

Full length article

Removing aliases in time-series photometry

D. Kramer^{a,b,*}, M. Gowanlock^{a,b}, D. Trilling^{b,a}, A. McNeill^{b,c}, N. Erasmus^d^a School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ 86011, USA^b Department of Astronomy and Planetary Science, Northern Arizona University, Flagstaff, AZ 86011, USA^c Department of Physics, Lehigh University, 16 Memorial Drive East, Bethlehem, PA 18015, USA^d South African Astronomical Observatory, Cape Town, 7925, South Africa

ARTICLE INFO

Article history:

Received 30 October 2022

Accepted 24 April 2023

Available online 2 May 2023

Keywords:

Time series

Lomb–Scargle

Light curves

Aliasing

LSST

ZTF

ABSTRACT

Ground-based, all-sky astronomical surveys are imposed with an inevitable day–night cadence that can introduce aliases in period-finding methods. We examined four different methods – three from the literature and a new one that we developed – that remove aliases to improve the accuracy of period-finding algorithms. We investigate the effectiveness of these methods in decreasing the fraction of aliased period solutions by applying them to the ZTF and the SSPDB asteroid datasets. We find that the VanderPlas method had the worst accuracy for each survey. The mask and our newly proposed window method yields the highest accuracy when averaged across both datasets. However, the Monte Carlo method had the highest accuracy for the ZTF dataset, while for SSPDB, it had lower accuracy than the baseline where none of these methods are applied. Where possible, detailed de-aliasing studies should be carried out for every survey with a unique cadence.

© 2023 Elsevier B.V. All rights reserved.

1. Introduction

Determining periodic behavior among astrophysical sources is useful for describing their physical properties. For example, the internal strength of an asteroid can be determined using, among other observed properties, its rotation period (McNeill et al., 2018) and light curves have been used to categorize different stellar types (Barbara et al., 2022).

Ground-based telescopic surveys that produce sparse data inevitably have signals in the object's periodogram – typically at or near 12 h, 24 h, 48 h, and 96 h – related to the day–night cycle of the Earth. These cadences cause aliasing, an effect where there are peaks in a periodogram that are not at the real period of an observed object.

If the output of a period-finding method and its corresponding light curve is visually examined, an astronomer can potentially make a judgment if a derived period is an alias or not. With large-scale surveys, like the Legacy Survey of Space and Time (LSST), too many objects will be observed for humans to manually confirm each derived period. This requires automating both deriving the periods and determining if the period is correct.

The problem of period-finding at scale will become more acute when LSST is producing data, as more than 100 million periodic sources are expected in the LSST catalog (Abell et al., 2009). If,

for example, 20% of period solutions are aliases, then 20 million sources would have incorrectly derived periods. The incorrect periods would either (1) be naively included in a catalog, (2) have solutions at common alias periods discarded, or (3) have to be examined with some verification algorithm. Any method that reduces the total number of alias results improves the three outcomes by (1) reducing the number of bad solutions, improving the overall accuracy, (2) increasing the overall number of correct solutions, or (3) decreasing the amount of computation to verify/score the periods.

While this is an important problem for ongoing and upcoming surveys, there has been little progress on de-aliasing methods for sparse data. For instance, Carbonell et al. (1992) have examined the de-aliasing properties of past methods, notably CLEAN (Roberts et al., 1987). However, those methods only work on time series data with uniform observations during the night. Since modern ground-based surveys generate sparse data that is not uniform, these methods are unsuitable for these surveys. Recently, several papers have used some de-aliasing techniques using large scale survey data (Erasmus et al., 2021; Coughlin et al., 2021; Heinze et al., 2018), but none tested the improvement their method had over no de-aliasing.

In this paper, we analyze the performance of four approaches to de-aliasing, with performance and period-finding accuracy in mind.

The paper is organized as follows: Section 2 discusses the Lomb–Scargle periodogram, Section 3 gives an overview of the datasets used, Section 4 gives an overview of the four methods

* Corresponding author at: School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ 86011, USA.

E-mail address: drk98@nau.edu (D. Kramer).

that were tested, Section 5 discusses the results of the four methods, Section 6 provides our discussion of the results of the paper, and Section 7 presents our conclusions and some future work on this and related problems.

2. Period finding algorithms

For testing the accuracy of the different de-aliasing methods, a period finding algorithm is needed. In this section, we will discuss four different methods, Lomb–Scargle (L-S, [Lomb 1976](#), [Scargle 1982](#)), SuperSmoother ([Friedman, 1984](#)), Conditional Entropy ([Graham et al., 2013](#)), and Bootstrap χ^2 ([Durech et al., 2022](#)), and why L-S is used as the method for this analysis.

2.1. Lomb–Scargle

L-S is a period-finding method first developed by [Lomb \(1976\)](#) and later improved by [Scargle \(1982\)](#). It is one of the most popular period-finding algorithms used in astronomy. The general approach is to calculate the periodogram power, essentially a measure of the goodness of the solution, through Eq. (1) as follows.

$$P(f) = \frac{1}{2} \left\{ \frac{\left(\sum_n g_n \cos(2\pi f[t_n - \tau]) \right)^2}{\sum_n \cos^2(2\pi f[t_n - \tau])} + \frac{\left(\sum_n g_n \sin(2\pi f[t_n - \tau]) \right)^2}{\sum_n \sin^2(2\pi f[t_n - \tau])} \right\} \quad (1)$$

$$\tau = \frac{1}{4\pi f} \tan^{-1} \left(\frac{\sum_n \sin(4\pi f t_n)}{\sum_n \cos(4\pi f t_n)} \right)$$

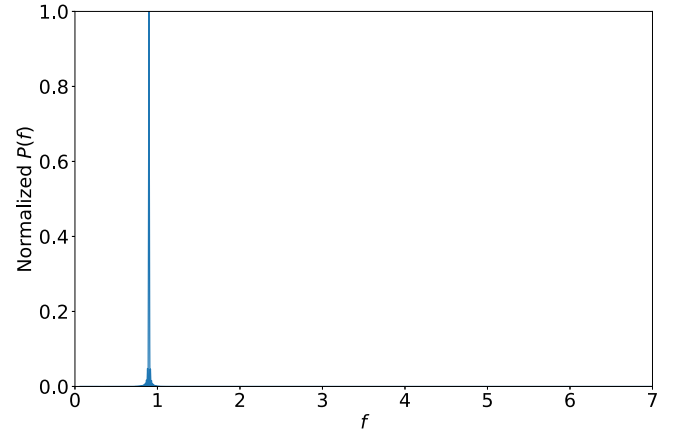
Here, $P(f)$ is the power for an angular frequency f , g_n is the observed telescopic magnitude of observation n and t_n is the time of observation n . Higher powers indicate a greater likelihood that f is the angular frequency of the observed object. Note that the LSP has a time complexity of $O(nm)$, where n is the number of observations used and m is the number of frequencies examined.

A periodogram is typically computed by calculating powers for a range of frequencies (or periods) in $[f_{\min}, f_{\max}]$ that are sampled over a uniform frequency space. Ideally, there would only be a peak in the periodogram at the angular frequency f that corresponds to the object's physical rotation state, with all other values of $P(f)$ having a value of zero. In reality, because of uncertainties in the measurements, a non-uniform cadence, and aliasing, the periodogram is often noisy with multiple peaks, so determining the correct peak is not always straightforward. [Fig. 1](#) shows how a periodogram generated with a uniformly sampled sine wave with a small Δt compares to a randomly sampled sine wave's periodogram. The randomly sampled sine wave's periodogram is noisy around $P(f) = 0$ while the uniformly sampled sine wave only has a peak at the sine curve's frequency.

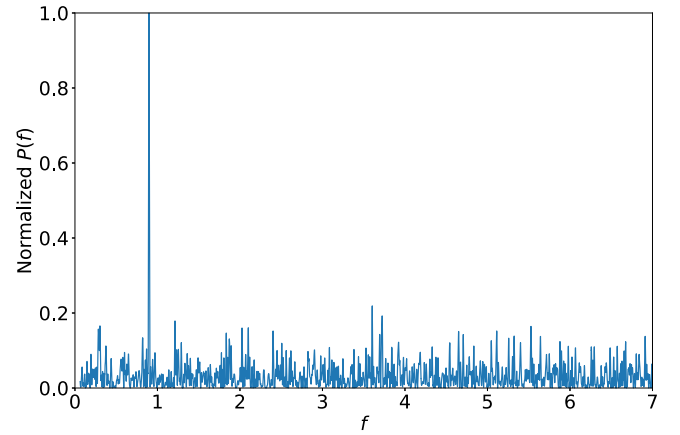
2.2. Other algorithms

Other period finding algorithms exist, like SuperSmoother ([Friedman, 1984](#)), Conditional Entropy ([Graham et al., 2013](#)), Bootstrap χ^2 ([Durech et al., 2022](#)), among others. With all of these algorithms, there is a trade-off between accuracy, speed, space, and light-curve shape flexibility. We summarize the three algorithms above as follows:

- *SuperSmoother* is especially useful for periodic signals that are not sinusoidal ([Becker et al., 2011](#); [Huber et al., 2005](#)), has a time complexity and space complexity of $O(nm)$, and it is susceptible to aliasing ([Gowanlock et al., 2022](#)).



(a) Uniformly sampled sine wave.



(b) Randomly sampled sine wave.

Fig. 1. The periodograms for a uniformly sampled sine wave and a random sample of it. The x-axis is the angular frequency and the y-axis is the normalized power. The sine wave was given an angular frequency of $2\pi/7$ (period of 7) and an amplitude of 0.3. The pure sine wave used 1000 points and the sampled sine wave used 0.5% of those points, for about 50 total points used.

- *Conditional Entropy* also has time complexity of $O(nm)$, is less accurate than L-S for fast periods, and is susceptible to aliasing ([Coughlin et al., 2021](#)).
- *Bootstrap χ^2* is the slowest, with a time complexity of $O(snm)$, where s is the number of bootstrap samples, but it is the most accurate, with aliasing having the smallest effect on this method compared to the others described above ([Durech et al., 2022](#)).

L-S has the same time complexity as SuperSmoother and Conditional Entropy, all three being smaller than Bootstrap χ^2 . As SuperSmoother has a space complexity of $O(nm)$ while L-S and Conditional Entropy have a space complexity of $O(n+m) \approx O(m)$, SuperSmoother was not used. Finally, since L-S and Conditional Entropy are similar, and L-S is the standard algorithm used for deriving the periods in astronomy, we elect to use L-S in this paper. However, any period finding algorithm that produces a periodogram is capable of implementing the methods described in Section 4.

3. Data used

The upcoming LSST – to be carried out with the Vera C. Rubin Observatory that is presently under construction in Chile – will revolutionize many fields of astronomy (Ivezić et al., 2019). LSST will generate sparse photometry on the around 500–1000 measurements over ten years for some 40 billion astronomical sources. The vast majority of these will be sidereal objects (stationary on the sky); around 5 million of them will be moving objects. Many of these LSST-observed sources will be variable, with regular periods, so it is of interest to develop and implement accurate period-finding algorithms that can operate at the vast LSST scale.

The ongoing all-sky survey being carried out by the Zwicky Transient Facility (ZTF; Bellm et al. 2018) acts as a kind of LSST precursor. ZTF is carrying out a public survey that is very LSST-like in terms of cadence, data type, and data accessibility, but at something like one-tenth the LSST scale.

The work presented here has its origin in Solar System science and asteroid period finding but is relevant for period-searching for any kind of astrophysical source in either ZTF or LSST data. For ZTF, we used asteroid data from SNAPShot1 (Trilling et al., 2023), which used ZTF observations from 2018–07–19 to 2020–05–19, using only numbered asteroids with more than 50 observations with Real-Bogus scores ≥ 0.55 (Duez et al., 2019); the Real-Bogus cut eliminated about 9% of all observations.

There is no actual LSST data yet as science operations will commence in 2024, so we used the LSST Solar System Products Data Base (SSPDB; Juric et al. 2021), a complete simulation of asteroid observations over the 10 year nominal lifetime of LSST, as our LSST testbed. However, the objects in this synthetic database are all assumed to be spherical and thus, unlike almost all real asteroids, do not show shape-induced periodic variability in the lightcurves. We therefore assigned lightcurve amplitudes and rotation periods to each SSPDB object, as described in Appendix A.

4. Methods

In this section, we present our de-aliasing analysis on the ZTF and SSPDB asteroid datasets. We use three methods from the literature – masking; MC; and VP – and present our new approach, the window method. Broadly speaking, the MC method tries to remove aliasing through subsampling (with multiple trials), whereas the other methods attempt to remove signals at the expected alias periods.

One method that will not be tested is the False Alarm Probability (Baluev, 2008). False Alarm Probability is sometimes incorrectly used as a proxy for the goodness of a L-S solution. However, False Alarm Probability is a way of calculating the probability $p = (\text{power}|\text{noise})$: the probability that a given periodogram power is part of the noise of the periodogram (Baluev, 2008; VanderPlas, 2018). If there is an alias peak with a larger $P(f)$ than the real peak, then it would have a lower False Alarm Probability. Because False Alarm Probability is not testing for the authenticity of a signal related to the physical period in the system being monitored, it is not a useful method for determining if a peak is a real period or an alias.

4.1. Masking

Masking is the most straightforward approach presented in this paper: solutions near the known alias solutions are simply rejected. The alias periods or a small range around the alias can be removed (masked out) from the periodogram so the real period's peak would therefore have the largest remaining $P(f)$. Usually,

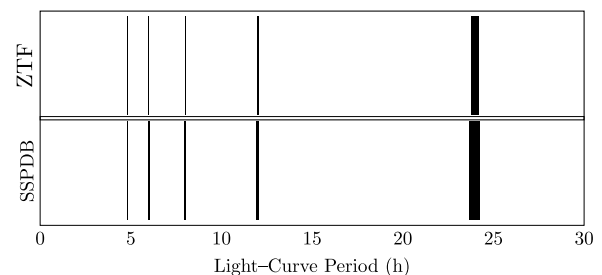


Fig. 2. Visualization of the mask ranges from Table 1. The black bars indicate the masked regions where all solutions are rejected. ZTF and SSPDB have similar, but not exactly the same, ranges.

there are multiple aliases, so several masks are needed to remove them.

This method was used in Erasmus et al. (2021), which presented asteroid photometry from Asteroid Terrestrial-impact Last Alert System and ZTF where they masked out periods of {8, 12, 16, 24, 48} h. The remaining periodogram peaks with the largest power were then found to be those representing super-slow rotation periods. Erasmus et al. (2021) showed that the masking method is a viable way of removing alias period solutions so we incorporate this method into our analysis. This method was also used in Coughlin et al. (2021), which used a similar method to derive the mask ranges as described below. The Coughlin et al. (2021) mask ranges, in rotation period space, are [(0.5, 0.5), (0.51, 0.51), (0.52, 0.52), (5.93, 6.08), (7.87, 8.14), (11.71, 12.31), (22.86, 25.26), (46.15, 50.0), (600.0, 800.0)] hours.

Table 1 and Fig. 2 show the period ranges that are masked out for each dataset. The mask ranges were derived through the following steps:

1. Generate a histogram of the derived periods.
2. Create a new mask range by selecting the bin with the highest number of objects. For example, if the bin with the most objects is [23.9 h, 24.1 h], then it would be added as a mask range.
3. Re-derive the periods with this new mask range.
4. Repeat steps 1 through 3 consecutively, adding new masks. Then select the mask ranges that provide the maximum match percentage to a database of accurate periods for the objects. For example, Fig. 3 shows percent match against the Light Curve Database (LCDB; Warner et al. 2021) as a function of the number of masks for ZTF.

The ranges for both ZTF and SSPDB are similar (Table 1 and Fig. 2), but not all surveys will have these same ranges because they will have different cadences and different observational errors. Despite this, we expect that all ground-based surveys will have the 24 h alias.

Advantage: The primary advantage of this approach is that it is the fastest to compute. The time complexity for the method is $O(nm)$, where n is the number of data points used and m is the number of frequencies examined. If the masks are pre-computed, the run time is slightly faster compared to a normal LSP because, with the same frequency range and Δf , fewer frequencies (m) would have to be examined as the frequencies in the mask ranges would be excluded.

Disadvantage: This method has the disadvantage that true periods that are within one of the masked ranges will never be identified. The LCDB contains a large number of curated asteroid rotation periods and only about 1% – 2% of objects in the LCDB have rotation periods that are in the masked ranges of Table 1, so only a low number of objects would be impacted.

Table 1

The period ranges (exclusive) for each of the datasets to be excluded from the (light-curve) period space. Note that these ranges are light curve periods, not rotation periods, where an object's light curve period is half its rotation period.

Dataset	Mask Ranges
ZTF	(4.789, 4.814), (5.989, 6.014), (7.972, 8.014), (11.947, 12.039), (23.764, 24.164)
SSPDB	(4.789, 4.839), (5.939, 6.039), (7.939, 8.039), (11.889, 12.039), (23.639, 24.239)

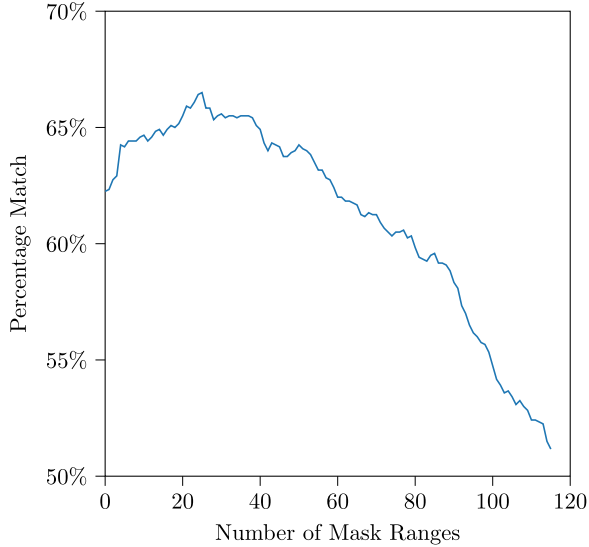


Fig. 3. The match percentage for each new mask range. The x-axis is the number of 0.05 h size mask ranges. The y-axis is the match percentage.

4.2. Monte Carlo (MC)

Since aliases are generated by the observing cadence intervals, randomly subsampling the observations, across many trials, might suppress the signal from the aliases. An overall periodogram for each asteroid can then be found either by finding the most common peak power from each trial's periodogram, or by summing over all the periodograms. Some attention must be paid to ensure that the time baseline of the subsample (T_{sub}) is greater than P_{max} , the period corresponding to the minimum frequency examined, otherwise, L-S will derive the period of a partial light-curve, causing an incorrect period to be derived. For large scale surveys, like ZTF and LSST, that produce sparse photometry over years, this is not an issue as T_{sub} will always be greater than P_{max} for any reasonable P_{max} .

For the calculations in this paper, we used 100 selection samples/trials (i) per object with 50 random observations (n). SSPDB objects had to have at least 200 observations while ZTF objects only had to have 90 observations so there would be enough objects to get a large statistical sample. These parameters were found via an ad hoc process of changing the minimum number of observations required, the number of observations per subsample, and the number of MC trials to find a high match percentage.

Finally, we sum the un-normalized periodograms over all trials to yield the overall periodogram. The highest peak from this periodogram is used as the derived period.

Advantage: This method has the advantage that, compared to the other methods, it is the most oblivious to aliases like those in Table 1. Unlike the masking method, the MC method could derive a correct rotation period that happens to be an alias.

Disadvantage: This method also requires more observations than the other methods discussed because in order to produce

a credible periodogram, the number of observations in each subsample needs to be sufficiently large while also being significantly smaller than the total number of data points. This is needed so that pairs of observations that are, for example, 24 h apart are excluded from the same subsample.

This method is also the slowest, with a time complexity of $O(nmi)$, where n is the number of data points used for each iteration, m is the number of frequencies checked, and i is the number of selection samples.

4.3. Window

The window function derives aliases caused by the observational cadence. It takes the temporal data from an object's observations to produce a periodogram-like output where a frequency having a high power corresponds to high aliasing (VanderPlas, 2018).

$$P_w(f; \{t_n\}) = \left| \sum_{n=1}^N e^{-2\pi i f t_n} \right|^2 \quad (2)$$

Eq. (2) shows the window function, but VanderPlas (2018) showed that a LSP can be used as an approximation of the window function if in Eq. (1), $g_n = 1$; this simulates a completely spherical, homogeneous object. Therefore, any signal present in the window periodogram would be aliasing caused by the underlying cadence.

Fig. 4 provides an example of how an object's LSP and window periodogram relate. There are three key observations from Fig. 4: (1) there are peaks in the LSP that do not correspond to the real period, (2) all of the significantly strong peaks that are in the window periodogram also appear in the LSP, and (3) there exist peaks in the LSP that are not the real period or in the window periodogram; these are pseudo-aliases. The example pseudo-alias marked in Fig. 4 is $((1/P) - (1/24 \text{ h}))^{-1}$ where P is the real period for the object (Durech et al., 2022).

Previous algorithms that use the window function, like deconvolution and CLEAN, do not work for removing aliases (VanderPlas, 2018) because they assume that the strongest peak in the periodogram is the peak that corresponds to the real period. Here we present a new way to use the window function to remove aliases, where the pseudocode can be found in Algorithm 1. This differs from deconvolution and CLEAN because they rely on deconvolution for their alias removal while Algorithm 1 does not.

Algorithm 1 takes an object's time of observation $T = \{t_1, t_2, \dots, t_n\}$ (time), observed magnitudes $G = \{g_1, g_2, \dots, g_n\}$ (mag) where n is the number of observations, and the frequency grid on $[f_{\text{min}}, f_{\text{max}}]$ (freqs) as arguments. The LS and Window functions on lines 2 and 3 calculate the L-S and window periodograms respectively. The findPeaks methods on lines 4 and 5 find the peaks in those periodograms and return the peak frequencies and their powers. Line 6 sorts the LSP peaks by their power in descending order. Line 9 loops until the current LSP peak is not contained in the window peaks. If the LSP peak is in the set of window peaks, the index of the current peak is incremented. Line 12 tests if all of the detected LSP peaks have been compared against the set of window peaks. If they have, then the peak with the highest power is used as the correct peak (although this is

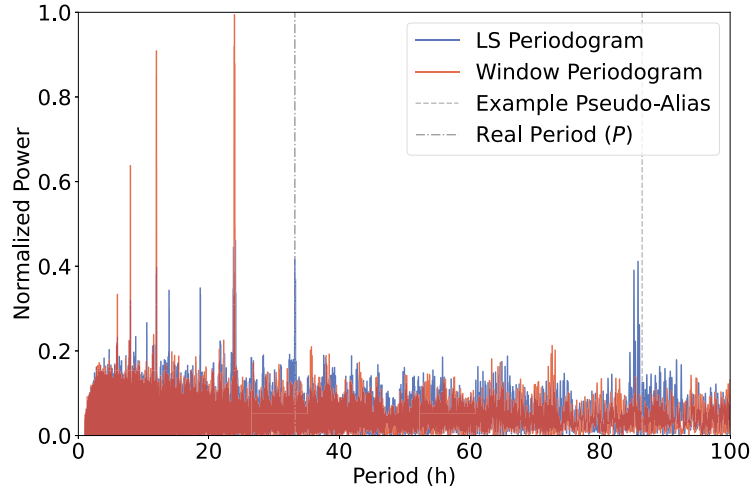


Fig. 4. The LSP and window periodogram for an object from SSPDB. The LSP is in red and the window periodogram is in blue. The given period for the object is marked by the dash-dot-dash line and its peak is only present in the LSP. The dashed line marks an example pseudo-alias which is also only present in the LSP. The peaks in the window periodogram are the aliases which are also all present in the LSP.

Algorithm 1 Our method of using the peaks in the window periodogram to check if a peak in the LSP is an alias.

```

1: procedure WINDOWMETHOD(time, mag, freqs)
2:   pgramLS  $\leftarrow$  LS(time, mag, freqs)
3:   pgramWindow  $\leftarrow$  Window(time, freqs)
4:   LSPeaks  $\leftarrow$  findPeaks(pgramLS)
5:   WindowPeaks  $\leftarrow$  findPeaks(pgramWindow)
6:   sort(LSPeaks)
7:   NPeaks  $\leftarrow$  length(LSPeaks)
8:   index  $\leftarrow$  0
9:   while index < NPeaks and LSPeaks[index]  $\notin$  Window-
     Peaks do
10:    index  $\leftarrow$  index+1
11:   end while
12:   if index  $\geq$  length(LSPeaks) then
13:     correctPeak  $\leftarrow$  LSPeaks[0]
14:   else
15:     correctPeak  $\leftarrow$  LSPeaks[index]
16:   end if
17:   return correctPeak
18: end procedure

```

probably an alias), otherwise, the peak at the current index is used as the correct peak. The correct peak is then returned. A python implementation is available as a [GitHub gist here](https://gist.github.com/drkr98/6b15633e8fc43e8daf6b628548006376).¹

Advantage: This method has a distinct advantage over the masking method. The window function exposes aliases on a per-object basis, whereas the masking method uses a single set of masks for all objects in a catalog. Consequently, the window method may enable finding correct periods that are typical aliases (e.g., those defined in Section 1), whereas the masking method excludes all of these periods.

Disadvantage: Using Algorithm 1, this method has the same time complexity for L-S as the other methods, which is $O(nm)$. It also has an additional time complexity for the peak finding algorithm, Ψ , which we describe below. However, since this method requires two LSP to be calculated, it is at least twice as many operations as the LSP. The implementation for the window periodogram peak finding function considered any power greater than 5σ from 0, where σ is the standard deviation of all the

Table 2

The parameters for the window/VanderPlas methods. “L-S Min. Peak Height” is the minimum power of a peak, relative to the highest power, to still be considered a peak. For example, for LSST, if the highest peak’s power was 0.8, then the peak power cutoff would be 0.08. “L-S Peak Order” is the “order” argument to the `argre1max` function in `scipy/cuSignal`.

	L-S Min. Peak Height	L-S Peak Order
ZTF	1/7	10 000
SSPDB	1/10	3000

powers in the window periodogram, a peak. The LSP peaks were found using the `argre1max` function in `SciPy/cuSignal` (Virtanen et al., 2020; RAPIDS Development Team, 2018), which has a time complexity Ψ of $O(om)$ where m is the number of frequencies examined and o is the order parameter. The order parameter for `argre1max` and the minimum peak height, relative to the strongest peak, is located in Table 2.

4.4. VanderPlas (VP) method

VanderPlas (2018) describes a method for removing aliases that also utilizes the window function. The difference between this method and the window method is that it compares more peaks between the LSP and window periodogram and in particular it considers pseudo-aliases.

The VP method steps are as follows, where $f_{\text{peak}} = \max(P(f))$:

1. Check if there are any peaks in the LSP at f_{peak}/m , where $m \in \{2, 3\}$ which checks if the found peak is an integer multiple of the real peak.
2. Check for peaks at $f_{\text{peak}} \pm n\delta f$, where $n \in \{1, 2\}$ and where δf is the frequency having the highest window periodogram power.
3. Manually check the highest peaks and fit a model to find the best one. We ignore this step as this requires human intervention, which is not feasible for large-scale survey data.

This method has the same time complexity as the window method at $O(nm) + \Psi$, where Ψ , when using `argre1max`, is $O(om)$ and it shares its **advantages** and **disadvantages** since the methods are similar. The implementation also uses the same parameters as the Window method, located in Table 2.

¹ <https://gist.github.com/drkr98/6b15633e8fc43e8daf6b628548006376>

Table 3

The percentage match between the derived period and the real period for the three surveys and the percentage change from the baseline. The “None/Baseline” row is when none of the methods were used and the highest LSP power was used as the derived period. SSPDB used a sample of 10 000 objects.

Method	Percentage match		Percentage change	
	ZTF	SSPDB	ZTF	SSPDB
None/Baseline	64.8%	57.9%	–	–
Mask	65.8%	74.5%	1.56%	28.6%
MC	70.1%	53.5%	8.08%	–7.61%
Window	69.0%	71.5%	6.39%	23.4%
VP	14.0%	8.91%	–78.4%	–84.6%

5. Results

All L-S and window functions used the implementations from SciPy/cuSignal, using 5×10^6 frequencies on a uniform period grid on [1 h, 150 h].

As a baseline, Fig. 6 shows the period distribution when no dealiasing is applied and Figs. 7–10 show the period distributions for the four methods presented above.

Each of the distributions has spikes at aliases, meaning that none of the approaches are completely successful at removing all aliases because we assume that the true period distributions for our population are continuous.

Our next step is to compare the derived results with known values. For ZTF period solutions, the known values are provided by the LCDB (Section 4.1), which contains few alias solutions because of human curating and/or observational cadence that do not have aliasing at the derived periods. There are 2544 objects that are in both the LCDB and ZTF; all of them were used for the calculations. For the SSPDB, since we assigned rotation periods for every object, verifying the correct periods in aliased and de-aliased solutions for this case is straightforward.

The percentage match between the derived period and the real period is shown in Table 3. Objects were considered to have matching periods if their period, half their period, or double their period was within 10% of the LCDB (for ZTF) or assigned period (for SSPDB). The baseline method for both surveys had a match rate of over 50%, so most objects have their correct period derived. Once the methods were applied, the masking and window methods increased the match percentage over the baseline for both surveys while the MC method only increased the match percentage for ZTF, and the VP method decreased the match percentage for both surveys.

6. Discussion

We begin by examining ZTF and excluding the VP method (which will be described later). The best match percentage using ZTF is with the MC method; however, all of the methods are within $\approx 6\%$ of each other, including the baseline. At a minimum, if de-aliasing is needed for a ZTF-like survey, the masking method should be implemented as it incurs no extra cost. If computational cost is not prohibitive, then the MC should be implemented. In all cases (except where VP is implemented), our results suggest that around two-thirds of all reported solutions are likely to be correct.

For SSPDB, the variation among non-VP approaches is greater than for ZTF, ranging from barely better than 50% (MC) to almost 75% accurate (masking). Notably, the MC method performs worse than the baseline. We hypothesize that the long baseline of the SSPDB observations leads to the points in each subsample being

too temporally distant² for L-S to reliably derive the correct period.

The VP method was the worst performing of the methods as it seems to incorrectly select aliases/pseudo-aliases, causing the method to “overcorrect”. Fig. 10 shows that there are regions where no periods are detected and periods tend to be derived at longer periods, meaning that the correct periods get derived a small percentage of the time. With the VP and window methods being similar, we hypothesize that the reason the VP method is worse is that the pseudo-alias check causes an overcorrection of the periods.

One may wonder whether the match percentages found here could be improved since all of the methods still have derived periods at or near aliases. It is possible that better parameters for the methods could be found with a more exhaustive search of the parameter space, like those in Table 2. However, such a search is impractical due to the large volume of data in the catalogs.

For the masking method, since it is static and therefore unable to react to changes in aliases, the masks might have to be re-derived after the survey starts if an inaccurate simulation was used or if the observational cadence of the survey changes during the survey.

One important conclusion is that the best de-aliasing approach is not the same for ZTF and SSPDB, which implies that a study like this should be carried out for every large-scale survey. If this is impractical and a single uniform approach is preferred, we identify the masking process as the most effective, though this conclusion is based only on the two datasets considered here.

7. Conclusions and future work

We used two sets of survey data, one real and one synthetic, to test four de-aliasing techniques for period solutions from all-sky surveys.

We find that the masking method provides the overall best results and should be chosen for any given survey. This method has a relatively low time complexity. The masks for ZTF and SSPDB, and therefore LSST, have been generated and are presented in Table 1. The window method would also be a good choice to apply if the computational performance loss compared to the masking method is not important as it provides aliases for each object individually. However, we note that results may vary from survey to survey, and the best approach is to carry out individualized studies, such as this one.

This paper leads to several lines of future investigation:

- Improving the window method so it provides a higher match rate.
- Use several of the methods together to see if that improves the match rate.
- Develop a GPU version of the MC method in order to decrease its expensive computation time.
- Develop a method that removes the pseudo-aliases.
- Test how differences in a survey’s simulated data and its real data change its mask ranges.
- Develop and test methods for determining a “confidence” in a periodogram result in order to better gauge if a derived period result is correct.

² This is only an issue for the MC method. Having a long baseline of observations is beneficial for the other methods.

CRediT authorship contribution statement

D. Kramer: Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft, Writing – review & editing. **M. Gowanlock:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **D. Trilling:** Conceptualization, Data curation, Writing – original draft, Writing – review & editing. **A. McNeill:** Writing – original draft, Writing – review & editing. **N. Erasmus:** Methodology, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We acknowledge many useful conversations with Nat Butler and John Kecicioglu and with Tom Matheson and the ANTARES team. This work has significantly benefited from all of their expertise.

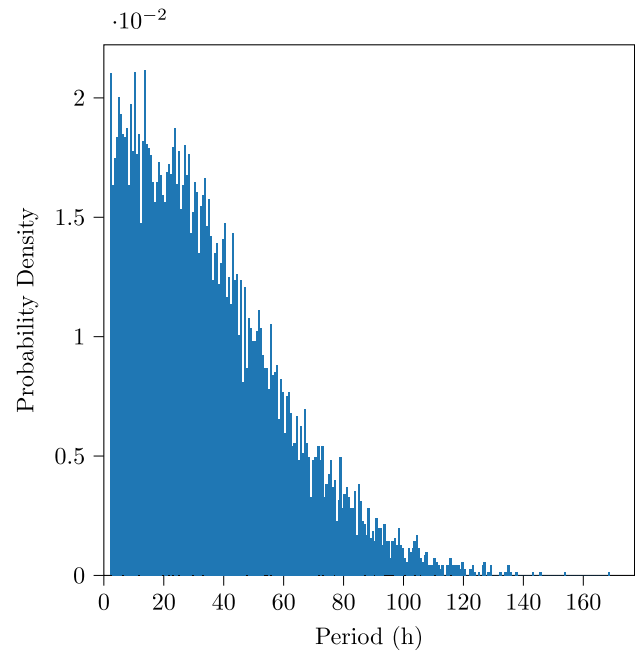
This work has been supported in part by the Arizona Board of Regents, Regents' Innovation Fund.

ZTF is a public-private partnership, with equal support from the ZTF Partnership and from the U.S. National Science Foundation through the Mid-Scale Innovations Program (MSIP). The ZTF partnership is a consortium of the following universities and institutions (listed in descending longitude): TANGO Consortium of Taiwan; Weizmann Institute of Sciences, Israel; Oskar Klein Center, Stockholm University, Sweden; Deutsches Elektronen-Synchrotron & Humboldt University, Germany; Ruhr University, Germany; Institut national de physique nucléaire et de physique des particules, France; University of Warwick, UK; Trinity College, Dublin, Ireland; University of Maryland, College Park, USA; Northwestern University, Evanston, USA; University of Wisconsin, Milwaukee, USA; Lawrence Livermore National Laboratory, USA; IPAC, Caltech, USA; Caltech, USA. We thank the anonymous reviewer for their insightful comments and helpful feedback on our manuscript.

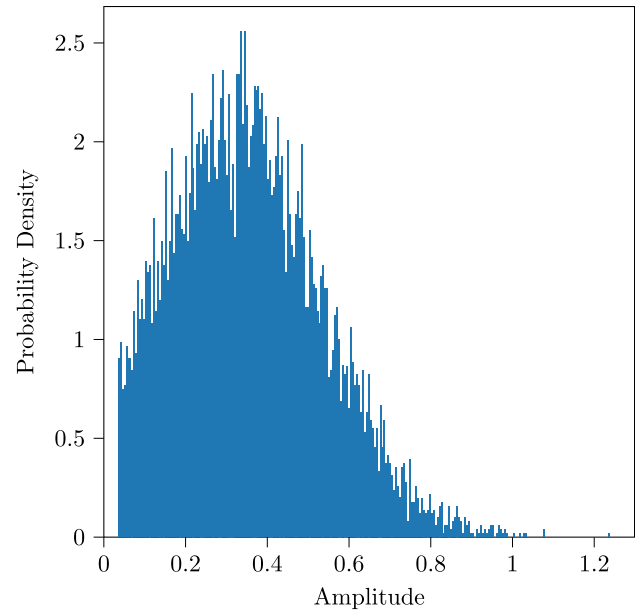
Appendix A. LSST generation

For each object from the LSST Synthetic Moving Objects Database, the following cuts/steps were taken in generating a periodic signal in their data

1. Import all objects with at least 30 observations in at least two filters.
2. Convert the apparent magnitudes to absolute magnitudes using the “filterg12” value as that filter’s G . SSPDB uses the Bowell HG system (Bowell et al., 1989) not the HG12 system (Muinonen et al., 2010) even though the field is called “filterg12”. If that filter’s “g12” data was NaN, then the average of the rest of the filter “g12” values were used. If all the filter’s “g12” values are NaN, then a value of $G = 0.15$ was used for all filters.
3. A period and amplitude were generated for each object and a sine wave with those properties was added to the objects derived absolute magnitudes. Both were generated using the parameters located in Table 4 and the resulting distributions are shown in Fig. 5. The truncated normal dis-



(a) The period distribution used for SSPDB.



(b) The amplitude distribution used for SSPDB.

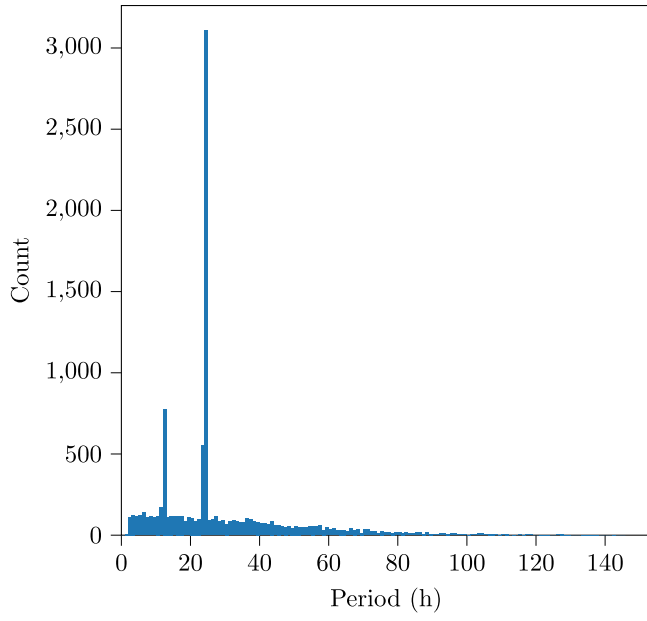
Fig. 5. The period and amplitude distributions for SSPDB. The x-axis is the period/amplitude and the y-axis is the probability density of each bin (the area under the histogram is 1).

Table 4

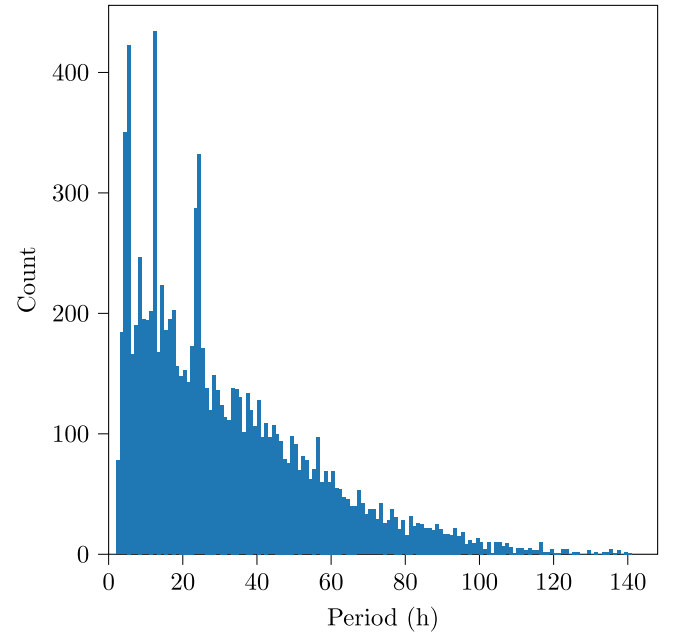
The distributions and parameters used for the period and amplitude generation. Plots for these distributions are available in Fig. 5. The distributions were approximated from the LCDB data (Section 4.1).

Distribution	Period		Amplitude	
	Truncated normal		Gamma	
Parameters	a	−0.1	a	0.48
	b	50	$\frac{1}{\beta}$	0.18
	μ	6		
	σ	40		

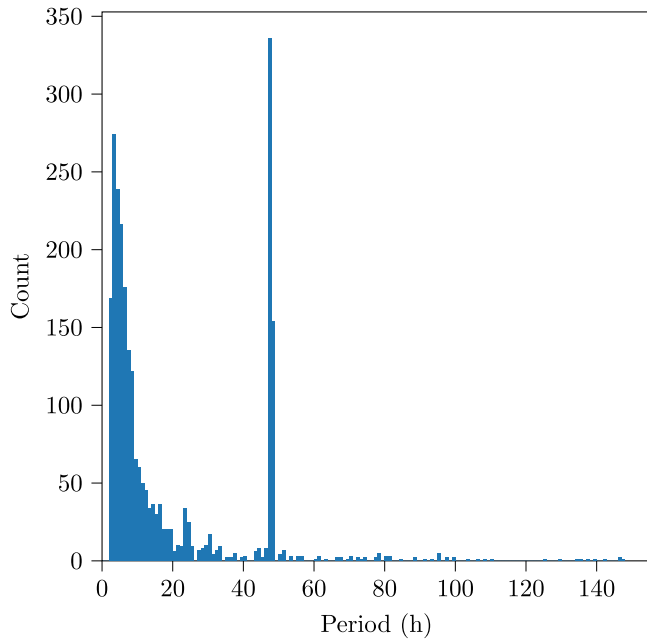
tribution used SciPy’s `truncnorm` function and the gamma distribution used SciPy’s `gamma` function.



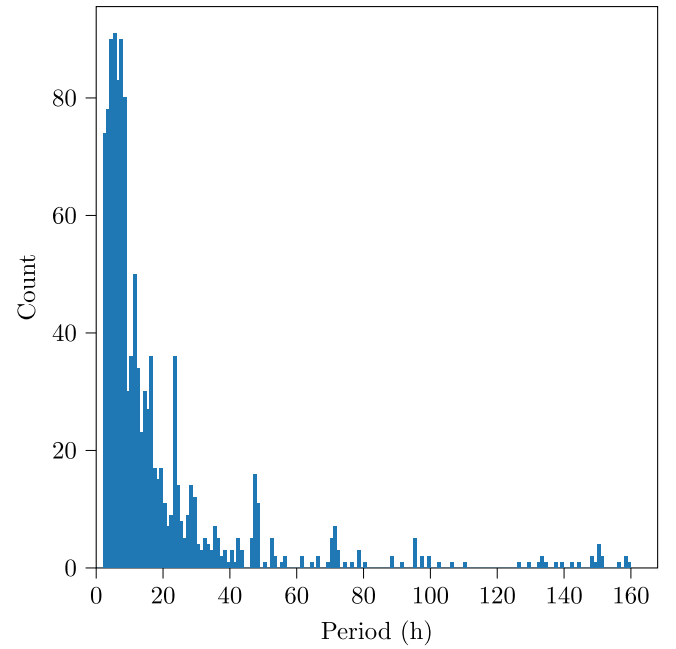
(a) SSPDB



(a) SSPDB



(b) ZTF



(b) ZTF

Fig. 6. The derived period distribution for the given survey using the no method (the base derived period distribution).

Fig. 7. The derived period distribution for the given survey using the mask method.

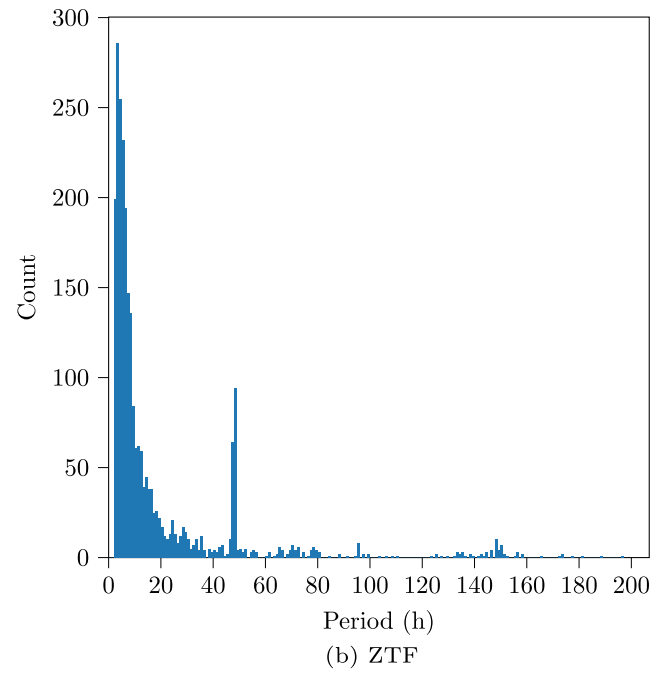
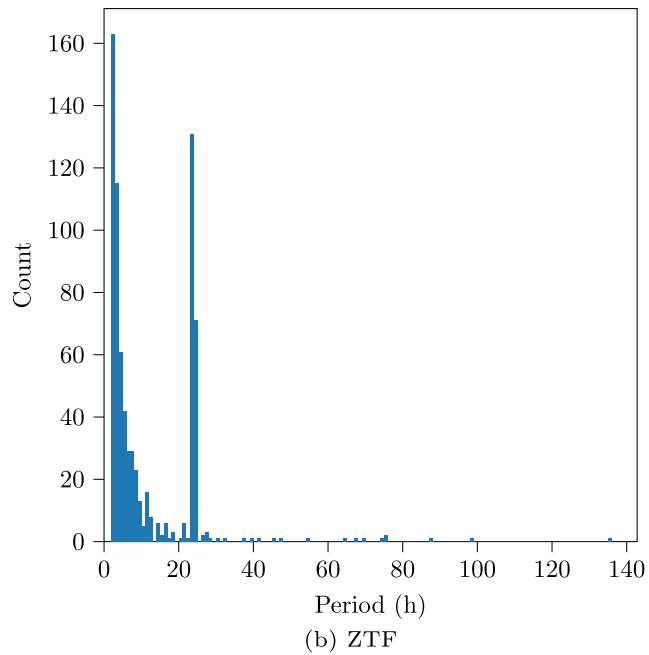
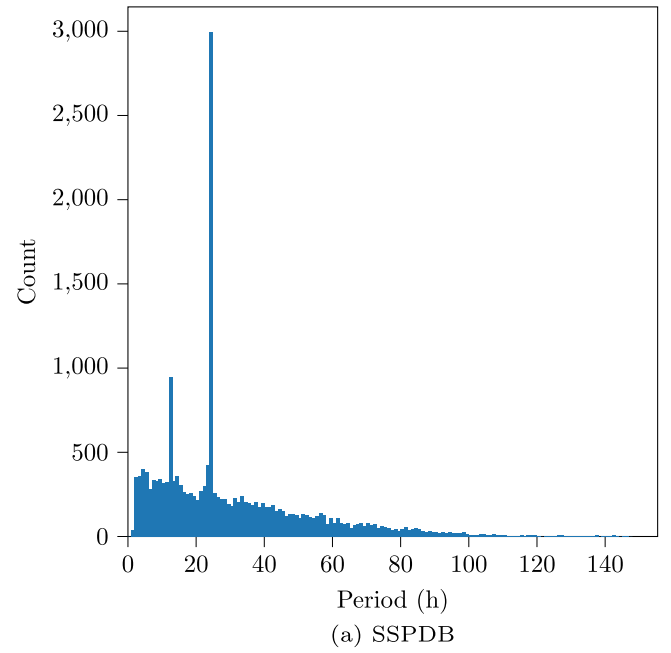
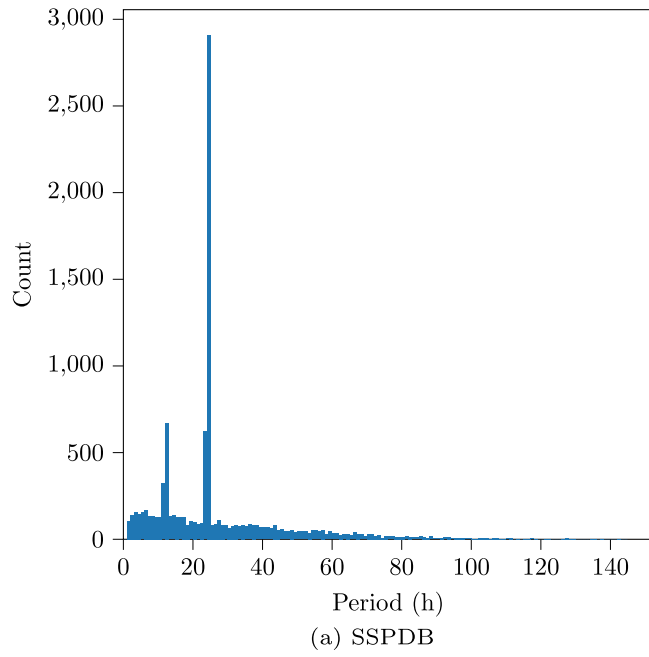


Fig. 8. The derived period distribution for the given survey using the Monte Carlo method.

Fig. 9. The derived period distribution for the given survey using the window method.

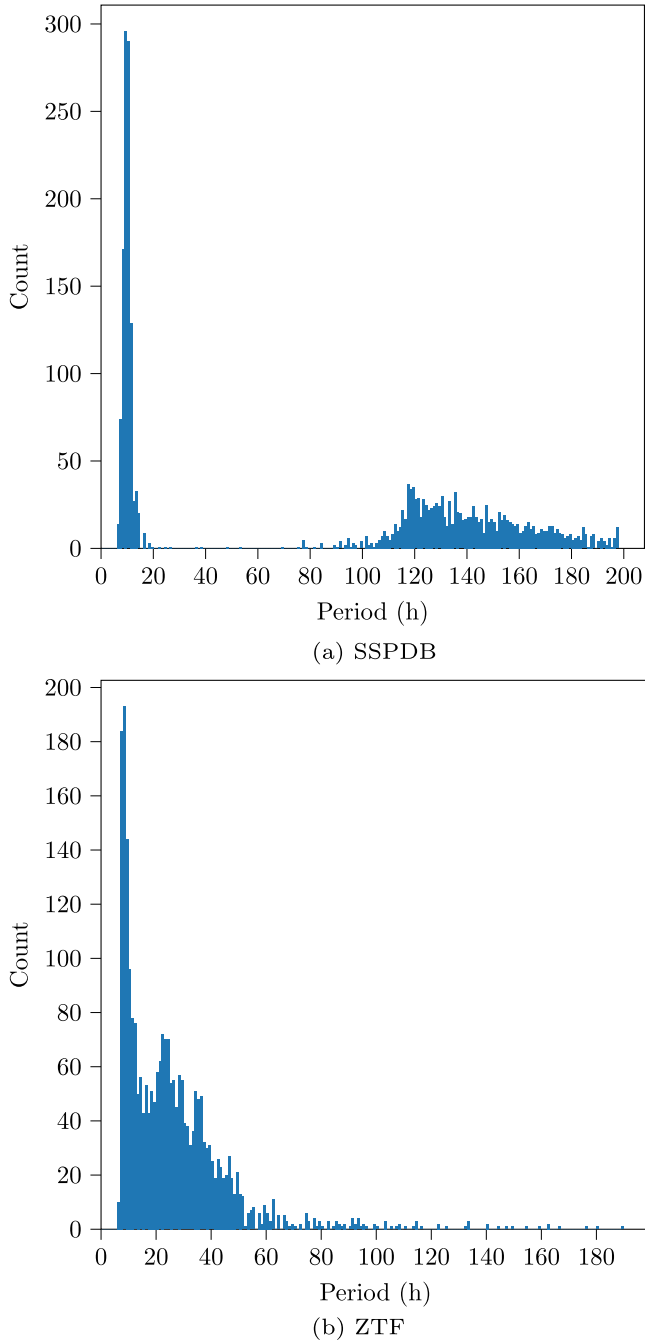


Fig. 10. The derived period distribution for the given survey using the VanderPlas method.

4. For each filter, the mean of the absolute magnitudes, μ_H , was calculated. $H - \mu_H$ for each filter was then concatenated, resulting in a single band of data with $\mu_H = 0$.
5. The observation data along with the assigned period and amplitude values were stored.

Appendix B. Derived period distributions

See Figs. 6–10.

References

LSST Science Collaboration, Abell, P.A., Allison, J., Anderson, S.F., Andrew, J.R., Angel, J.R.P., Armus, L., Arnett, D., Asztalos, S.J., Axelrod, T.S., Bailey, S.,

Ballantyne, D.R., Bankert, J.R., Barkhouse, W.A., Barr, J.D., Barrientos, L.F., Barth, A.J., Bartlett, J.G., Becker, A.C., Becla, J., Beers, T.C., Bernstein, J.P., Biswas, R., Blanton, M.R., Bloom, J.S., Bochanski, J.J., Boeshaar, P., Borne, K.D., Bradac, M., Brandt, W.N., Bridge, C.R., Brown, M.E., Brunner, R.J., Bullock, J.S., Burgasser, A.J., Burge, J.H., Burke, D.L., Cargile, P.A., Chandrasekharan, S., Chartas, G., Chesley, S.R., Chu, Y.-H., Cinabro, D., Claire, M.W., Claver, C.F., Clowe, D., Connolly, A.J., Cook, K.H., Cooke, J., Cooray, A., Covey, K.R., Culliton, C.S., de Jong, R., de Vries, W.H., Debattista, V.P., Delgado, F., Dell'Antonio, I.P., Dhital, S., Di Stefano, R., Dickinson, M., Dilday, B., Djorgovski, S.G., Dobler, G., Donalek, C., Dubois-Felsmann, G., Durech, J., Eliasdottir, A., Eracleous, M., Eyer, L., Falco, E.E., Fan, X., Fassnacht, C.D., Ferguson, H.C., Fernandez, Y.R., Fields, B.D., Finkbeiner, D., Figueroa, E.E., Fox, D.B., Francke, H., Frank, J.S., Frieman, J., Fromenteau, S., Furqan, M., Galaz, G., Gal-Yam, A., Garnavich, P., Gawiser, E., Geary, J., Gee, P., Gibson, R.R., Gilmore, K., Grace, E.A., Green, R.F., Gressler, W.J., Grillmair, C.J., Habib, S., Haggerty, J.S., Hamuy, M., Harris, A.W., Hawley, S.L., Heavens, A.F., Hebb, L., Henry, T.J., Hileman, E., Hilton, E.J., Hoadley, K., Holberg, J.B., Holman, M.J., Howell, S.B., Infante, L., Ivezić, Z., Jacoby, S.H., Jain, B., R. Jedicke, Jee, M.J., Garrett Jernigan, J., Jha, S.W., Johnston, K.V., Jones, R.L., Juric, M., Kaasalainen, M., Styliani, Kafka, Kahn, S.M., Kaib, N.A., Kalirai, J., Kantor, J., Kasliwal, M.M., Keeton, C.R., Kessler, R., Knezevic, Z., Kowalski, A., Krabbenham, V.L., Krughoff, K.S., Kulkarni, S., Kuhlman, S., Lacy, M., Lepine, S., Liang, M., Lien, A., Lira, P., Long, K.S., Lorenz, S., Lotz, J.M., Lupton, R.H., Lutz, J., Macri, L.M., Mahabal, A.A., Mandelbaum, R., Marshall, P., May, M., McGehee, P.M., Meadows, B.T., Meert, A., Milani, A., Miller, C.J., Miller, M., Mills, D., Minniti, D., Monet, D., Mukadam, A.S., Nakar, E., Neill, D.R., Newman, J.A., Nikolaev, S., Nordby, M., O'Connor, P., Oguri, M., Oliver, J., Olivier, S.S., Olsen, J.K., Olsen, K., Olszewski, E.W., Oluseyi, H., Padilla, N.D., Parker, A., Pepper, J., Peterson, J.R., Petry, C., Pinto, P.A., Pizagno, J.L., Popescu, B., Prsa, A., Radcka, V., Raddick, M.J., Rasmussen, A., Rau, A., Rho, J., Rhoads, J.E., Richards, G.T., Ridgway, S.T., Robertson, B.E., Roskar, R., Saha, A., Sarajedini, A., Scannapieco, E., Schalk, T., Schindler, R., Schmidt, S., Schmidt, S., Schneider, D.P., Schumacher, G., Scranton, R., Sebag, J., Seppala, L.G., Shemmer, O., Simon, J.D., Sivertz, M., Smith, H.A., Allyn Smith, J., Smith, N., Spitz, A.H., Stanford, A., Stassun, K.G., Strader, J., Strauss, M.A., Stubbs, C.W., Sweeney, D.W., Szalay, A., Szkody, P., Takada, M., Thorman, P., Trilling, D.E., Trimble, V., Tyson, A., Van Berg, R., Vanden Berk, D., VanderPlas, J., Verde, L., Vrsnak, B., Walkowicz, L.M., Wandelt, B.D., Wang, S., Wang, Y., Warner, M., Wechsler, R.H., West, A.A., Wiecha, O., Williams, B.F., Willman, B., Wittman, D., Wolff, S.C., Wood-Vasey, W.M., Wozniak, P., Young, P., Zentner, A., Zhan, H., 2009. LSST science book, version 2.0. arXiv e-prints [arXiv:0912.0201](https://arxiv.org/abs/0912.0201).

Baluev, R.V., 2008. Assessing the statistical significance of periodogram peaks. *Mon. Not. R. Astron. Soc.* 385 (3), 1279–1285. doi:[10.1111/j.1365-2966.2008.12689.x](https://doi.org/10.1111/j.1365-2966.2008.12689.x), [arXiv:0711.0330](https://arxiv.org/abs/0711.0330).

Barbara, N.H., Bedding, T.R., Fulcher, B.D., Murphy, S.J., Van Reeth, T., 2022. Classifying Kepler light curves for 12,000 A and F stars using supervised feature-based machine learning. *Mon. Not. R. Astron. Soc.* doi:[10.1093/mnras/stac1515](https://doi.org/10.1093/mnras/stac1515), [arXiv:2205.03020](https://arxiv.org/abs/2205.03020).

Becker, A.C., Bochanski, J.J., Hawley, S.L., Ivezić, Ž., Kowalski, A.F., Sesar, B., West, A.A., 2011. Periodic variability of low-mass stars in sloan digital sky survey stripe 82. *Astrophys. J.* 731 (1), 17. doi:[10.1088/0004-637X/731/1/17](https://doi.org/10.1088/0004-637X/731/1/17), [arXiv:1102.1387](https://arxiv.org/abs/1102.1387).

Bellm, E.C., Kulkarni, S.R., Graham, M.J., Dekany, R., Smith, R.M., Riddle, R., Masci, F.J., Helou, G., Prince, T.A., Adams, S.M., Barbarino, C., Barlow, T., Bauer, J., Beck, R., Belicki, J., Biswas, R., Blagorodnova, N., Bodewits, D., Bolin, B., Brinell, V., Brooke, T., Bue, B., Bulla, M., Burruss, R., Cenko, S.B., Chang, C.K., Connolly, A., Coughlin, M., Cromer, J., Cunningham, V., De, K., Delacroix, A., Desai, V., Duev, D.A., Eadie, G., Farnham, T.L., Feeney, M., Feindt, U., Flynn, D., Franckowiak, A., Frederick, S., Fremling, C., Gal-Yam, A., Gezari, S., Giomi, M., Goldstein, D.A., Golkhou, V.Z., Goobar, A., Groom, S., Hacopian, E., Hale, D., Henning, J., Ho, A.Y.Q., Hover, D., Howell, J., Hung, T., Huppenkothen, D., Imel, D., Ip, W.H., Ivezić, Ž., Jackson, E., Jones, L., Juric, M., Kasliwal, M.M., Kaspi, S., Kaye, S., Kelley, M.S.P., Kowalski, M., Kramer, E., Kupfer, T., Landry, W., Laher, R.R., Lee, C.D., Lin, H.W., Lin, Z.Y., Lunnan, R., Giomi, M., Mahabal, A., Mao, P., Miller, A.A., Monkevitz, S., Murphy, P., Ngeow, C.C., Nordin, J., Nugent, P., Ofek, E., Patterson, M.T., Penprase, B., Porter, M., Rauch, L., Rebbapragada, U., Reiley, D., Rigault, M., Rodriguez, H., van Roestel, J., Rusholme, B., van Santen, J., Schulze, S., Shupe, D.L., Singer, L.P., Soumagnac, M.T., Stein, R., Surace, J., Sollerman, J., Szkody, P., Taddia, F., Terek, S., Sistine, A.V., van Velzen, S., Vestrand, W.T., Walters, R., Ward, C., Ye, Q.Z., Yu, P.C., Yan, L., Zolkower, J., 2018. The Zwicky transient facility: System overview, performance, and first results. *Publ. Astron. Soc. Pac.* 131 (995), 018002. doi:[10.1088/1538-3873/aaecbe](https://doi.org/10.1088/1538-3873/aaecbe).

Bowell, E., Hapke, B., Domingue, D., Lumme, K., Peltoniemi, J., Harris, A.W., 1989. Application of photometric models to asteroids. In: Binzel, R.P., Gehrels, T., Matthews, M.S. (Eds.), *Asteroids II*. pp. 524–556.

Carbonell, M., Oliver, R., Ballester, J.L., 1992. Power spectra of gapped time series - A comparison of several methods. *Astron. Astrophys.* 264 (1), 350–360.

Coughlin, M.W., Burdge, K., Duev, D.A., Katz, M.L., van Roestel, J., Drake, A., Graham, M.J., Hillenbrand, L., Mahabal, A.A., Masci, F.J., Mróz, P., Prince, T.A.,

- Yao, Y., Bellm, E.C., Burruss, R., Dekany, R., Jaodand, A., Kaplan, D.L., Kupfer, T., Laher, R.R., Riddle, R., Rigault, M., Rodriguez, H., Rusholme, B., Zolkower, J., 2021. The ZTF source classification project: II. Periodicity and variability processing metrics. *Mon. Not. R. Astron. Soc.* 505 (2), 2954–2965. doi:10.1093/mnras/stab1502, arXiv:2009.14071.
- Duev, D.A., Mahabal, A., Masci, F.J., Graham, M.J., Rusholme, B., Walters, R., Karmarkar, I., Frederick, S., Kasliwal, M.M., Rebbapragada, U., Ward, C., 2019. Real-bogus classification for the Zwicky Transient Facility using deep learning. *Mon. Not. R. Astron. Soc.* 489 (3), 3582–3590. doi:10.1093/mnras/stz2357.
- Erasmus, N., Kramer, D., McNeill, A., Trilling, D.E., Janse van Rensburg, P., van Belle, G.T., Tonry, J.L., Denneau, L., Heinze, A., Weiland, H.J., 2021. Discovery of superslow rotating asteroids with ATLAS and ZTF photometry. *Mon. Not. R. Astron. Soc.* 506 (3), 3872–3881. doi:10.1093/mnras/stab1888.
- Friedman, J.H., 1984. A variable span smoother.
- Gowanlock, M., Butler, N.R., Trilling, D.E., McNeill, A., 2022. GPU-enabled searches for periodic signals of unknown shape. *Astron. Comput.* 38, 100511. doi:10.1016/j.ascom.2021.100511.
- Graham, M.J., Drake, A.J., Djorgovski, S.G., Mahabal, A.A., Donalek, C., 2013. Using conditional entropy to identify periodicity. *Mon. Not. R. Astron. Soc.* 434 (3), 2629–2635. doi:10.1093/mnras/stt1206.
- Heinze, A.N., Tonry, J.L., Denneau, L., Flewelling, H., Stalder, B., Rest, A., Smith, K.W., Smartt, S.J., Weiland, H., 2018. A first catalog of variable stars measured by the asteroid terrestrial-impact last alert system (ATLAS). *Astron. J.* 156 (5), 241. doi:10.3847/1538-3881/aab47f, arXiv:1804.02132.
- Huber, M.E., Nikolaev, S., Cook, K.H., Rest, A., Smith, C., Olsen, K., Suntzeff, N., Aguilera, C., Stubbs, C., Garg, A., Challis, P., Becker, A., Covarrubias, R., Miceli, A., Welch, D., Miknaitis, G., Prieto, J.L., Clocchiatti, A., Minniti, D., Morelli, L., 2005. Mining the variables out of the SuperMACHO dataset. In: *American Astronomical Society Meeting Abstracts*, Vol. 207. p. 122.16.
- Ivezić, J., Kahn, S.M., Tyson, J.A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S.F., Andrew, J., et al., 2019. LSST: From science drivers to reference design and anticipated data products. *Agron. J.* 873, 111. doi:10.3847/1538-4357/ab042c, arXiv:0805.2366.
- Juric, M., Eggl, S., Jones, L., Moeyens, J., Bellm, E., Ivezić, Z., Lust, N., Stetzler, S., Cornwall, S., Berres, A., Chernyavskaya, M., Project, R.O.C., 2021. Being ready at first light: 10yr simulations of the legacy survey of space and time (LSST). *Bull. AAS* 53 (7), <https://baas.aas.org/pub/2021n7i101p06>.
- Lomb, N.R., 1976. Least-squares frequency analysis of unequally spaced data. *Astrophys. Space Sci.* 39 (2), 447–462. doi:10.1007/BF00648343.
- McNeill, A., Trilling, D.E., Mommert, M., 2018. Constraints on the density and internal strength of 11/oumuamua. *Astrophys. J. Lett.* 857 (1), L1. doi:10.3847/2041-8213/aab9ab, arXiv:1803.09864.
- Muironen, K., Belskaya, I.N., Cellino, A., Delbò, M., Lévassieur-Regourd, A.C., Penttilä, A., Tedesco, E.F., 2010. A three-parameter magnitude phase function for asteroids. *Icarus* 209 (2), 542–555. doi:10.1016/j.icarus.2010.04.003.
- RAPIDS Development Team, 2018. RAPIDS: Collection of libraries for end to end GPU data science. URL: <https://rapids.ai>.
- Roberts, D.H., Lehar, J., Dreher, J.W., 1987. Time series analysis with CLEAN. I. Derivation of a spectrum. *Astron. J.* 93, 968–989. doi:10.1086/114383.
- Scargle, J.D., 1982. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *Agron. J.* 263, 835–853. doi:10.1086/160554.
- Trilling, D.E., Gowanlock, M., Kramer, D., McNeill, A., Donnelly, B., Butler, N., Kececioglu, J., 2023. The solar system notification alert processing system (SNAPS): Design, architecture, and first data release (SNAPShot1). *Astron. J.* 165 (3), 111. doi:10.3847/1538-3881/acac7f.
- VanderPlas, J.T., 2018. Understanding the Lomb–Scargle periodogram. *Astrophys. J. Suppl. Ser.* 236 (1), 16. doi:10.3847/1538-4365/aab766.
- Đurech, J., Vávra, M., Vančo, R., Erasmus, N., 2022. Rotation periods of asteroids determined with bootstrap convex inversion from ATLAS photometry. *Front. Astron. Space Sci.* 9, 809771. doi:10.3389/fspas.2022.809771.
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. Scipy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods* 17, 261–272. doi:10.1038/s41592-019-0686-2.
- Warner, B.D., Harris, A.W., Pravec, P., 2021. Asteroid Lightcurve Database (LCDB) Bundle V4.0. NASA Planetary Data System, doi:10.26033/J3XC-3359, URL: https://sbn.psi.edu/pds/resource/doi/lcdb_4.0.html.