8