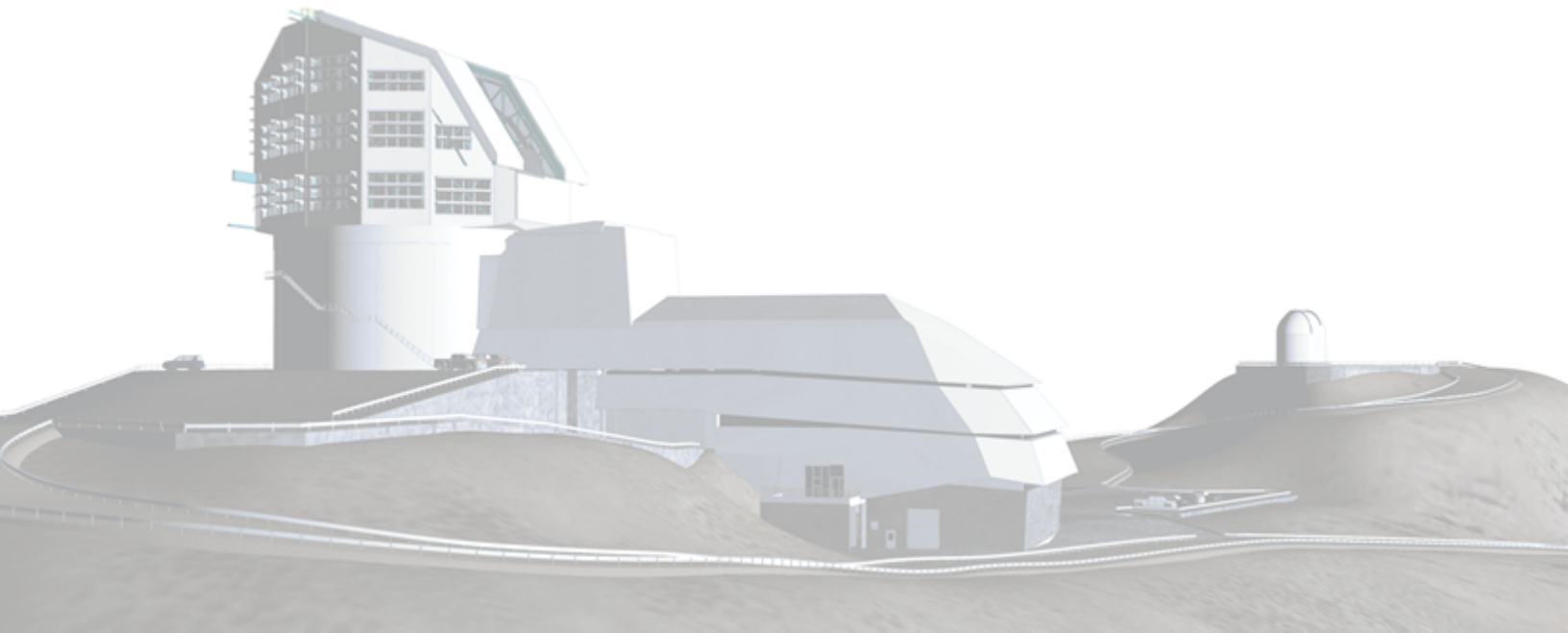

Vera C. Rubin Observatory
Data Management

Periodicity Analysis in Alert Production

Anastasios Tzanidakis, Eric Bellm

DMTN-221

Latest Revision: 2022-10-27

Abstract

The baselined timeseries features to be computed in Alert Production include a Lomb-Scargle periodogram. We assess the computational and scientific performance of several configurations on simulated alert data.

Change Record

Version	Date	Description	Owner name
1	YYYY-MM-DD	Unreleased.	Anastasios Tzanidakis, Eric Bellm

Document source location: `https://github.com/lsst-dm/dmtn-221`

Contents

Periodicity Analysis in Alert Production

1 Note

Please note that this DM-technote is currently work in progress, and publicly available. The final findings will be pushed to the DMTN-221 repository in late November of 2022 by Anastasios Tzanidakis.

2 Motivation

The characterization of periodicity from time-series is a fundamental constraint to numerous astrophysical applications. Phenomenologically, estimating the periodicity and its significance can shed light on stellar pulsation theory (Antonello & Pastori, 1981), distance estimation and mapping of the Galaxy through the period-luminosity relationship (Skowron et al., 2019), constraint fundamental parameters of stellar binaries (Farinella et al., 1979) and stellar rotation (Walkowicz & Basri, 2013). In the recent decade, the use of periodicity has also been extensively used as a feature to classify variable phenomena (Richards et al., 2011) across the HR diagram.

Previous work from Oluseyi et al. (2012) demonstrate the concept of injection-recovery test for synthetically generated RR Lyrae templates from SDSS using the op_sim version 1.29 cadence strategy. However, no study to date has demonstrated the recovery period distribution of more complex periodic phenomena (i.e eclipsing binaries, quasi-periodic AGN). Here we perform an end-to-end, injection and recovery analysis for multi-band time series to asses the statistical significance of period finding in the context of AP.

3 Synthetic Light Curves

In this study we use the training set Extended LSST Astronomical Time-series Classification Challenge (ELAsTiCC) light curves with a 12 month history that will best mimic the calculations performed for AP. In short, ELAsTiCC is an end-to-end real-time pipeline for generating mock photometric alerts at the expected Legacy Survey of Time and Space (LSST) rate. The photometric alerts will be distributed in real-time to brokers to benchmark classification algorithms.

The ELAsTiCC uses the v.1.7 cadence strategy from op_sims. A single detection is based on the DIA performance from DC2. Each light curve include photometric noise using Poisson noise that includes the equivalent area, background noise per unit area, sky and CCD read noise. Most importantly, the ELAsTiCC light curves simulate force photometry alerts which we recommend all time series features to be calculated upon.

Here we present two classes of periodic sources: RR Lyrae (RRL) and eclipsing binaries (EB). Both RRL's and EB's represent both the extrema of simple and complex periodic sources that will throughly test the underlying ability of the periodogram to find the correct period. Currently, the eclipsing binary sample from ELAsTiCC that was generated from open-source Python package (ELISa). the only contains a few hundred simulated binaries. To maximize the use different eclipsing binary stars sampled at different phases, each light curve over a three year photometric history is sampled at some random time and ensures that the maximum baseline is less than 12 months. In Figure 1 we show the typical multi-band AP light curve sampled from ELAsTiCC with uncertainties. A detailed Jupyter notebook interacting and fetching the ELAsTiCC light curves can be found here.

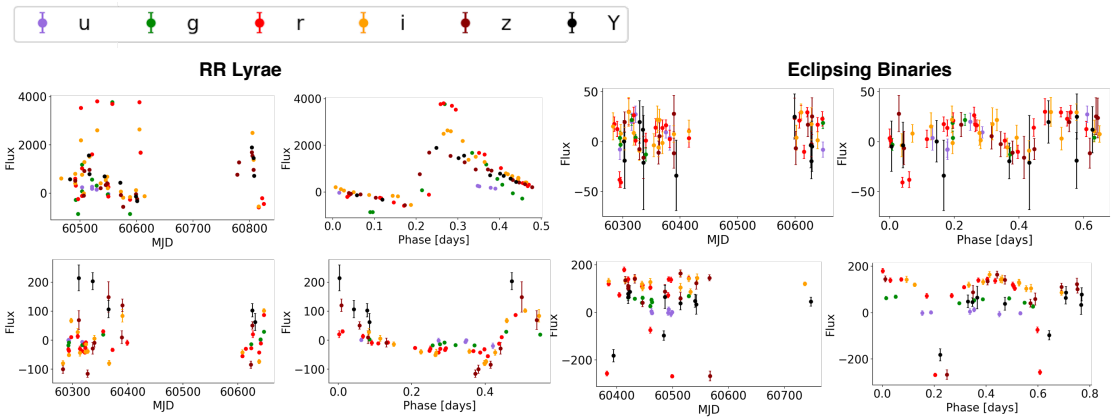

FIGURE 1: The above panels demonstrate the multi-band light curves simulated from ElasTiCC using RR lyrae (left panel) and eclipsing binaries (right panel). The first column of each panel shows the observed light curves, and the second the phase folded light curves at the correct period after 12 months.

4 Injection-Recovery Testing

We consider two models of periodic phenomena: RR lyrae and eclipsing binaries ranging from a periods of 0.2 days to 10 days. All computations of the floating-mean Lomb-Scargle periodogram are using the open-source Python package 'gatspy' VanderPlas & Ivezić (2015).

4.1 Single Band Lomb-Scargle Periodogram

We first consider the case of single-band detections from AP. Here we aim to answer if the 12 month AP history in a single-band light curve is enough to recover periods. We consider the canonical floating mean to estimate the periodogram of 10,000 alert light curves in the *r*-band for three Fourier component modes. In Figure 2 we show the results of this analysis for RRL's (rop row) and EB's (bottom row). For each light curve model, we randomly sample a light curve and compute the Lomb-Scargle periodogram using a heuristic period grid with an oversampling factor of 2 and Nyquist frequency of 30 (∼0.03 days) to account for the fast periodic phenomena. We extrapolated the highest power period from the periodogram and compare it to the true period. In Figure 2 we also include the number of *r*-band detections for each light curve. As expected, we did not find any preference of an optical single-band photometric filter that outperformed the others. While the RR Lyrae exhibit at larger Fourier modeling modes more aliasing, overall the recovery fraction of correctly identified periods is larger compared to the EBs. For both light curve models, we consistently found in each model configuration that the average correctly identified period coincides with a larger number of detections. Overall we find consistent results for the RR Lyrae done by a similar analysis by VanderPlas & Ivezić (2015). The true challenge of this approach comes from the more complex light curves such as the EBs.

- Injected vs recovered period plots eclipsing binaries and RRL
- Fraction of correctly recovered periods using single-band Lomb-Scargle (for N=1...3 Fourier components)

Even with a low completeness, computing the Lomb-Scargle periodogram on single-band time series alerts, this will likely outperform any other known survey due to the large survey volume. Single-band Lomb-Scargle implementations are also faster by a factor of ∼5 due to the lower average number of detections per filter. Given that only roughly 17 percent of each bandpass filter will be available, the single-band Lomb-Scargle misses by a higher margin the underlying true period. We now turn our attention to the multi-band Lomb-Scargle.

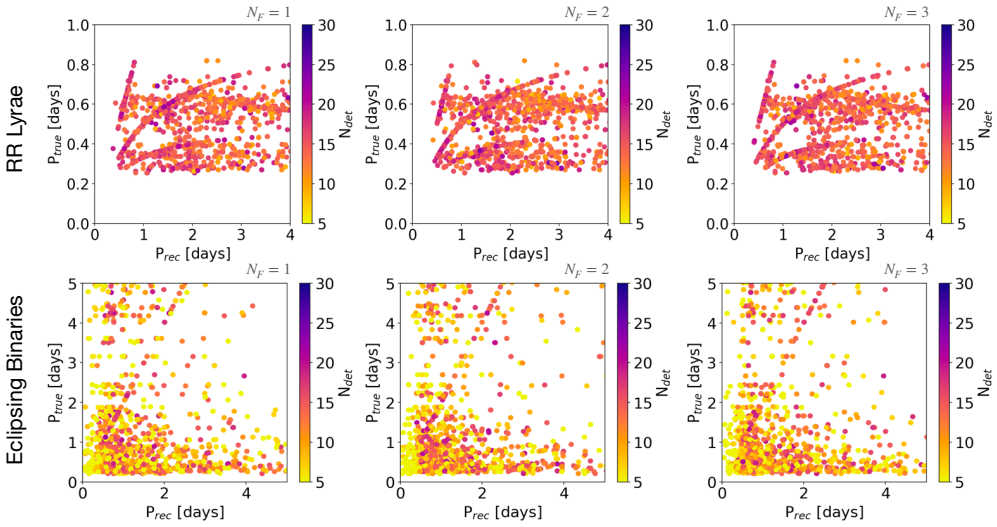

FIGURE 2: We show the injected (true period) versus recovered highest peak in the periodogram for a sample of 1000 r-band detections from RR lyrae and eclipsing binaries. Each column shows the period recovery for N=1 to N=3 Fourier components when computing the Lomb-Scargle periodogram.

4.2 Multi-Band Lomb-Scargle Periodogram

Next, we test the multi-band Lomb-Scargle periodogram implementation adapted on *gatspy* to our ELAsTiCC light curves. Given the rich multi-band alert detections LSST will deliver, we expect that that conventional period finding algorithms will perform better. Generally, the multi-band Lomb-Scargle periodogram we will be considering for this study can be written out with two components:

$$y = theta_0 + \sum_{n=1}^{M_{base}} \left[\theta_{2n-1} \sin(n\omega t) + \theta_{2n} \cos(n\omega t)\right] \theta_0^{(k)} + \sum_{n=1}^{M_{band}} \left[\theta_{2n-1}^{(k)} \sin(n\omega t) + \theta_{2n}^{(k)} \cos(n\omega t)\right] \quad (1)$$

First is the base Fourier component that describes the overall **shared** variability, and second is the per band component that is modeled by the residuals of the base component with each photometric filter. In practice, a larger configuration of base and band components can lead to the estimation of more complex light curves. *Gatspy* automatically determines through the use of least-squares minimization the scalar parameters from the above equation.

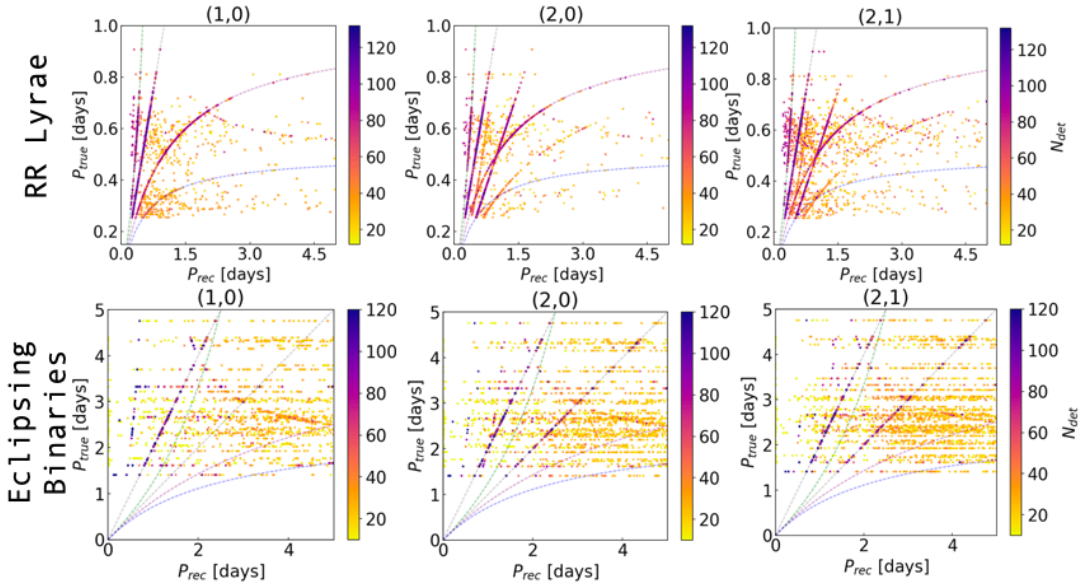

FIGURE 3: Multi-band injection-recovery test for 5000 RRL (top row) and EB (bottom row). Each column title includes the number of Fourier base and band terms used to compute the Lomb-Scargle periodogram. Additionally we color code each recovered period by the number of detections. The overlaid curved lines represent the n=1 aliasing, while the straight lines represent the n=1,2 harmonics.

4.3 Peak Significance Metric

In this section we discuss possible metrics for peak significance. One challenge when computing the periodogram for billions of sources is the inevitability of spurious period detections, effects of harmonics, and aliasing. One diagnostic used in the literature is the False Alarm Probability (FAP) of a given peak in the power spectrum. The FAP measures the probability uncertainty associated with the selected peak under the null hypothesis.

4.4 Timing Analysis

In this final section we discuss the average run time per method mentioned above.

A References

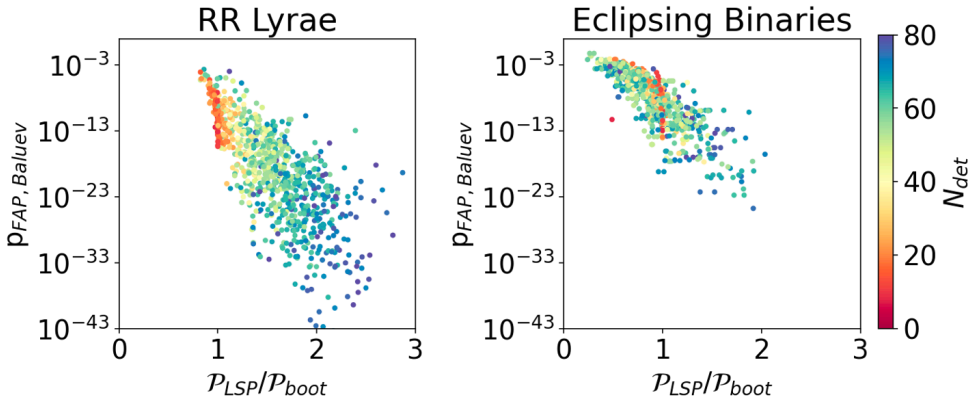

FIGURE 4: Multi-band FAP rate computed via the Baluev approximation compared to a 2σ bootstrap approach for 5000 RRL and EB alert light curves. For each light curve we also color code the total number of $ugrizy$ detections.

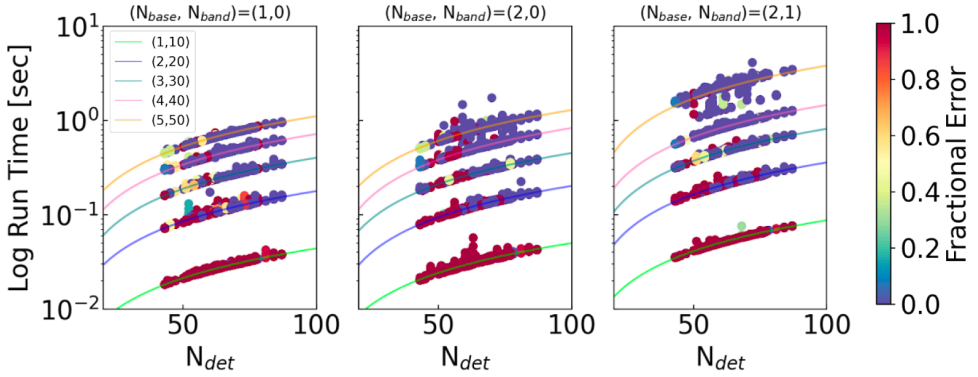

FIGURE 5: Average run time per number of $ugrizy$ light curve detections. Each column represents the model complexity of each multi-band Lomb Scargle periodogram. Within each panel, we use a heuristically determined period grid by varying the oversampling factor and the the Nyquist frequency. Finally, for each successive run we color code the fractional error compared from the true period.

References

Antonello, E., Pastori, L., 1981, PASP, 93, 237, doi:10.1086/130812, ADS Link

Farinella, P., Luzny, F., Mantegazza, L., Paolicchi, P., 1979, ApJ, 234, 973, doi:10.1086/157581, ADS Link

Oluseyi, H.M., Becker, A.C., Culliton, C., et al., 2012, AJ, 144, 9, doi:10.1088/0004-6256/144/1/9, ADS Link

Richards, J.W., Starr, D.L., Butler, N.R., et al., 2011, ApJ, 733, 10 (arXiv:1101.1959), doi:10.1088/0004-637X/733/1/10, ADS Link

Skowron, D.M., Skowron, J., Mróz, P., et al., 2019, Science, 365, 478 (arXiv:1806.10653), doi:10.1126/science.aau3181, ADS Link

VanderPlas, J.T., Ivezić, Ž., 2015, ApJ, 812, 18 (arXiv:1502.01344), doi:10.1088/0004-637X/812/1/18, ADS Link

Walkowicz, L.M., Basri, G.S., 2013, MNRAS, 436, 1883 (arXiv:1309.2159), doi:10.1093/mnras/stt1700, ADS Link

B Acronyms

Acronym	Description
AGN	active galactic nuclei
AP	Alert Production
CCD	Charge-Coupled Device
DC2	Data Challenge 2 (DESC)
DIA	Difference Image Analysis
DM	Data Management
DMTN	DM Technical Note
EB	ExaByte
HR	Human Resources
LSST	Legacy Survey of Space and Time (formerly Large Synoptic Survey Telescope)
SDSS	Sloan Digital Sky Survey