Simulations of multiband Lomb-Scargle-derived variable star periods

K. Tisanić[1]*; L. Palaversa [1]

[1] Ruđer Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia

April 15, 2021

Abstract

We have employed Monte Carlo simulations for estimating the smallest number of measurements per band needed to reliably estimate a pulsating periodic variable star's period for the upcoming Vera C. Rubin observations. We use a sample of variable stars detected and classified in the Gaia Data Release 2 with counterparts in the Zwicky Transient Facility DR4 to generate synthetic light curves in the g, r and i bands. The synthetic light curves are used to simulate the effects of different light curve sampling. We conclude that for the simulated variable star types, there is a 50% chance of reaching a 0.1% relative error with 10 measurements per band, and 99.7% chance of reaching 1% relative error for 15-20 measurements for band, except for type II Cepheids and Miras.

1 Science Case

The goal of this analysis is to determine the minimum number of observations required to obtain the correct period of a pulsating variable star: a fundamental parameter for distance estimation, classification and stellar evolution studies. The procedure consists of two stages: building up a library of "synthetic" light curves and simulations of period determination through variation of the number of observations drawn from the synthetic light curves.

The synthetic phased light curve library is built up from a crossmatch of Gaia DR2 variable stars catalog and the Zwicky Transient Facility (ZTF DR3, cone radius 3 arcsec yielding 54535 matches in total). Gaia variable stars catalog was selected as it is one of the largest all-sky catalogs with reliable classification of variable stars spanning a wide range of types, brightness, ages and environments. ZTF, on the other hand, provides well-sampled light curves in gri bands similar to those of the R-LSST. The cross-matched sample contained a total of:

*ktisanic@irb.hr

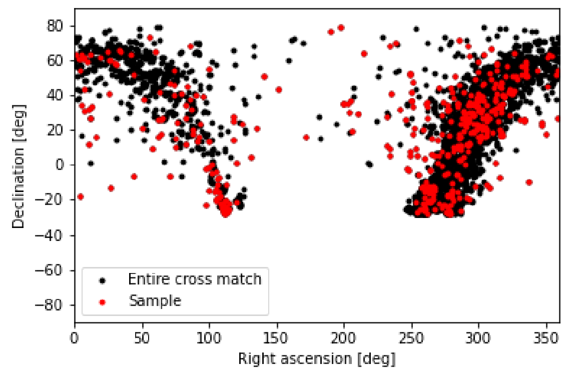

Figure 1: Positions of the Gaia DR2 and ZTF crossmatched stars (black points) and the positions of stars in the chosen subsample (red points).

- 274 Cepheids (21 classical, 203 type-II and 50 anomalous)
- 100 Type ab RRL
- 100 Long period variables (we expect that most are Miras)

Additionally we note that the Cepheid sample was heavily biased against classical Cepheids (only 50 classical Cepheids out of 274). Furthermore, since there were many more Miras and RR-Lyrae in the crossmatch, we chose a representative sample of 100 Miras and 100 ab RR-Lyrae with light curves that covered the different phenomenological subtypes (shapes) of light curves in order not to create a bias towards a particular light-curve shape or variable star type.

For the cross-matched stars, with coordinates shown in Fig. 1, we ran the single-term multiband Lomb-Scargle (LS) periodogram, implemented using the `gatspy` package (VanderPlas & Ivezić, 2015), and verified the result by comparison to the Gaia DR2 periods. In the cases where there was a difference, a manual inspection was performed and the correct period was selected. Additionally, the phased light curves were modeled using the LS periodogram evaluated at the correct periods, using a 0.1-10 times Gaia period search range and the default gatspy sampling width based on the observed light curve time interval. We chose a sample of phase points, converted each individual phase point to observation time, evaluated the LS periodogram at these observation times, and added noise to the simulations using heteroscedastic magnitude errors.

The heteroscedastic magnitude errors have been simulated by drawing samples from a fitted log-normal distribution to the relative magnitude errors. The log-normal distribution does not appear to be the best-fitting distribution, but should be considered a worst-case scenario since the distribution of relative error seems to be more leptokurtic than the log-normal distribution. This distribution was chosen because it reproduces the skewness of the initial dataset when converted from the relative error to the magnitude error.

This procedure allowed us to create a desired sampling of the *gri* light curve that would be sufficiently similar to the future R-LSST light curves and does not depend on theoretical assumptions. We do note that the sample used here is much shallower than the one that will be observed by the R-LSST and does not cover some important but specific regions such as the Magellanic clouds and the Gal. bulge. On the other hand, the sample used here is representative of the Halo and the Disk populations of the variable stars.

We use a log-relative period error as a statistic for estimating the reliability of the derived Lomb-Scargle best-fitting periods:

$$R = \log \left| \frac{P - P_{ZTF}}{P_{ZTF}} \right|, \tag{1}$$

where P is the period derived from each simulation, and P_{ZTF} is the period fitted to the ZTF data when constructing the library of LS parameters. An estimate of the percentiles of R for simulations grouped by number of observations per band, is shown in Table 1 and Fig. 2.

We find that there is no significant difference in the behavior of the 16th, 50th, 84th, and 99.7th percentiles of the R statistic between simulated types of variable stars, except for the 99.7th percentiles of type II Cepheids and Miras. We conclude that for simulated variable star types, there is a 50% chance of reaching a 0.1% relative error for 10 measurements per band, and 99.7% chance of reaching 1% relative error for 15-20 measurements for band, except for type II Cepheids and Miras. We find that there is no significant improvement above $20 - 25$ measurements per band for 16th, 50th, 84th percentiles.

We note that we have not carried out a detailed analysis of (very) noisy cases: those cases in which SNR of the observations is comparable to the amplitude of the variation. In those cases it may be indeed prudent to have more visits per pointing in order to allow period recovery. Indeed, low-amplitude variables may be the single most numerous category of the variable stars and its study was the turning point for development of asteroseismology (CoRoT, Kepler and Brite spacecraft) and understanding red giant variability through the past decade. We believe that LSST may provide material for lower precision ground-based "asteroseismology", particularly due to its long observing baseline.

2 Answers to questions

[A1] It is likely that a 10% decrease in the number of visits per pointing would not have a significant effect on the determination of periods variable stars over the planned 10-yr survey length. On the other hand, inclusion of the Galactic plane region(s) into the WFD would significantly increase the total number of observed, detected, classified and characterised variable stars of all types. Therefore, inclusion of the Galactic plane regions into the survey would definitely benefit the variable star community.

Table 1: Relative error of the periods derived from the simulated light curve library as a function of the number of simulated observations per band ('obs/band'). The P_{16}, P_{50}, P_{84} and $P_{99.7}$ columns show the 16th, 50th, 84th and 99.7the percentiles of the log-relative error. The statistics are calculated for the entire dataset ('All' column), and for different types of variable stars separately.

obs/band	All	Mira	Cepheids			RR Lyr	obs/band	All	Mira	Cepheids			RR Lyr
			anomalous	classical	type-II	ab				anomalous	classical	type-II	ab
	P_{16} [dex]							P_{84} [dex]					
5	-3.01	-3.35	-3.01	-2.70	-3.02	-2.93	5	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06
10	-3.71	-3.94	-3.68	-3.60	-3.71	-3.66	10	-2.67	-2.65	-2.73	-2.66	-2.66	-2.67
15	-3.83	-4.09	-3.79	-3.80	-3.83	-3.72	15	-2.81	-2.76	-2.87	-2.81	-2.80	-2.84
20	-3.90	-4.18	-3.90	-3.81	-3.90	-3.83	20	-2.90	-2.83	-2.97	-2.87	-2.89	-2.92
25	-3.96	-4.27	-3.96	-3.89	-3.95	-3.93	25	-2.97	-2.88	-3.04	-2.96	-2.96	-2.97
30	-4.02	-4.24	-3.98	-3.93	-4.02	-3.95	30	-3.01	-2.98	-3.08	-2.99	-3.00	-3.03
35	-4.04	-4.29	-4.01	-3.98	-4.04	-4.00	35	-3.05	-3.00	-3.10	-3.04	-3.03	-3.07
40	-4.07	-4.33	-4.06	-3.99	-4.07	-4.00	40	-3.08	-3.05	-3.14	-3.06	-3.06	-3.10
45	-4.09	-4.32	-4.08	-4.03	-4.09	-4.09	45	-3.12	-3.05	-3.17	-3.11	-3.10	-3.14
	P_{50} [dex]							$P_{99.7}$ [dex]					
5	-0.42	-0.58	-1.78	-0.19	-0.37	-0.48	5	0.63	0.62	0.58	0.63	0.61	0.64
10	-3.14	-3.22	-3.14	-3.06	-3.14	-3.10	10	-0.07	-0.08	-2.06	-1.91	-0.06	-0.06
15	-3.25	-3.42	-3.26	-3.20	-3.26	-3.20	15	-0.32	-0.14	-2.30	-2.26	-1.99	-0.13
20	-3.34	-3.47	-3.35	-3.26	-3.34	-3.30	20	-0.53	-0.06	-2.44	-2.38	-2.00	-0.13
25	-3.40	-3.56	-3.42	-3.33	-3.40	-3.38	25	-1.83	-0.05	-2.48	-2.45	-2.10	-0.21
30	-3.44	-3.60	-3.45	-3.38	-3.44	-3.41	30	-2.05	-2.07	-2.48	-2.53	-2.22	-1.58
35	-3.48	-3.63	-3.48	-3.44	-3.48	-3.45	35	-1.99	-1.89	-2.52	-2.49	-2.20	-1.41
40	-3.51	-3.67	-3.52	-3.45	-3.52	-3.49	40	-2.09	-2.17	-2.60	-2.54	-2.36	-1.76
45	-3.53	-3.65	-3.54	-3.48	-3.54	-3.53	45	-2.21	-2.07	-2.61	-2.58	-2.33	-1.85

[A2] Mini-surveys are indeed very interesting for the field as they could focus on a particular type of variability over a shorter period of time and utilize tailored cadence and high or specific sampling rates. In order to maximize the scientific returns we recommend that the mini-surveys observe high-density regions such as the Disk, the Bulge and the Magellanic clouds.

One particular example of a mini survey could be the search and characterisation of short-period variability (e.g. P<3h). Dedicating, for example, a single night (and perhaps a reduced set of filters and shorter exposures) to a high source-density region would also be beneficial for understanding of variable stars with longer periods.

[A3] The value of any given filter in the context of period determination is proportional to the number of observations made in it. Therefore, a decrease in the total number of observations is detrimental to the period recovery. An additional negative impact on the determination of periods would be an increase in the exposure length.

The flux of most of the sources of interest for the variable star community is smaller in the u-band than in other bands. Given the same exposure length, some of the sources saturated in other bands are not saturated in the u-band, and u-band observations can be used to determine periods for the sources saturated in the redder filters. If the u-band exposure were increased, this opportunity would be lost.

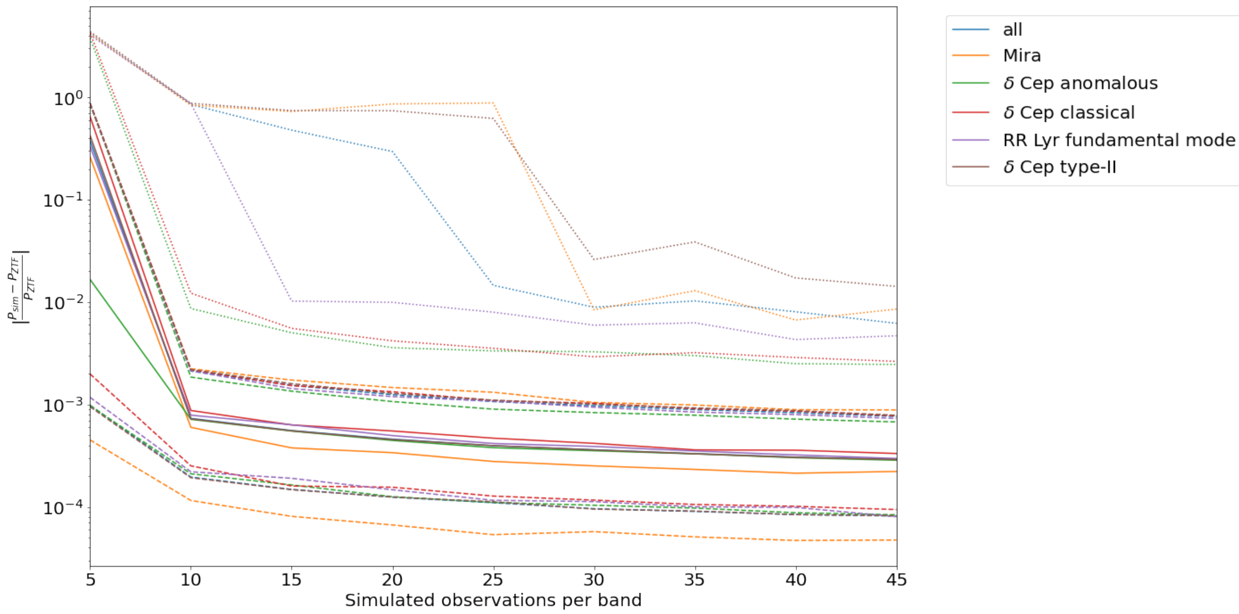

Figure 2: Relative error of the simulated variable star periods as a function of the number of simulated observations per band. Solid lines were computed using the the median of the log-relative error statistic, dashed lines show the corresponding 1σ percentile interval, while the dotted lines show the 99.7th percentile (3σ percentile).

[A4] Increasing the number of observations in redder filters (at the expense of bluer filters) in extincted regions would be beneficial to the determination of periods. The main line of reasoning for this conclusion is that the observations in bluer filters would be effectively lost in extincted regions (sources would be too faint). This would have a negative impact on period determination, as the main driver for successful period recovery is the number of epochs.

[A5] It is unlikely that this strategy would have a significant effect on the period discovery. We do note, however, that built-in uniformity in the survey cadence will have a very negative impact on the period recovery as it will create strong peaks in the window function that will be projected in the periodogram.

[A6] Rolling cadences should generally be preferred as they allow recovery of periods on varying timescales which is particularly important for short-period variability and sources with a low duty cycle such as eclipsing binary stars. Given the length of the survey and the projected total number of observations it is unlikely that rolling cadences would have a detrimental effect on long period variables.

[A7] Dithering patterns are not relevant for this science case.

References

VanderPlas, J. T., & Ivezić, Ž. 2015, ApJ, 812, 18