The Gaia-ASAS-SN Classical Cepheid Sample. I. Sample Selection

Laura Inno1,5, Hans-Walter Rix2, K. Z. Stanek3,6, T. Jayasinghe3, E. Poggio1,7, R. Drimmel1, and A. Rotundi1,6

1 Science and Technology Department, Parthenope University of Naples, Naples, Italy; laura.inno@uniparthenope.it
2 Max-Planck-Institut für Astronomie, 69117, Heidelberg, Germany
3 Department of Astronomy, The Ohio State University, 140 West 18th Avenue, Columbus, OH 43210, USA
4 Osservatorio Astrofisico di Torino, Istituto Nazionale di Astrofisica (INAF), I-10025 Pino Torinese, Italy

Received 2020 July 29; revised 2021 April 10; accepted 2021 April 12; published 2021 June 23

Abstract

We present a well-defined and characterized all-sky sample of classical Cepheids in the Milky Way, obtained by combining two time-domain all-sky surveys: Gaia DR2 (Gaia Collaboration et al. 2018) and All-Sky Automated Survey for Supernovae (ASAS-SN; Shappee et al. 2014). We first use parallax and variability information from Gaia to select 30,000 bright (G < 17) Cepheid candidates with $M_K < -1$. Then, we analyze their ASAS-SN V-band light curves, determining periods and classifying the light curves using their Fourier parameters. This results in 1900 likely Galactic Cepheids, which we estimate to be $\geq 90\%$ complete and pure within our adopted selection criteria. This is the largest all-sky sample of Milky Way Cepheids that has such a well-characterized selection function, needed for population modeling and for systematic spectroscopic follow-up foreseen with Sloan Digital Sky Survey-V. About 130 of these potential Cepheids have not been documented in the literature even as possible candidates.

 Unified Astronomy Thesaurus concepts: Cepheid variable stars (218); Milky Way Galaxy (1054)

Supporting material: extended figure, machine-readable tables

1. Introduction

Classical Cepheids are fundamental calibrators of the cosmic distance scale. They are also near-ideal tracers of the young Milky Way (<300 Myr), as luminosities, distances, masses, and ages can be inferred from their light curves. Despite this fundamental importance, classical Cepheids have been poorly mapped in our Galaxy until very recently, with only ~600 identified over the past decades (e.g., Laney & Stobie 1992; Berdnikov & Turner 2004; Jayasinghe et al. 2018, hereinafter J18; Matsunaga et al. 2018, and references therein). The advent of automated all-sky, time-domain surveys over the past years have brought about a quite dramatic change: new catalogs based on surveys such as the Asteroid Terrestrial-impact Last Alert System (ATLAS), All-Sky Automated Survey for Supernovae (ASAS-SN), Wide-field Infrared Survey Explorer (WISE), and the Zwicky Transient Facility (ZTF) have been released, resulting in the identification of about 500 (Heinze et al. 2018; Udalski et al. 2018, hereinafter U18), 200 (J18), 1300 (Chen et al. 2019, hereinafter Chen19 and 429 (Chen et al. 2020, hereinafter Chen20) new potential classical Cepheids. Smaller, dedicated surveys can offer higher sample completeness and purity, especially in highly extincted regions at low Galactic latitude: for example, the near-IR Visible and Infrared Survey Telescope for Astronomy (VISTA) survey Via Lactea reported the discovery of 200 Cepheids beyond the Galactic center (Dekány et al. 2015a, 2015b, 2019); and the InfraRed Survey Facility (IRSF)-SIRIUS survey identified 25 obscured Cepheids in the inner Galaxy (Matsunaga et al. 2015; Tanioka et al. 2017; Inno et al. 2019).

Gaia DR2 (Gaia Collaboration et al. 2018) led to a catalog of variable stars with time-series photometry, which includes ~1200 classical Cepheids within a Galactic latitude of $\pm 20^\circ$ (Clementini et al. 2019). Recently, Ripepi et al. (2019, hereinafter R19) have presented a new classification of all variable sources in the Gaia DR2 catalog, selecting a cleaner sample of ~680 fundamental mode (FU) and 147 first overtone mode (10V) Cepheids. Among those, ~250 had not been reported earlier. U18 and S20 recently released the Optical Gravitational Lensing Experiment (OGLE-IV) Collection of Variable Stars (OGLE-IV CVS), which includes about 1800 newly identified classical Cepheids. On the basis of this sample, and additional photometric data from the ATLAS and ASAS-SN surveys, Skowron et al. (2019a, 2019b, hereinafter S19b) were able to accurately map the structure of the young Galactic disk in 3D for the first time: they determined some structural parameters of the Galactic disk, such as the scale height, the flaring, and the warp, between a range of Galactocentric radii between 3 and 20 kpc.

One aspect common to all these catalogs is that they do not provide a quantitative characterization of their selection function, such as the directions and distance at which Cepheids have a quantifiable probability of entering the sample, as a function of their luminosity, with its implied mass and age, their variability amplitude, effective temperature (or color), and light-curve shape, given also the presence of possible dust reddening. Yet, this is an indispensable basis for any inference about the Cepheid’s spatial and age distribution and also their number density in the Galaxy.

In this work we set out to identify a sample of classical Galactic Cepheids that is (a) all-sky, (b) extensive (~2000 objects), and (c) has a well-defined, consistent, and reproducible selection function. We do this by combining data from two available all-sky optical time-domain surveys, Gaia and ASAS-SN. Specifically, we lay out a clear algorithmic method to identify an initial set of Cepheid candidates through a query of the Gaia DR2 main catalog that combines rms variability,
color variability, luminosity, and intrinsic color. We then analyze ASAS-SN time-resolved photometry (Shappee et al. 2014) to identify and classify the likely classical Cepheids through a Fourier analysis of their light curves. We consciously did not choose the largest possible classical Cepheid catalog as our goal, and consciously left out known Cepheids if they did not fulfill our selection criteria. This catalog will be used by Sloan Digital Sky Survey (SDSS)-V for systematic, all-sky follow-up of Galactic disk Cepheids.

The method is described in Section 2, while the characterization of the output catalog in terms of completeness and purity, and its comparison to catalogs in the literature, is provided in Sections 3–5, respectively.

2. The Selection of Galactic Classical Cepheids from Gaia DR2 Data and ASAS-SN Time Series

Our goal for this paper is to identify an extensive, large all-sky sample of Galactic classical Cepheids with a well-characterized selection function, and modest or small contamination. This is needed for systematic multi-epoch spectroscopic follow-up, which has just been initiated with SDSS-V (Kollmeier et al. 2017).

We do this by developing a selection function for application to Gaia DR2 × ASAS-SN data, which encapsulates most of the basic properties that could plausibly be used to identify Cepheids:

1. They are luminous, with absolute $K_S$-band magnitude of $M_K < -1$.
2. Their optical light curves vary periodically, with periods of days to about 2 months, and with a range of (optical or $G$ band) peak-to-peak amplitudes around a characteristic value of 0.5 mag.
3. They are pulsating variables, where luminosity changes go along with changes in the effective photosphere temperatures: they vary in color, as they vary in magnitude.
4. They are young ($\tau \lesssim 250$ Myr) and should move on disk-like orbits.
5. Their light-curve shapes depend on their period and on the nature of their dominant pulsation mode (FU or 1OV), which can be compactly characterized by a Fourier expansion (see, e.g., Simon & Lee 1981; Antonello & Poretti 1986; Mantegazza & Poretti 1992; Deb & Singh 2009; Inno et al. 2015).

We implement this by first applying a subset of these criteria to the Gaia DR2 data, resulting in an extensive, but highly contaminated, candidate list, and then analyzing the candidate’s ASAS-SN light curves for identification and classification.

2.1. Initial Candidate Selection from Gaia DR2

The set of Cepheid candidates across the entire sky can be obtained straightforwardly by a single query of the Gaia data archive. We now describe the rationale behind each element of this query.

The selection function should reflect that Cepheids are luminous, for which we adopt $M_K^\text{lim} = -1$. But for many candidates the Gaia parallaxes will be marginal. Therefore, we require that $M_K^\text{lim} < -1$, even if the true parallax were 1σ larger than the best estimate. This leads to

$$\sigma_\varpi < 10^{0.2(10.0 - \ln_8 - M_K^\text{lim})},$$

where $n_8 \equiv \frac{\text{phot}_g\_\text{mean} \_\text{mag} \cdot 1.72(\text{bp}_r) + 0.07(\text{bp}_r)}{\text{phot}_g\_\text{mean} \_\text{flux} \_\text{error}}$ is an empirically determined prediction of $m_K$, based on Gaia colors alone; this approximation appears good to 0.1 mag even once optical extinction values reach several magnitudes. This effects a reddening-insensitive estimate of the apparent and ultimately absolute magnitude, yet avoids involving a second catalog (e.g., Two Micron All Sky Survey) in the selection.

While Gaia DR2 is generally not a time-domain data release, limited information about variability is encoded in the data products. Specifically, we use $\Delta G$

$$\Delta G \equiv \sqrt{\frac{\text{phot}_g\_\text{mean} \_\text{flux} \_\text{error}}{\text{phot}_g\_\text{mean} \_\text{flux}}}$$

as a proxy for rms variability in the Gaia $G$ band. This variability measure was proposed by Deason et al. (2017) to identify RR Lyrae. One can also construct analogous quantities $\Delta_B P$ and $\Delta_R P$ to characterize the variability in the $B$ and $R P$ bands. These quantities describe the cross-epoch variance among the individual magnitude determinations that were used in DR2. They reflect the combination of magnitude measurement uncertainties and of the intrinsic variability. For $G < 17.5$, the photon-noise contribution to $\Delta G$ is $\lesssim 0.03$ mag, as can be demonstrated from the plethora of non-varying sources. Substantially larger values $\Delta G$ must then reflect intrinsic variability. This variance is typically determined from $\gtrsim 20$ epochs across 18 months; for most Cepheids it may therefore reflect a good approximation to the rms light-curve variability. For randomly sampled sinusoidal light curves, $\Delta G$ is $1/4$ of the peak-to-peak variation. So, our adopted selection of $\Delta G \gtrsim 0.06$ will select Cepheid light curves with peak-to-peak amplitudes $\gtrsim 0.24$ mag.

It appears that $\Delta G$ is a very effective way to select variable sources in Gaia. But of course, such a crude rms measure cannot distinguish between light curves that vary because of pulsation, eclipses, secular changes, or other reasons. Among luminous stars, eclipsing binaries and long-period variables are the main contaminants. But we can now exploit the fact that the effective temperatures of Cepheids, typically 5500–6800 K, vary during the pulsation (Fukue et al. 2015); that means that the variations in a bluer (optical) passband should be higher. The quantity $\Delta_r \frac{\Delta P}{\Delta r}$ characterizes how much more sources vary in the blue versus the red. In Figure 1, we show $\Delta_r \frac{\Delta P}{\Delta r}$ for an extensive set of luminous variable objects. Figure 1 shows therefore that the distribution of $\Delta_r \frac{\Delta P}{\Delta r}$ is remarkably tri-modal and that the three different groups (in $\Delta_r \frac{\Delta P}{\Delta r}$) map very well onto the independent classification by J18: eclipsing systems, where the flux variation arises from viewing geometry variations, show very little color dependence and $\Delta_r \frac{\Delta P}{\Delta r} \approx 1$; luminous, long-period variables (usually very red) near the top of the giant branch show distinctly stronger variations in BP. Bluer pulsating stars, including Cepheids, lie in between at $\Delta_r \frac{\Delta P}{\Delta r} \approx 1.4$. This shows that $\Delta_r \frac{\Delta P}{\Delta r}$ can be very effective in eliminating contaminants, and we adopt $1.15 < \Delta_r \frac{\Delta P}{\Delta r} < 1.85$ for candidate selection.
Finally, we select Cepheid candidates to be at low Galactic latitudes, $|b| < 20^\circ$ (property “4,” above) and to have $G < 17$; the latter selection criterion is motivated by the availability of useful ASAS-SN light curves for subsequent classification.

In addition, we apply an astrometric data quality, $\sqrt{\text{astrometric\_chi2\_al}/(\text{astrometric\_n\_good\_obs\_al} - 5)} < 2.5$, as recommended by Gaia DR2. Note that this is the only selection criterion that is not immediately related to a physical property of the objects themselves.

The Gaia DR2 SQL query encapsulating this selection function is listed in Appendix A.

2.2. Identification of Classical Cepheid from ASAS-SN Light Curves

The above selection criteria defined a pool of 29,737 Cepheid candidates. As the initial selection neither assures that the light-curve variations are periodic, nor that they have the shape expected for Cepheids, we may expect that the large majority of these candidates are contaminants. We then retrieved multi-epoch $V$-band photometry from the ASAS-SN automatic survey (Shappee et al. 2014), down to a limiting magnitude of $V \lesssim 18$ mag, extracted as described in J18 using image subtraction (Alard & Lupton 1998) and aperture photometry on the subtracted images with a 2 pixel radius aperture. We corrected the zero-point offsets between the different cameras as described in J18. The photometric errors were recalculated as described in Jayasinghe et al. (2019), hereinafter J19 and for individual epochs they range from 0.02 mag at $V = 11$ mag to 0.16 mag at $V = 18$ mag.

A fraction of $\lesssim 6\%$ of candidates had to be excluded from subsequent analysis because their light curves had less than 40 photometric measurements, excluding upper limits.

We thus analyzed the light curves of $\sim 30,000$ candidates to see which of them matched expectations for light-curve shapes of classical Cepheids, first finding their periods then analyzing the shape of their folded light curves. We started out with two independent period-finding analyses: a Lomb–Scargle (LS; Lomb 1976; Scargle 1982) periodogram, with light curves folded by peak period of the power spectrum; and we used a direct template-fitting approach to find the best period, described in Inno et al. (2015) and Kains et al. (2019) (further details on the fitting procedure, the code, and some examples are given in Appendix B and on the GitHub repository9). The main difference between the LS approach and the template approach is that the former assumes a generic sinusoidal shape for the light curve, while the latter uses empirical light curves from observed Cepheids, and because it makes a stricter assumption on the light-curve shape, can lead to wrong results for different kinds of variable sources. Thus, the template-fitting technique will give correct periods for light-curve shapes similar to the ones of Cepheids and incorrect results for other kinds of variables or noisy light curves.

The difference between the most likely periods found with these two approaches was larger than 10% for 49% of the cases. In particular, periods of 0.5, 1, 2, and 29.4 days were found for 20% of the objects, presumably due to aliasing. Such a high fraction of discrepant period estimates need not be surprising or disconcerting, as we do not have direct a priori information whether the variability is even (short term, <1 yr) periodic. By removing all candidates with discrepant period estimates from the initial candidate list, we end up with a pool of $\sim 16,000$ candidates with presumably periodic light curves and with typically $230 \pm 100$ (standard deviation) ASAS-SN epochs.

Among them, classical Cepheids should form a distinctive family of light-curve shapes. We characterized these shapes by fitting the light curves with a seventh-order Fourier series, assuming the best period from the template-fitting method, which is expected to be more robust against possible outliers.

All the ASAS-SN folded light curves for the Cepheids we selected are shown in the electronic version of Figure 9, along with their Fourier-domain representation. The results of such an analysis can be usefully quantified and illustrated by the ratio of the second and third Fourier amplitude to that of the first, $R_{23}$ and $R_{31}$, along with that the corresponding phase differences, $\phi_{23}$ and $\phi_{31}$.

9 https://github.com/laurainno

Figure 1. Selection of the Cepheid candidates by their Gaia variability and color variability. The figure shows the $G$-band variability (see Equation (2)) vs. the color-dependent variability $\Delta G$. This separates quite neatly eclipsing binaries, pulsators, and cool, luminous giants, as indicated by the color coding according to the classification provided by J18. We use both $\Delta G$ and $\Delta G$ in selecting Cepheid candidates, and specifically the range indicated by the dashed region.
Figure 2. Fourier-domain characterization of ASAS-SN photometry light curves for the candidate Cepheids, selected from Gaia DR2 (see Section 2.1). The four panels show as a function of the pulsation period: the ratio of the second (top left) and third (top right) Fourier amplitude to the first amplitude and the corresponding phase shifts between these Fourier modes in the bottom panels. The dashed lines illustrate the Cepheid-selection criteria in this parameter space that we devised and adopted to select classical Cepheids pulsating in their FU (magenta) and their 1OV (cyan). The Cepheids selected thereby are shown as cyan (FU), and magenta (1OV) dots; the open symbols indicate sources in common with the OGLE catalogs (U18; S20), as denoted by the inset. Note that the distinct clumps of objects at periods shorter than 1 day are RR Lyrae variables (∼5000 already known and ∼2000 new objects), which are likely to be the more abundant contaminants in our selection for short-period Cepheids.

Table 1
Selection of Classical Cepheids in Light-curve Shape Space

| Fundamental Pulsators | 1OV Pulsators |
|-----------------------|---------------|
| 1.0 ≤ P ≤ 70 (P ≥ 20 AND R21 ≤ 0.6) OR ((10 ≤ P ≤ 20) AND (R21 ≤ 0.4)) OR ((P < 10) AND (R21 ≤ 0.8)) (R21 ≥ 0.5 log P − 0.6)) AND (R31 ≥ (0.4 − 0.4 log P)) | 0.35 ≤ P ≤ 7.5 (P ≥ 4 AND R21 ≤ 0.25) OR (P < 4 AND R21 ≤ (0.45 − 0.45 log P)) |
| R31 ≤ 0.5 AND (P > 31 AND R31 ≥ 0.06) OR P ≤ 31 | P > 0.8 OR (P ≤ 0.8 AND R31 ≤ 0.3) |
| (15 ≤ P ≤ 30) AND (R31 ≥ 0.08) OR P > 30 OR P < 15 | R31 ≤ 0.25 AND ϕ21 ≤ 6.2 |
| (P < 9 AND ϕ21 > 3.7) OR P > 9 | P > 0.7 OR (P ≤ 0.7 AND ϕ21 ≤ 3.8 AND ϕ21 > 2.7) |
| (P < 7 OR P > 12) AND ϕ21 > 5.6 OR 7 < P < 12 | P > 1.2 OR (P < 1.2 AND ϕ21 ≤ 4.8) |
| (P > 12 AND ϕ31 > 4.0) OR P < 12 | |
| (P ≤ 13 AND ϕ31 < 5.0 AND ϕ31 > 2.0) OR P > 13 | |

Note.

* Period measured in days.

The distribution of the resulting Fourier parameters are shown in Figure 2 (gray crosses) and form a distinctly multimodal distributions. Different parts of this distribution separate classical Cepheids from other variables, in particular, from RR Lyrae. And among classical Cepheids these diagrams neatly separate those oscillating in the FU from those in the 1OV, as the pale green and lavender symbols classified by U18 and S20 from OGLE I-band light curves show.

This diagram allows us now to identify potential FU Cepheids (dark purple) and 1OV Cepheids (dark green) among the totality of our candidates (gray dots). We formalize this classification by the cuts shown in Figure 2 and listed in detail in Table 1.
Table 2

Gaia DR2 ID and Parameters from our Analysis of all Cepheids in the I21 Sample

| Gaia DR2 source_id | $P_T^a$ (days) | $P_T^b$ (days) | $N_{epochs}$ | $< V >$ (mag) | $A_1$ | $R_{21}$ | $R_{31}$ | $\Delta$, $\phi_21$ | $\phi_31$ | $\Delta\phi$ (mag) | Mode |
|-------------------|----------------|---------------|--------------|--------------|------|--------|--------|----------------|--------|----------------|------|
| 175811260743708672 | 2.1136 | 2.114 | 126 | 13.805 | 0.309 | 0.507 | 0.332 | 4.375 | 2.615 | 0.247 | F |
| 180464393957209504 | 1.08318 | 1.083 | 381 | 14.295 | 0.200 | 0.224 | 0.107 | 4.501 | 2.551 | 0.121 | 1.525 | IO |
| 181626085307496192 | 1.66809 | 1.668 | 122 | 14.345 | 0.162 | 0.525 | 0.344 | 4.359 | 2.481 | 0.170 | 1.590 | F |
| 182682513979002528 | 3.33083 | 3.331 | 316 | 12.921 | 0.141 | 0.078 | 0.039 | 3.594 | 1.235 | 0.068 | 1.532 | IO |
| 186678245267045376 | 2.41531 | 2.415 | 153 | 14.828 | 0.174 | 0.524 | 0.277 | 4.186 | 2.478 | 0.129 | 1.615 | F |
| 187322782037490048 | 9.40180 | 4.910 | 153 | 13.877 | 0.223 | 0.418 | 0.178 | 4.963 | 3.42 | 0.129 | 1.443 | F |
| 18872423459584256 | 18.19102 | 18.21 | 256 | 12.061 | 0.360 | 0.165 | 0.122 | 4.942 | 2.124 | 0.185 | 1.489 | F |
| 189726984845704080 | 5.2599 | 5.261 | 189 | 12.478 | 0.351 | 0.455 | 0.205 | 4.882 | 3.277 | 0.187 | 1.482 | F |
| 190508157519690112 | 1.82804 | 1.828 | 266 | 12.548 | 0.206 | 0.215 | 0.075 | 4.754 | 3.579 | 0.111 | 1.656 | IO |
| 193737217573544464 | 3.85858 | 3.859 | 163 | 9.761 | 0.276 | 0.327 | 0.204 | 4.711 | 2.834 | 0.122 | 1.274 | F |
| 197337185858157440 | 4.40485 | 4.405 | 267 | 12.123 | 0.340 | 0.389 | 0.17 | 4.613 | 2.75 | 0.222 | 1.756 | F |
| 1998681686717279616 | 2.65975 | 2.680 | 164 | 13.541 | 0.293 | 0.495 | 0.266 | 4.579 | 2.758 | 0.147 | 1.444 | F |

Notes.

a Period obtained with the LS method.
b Period obtained with the template-fitting procedure.

Remarkably, our well-defined, all-sky preselection of bright ($G < 17$) candidates, followed by an algorithmic light-curve classification results in a very large sample: a total of $\sim1000$ potential classical Cepheids, of which 1269 are FU and 622 are 10V (note that we do not classify mixed mode pulsators). About 500 of them are new identifications (of which 25% without any previous record). We provide the Gaia source_id, the Gaia coordinates, and all the parameters obtained from our analysis for this sample in Table 2, which is available in its entirety in the online version of the paper.

For the rest of the paper, we refer to this sample as I21.

3. Completeness of the I21 Sample

As mentioned in the Introduction, we deem understanding the selection function, or completeness and purity of this I21 sample paramount.

We have reason to believe that the purity of the sample is high ($\geq\%80$; see Section 5.3), or that the contamination of the sample is relatively low, as the light-curve shapes are of high discriminating value, and as the Gaia parallaxes are effective at eliminating low-luminosity (or more nearby) pulsators as contaminants; we will therefore focus first on completeness here.

Before detailing this aspect, it is worth taking a step back to spell out an operative definition of completeness. Most efforts to define a sample of astronomical sources set out to find objects within a well-defined set of criteria. Such samples should not contain objects well outside of these criteria, and it would make little sense to attribute their absence to incompleteness. For example, the U18 sample contains no OGLE Cepheids in parts of the sky where OGLE did not observe. But this only implies that the global census of Cepheids in the Galaxy is incomplete, not that the U18 sample is highly incomplete.

To explore the likely completeness of I21, we take S19b as the comparison; ideally, one would like to have a gold standard sample with 100% completeness and 0% contamination, which is not practicable, but we take S19b as an approximation to that.

We illustrate this comparison with a Venn diagram (Venn 1881), shown in Figure 3, which provides a structured way of answering: How many of the Cepheids in S19b are not in I21 and why?
Figure 3. Venn diagram for the Cepheids in the S19b catalog that were not selected by the two-step selection process of I21: the initial Gaia DR2 query to obtain candidates, followed by the light-curve analysis. Among the criteria considered in GDR2 candidate selection are the $G$-band magnitude range (blue), variability amplitude (lilac), the $K$-band absolute magnitude (rose), and the astrometric data quality (yellow). The quality cuts and requirements in the light-curve analysis are indicated in pale orange; the sub-aspects of the light-curve analysis are specified in the right panel. They denote the reasons for not being included in I21: no photometry available from the ASAS-SN survey, different periods found with the two period-finding methods or period found beyond the chosen limits (0.3–70 days), manifest evidence of period aliasing, and failing to pass our cuts for the light-curve shape Fourier analysis. As discussed in Section 3, completeness must be in reference to the objects we set out to identify. Therefore, objects in S19b that are outside the candidate search parameters of I21 do not contribute to I21s incompleteness. What contributes, are objects missing because of data quality aspects: the dominant factor is the light-curve analysis being the main factor.

Table 3

| S19b Sample (S19b) | ASAS-SN Sample (J18, J19) |
|--------------------|---------------------------|
| All Cepheids with Gaia DR2 entries and within $|b| \leq 20^\circ$ | 2485 | 1035 |
| $G$-band magnitude: $6 \leq G < 17$ | 2056 | 1033 |
| Absolute $K$-band magnitude: $\varpi + \sigma_0 < 10^{16.15 + 2.5 \cdot (M - M_{lim})}$ | 1907 | 1023 |
| Variability: $0.06 \text{mag} < \Delta \varpi < 2.5 \text{mag} & 1.15 < \frac{\text{astrometric}_\text{chi2}_\text{al}}{\text{astrometric}_\text{n_good}_\text{obs}_\text{al} - 5} < 1.85$ | 1607 | 852 |
| Astrometric quality,$\sqrt{\frac{\text{astrometric}_\text{chi2}_\text{al}}{\text{astrometric}_\text{n_good}_\text{obs}_\text{al} - 5})} < 2.5$ | 1811 | 990 |
| All Gaia DR2 cuts | 1429 | 827 |
| ASAS-SN photometric time series analyzed | 1410 | 824 |
| Period from LS and template fitting within 10% | 1212 | 815 |
| Fourier parameters within selected ranges in Table 1 | 1112 | 743 |
| Relative completeness of I21 | 79% | 90% |

At first glance this comparison may seem sobering with respect to completeness: of the 2483 objects in S19b only 1112 are also in I21, and 1371 of them seem to be missing. But Table 3 and Figure 3 explain why that is: 1054 of the 1371 are not in I21 because they do not satisfy one or more of I21’s initial selection criteria: the $G$-band magnitude range (blue), variability amplitude (lilac), the $K$-band absolute magnitude (rose), and the astrometric data quality (yellow), with the colors referring to the areas of the Venn diagram in Figure 3. So, only $\leq60%$ of these S19b Cepheids have intrinsic astrophysical properties that I21 was looking for. However, 317 of the S19b Cepheids are not in I21, even though their physical properties place them within I21’s selection criteria. Almost all of them are missing because they were excised during the light-curve analysis (see Section 2.2); they were eliminated for a number of reasons, detailed in the right panel of Figure 3, with inconsistent period estimates among our two approaches as the leading reason.

These missing 317 Cepheids constitute true incompleteness. If we were to presume that the S19b sample is perfectly complete and pure, this implies a completeness for the I21 sample of $\sim78\%$ (1112 of 1429 objects). If the S19b sample has any contamination, the completeness may increase slightly.

In the next section we address instead the question of the purity by comparing the I21 sample with both S19b and the pre-existent classification made on the basis of the same time-series data, i.e., the ASAS-SN catalog of variable stars (J18, J19).

4. Purity of the I21 Sample

Since we used photometric time-series data from the ASAS-SN survey, it is only natural to compare our new classification with the one provided by the survey (and available at asas-sn.osu.edu/variables) for the objects in common. The overlap between the two selections include only 71% of our new sample, namely, 1354 objects, of which 55% have been classified also as classical Cepheids by J18 and J19. This leaves us with 610 candidate Cepheids in the I21 with a different classification in the ASAS-SN database. However, among those, 314 are indeed already classified as Cepheids by S19b or by S20, and therefore we will not discuss further this subset on the assumption that the OGLE
sample is our best approximation of a gold standard, and it does not include (or includes little) contamination.

We therefore end up with only 296 objects with a previous different classification, which would lead to a contamination of $\sim 15\%$. However, we can determine that purity of the I21 sample is actually significantly higher.

In fact, we find that the pulsation period reported in the ASAS-SN database is different for 36\% of the objects in common (for half of which the ASAS-SN period being double the period we found), and a wrongful determination of the period can easily lead to a misclassification. By looking at the folded light curves, we verified that the new period is indeed the correct one for all the cases. However, we find that the objects classified as nonperiodic in ASAS-SN, show noisy light curves and therefore we consider them as possibly misclassified objects (21). For the remaining ones (including the 188 with the same period), we measure the skewness and kurtosis of their light-curve shapes, in order to quantify how fast is the rising branch with respect to the decreasing branch, and then visually inspected all light curves individually.

By examining both these parameters’s distribution and the light curves by eye, we were able to confirm our new classification for 78\% of the cases, with contaminants being mostly Type 2 Cepheids and eclipsing binaries (the latter could be potentially removed by a clean cut on skewness $\leq 0.25$).

Therefore, the contaminants amounts to 60 objects and hence the purity of the overlapping sample is $\geq 90\%$, and the relative completeness within our selection criteria is 90\%.

Finally, we compare the I21 overlap with the S19b sample and with the ASAS-SN sample to assess their relative purity (933 objects). We find that 33\% of the objects in common between the three samples are classified as Cepheids by I21 and S19b but as different variables by ASAS-SN. We also searched the OGLE database for a match with the 641 objects in I21 with an ASAS-SN different classification: among the 167 matches, 54 objects were classified differently also in OGLE. Instead, by comparing the objects classified as Cepheids in both the ASAS-SN and I21 with the OGLE entire database, we found only 10 objects classified differently.

In the next section, we address the purity and completeness of the I21 sample into the context of all Galactic Cepheid samples, many of which have some extensive overlap with I21.

Nonetheless, a remarkable 126 potential Cepheids in our I21 sample are without prior published record, as we will show in the next discussion.

5. The I21 Sample in Context: Comparison with Published Samples

In this section we put the I21 sample into the context of all previously published samples, and verify if the relative completeness and purity discussed above remains consistent. The basic motivation behind this step is that, as mentioned before, there is not a gold standard for the classical Cepheid sample in the Milky Way to date, and while we believe S19b can be adopted as a good approximation, we want to check if the completeness that we estimated for the I21 sample ($\sim 78\%$) would either decrease or increase when considering all the public samples available. Also, we want to determine how many of the Cepheids we identified are truly new discoveries.

We address the first objective in Sections 5.2 and 5.3, and the second one in Section 5.4, but first we need to compile a master catalog from the available public databases to be used as the reference catalog, as we do in the following Section.

5.1. The Reference Catalog of Galactic Classical Cepheids Compiled from the Literature

In order to include all the classical Cepheids (or published candidates) currently known in the Milky Way, we assembled a master catalog from the following public databases:

1. OGLE: The OGLE Collection of Galactic Cepheids, as published by U18, which includes 1,488 detection from OGLE-III and OGLE-IV (U18), plus 300 recently added by S20, plus the additional 1,142 objects collected by Skowron et al. (2019a) and S19b from the following catalogs: GCVS (702), ATLAS (168), ASAS (106), ASAS-SN (115), WISE (nine), the VSX11 (seven), and specific papers: Matsunaga et al. (2016) and Tanioka et al. (2017);

2. Gaia: The Gaia Catalog of classical Cepheids, as recently reclassified by R19, which includes 684 stars (520 FU, 147 10V, 17 multimode);
3. ASAS-SN: The ASAS-SN catalog of classical Cepheids cross matched to Gaia DR2, which includes 2126 members, of which 1044 are within $|b| \leq 20^\circ$ (812 FU and 232 1OV);

4. SIMBAD: $\sim$635 objects classified as *deltaCep* in the SIMBAD database;

5. WISE: $\sim$2100 Cepheids, including 779 new candidates identified from the WISE time-series data by Chen19, all assumed FU pulsators.

6. ZTF: $\sim$1262 Cepheids, including 451 new candidates identified from the ZTF Survey by Chen20. Chen20 all assumed FU pulsators.

7. VSX: $\sim$238 star out of the 3143 classified as DCEP, DEPCS, DCEP, DCEP(B), and DCEPS(B) in the VSX database within $20^\circ$ from the Galactic plane but not already included in the previous catalogs.

The membership among almost all these catalogs has some, often substantial, overlap. But there is no published sample that mostly encompasses any of the other samples listed. This is a somewhat bewildering situation, which we try to illustrate in a Venn diagram, shown in Figure 4. This figure shows that there are, e.g., $\geq$1100 Cepheids that are only in the OGLE sample, $\geq$1000 that are only in the WISE sample, $\geq$700 that are only in the ASAS-SN sample, and $\geq$1000 that are only in the SIMBAD, VSX, or ZTF samples (labeled as “others”). It is beyond the scope of the present paper to sort out and discuss comprehensively the different overlap fractions in these catalogs.

But Figure 4 can serve to illustrate an elementary but central point in the present context: the completeness (or purity) can only be sensibly discussed when the selection criteria are clearly stated; and then completeness should refer to the fraction of objects included, whose properties fall within the stated selection criteria.

For example: a sample based purely on OGLE data will obviously be 100% incomplete in parts of the sky not covered by OGLE. Similarly, a Gaia based Cepheid catalog will obviously be 100% incomplete in objects too reddened to be visible in the optical; any magnitude-limited catalog will be—by construction—fully incomplete in objects of fainter magnitudes, etc. Therefore completeness is only a useful concept when phrased (implicitly or explicitly) as complete within the stated selection criteria. It is in this sense that we now put our I21 sample into context.

Thus, we build a comprehensive compilation of Cepheids (or candidates) known in the Milky Way, by removing all duplicated sources, and adopting the pulsation period and mean luminosities (when available) by giving precedence to the catalog order in the above list. Moreover, we also crossmatch the catalog with the OGLE database in order to remove possible contaminants ($54$ objects with a different classification: $30$ in J18, six in R19, nine in Chen19, and nine in the literature) This leads to a final compilation of the literature catalog that includes 4147 classical Cepheids (or published candidates) in the Milky Way (see Table 4), which we will now take as a reference catalog to check the completeness of the new I21 sample as done for the S19b and ASAS-SN catalogs in the previous section.

### Table 4

| Mode   | OGLE | Gaia | ASAS-SN | Lit$^a$ | WISE | ZTF | Gaia-ASAS-SN (I21) | Total$^b$ |
|--------|------|------|---------|--------|------|-----|------------------|----------|
| FU     | 1840 | 520  | 812     | 971    | 2103 | 1238 | 1269             | 3280     |
| 1OV    | 706  | 147  | 232     | 251    | ...  | ...  | ...              | 805      |
| MULTI  | 239  | 17   | ...     | 61     | ...  | ...  | ...              | 249      |

| Subset within Gaia DR2 selection |
|----------------------------------|
| FU/1OV/MULTI                     |
| Initial data set                 |
| FU/1OV/MULTI                     |
| Reference catalog                |
| FU/1OV/MULTI                     |
| Merged catalog including I21     |
| FU/1OV/MULTI                     |
| Overlap with I21                 |
| FU/1OV/MULTI                     |
| Subset within Gaia DR2 selection |
| FU/1OV/MULTI                     |
| Initial catalogs with original classification (duplicated objects included). |
| d We assume FU pulsation for the Cepheids in Chen20. |

Notes.

$^a$ Includes VSX-cCs and SIMBAD-cCs.

$^b$ Without repetition.

$^c$ Initial catalogs with original classification. 

$^d$ We assume FU pulsation for the Cepheids in Chen20.

Figure 5. Venn diagram for the Cepheids in the literature catalog that are not in the I21 sample.
In fact, the I21 sample contains ∼1900 presumed classical Cepheids, of which ∼502 are in no other Cepheids catalog (see right panel of Figure 4). This leaves about 1390 objects overlapping with other catalogs, which is only ∼40% of the ∼2500 objects encompassed by those other catalogs. This could in principle be a consequence of I21 incompleteness or of the contamination of the other catalogs.

In Figure 5, we show a Venn diagram of the objects contained in other catalogs, but not in I21; this could be because objects failed to pass varying combinations of the initial selection or light-curve classification steps in I21. Figure 5 illustrates that not passing the G-band magnitude limit, not varying enough or not being luminous enough are the most common reasons. However, I21 also used some data quality cuts, i.e., criteria that do not refer to observables proprieties of the target objects, in particular, the astrometric data quality. But, Figure 5 illustrates that this eliminates only a small fraction of objects (∼5%). This implies that the I21 sample is highly complete (∼90%) within its stated target selection criteria,

5.2. Completeness of I21 with Respect to the Reference Catalog

In order to better characterize the completeness of the sample in terms of the different stellar parameters included in our selection (luminosity, color, variability etc.), we can study their distribution with respect to the reference catalog. In Figure 6, we show the distribution of the I21 sample as a function of the Gaia G-band magnitude and its relative completeness (right panel) compared to the catalogs mentioned above (top row, see labels). By looking at this figure, we can see that the I21 sample is indeed more complete than the reference catalog (purple line) between 10 and 17 G-mag, while the S19b sample alone is highly complete (90%–100%) in the range brighter than 8 mag and complete to ∼70% for fainter Cepheids.

If we now add the newly identified Gaia-ASAS-SN Cepheids to the reference catalog, creating a final merged catalog of about 4721 Cepheids, of which 2504 within the Gaia DR2 parameter range that we are exploiting (green line in Figure 6), the completeness of the reference catalog drops by ∼20%–40% in the magnitude range between 12 and 17 mag, while the I21 is indeed the most complete.

5.3. Purity of the I21 Sample by Comparison with the Galactic Reference Catalog

As mentioned above, there is a substantial overlap of the I21 sample and previously known catalogs. In this section, we
**Table 5**

Classification of the 121 Cepheids from Other Catalogs

| Source_id   | Source  | Star  | Mode | $P$  | Ref | ASAS-SN Other | ASAS-SN Other Type | OGLE Other | OGLE Other Type |
|-------------|---------|-------|------|------|-----|---------------|-------------------|------------|----------------|
| 153236080001040384 | ASAS-SN | J051348.97 +311429.6 | IO   | 1.0832 |     |                |                   |            |                |
| 180463935927095040 | ASAS-SN | J050149.92 +371627.8 | F    | 2.4154 |     | J050149.92     | 2.41563          | CWB        |                |
| 189726984845004800 | GCVS   | J052240.06 +414302.0 | IO   | 1.828  |     |                |                   |            |                |
| 190501857519900112 | ASAS-SN | J05074.037490048     | F    | 5.107  |     |                |                   |            |                |
| 190537321757345446 | GCVS   | J052240.06 +414302.0 | IO   | 1.828  |     |                |                   |            |                |
| 197337185581574400 | GCVS   | J05074.037490048     | F    | 4.000  |     |                |                   |            |                |
| 198616867172961600 | GCVS   | J05074.037490048     | F    | 4.000  |     |                |                   |            |                |
| 20000061115820797440 | GCVS   | J05074.037490048     | F    | 4.000  |     |                |                   |            |                |
| 20004494970215490560 | ATLAS  | J05074.037490048     | F    | 4.000  |     |                |                   |            |                |
| 200256084290032256 | ZTF    | ZTF044914.01         | U    | 2.1893 |     |                |                   |            |                |
| 202150768065410944 | GCVS   | J052240.06 +414302.0 | IO   | 1.828  |     |                |                   |            |                |
| 203413602509201792 | GCVS   | J052240.06 +414302.0 | F    | 10.0446|     | J051393.18     | NON-P            | VAR        |                |
| 204510644535613184 | Khruslov & Kusakin (2016) | GSC02901-00089 | IO2O | 0.5338 |     | J045941.55     | NON-P            | VAR        |                |
| 204845480185040512 | ZTF    | ZTF044914.01         | U    | 2.6796 |     | J045215.80     | 0.6898208        | RRAB        |                |
| 206181318094220416 | ZTF    | ZTF044914.01         | U    | 1.228  |     | J045215.80     | 0.6898208        | RRAB        |                |
| 206211859601169792 | GCVS   | J05074.037490048     | F    | 3.4072 |     |                |                   |            |                |
| 206577210999441536 | GCVS   | J05074.037490048     | F    | 13.8414 |    |                |                   |            |                |
| Source_id       | Source  | Star        | Mode | P      | Ref   | ASAS-SN Other Name | ASAS-SN P | ASAS-SN Other Type | OGLE Other Name | OGLE Other Type |
|-----------------|---------|-------------|------|--------|-------|--------------------|-----------|--------------------|----------------|----------------|
| 20772317998325632 | ASAS-SN | J052333.55+444536.5 | F    | 2.4769 | S19b  | ...                | ...       | ...                | ...            | ...            |
| 228673477705878400 | CSS     | J041331.0+411907 | F    | 2.3127 | S19b  | ...                | ...       | ...                | ...            | ...            |
| 228741784865956224 | GCVS    | SX Per      | F    | 4.2901 | S19b  | ...                | ...       | ...                | ...            | ...            |
| 229602496313775872 | GCVS    | GP Per      | F    | 2.0416 | S19b  | ...                | ...       | ...                | ...            | ...            |
| 232773865804463232 | ATLAS   | J065.6485+45.5819 | IO   | 2.7754 | S19b  | ...                | ...       | ...                | ...            | ...            |
| 236511827381118592 | ...     | ...         | ...  | ...    | ...   | ...                | ...       | ...                | ...            | ...            |
| 247358200355952256 | ATLAS   | J062.4556+49.7355 | F    | 3.5932 | S19b  | J034523.42        | NON-P     | VAR                | ...            | ...            |
| 248494232082353152 | GCVS    | MM Per      | F    | 4.1184 | S19b  | +480501.2         | ...       | ...                | ...            | ...            |

Notes. Here we indicated only the first 10 columns of the table, while columns: “ZTF Other Name,” “ZTF Other Type,” and “Possible_Contaminant” are missing from the header, but the extended version of the table is fully available in electronic form.

a Source of photometric data.
b Name of the star from literature sources.
c Reference of the catalog from which period and mode are adopted.
d Name for objects nonclassified as Cepheids in ASAS-SN.
e Name for objects nonclassified as Cepheids in OGLE.

(This table is available in its entirety in machine-readable form.)
describe the overlap in detail, in order to provide quantitative estimates of our selection purity.

Among the sources we selected, there are

1. 352 already included in the U18+S20 catalog, and we found the same period for 99% of them. For 14 Cepheids, we identify a different pulsation mode, and by visually inspecting the light curves, we found that our classification seems to be correct for about 50% of the cases.

2. 592 in the Gaia Catalog of Variable Stars, and we found the same period for 99%, but a different classification for 15% of them.

3. 1354 objects in the classification by J18 and J19, of which 744 also classified as classical Cepheids. Since we have already extensively discussed the overlap between these two samples in Section 4, we will only remind the reader here that, on the basis of the light-curve analysis for the objects in common, we estimated a much lower contamination, around the 7%. We also found an additional 1070 objects classified as Cepheids by us and by other literature catalogs, but with a different classification in the ASAS-SN catalog.

4. 965 in the catalog of about 1400 Cepheid candidates identified in the WISE data by Chen19. However, the pulsation period we find is different for 14% of the Cepheids, and we find a different pulsation mode for 13% of them. Moreover, we find that among the 1209 objects in common between the Chen19 sample and the OGLE sample, 13% have a different pulsation period and 13% are also classified with a different pulsation mode by OGLE.

5. 583 objects overlap with the ZTF sample by Chen20, and 82% of them are also classified as classical Cepheids, while 104 objects are classified differently (among which 24 with a different period).

6. 1358 objects with a record in the SIMBAD database, of which 27% classified as non-Cepheids. However, 20% of the objects in the S19b catalog are also classified as non-Cepheids in SIMBAD, meaning that the SIMBAD classification could be out of date with respect to more modern studies of the same variables, and it should not be taken into account when estimating the purity of the sample.

In Table 5, we provide for each potential Cepheid in I21 the name, data source, classification, and classification reference from the literature for the objects that are also classified as Cepheids, while for the objects with also a different classification we provide the classification from the ASAS-SN, OGLE, or ZTF catalogs. Moreover, the 147 objects that we confidently identified as non-Cepheid variables from examining their light curves, are also flagged as possible contaminants (Possible_Contaminants = 1), but we do not flag noisy light curves, which did not allow us to perform a clear identification. As mentioned before, we expect the contamination to be slightly larger (~9%, see the next paragraph).

5.4. New Potential Cepheids in I21

By removing all objects in common with other catalogs of Cepheids, we end up with 502 potentially new Cepheids. Among those, 253 have been classified as different pulsating stars by J18, 99 (including 53 in common with J18) by OGLE (we already discussed these objects in Section 4) and 79 (of which 47 in common with the previous two) by Chen20. Therefore, 502 are potential new Cepheids: 331 with a previous different classification, 126 new discoveries, and 45 with a record either in SIMBAD or the VSX database (of which four as Cepheids).12 In order to validate the new discoveries and new classifications, we...

---

12 Please note that these numbers can change over time, as new entries can be added to the ASAS-SN, SIMBAD, or VSX database at anytime.
we visually inspected the light curves and confirmed our classification for 65% of them, with most of the unconfirmed classifications being either eclipsing binaries and RR Lyrae (flagged as Possible Contaminants in Table 5) or light curves too noisy for providing a good classification. Therefore, the final number of contaminants we were able to identify is 9%, confirming an overall purity higher than 90%.

In summary, the Gaia-ASAS-SN catalog includes a sample of 1891 potential Cepheids that are complete and pure up to 90% (9 mag ≤ G < 17 mag), of which 126 do not have any previous detection in the aforementioned catalogs and 331 previously classified into different variability types. The distribution in galactic coordinates of this sample is shown in Figure 7.

6. Conclusion

We laid out and applied a straightforward approach for assembling a large (~1900), all-sky sample of classical Cepheids in the Milky Way, combining Gaia and ASAS-SN data. Our priority was to arrive at such a sample with a well-defined, reproducible, and modelable selection function (rather than the most extensive compilation of such objects). Our sample, I21, is currently the only one that satisfies such selection-function quality criteria and has a completeness of ~90% (9 mag ≤ G < 17 mag).

A well-defined selection function is particularly important if one wants to use Cepheids as tracers of the young star population across the Milky Way (e.g., Paper II, L. Inno et al. 2021, in preparation). But it is also important when selecting targets for ongoing and future spectroscopic surveys, such as SDSS-V (Kollmeier et al. 2017) or 4MOST. The selection parameters for this sample have been geared specifically toward SDSS-V, which has already started taking data on this sample. In fact, SDSS-V aims to only observe samples of objects with algorithmically quantified selection function, making samples like this indispensable.

We are extremely grateful to C.S. Kochanek for his thoughtful reading of an early version of this manuscript. We thank the anonymous referee for providing valuable comments. L.I. acknowledges the “PON Ricerca e Innovazione: Attraction and International Mobility (AIM)” program for support. This work was partially supported by the Sonderforschungsbereich SFB 881 (subproject A9) of the German Research Foundation (DFG). E.P. acknowledges support from the Centre national d’études spatiales (CNES). This work has made use of data from the European Space Agency (ESA) mission Gaia (https://www.cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC; https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular, the institutions participating in the Gaia Multilateral Agreement.

Appendix A

Gaia Query

SELECT top 500,000 *, sqrt(phot_g_n_obs)/phot_g_mean_flux_over_error as variability FROM gaiadr2.gaia_source where phot_g_mean_mag < 17 and parallax + parallax_error < power(10., (10. - (phot_g_mean_mag - 1.72 * (bp_rp) + 0.07 * (bp_rp - bp rp)) - 1.)) / 5. and sqrt(astrometric_chi2_al/astrometric_n_good_obs_al - 5) < 2.5 and sqrt((phot_g_n_obs)/phot_g_mean_flux_over_error) > 0.06 and (sqrt((phot_bp_n_obs)/phot_bp_mean_flux_over_error)) / (sqrt((phot rp_n_obs)/phot rp_mean_flux_over_error)) < 1.85 and (sqrt((phot_bp_n_obs)/phot_bp_mean_flux_over_error)) / (sqrt((phot rp_n_obs)/phot rp_mean flux_over_error)) > 1.15 and abs(b) < 20

Appendix B

Template Fits and Power Spectrum

B.1. Template-fitting Method

We used the Python packages gatspy13 in order to obtain the LS periodogram (VanderPlas & Ivezić 2015) of the light curves for each source, then we fold the light curve according to each of the three top periods found by the algorithm and finally we use SciPy to perform a fit of the folded light curve by means of an empirically calibrated templates. The procedure then selects the period at which the template fitting produces the lowest residuals (chi-squared minimization). For further details, see Figure 8.

B.2. LS-based Method

We used the Python package Astropy (Astropy Collaboration et al. 2013, 2018) in order to obtain a LS periodogram (we adopted the Nyquist factor = 300) and then identified the period corresponding to the frequency of the peak. For further details, see Figure 8.

13 https://zenodo.org/badge/DOI/10.5281/zenodo.14833.svg
Appendix C

Light Curve Atlas

Figure 8. Output of our light-curve fitting code for four different objects with Gaia-ID as labeled in the top of each plot. For each object in each of the panels ((a)–(d)), we show the output form the template-fitting procedure (left) and from the application of the LS periodogram (right). The light curves shown in the top subpanel is folded according to best value of the period found by each procedure. In the case of the template fitting, this is done by performing a chi-squared minimization of the data with the respect to the template curve (green dashed line) for values of the period shown by the orange line in the power spectrum shown below. The LC folded accordingly, is then fitted with a Fourier series (blue dashed line). The second method shown in the left is simply based on finding the period corresponding to the highest peak in the LS periodogram shown below (cyan line, frequency space). Cases (a) and (b) show two examples of objects for which the two methods obtain the same value for the period. Case (c) shows an example where the template-fitting method finds the correct period despite the presence of outliers, while the simple periodogram finds aliasing around 1 day. Case (d) shows a case in which the template fitting finds aliasing, while the LS methods finds the correct period.
Figure 9. Period-folded ASAS-SN light curves of Cepheids in our I21 sample sorted with decreasing period. The red lines show the seventh-order Fourier fit, whose coefficients are used in the light-curve classification (see Figure 2), while the blue dashed line shows the best-fitted template. The reference catalog in which each object is also classified as a Cepheid is given in the label: Ref. If a different classification is also available from other catalogs (i.e., S19b; S20; Chen19; ZTF), it is also reported in the label.

(An extended version of this figure is available.)
ORCID iDs

Laura Inno @ https://orcid.org/0000-0002-0271-2664
Hans-Walter Rix @ https://orcid.org/0000-0003-4996-9069
T. Jayasinghe @ https://orcid.org/0000-0002-6244-477X
E. Poggio @ https://orcid.org/0000-0003-3793-8505
R. Drimmel @ https://orcid.org/0000-0002-1777-5502
A. Rotundi @ https://orcid.org/0000-0001-5467-157X

References

Alard, C., & Lupton, R. H. 1998, ApJ, 503, 325
Antonello, E., & Poretti, E. 1986, A&A, 169, 149
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Berdnikov, L. N., & Turner, D. G. 2004, A&AT, 23, 253
Chen, X., Wang, S., Deng, L., et al. 2019, NatAs, 3, 320
Chen, X., Wang, S., Deng, L., et al. 2020, ApJS, 249, 18
Clementini, G., Ripepi, V., Molinaro, R., et al. 2019, A&A, 622, A60
Deason, A. J., Belokurov, V., Erkal, D., et al. 2017, MNRAS, 467, 2636
Deb, S., & Singh, H. P. 2009, A&A, 507, 1729
Dékány, I., Hajdu, G., Grebel, E. K., et al. 2019, ApJ, 883, 58
Dékány, I., Minniti, D., Hajdu, G., et al. 2015b, ApJL, 799, L11
Dékány, I., Minniti, D., Majaess, D., et al. 2015a, ApJL, 812, L29
Fukue, K., Matsunaga, N., Yamamoto, R., et al. 2015, ApJ, 812, 64
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Heinze, A. N., Tomry, J. L., Denneau, L., et al. 2018, AJ, 156, 241
Inno, L., Matsunaga, N., Romaniello, M., et al. 2015, A&A, 576, A30
Inno, L., Urbaniaja, M. A., Matsunaga, N., et al. 2019, MNRAS, 482, 83
Jayasinghe, T., Kochanek, C. S., Stanek, K. Z., et al. 2018, MNRAS, 477, 3145
Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019, MNRAS, 486, 1907
Kains, N., Calamida, A., Rejkuba, M., et al. 2019, MNRAS, 482, 3058
Khruslov, A. V., & Kusakin, A. V. 2016, arXiv:1605.01313
Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al. 2017, arXiv:1711.03234
Laney, C. D., & Stobie, R. S. 1992, A&AS, 93, 93
Lomb, N. R. 1976, Ap&SS, 39, 447
Mantegazza, L., & Poretti, E. 1992, A&A, 261, 137
Matsunaga, N., Bono, G., Chen, X., et al. 2018, SSRv, 214, 74
Matsunaga, N., Feast, M. W., Bono, G., et al. 2016, MNRAS, 462, 414
Matsunaga, N., Fukue, K., Yamamoto, R., et al. 2015, ApJL, 799, 46
Ripepi, V., Molinaro, R., Musella, I., et al. 2019, A&A, 625, A14
Scargle, J. D. 1982, ApJ, 263, 835
Shappee, B., Prieto, J., Stanek, K. Z., et al. 2014, AAS Meeting Abstracts, 223, 236.03
Simon, N. R., & Lee, A. S. 1981, ApJ, 248, 291
Skowron, D. M., Skowron, J.,orable, P., et al. 2019a, Sci, 365, 478
Skowron, D. M., Skowron, J., Mróz, P., et al. 2019b, A&A, 69, 305
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2020, A&A, 70, 101
Tanioka, S., Matsunaga, N., Fukue, K., et al. 2017, ApJ, 842, 104
Udalski, A., Soszyński, I., Pietrukowicz, P., et al. 2018, A&A, 68, 315
VanderPlas, J. T., & Ivezić, Ž. 2015, ApJ, 812, 18
Venn, J. 1881, Symbol Logic (London: Macmillan)