Draft version March 9, 2023
Typeset using LaTeX **twocolumn** style in AASTeX63

ZTF IC 10 variable catalog

[1] and [1]

[1] *New York University Abu Dhabi*

(Dated: March 9, 2023)

ABSTRACT

[ChatGPT abstract]We present the ZTF IC 10 variable catalog, which is a comprehensive catalog of variable sources in IC 10, a nearby dwarf irregular galaxy. The catalog is based on the ZTF (Zwicky Transient Facility) survey data, which covers a large area of the sky and provides a high cadence of observations. We classify the sources in the catalog into three categories: non-variable, non-periodic variable, and periodic variable. We compare the ZTF IC 10 variable catalog with several other catalogs of variable sources in IC 10, including those based on previous surveys and those derived from theoretical models. Our analysis shows that the ZTF IC 10 variable catalog provides a significant improvement in terms of completeness and accuracy compared to previous catalogs. We also investigate the properties of the variable sources in IC 10, including their spatial distribution, variability amplitude, and period distribution. Our study provides valuable insights into the nature and evolution of variable sources in dwarf irregular galaxies, and serves as a useful resource for future studies of IC 10 and other similar systems. The ZTF IC 10 variable catalog is publicly available and can be accessed through the ZTF data portal.[end ChatGPT]

[My modification to ChatGPT]We present the ZTF IC 10 variable catalog, which is a comprehensive catalog of variable sources in IC 10, a nearby dwarf irregular galaxy. The catalog is based on the ZTF (Zwicky Transient Facility) survey data, which covers a large area of the sky and provides a high cadence of observations. We classify the sources in the catalog into three categories: non-variable, non-periodic variable, and periodic variable. We cross match our ZTF IC 10 variable catalog with several other catalogs of variable sources in IC 10, including those based on previous surveys or from different classification methods. Our analysis shows that our ZTF IC 10 variable catalog provides a significant improvement in terms of completeness and accuracy compared to previous catalogs. We also investigate the properties of the variable sources in IC 10, including their population demographics, magnitude distribution, and color-magnitude diagram. Our study provides valuable insights into the nature and evolution of variable sources in dwarf irregular galaxies, and serves as a useful resource for future studies of IC 10 and other similar systems. The ZTF IC 10 variable catalog is publicly available and can be accessed through [somewhere].

1. CHATGPT INTRODUCTION (I BELIEVE SOME OF THE CITATIONS ARE MADE-UP/FAKE)

The study of variable sources in nearby galaxies is important for a variety of reasons. Variable sources provide important clues to the underlying physical mechanisms that drive their variability, as well as insights into the properties and evolution of the Universe. For example, classical Cepheids are widely used as distance indicators, allowing us to measure the distances to nearby galaxies (e.g., Freedman & Madore 2010). In addition, variable sources can provide a wealth of information about the astrophysical processes that take place in these systems, such as mass transfer in binary systems, stellar pulsations, and supernova explosions (e.g., Hubble 1926; Strom 1987).

IC 10 is a nearby irregular galaxy in the Local Group that has been the subject of several studies focused on its variable sources. IC 10 is particularly interesting because it is a relatively nearby galaxy with a high rate of star formation, and it contains a diverse range of variable sources (Yang et al. 2011). The study of variable sources in IC 10 can provide insights into the properties and evolution of star-forming galaxies, as well as the astrophysical processes that drive the variability of these sources.

The Zwicky Transient Facility (ZTF) is one of several large-scale surveys that aim to identify and characterize variable sources in the sky. ZTF is a time-domain survey

that uses a dedicated 48-inch telescope and a wide-field camera to observe the sky in multiple filters (Bellm et al. 2019). ZTF has already produced a wealth of data on variable sources, including the identification of over 2500 classical Cepheids in the Milky Way (Masci et al. 2019).

Previous work has shown that ZTF data can be used to identify and classify variable or periodic sources in a variety of settings (e.g., Graham et al. 2019; De et al. 2020). In particular, Hoffmann et al. (2018) used ZTF data to identify and classify variable sources in IC 10, finding a total of 1348 periodic sources. However, their study was limited to sources with periods shorter than 50 days, and did not include non-periodic variable sources.

In this paper, we present a catalog of variable sources in IC 10, based on ZTF data spanning several years. Our catalog includes both periodic and non-periodic sources, and covers a wide range of periods and variability types. We use a machine learning algorithm to classify the sources in our catalog, and we compare our results to previous studies of variable sources in IC 10. Our catalog will serve as a valuable resource for future studies of the properties and evolution of variable sources in IC 10, and will help to shed light on the underlying physical mechanisms that drive their variability.

2. INTRODUCTION

A nearby galaxy called IC 10 is made of a large and diverse collection of massive stars, such as Luminous Blue Variables, Wolf-Rayet stars, High-Mass X-ray binaries. Studying the transients in IC 10 will help revealing how massive stars affect their environments. [need to make it longer. some of paragraphs written by ChatGPT seems nice.]

In this work, Zwicky Transient Facility (ZTF) r band and g band data will be used to identify optical transients. Many work has been done to identify transients using ZTF data. Chen et al. (2020) published a ZTF catalog of periodic variable stars. They discovered 781,602 periodic variables using Lomb-Scargle periodogram based false-alarm probability, and further classified these periodic variable into 11 class labels according to their distribution across many parameters like period, phase difference, amplitude, amplitude ratio, and absolute magnitude. Cheung et al. (2021) further proposed a new model to better classify these periodic variables in Chen's catalog using a convolutional variational autoencoder and hierarchical random forest. However, in Chen's catalog, only one periodic variable star falls in the field of IC 10. There should be more than only one interesting sources inside IC 10, therefore here we closely examine all ZTF sources within the IC 10 region to come up with a finer collection of periodic variable stars, as well as non-periodic variables.

In Section 3 we will introduce the data used in this work. Section 4 details our classification scheme. In Section 5 we present interesting statistics of our catalog in various aspects. In Section 6 and 7 we reference our catalog to many other catalogs and classification methods. [to be continued]

3. DATA

A brief summary of data used in this work is presented in Table 1.

[Modified from ChatGPT]The Zwicky Transient Facility (ZTF) is a wide-field optical time-domain survey that began operations in March 2018 (Bellm et al. 2019). ZTF aims to discover and study transient and variable astrophysical sources, such as supernovae, active galactic nuclei, and asteroids, among others. The survey uses a 47 square degree field of view and a cadence of 3 nights per month to cover the northern sky visible from California. ZTF data (Masci et al. 2018) are publicly available through the Infrared Science Archive (IRSA)[1]. The data include images, catalogs, and time series photometry of millions of celestial sources.

ZTF has observed IC 10 roughly everyday during the past few years, making it a perfect dataset to work on. Here we use the ZTF DR 15 data in g band and r band, since they are two most sensitive ZTF bands. ZTF g band has a filter central wavelength of 4803 Å and FWHM of 1321 Å, ZTF r band has a filter central wavelength of 6434 Å and FWHM of 1557 Å. We do a radial query of $225''$ around the center of IC 10 (RA = 00 : 20 : 23.16, DEC = +59 : 17 : 34.7 (Cotton et al. 1999)) to cover the optical size of IC 10 (Huchra et al. 1999). 2407 and 1334 ZTF labeled lightcurves are found in r and g band, respectively. The lightcurves cover roughly a four-year span starting from March 2018. Only clear observations without any flags are used in this work.

4. METHODOLOGY

For every lightcurve, we first determine whether it is a variable or a non-variable by fitting it to a constant line and looking at the survival function $S_k(\chi^2)$ of the constant line model. For variables, we apply Lomb-Scargle analysis and determine whether it is a periodic variable or non-periodic variable by cutting on the false alarm probability (FAP), along with de-aliasing cuts Q_P and C_T. In the end, we group OIDs (ZTF source label)

[1] https://irsa.ipac.caltech.edu/Missions/ztf.html

Filter info	ZTF g band: $\lambda_{cen} = 4803$ Å, FWHM = 1321 Å ZTF r band: $\lambda_{cen} = 6434$ Å, FWHM = 1557 Å
IC 10 field	$225''$ radius circle around RA = 00 : 20 : 23.16, DEC = +59 : 17 : 34.7
Dates covered	March 2018 - Nov 2022, 58194 - 59892 in MJD

Table 1. Summary of data used in this paper

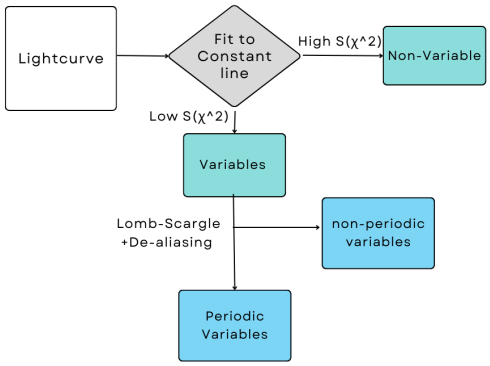

Figure 1. Visualization of our methodology

with position less than $1''$ apart as the same source, and resolve the resulting possible degeneracy in variability according the the number of observations. A summary of our criteria can be find in flowchart (Figure 1) and Table 2.

4.1. *Variability cut* $S_k(\chi^2)$

A non-variable source by definition has constant magnitude with time. By fitting the lightcurve with a line of constant value and look at how good the fit is, we can very effectively separate possible variable sources from non-variable sources. For a lightcurve with n observations of magnitudes m_i and errors in magnitude σ_i , the chi-square χ^2 of the best fitting constant magnitude m_C is

$$\chi^2 = \sum_{i=1}^{n} \frac{(m_i - m_C)^2}{\sigma_i^2} \tag{1}$$

For a non-variable source, changes in m is due to statistical error, but not intrinsic variations, therefore a constant magnitude m_C fit should give $\chi^2 \approx n$. By chance, the χ^2 of a constant source can be larger than n. Such chance can be estimated by the cumulative distribution function (CDF) of a χ^2 distribution. We hence measure the probability that an observed lightcurve matches the constant line model via survival function $S_k(\chi^2)$, the chance that a constant line fits the observed lightcurve,

$$S_k(\chi^2) = 1 - \mathrm{CDF}_k(\chi^2) \tag{2}$$

where $\mathrm{CDF}_k(\chi^2)$ is the cumulative density function of a χ^2 distribution with degree of freedom k, when $k = n-1$. We define variable sources to be sources with 99% or higher chance being variable, that is, sources with

$$S_k(\chi^2) \leq \frac{1}{100 \times N_{lc}} \tag{3}$$

where N_{lc} is the total number of lightcurves in the corresponding band. The rest are classified as non-variable sources. The distribution of $S_k(\chi^2)$ and its cutoff can be visualized in Figure 2.

4.2. *Classify variables with periodicity*

$S_k(\chi^2)$ distinguishes variables and non-variables. Lomb-Scargle periodogram based criteria FAP, Q_P, and C_T further classifies variables into periodic variables and non-periodic variables.

4.2.1. *FAP*

FAP, false alarm probability, is a quantity that is very widely used to look for periodic variables. FAP originates from the Lomb-Scargle periodogram, a commonly used tool to characterize periodic signals in unevenly spaced observations named after Lomb (1976) and Scargle (1982). A periodogram calculates the power on a frequency/period grid, and FAP gives the probability of measuring a peak at certain or higher power assuming Gaussian noise and non-periodic data. A smaller FAP hints that the sources is more likely to be periodic.

Here we use Python package *Astropy* (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018) to calculate the Lomb-Scargle periodogram, and then *Astropy* further approximates the false alarm probabilities at Lomb-Scargle periodogram peak following Baluev (2008). The distribution of FAP in r and g band is shown in Figure 3. To get sources with 99% or higher probability of having a non-false alarm, we set FAP cutoffs at $\frac{1}{100 \times N_{lc}}$, where N_{lc} is the total number of lightcurves in the corresponding band. Specifically, we keep r band sources with FAP $< 4.15 \times 10^{-6}$ and g band sources with FAP $< 7.36 \times 10^{-6}$ as "low FAP" sources. Although FAP is one of the most effective criteria to detect periodic signals, it is vulnerable to aliasing. The nightly/quarter/yearly observation window pattern can create fake peaks and contaminate the real peak on the periodogram. The following two criteria are mainly devoted to the de-aliasing of low FAP sources.

	Definition	Cutoff in r band	Cutoff in g band
$S_k(\chi^2)$	Survival function of the constant magnitude model (Variables/All lightcurves in band)	$S_k(\chi^2) < 4.15 \times 10^{-6}$ (250/2407)	$S_k(\chi^2) < 7.36 \times 10^{-6}$ (117/1334)
FAP	False alarm probability from Lomb-Scargle (Low FAP/Variables)	FAP $< 4.15 \times 10^{-6}$ (86/250)	FAP $< 7.36 \times 10^{-6}$ (29/117)
Q_P	$P_{\rm LS}/P_{\rm window}$ (Low FAP, High Q_P/Low FAP)	$Q_P > 1$ (53/86)	$Q_P > 1$ (18/29)
C_T	$\lvert T_{\rm LS} - T_{\rm window}\rvert/T_{\rm window}$ (Periodic variables/Low FAP, High Q_P)	$C_T > 0.01$ $C_T < 0.49,\ C_T > 0.91$ (40/53)	$C_T > 0.01$ $C_T < 0.49,\ C_T > 0.91$ (10/18)
	Merge nearby lightcurves into the same source (Periodic variable sources/Periodic variable lightcurves) (All sources in band/All lightcurves in band)	$d < 1''$ (32/40) (1516/2407)	$d < 1''$ (9/10) (864/1334)

Table 2. Periodic variable criteria. The fractions in parentheses denotes the (subcategory/parent catetory) the cut acts on, and the number (before/after) the cut.

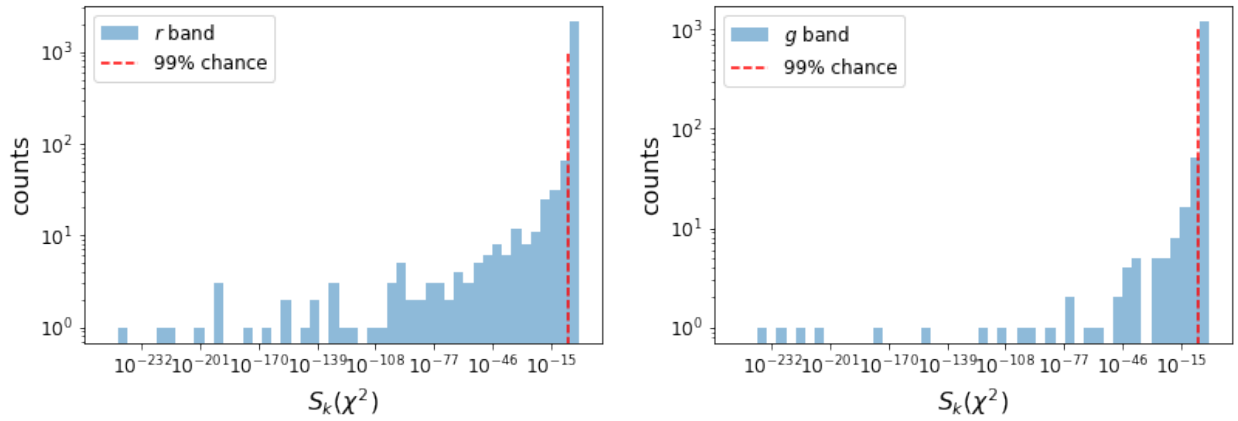

Figure 2. Distribution of $S_k(\chi^2)$ for all sources in ZTF r and g band. Sources left to the red dotted line has 99% or more chance to be variable and is classified as variable.

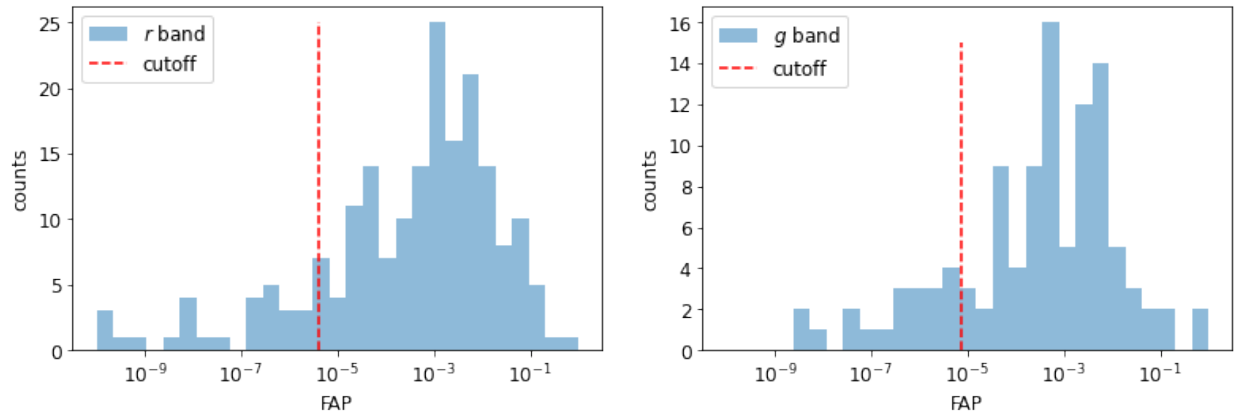

Figure 3. Distribution of FAP for variables in ZTF r and g band. We keep sources left to the red dotted line as low FAP sources.

4.2.2. Q_P

Q_P is the ratio of powers,

$$Q_P = P_{\rm LS}/P_{\rm window} \tag{4}$$

where $P_{\rm LS}$ is the Lomb-Scargle periodogram peak power, and $P_{\rm window}$ is the window function power. Window function refers to recalculating the Lomb-Scargle periodogram of a lightcurve by changing all magnitude to a constant value while keeping observation window the same. Window function power is the power at the period of the original Lomb-Scargle periodogram peak. $Q_P > 1$ indicates the period or frequency we found is more significant in the lightcurve of interest than in a straight line given same spacing in observations. An intrinsically non-periodic source could have a low FAP due to aliasing, but it not likely to have a $Q_P > 1$. Therefore such a cut will help removing aliasing. The distribution of Q_P for low FAP sources in both bands is shown in Figure 4.

4.2.3. C_T

C_T is the relative change in periods,

$$C_T = |T_{\rm LS} - T_{\rm window}|/T_{\rm window} \tag{5}$$

where $T_{\rm LS}$ is the period at Lomb-Scargle periodogram peak power, $T_{\rm window}$ is the period at window function peak power. C_T tells how close the period we found is to the aliasing period. We thus rule out sources with $C_T < 0.01$, i.e. period and aliasing period only differ by 1%. As Figure 5 shows, there is a suspicious spike between $C_T = 0.5$ and 1 in both bands. Due to the fact that ZTF make observations on a daily basis, most of sources have window function period $T_{\rm window} = 1\ day$. Aliases reappears following $\frac{1}{T_{\rm alias}} = \frac{1}{T_{\rm true}} + \frac{n}{T_{\rm window}}$ for integer values of n. In cases with no true period ($T_{\rm true} \to \infty$) and $T_{\rm window} = 1\ day$, we get $T_{\rm alias} = 1, 1/2, 1/3, 1/4, ...$ These aliasing period will return $C_T = 1/2, 2/3, 3/4, ...$, exactly where the spike is located. To resolve such aliasing, we rule out sources with $0.49 < C_T < 0.91$, accounting n up to 10, since higher n will have negligible aliasing effect. To prevent removing real periodic sources with periods lower than 0.5 days, we preserve sources that has extremely low FAP (FAP $< 10^{-30}$).

4.2.4. *Group nearby OIDs*

Criteria outlined above classifies all the lightcurves into three categories: non-variable, non-periodic variable, and periodic variable. These three categories is visualized in Figure 6 in terms of number of observations and mean magnitude. While we see a clear trend where brighter objects are observed more frequency as expected, there are two distinct line of distributions, one with higher number of observations, the other with lower number of observations. This is since ZTF's automated labeling scheme sometimes tags the same source with multiple OIDs. We group OIDs that are less than $1''$ apart as one source, and notice that in most cases, each source has two OIDs associated with it, one with < 100 number of observations, the other with higher number of observations. This explains the two distinct lines shown in Figure 6. In our catalog, for a sources with multiple OIDs, the variability and periodicity is determined by OID with highest number of observations, since it is easier to capture a variable pattern with more observations. If a source's highest number of observation OID do not shown any variability, but has an OID with lower number of observations that is variable, a variable suspect flag will be raised in the catalog, although such case is very rare (1 in r band and 4 in g band). We do not place a periodic suspect flag since any periodicity found in OID with smaller number of observations but not in OID with higher number of observations is unlikely to be true. Note that after the $1''$ grouping, most of the grouped OIDs should truly be the same source, but there are chances when they are actually different sources that are very close to each other. To account for this issue, we match our catalog with Gaia, which has a higher angular resolution that can resolve sources within $1''$. A more detailed discussion of Gaia counterparts can be found in Section 7.2.

5. CATALOG DEMOGRAPHICS

5.1. *General demographics*

A brief view of how many sources are variables, periodic variables, and non-periodic variables can be found in Table 3. Note here the number of "sources" is presented, but not number of "lightcurves", since some of the lightcurves belong to the same source, as mentioned in section 4.2.4. As shown in the table, there are more sources observed, or exclusively observed in r band than in g band, because sources in this work are typically ~ 1 magnitude (~ 2.5 times) brighter in r band than in g band, making r band more sensitive. Looking at the percentages, in both bands, roughly 11-12% of the sources show some variability, and 1-2% is periodic. These percentages roughly drops by half for "Both r and g" case. The only exception is that only 0.2% sources are periodic in both r and g band, but such lack of r and g common periodic sources could partially due to nearly half (15 out of 32) r band periodic source are never observed in g band. Besides, for those exclusively observed in g band, there seems to be an unexpected high number of variable sources, which might be worth investigating.

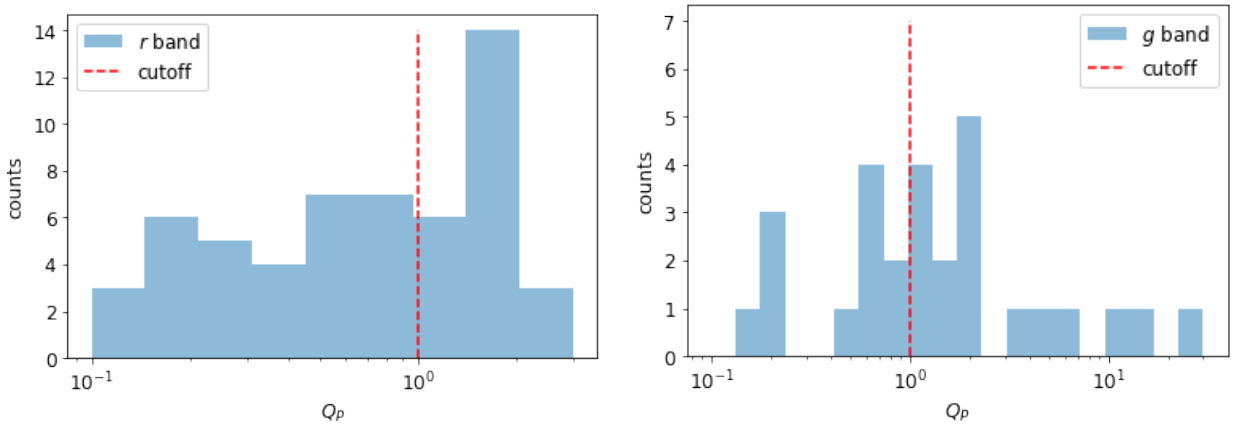

Figure 4. Distribution of Q_P for low FAP sources in ZTF r and g band. We keep sources right to the red dotted line, remove the ones on the left.

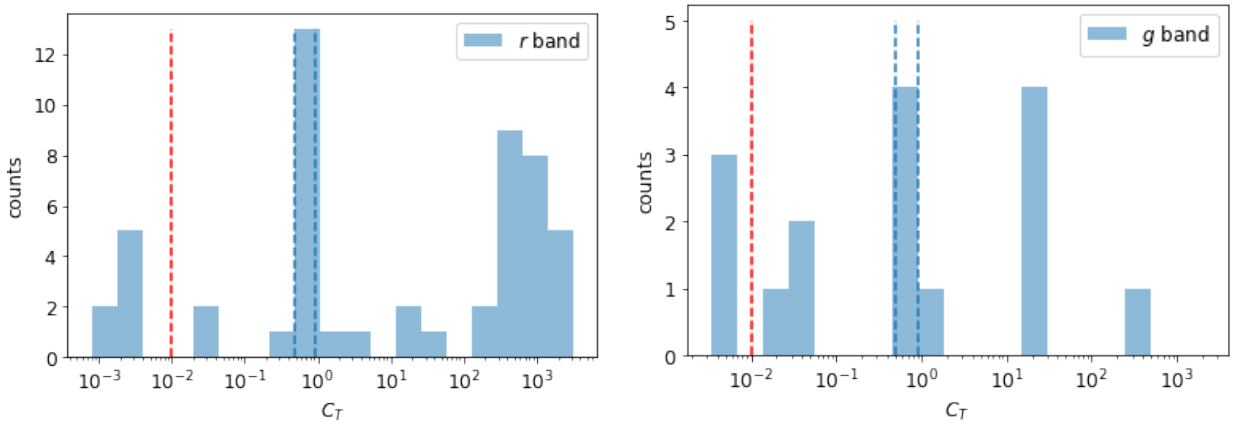

Figure 5. Distribution of C_T for low FAP, high Q_P sources in ZTF r and g band. We cut sources left to the red dotted line, and sources in between blue dotted lines, with exception of extremely low FAP.

	r band	g band	Both r and g	Only detected in r	Only detected in g
All Sources	1516 (100%)	864 (100%)	821 (100%)	695 (100%)	43 (100%)
Variables	187 (12.3%)	97 (11.2%)	42 (5.1%)	84 (12.1%)	18 (41.9%)
Periodic Variables	32 (2.1%)	9 (1.0%)	2 (0.2%)	15 (2.2%)	0 (0%)
Non-Periodic Variables	155 (10.2%)	88 (10.2%)	35 (4.3%)	69 (9.9%)	18 (41.9%)

Table 3. General statistics of the catalog. Column name "r/g band" means the sources are observed in r/g band, and is variable/periodic/non-periodic in r/g band; "Both r and g" means the sources are observed in both r and g band, and is variable/periodic/non-periodic in both r and g band; "Only detected in r/g" means the sources are only observed in r/g band, never seen in the other(g/r) band, and is variable/periodic/non-periodic in r/g band. The percentages mark the percentage inside each column.

5.2. *Distribution of magnitudes*

Periodic variables and non-periodic variables can be variable in very different ways: periodic sines, periodic eclipsing dips, sudden flares, step-wise change of magnitudes... Although looking at the lightcurves directly is probably the most intuitive and accurate way to figure out where the variability comes from, it would be unrealistic to manually check all the lightcurves. Instead, the shape of the distribution of magnitudes for each lightcurve would imply the actual shape of the lightcurve. Ideally, when the magnitudes are evenly sampled from every phase of a lightcurve, we would expect for a non-varying star, the distribution of magni-

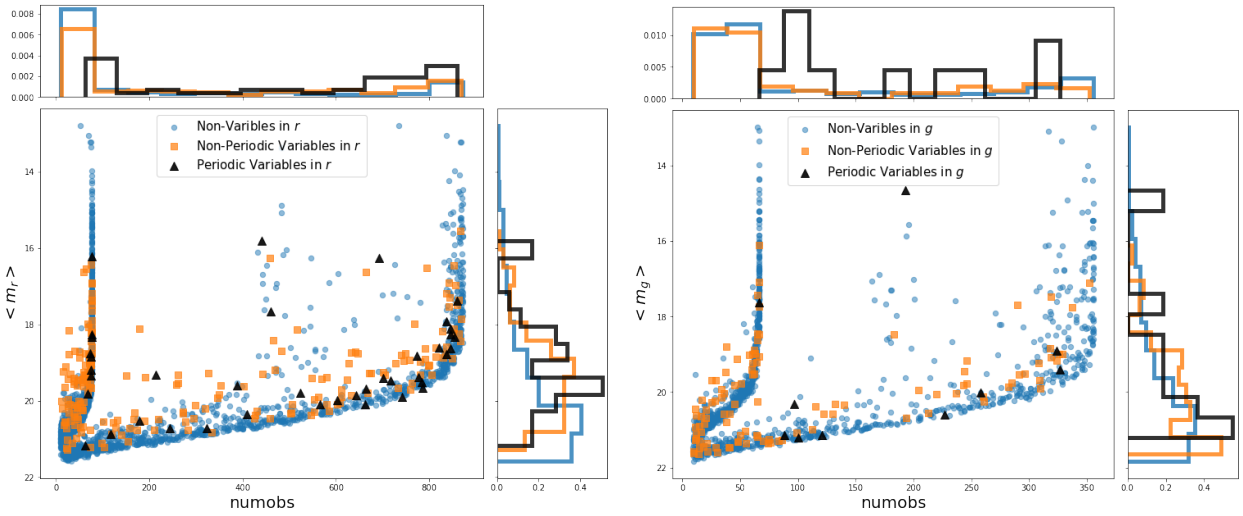

Figure 6. Number of detections verse the mean magnitude of each lightcurve labeled in ZTF OIDs. Top and right panels show the normalized distribution of x and y axis quantities respectively. Objects with lower magnitude has higher numbers of observations simply because brighter sources are more easily detected. However, there are two separate lines of such trend because one real source usually have multiple (two in most cases) ZTF OIDs.

tudes to be a Gaussian with standard deviation equal to the error in magnitude measurements, for a sine-like periodic variable, the distribution of magnitudes would be close to a plateau plus Gaussian noise, for eclipsing periodic variable, the distribution magnitudes would be close to a tilted Gaussian. Practically, we looked at typical histogram of magnitudes and found that most of them falls into either a one Gaussian, two Gaussian (close to the plateau + Gaussian noise case), and Gumbel distribution (close to tilted Gaussian). By looking at which of the three distributions fits the distribution of magnitudes relatively better, we can have some idea about what the actually lightcurve looks like. Note that the shape of magnitudes distribution is a nice indicator of the lightcurve's shape, but having certain shape of magnitudes distribution does not decisively guarantee a particular shape of the lightcurve. Magnitudes distribution heavily depends on how the magnitude measurements are distributed among different phases of the lightcurve, if all phases are completely covered by the measurements, and how the magnitudes are binned when plotting the histogram (here for all sources their magnitudes are placed in 10 evenly spaced bins).

Distribution of magnitudes are fitted into one Gaussian, two Gaussian, and and Gumbel distribution, and the three models are compared using chi-square error and F-test. One Gaussian and Gumbel distribution both have three fitting parameters, and they are compared with each other by chi-square error, $\chi^2_k = \sum_i \frac{M_{measured}-M_{model}}{k}$, where k is the degree of freedom. Two Gaussian, having 6 fitting parameters are treated with extra care via F-test. P-value is calculated to quantify how much two Gaussian is statistically better than one Gaussian or Gumbel distribution. Only if for a magnitude distribution, two Gaussian is statistically 90% better than the chi-square error winner between one Gaussian and Gumbel, the magnitude distribution is tagged by `2 Gaussian` shape, otherwise it is tagged by `1 Gaussian` or `Gumbel` depends one which has a smaller chi-square error. Besides, we also ask the location of `2 Gaussian`'s two peaks to be more separated than the sum of two standard deviations, and the standard deviations themselves can not be smaller than the mean error of magnitude measurements, so that the `2 Gaussian` has two real separated peaks, instead of being a pseudo one Gaussian.

The shape of magnitude distribution is included in the catalog, and Table 4 demonstrates how shape of magnitude distribution varies depending on if the source is variable or periodic. Among non-variable sources, `1 Gaussian` dominates the population as expected, since that is how non-variable sources defined. It seems suspicious that there are some non-variable sources that appears to be `Gumbel` and `2 Gaussian`. From Figure 7 we can see that most of these suspicious sources tends to have low number of observations such that the distribution of magnitude has less statistical importance. They are also fainter, therefore their error in magnitude could be larger, resulting a high $S_k(\chi^2)$, thus classified as non-variable.

For periodic variables, however, we can see a clear increase of `2 Gaussian` shapes in r band, since this is the expected shape for typical sine-like periodic lightcurves. Such increase is not as obvious in g band, probably due

to the extremely small sample size (only 9). Looking across different bands, the shape percentages are generally similar for both r and g band under a same category, as expected.

The shape information can hint us that `2 Gaussian` periodic variables are most likely to be sinusoidal periodic sources, while `Gumbel` and `1 Gaussian` periodic variable are more likely to be periodic eclipsing sources, or sinusoidal periodic sources with a gentler change in magnitude. `2 Gaussian` non-periodic variables could suggest that the sources are changing their magnitude for a significant amount of duration, while for `Gumbel` and `1 Gaussian` non-periodic variables, such change in magnitude might last for a shorter and shorter time.

5.3. *Color-magnitude Diagrams*

Figure 8 shows the magnitude and color distribution across sources of different variability. For reference, the distance modulus (m-M) of galaxy IC 10 is generally agreed to be around 24 (Sakai et al. 1999; Ovcharo & Nedialkov 2005; Kim et al. 2009). All these observed sources are very bright, the bottom bars in both bands are most likely to be giants, and the upper ones are more likely to be supergiants. Variable sources are well separated in color and brightness, leaving us a complete and diverse variable star sample to study. With the help of top and right panels in Figure 8, we do observe different distribution of color and magnitude for non-variables, non-periodic variables and periodic variables. This is expected since these three groups of sources should have different physics that leads to different features in their color and magnitude. In terms of magnitude, variable sources tends to be a bit brighter than non-variable ones, since brighter objects gets clearer observations thus more easily to be identified as variable. In terms of colors, non-periodic variables tends to be bluer (smaller g-r), and periodic variables are redder.

Table 5 gives a more detailed view on how much the color and magnitude is changing for these sources. Variables generally have greater changes in magnitude and color than non-variables do, as expected. To have a sense of the actual change of brightness of these stars with typical values of $< \Delta_{max} m >$ and $< \sigma_m >$, 1 magnitude smaller is 2.5 times brighter, and a 0.2 magnitude change is roughly 20% change in brightness.

6. CLASSIFICATION VERIFICATION

6.1. *Chen et al.*

As mentioned in the introduction, Chen et al. (2020)'s catalog of periodic variable stars contains one source within the field of IC 10. This particular periodic variable source is re-discovered in this work. The lightcurve is shown in Figure 9 It is identified as periodic variable in r band, while in g band it does not pass our survival function test. If we only use FAP to determine periodicity, as Chen et al. (2020) does, both r and g band lightcurves will survive our FAP cut. In addition, both r and g band lightcurves has `2 Gaussian` shape for magnitude distribution, will is the typical shape for periodic variables.

6.1.1. *UPSILoN, a machine learning classification*

To cross check the effectiveness of our methodology, we use a different approach to look for periodic variables. UPSILoN (AUtomated Classification of Periodic Variable Stars using MachIne LearNing) is a machine learning driven algorithm to identify and classify periodic variables in any optical survey developed by Kim & Bailer-Jones (2015). UPSILoN extracts key features of a lightcurve and then machine learning deploys these key feature to classify the lightcurve into classes including non-variable, δ Scuti, RR Lyrae, Cepheid, Type II Cepheid, eclipsing binary, long-period variable, as well as their subclasses. We run UPSILoN on all sources within IC 10 field. The detailed subclasses output of UPSILoN is included in our catalog.

For the purpose of cross-checking, we treat any source that is not classified as "non-variable" by UPSILoN, and is not flagged "suspicious" by UPSILoN as a periodic variable identified by UPSILoN. A overview of the result can be found in Table 6. Both this work and UPSILoN found more periodic stars in r band than in g band, and both found very few sources that are periodic in both bands, again since r band has more and brighter sources. A reasonable amount of sources are identified periodic under both methods, especially the two sources that are periodic in both r and g band. Figure 10 is a typical example when a source is identified as a periodic variable by UPSILoN but is not identified as periodic variable by the criteria outlines in this paper. This particular source does look promising on the folded lightcurve, however, it does not pass the survival function $S_k(\chi^2)$ cut. As shown in the lightcurve, the error in magnitude is comparable to the change in magnitude, thus a non-varying magnitude model can not be fully ruled out. The magnitude distribution shape for this source is `1 Gaussian`, which is in favor of a non-variable classification. On the other hand, Figure 11 is a typical example where we find a source as periodic variable but UPSILoN missed that source. This source definitely looks variable from its lightcurve, and shows some periodicity at around 238 days from its folded lightcurve. This source also has a `2 Gaussian` shaped magnitude distribution, which supports the periodic variable claim. We do not know the

	Total		1 Gaussian		Gumbel		2 Gaussian	
	r	g	r	g	r	g	r	g
Non-Variables	1329 (100%)	767 (100%)	974 (73.3%)	545 (71.1%)	303 (22.8%)	176 (23.0%)	52 (3.9%)	46 (6.0%)
Periodic Variables	32 (100%)	9 (100%)	14 (43.8%)	8 (88.9%)	4 (12.5%)	0 (0%)	14 (43.8%)	1 (11.1%)
Non-Periodic Variables	155 (100%)	88 (100%)	104 (67.1%)	50 (56.8%)	25 (16.0%)	24 (27.3%)	26 (16.8%)	14 (15.9%)

Table 4. The shape statistics of magnitude distributions. The percentages show the percentage in each row, with r and g band separated.

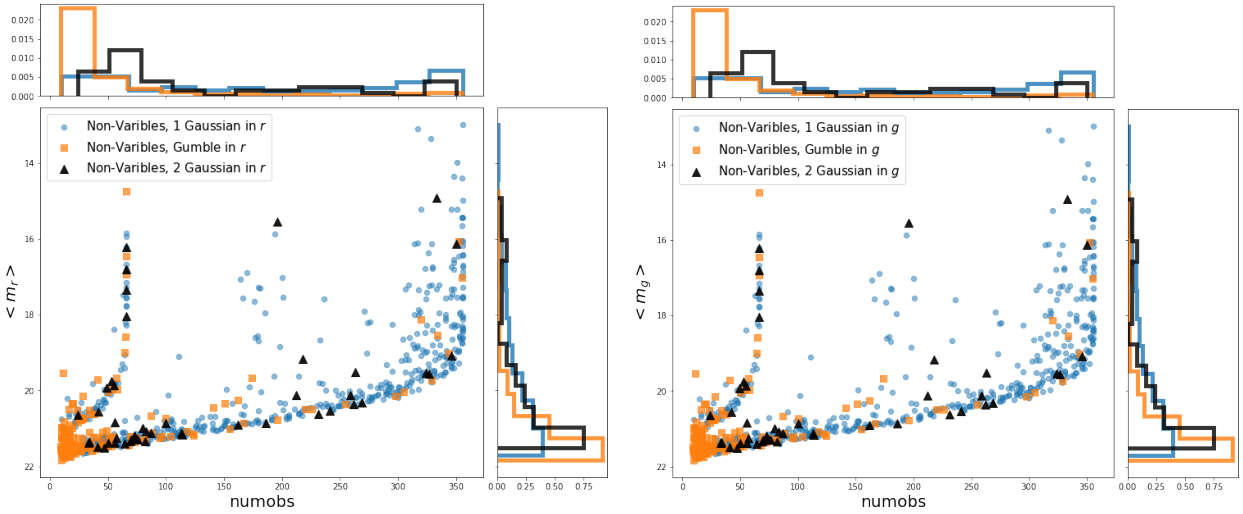

Figure 7. Most non-variable that has `Gumbel` and `2 Gaussian` shape magnitude distribution has low number of observations and is faint.

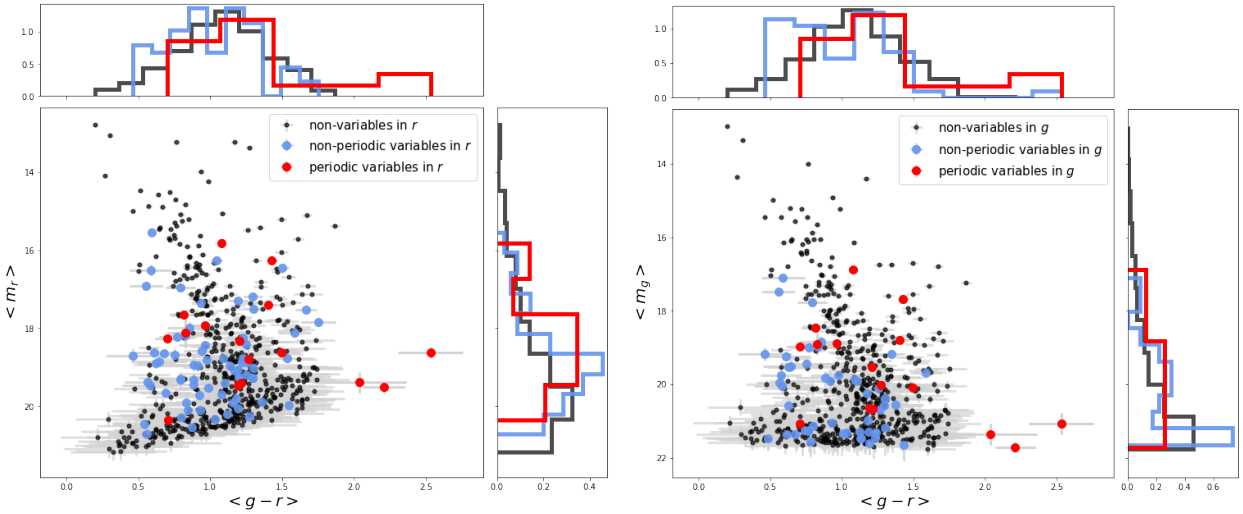

Figure 8. The color magnitude diagram. On horizontal axis is the mean g-r color, and on vertical axis is the mean r or g magnitude. Grey lines denotes 1σ in color and magnitude. Top and right panel shows the normalized distribution of x and y axis quantities respectively. All sources included in this plot are observed in both r and g band to have a meaningful g-r color.

	$< \Delta_{max} m >$	$< \sigma_m >$	$< \Delta_{max}(g-r) >$	$< \sigma_{(g-r)} >$
Non-variables in r	0.667	0.095	0.859	0.163
Non-Periodic variables in r	0.998	0.134	0.986	0.179
Periodic Variables in r	0.786	0.136	0.840	0.140
Non-variables in g	0.897	0.156	0.856	0.161
Non-Periodic variables in g	1.332	0.218	1.066	0.203
Periodic Variables in g	1.465	0.230	1.395	0.204

Table 5. A closer look at the change in magnitude and g-r color. The Columns from left to right in order gives the average maximum change in magnitude (in r or g band, depending on the row located), the average standard deviation in magnitude (in r or g band, depending on the row located), the average maximum change in color, and the average standard deviation in color.

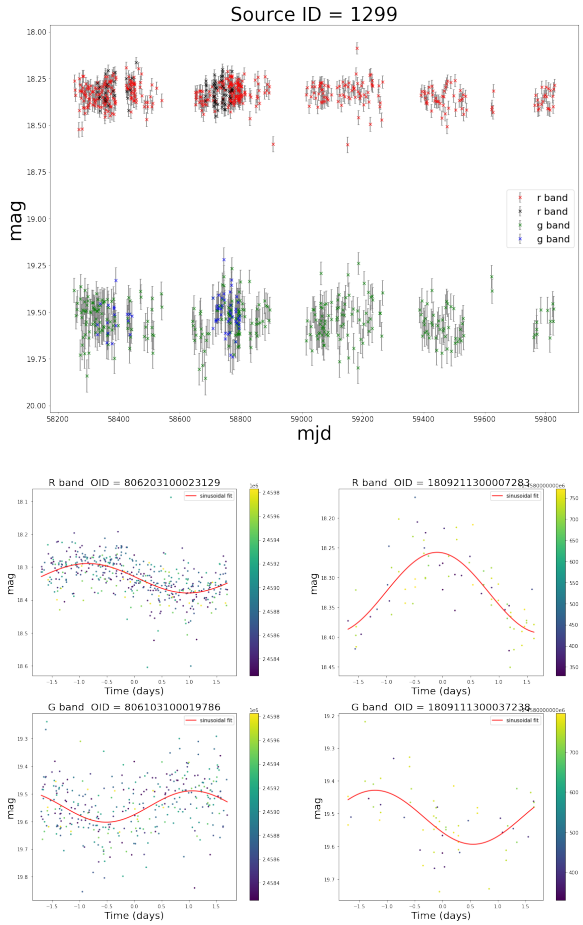

Figure 9. The periodic variable source found by Chen et al. in IC 10. This source is identified as variable in r by our algorithm, but did not pass the survival function test in g band due the relatively large error in g band magnitude measurements.

	This Work	UPSILoN	Overlaps
Periodic in r	32	35	23
Periodic in g	9	15	3
Periodic in both r and g	2	4	2

Table 6. Periodic sources found by machine-learning-driven UPSILoN compared to periodic sources found by the methodology outlined in this paper. An overlap is when both UPSILoN and our catalog mark the same source as periodic.

exact reason why UPSILoN rejected this source, but since the ZTF data used here and the data where UPSILoN is trained on inevitably have different characteristics such as band, observing window, instrument, and so on, it is not surprising that UPSILoN will not be always correct.

7. MATCHES WITH OTHER CATALOG

7.1. *Compare to SIMBAD database*

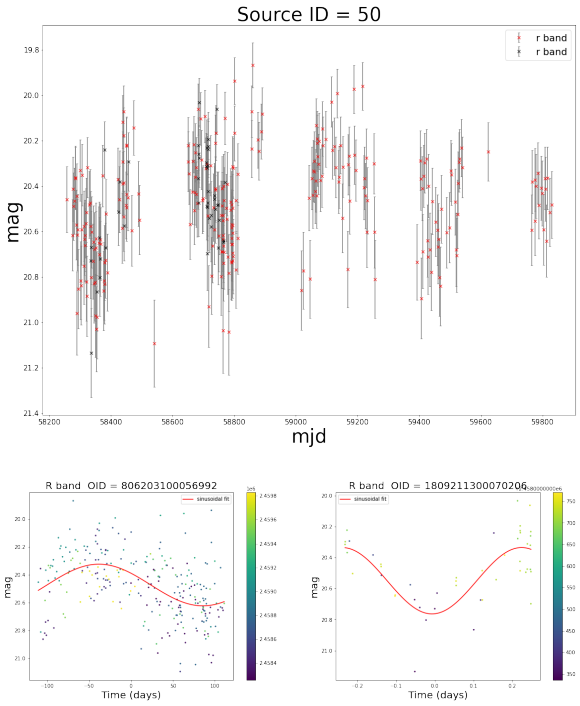

Figure 10. An example source that identified as periodic in r band by UPSILoN but not identified as periodic in r band by our catalog. Top panel shows the lightcurve with error in magnitude as error bar. Red and black represents two ZTF labeled OIDs, but the two OIDs essentially corresponds to the same source. Bottom panel are the folded lightcurve of the two OIDs, fitted with a sine function, and color-coded by mjd. This source is not identified as periodic variable by our catalog because it did not pass the survival function $S_k(\chi^2)$ cut, that is, the lightcurve is not variable enough to rule out a constant line model.

[Written by ChatGPT] SIMBAD, or Set of Identifications, Measurements, and Bibliography for Astronomical Data, is a comprehensive online database that provides information on celestial objects such as stars, galaxies, and planetary nebulae. Developed and maintained by the Centre de Données astronomiques de Strasbourg (CDS) in France, SIMBAD is a widely-used tool for astronomers and astrophysicists to access data related to various celestial objects (Wenger et al., 2000). The database includes physical characteristics such as positions, distances, magnitudes, and spectral types, as well as bibliographic information that includes references to scientific papers and articles (Zwicky, Herzog, & Wild, 1961). SIMBAD has been instrumental in a wide range of astronomical research, including the discovery of new objects and the characterization of known objects (Lépine et al., 2013). The database is constantly updated with new data and information, making it an essential resource for researchers working in the field of astronomy. [end of ChatGPT]

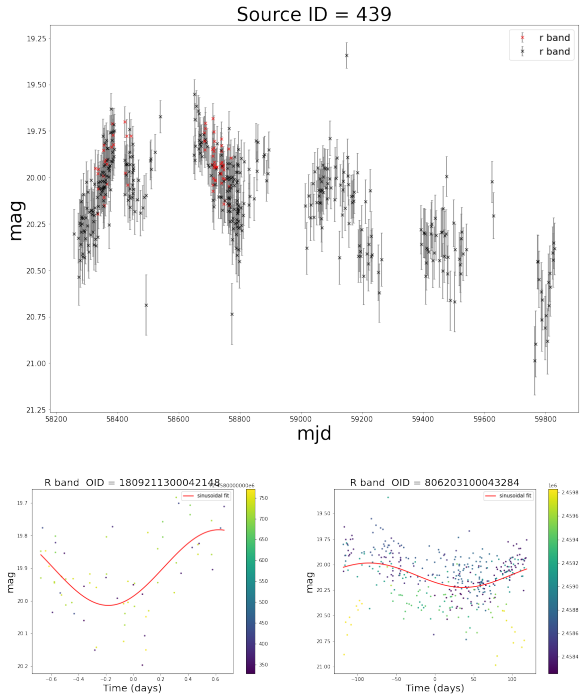

Figure 11. An example source that is identified as periodic in r band by our catalog but missed by UPSILoN. Top panel shows the lightcurve with error in magnitude as error bar. Red and black represents two ZTF labeled OIDs, but the two OIDs essentially corresponds to the same source. Bottom panel are the folded lightcurve of the two OIDs, fitted with a sine function, and color-coded by mjd.

A query of identified variable sources' locations is conducted in SIMBAD. The nearest (if there is any) object within 2 arcsec radius is treated as matched counterparts. An overview of the matched statistics is shown in Table 7,the detailed matched information including identifier name, identified type, and distance is include in our catalog. Among matched SIMBAD sources, we re-discovered many interesting labeled sources including Wolf-Rayet stars, emission-line stars, radio sources, HII regions, planetary nebula Candidates, X-ray sources, supernova remnants and carbon stars. For those that are observed long enough in both ZTF r and g band that allows us to compute the mean g-r color, their distribution on a color magnitude diagram is shown in Figure 12. Objects of the same type usually share similar physics and thus tends to cluster at certain places in the color-magnitude diagram. The object type classification of SIMBAD matched sources can hint us the classification of those unmatched sources, and guide us on where we to find interest sources on the color magnitude diagram.

7.2. *Compare to Gaia*

[Written by ChatGPT] Gaia is an ambitious mission launched by the European Space Agency (ESA) with the goal of creating a highly precise three-dimensional map of our Milky Way galaxy. It is named after the ancient Greek goddess Gaia, who personified the Earth, and its mission is to observe the positions, distances, motions, and other physical characteristics of approximately one billion stars in the Milky Way (Prusti et al., 2016). By doing so, Gaia aims to provide new insights into the structure, formation, and evolution of our galaxy, as well as to uncover the mysteries of dark matter and dark energy. The data gathered by Gaia is freely available to the scientific community and has already led to numerous groundbreaking discoveries in the field of astronomy (Brown et al., 2018). [end of ChatGPT]

A cross match of sources in our catalog and Gaia sources within 2″ radius distance is made, and a summary is shown in Table 8. Thanks to Gaia's higher angular resolution, this cross-matching to Gaia will tell us how many resolved real sources are under our same source label and avoid possible contamination. For ZTF sources with more than one Gaia matches within 2″, if the brightest (in terms of Gaia `G` band magnitude) Gaia match is not 100 times brighter (2.5 magnitudes lower) than the second brightest Gaia match, we place `multiflag=1` in the catalog to indicate possible contamination from multiple unresolved sources.

Gaia also measures the parallax, and parallax is a good indicator of how far the source is away from us. Therefore, we can distinguish nearby sources inside Milky way and the further sources within IC 10 galaxy. IC 10 is outside of Milky Way, and is around 950 kpc away (Massey & Armandroff 1995). Parallax measurements are highly impossible for such distances. Therefore we place a `Plxflag=1` for sources with a meaningful parallax (Plx>0, and Plx>e_Plx) to to indicate that these sources are likely not inside IC 10.

	This work	Simbad matches
Periodic variables in r	32	14
Periodic variables in r	9	3
Non-Periodic variables in r	155	43
Non-Periodic variables in g	88	26

Table 7. Number of SIMBAD matches within 2 arcsec radius.

	This work	Gaia matches
Periodic variables in r	32	28
Periodic variables in g	9	9
Non-Periodic variables in r	155	100
Non-Periodic variables in g	88	67

Table 8. Number of Gaia matches within 2 arcsec radius.

8. CATALOG COLUMNS

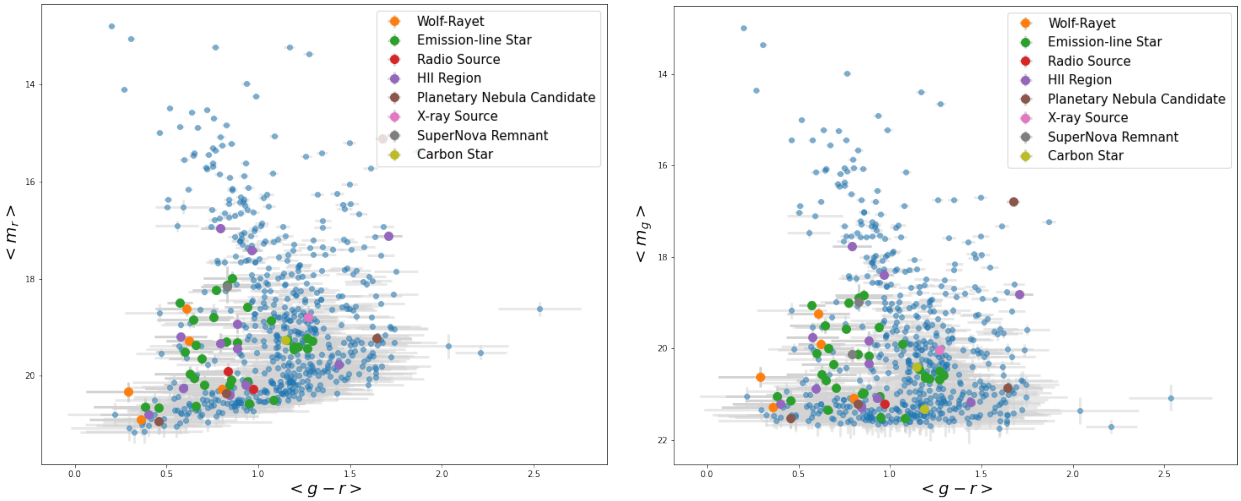

Figure 12. Sources labeled by SIMBAD on the color magnitude diagram. Blue dots are all the sources included in this catalog, colored dots are SIMBAD matches of different object types, and Grey lines denotes 1σ in color and magnitude.

All columns of our catalog are presented in Table 9. The catalog itself can be downloaded at [some place]. [The catalog itself, alone with the code to generate the catalog and all the plots in this paper, is located at [my github repo]].

9. CONCLUSION

ACKNOWLEDGMENTS

Based on observations obtained with the Samuel Oschin 48-inch Telescope at the Palomar Observatory as part of the Zwicky Transient Facility project. ZTF is supported by the National Science Foundation under Grant No. AST-1440341 and a collaboration including Caltech, IPAC, the Weizmann Institute for Science, the Oskar Klein Center at Stockholm University, the University of Maryland, the University of Washington, Deutsches Elektronen-Synchrotron and Humboldt University, Los Alamos National Laboratories, the TANGO Consortium of Taiwan, the University of Wisconsin at Milwaukee, and Lawrence Berkeley National Laboratories. Operations are conducted by COO, IPAC, and UW.

REFERENCES

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068

Baluev, R. V. 2008, MNRAS, 385, 1279, doi: 10.1111/j.1365-2966.2008.12689.x

Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002, doi: 10.1088/1538-3873/aaecbe

Chen, X., Wang, S., Deng, L., et al. 2020, The Astrophysical Journal Supplement Series, 249, 18, doi: 10.3847/1538-4365/ab9cae

Cheung, S.-H., Villar, V. A., Chan, H.-S., & Ho, S. 2021, A New Classification Model for the ZTF Catalog of Periodic Variable Stars, arXiv, doi: 10.48550/ARXIV.2112.04010. https://arxiv.org/abs/2112.04010

Cotton, W. D., Condon, J. J., & Arbizzani, E. 1999, ApJS, 125, 409, doi: 10.1086/313286

Huchra, J. P., Vogeley, M. S., & Geller, M. J. 1999, The Astrophysical Journal Supplement Series, 121, 287, doi: 10.1086/313194

Kim, D.-W., & Bailer-Jones, C. A. L. 2015, UPSILoN: AUtomated Classification of Periodic Variable Stars using MachIne LearNing, Astrophysics Source Code Library, record ascl:1512.019. http://ascl.net/1512.019

Kim, M., Kim, E., Hwang, N., et al. 2009, The Astrophysical Journal, 703, 816, doi: 10.1088/0004-637x/703/1/816

Lomb, N. R. 1976, Ap&SS, 39, 447, doi: 10.1007/BF00648343

SourceID	Source ID for this lightcurve/OID. OIDs with less than $1''$ have the same SourceID.
filter	ZTF filter, either zr for r band or zg for g band.
OID	ZTF Object ID. Each OID has a lightcurve and makes up a row in this paper's catalog.
RA	Right ascension, from ZTF.
DEC	Declination, from ZTF.
numobs	Number of observations by ZTF, or number of points of this lightcurve.
sf	$S_k(\chi^2)$, survival function of the constant magnitude model, to classify variable vs. non-variable.
FAP	False alarm probability from Lomb-Scargle, to classify periodic variable vs. non-periodic variable.
period	$T_{\rm LS}$, period at Lomb-Scargle periodogram peak power. Period of the source if the source is periodic.
period_w	$T_{\rm window}$, period at window function peak power. $C_T = \lvert T_{\rm LS} - T_{\rm window}\rvert / T_{\rm window}$ for de-aliasing.
power	$P_{\rm LS}$, Lomb-Scargle periodogram peak power.
power_w	$P_{\rm window}$, window function power at $T_{\rm LS}$. $Q_P = P_{\rm LS}/P_{\rm window}$ for de-aliasing.
mean	Mean magnitude.
std	Standard deviation of magnitudes.
min	Minimum magnitude.
max	Maximum magnitude.
gaussian1_err	Chi-square error of 1 Gaussian model for distribution of magnitudes.
gaussian2_err	Chi-square error of 2 Gaussian model for distribution of magnitudes.
gumbel_err	Chi-square error of Gumbel model for distribution of magnitudes.
g2_flag	If 2 Gaussian model has two real separated peaks, instead of being a pseudo one Gaussian.
Pg1g2	P-value for 2 Gaussian model over 1 Gaussian model.
Pgug2	P-value for 2 Gaussian model over Gumbel model.
shape	Shape of magnitude distribution. "1 Gaussian/ 2 Gaussian/ Gumbel".
upsilon_class	Classification label from UPSILoN.
upsilon_prob	Classification probability from UPSILoN.
upsilon_flag	Suspicious classification flag by UPSILoN.
upsilon_period	Period found by UPSILoN.
identifier	SIMBAD database identifier of the nearest (if there is any) match within $2''$.
dist	Angular distance between ZTF location and SIMBAD location of the nearest match.
type	SIMBAD object type of the nearest match.
Gaia_count	Number of Gaia objects within $2''$ radius.
multiflag	If there is possible contamination from multiple unresolved sources, see Section 7.2.
GaiaDist	Angular distance between ZTF location and Gaia location of the brightest match.
fluxratio	Gaia G band flux of brightest Gaia match over that of second brightest Gaia match.
Gmag	Gaia G band magnitude.
Plx	Parallax of the brightest Gaia match.
e_Plx	Error in parallax of the brightest Gaia match.
Plxflag	If the source is outside IC 10.
lc_var	Classification of this lightcurve/OID. P = periodic variable, V = non-periodic variable, N = non-variable.
var	Classification of this source, P/V/N.
var_flag	Variable suspect flag, see Section 4.2.4.

Table 9. Columns of the catalog.

Masci, F. J., Laher, R. R., Rusholme, B., et al. 2018, 131, 018003, doi: 10.1088/1538-3873/aae8ac

Massey, P., & Armandroff, T. E. 1995, AJ, 109, 2470, doi: 10.1086/117465

Ovcharo, E., & Nedialkov, P. 2005, Aerospace Research in Bulgaria, 20, 85

Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Sakai, S., Madore, B. F., & Freedman, W. L. 1999, The Astrophysical Journal, 511, 671, doi: 10.1086/306716

Scargle, J. D. 1982, ApJ, 263, 835, doi: 10.1086/160554