Empirical completeness assessment of the Gaia DR2, Pan-STARRS 1 and ASAS-SN-II RR Lyrae catalogues

Cecilia Mateu,1⭐ Berry Holl,2,3 Joris De Ridder,4 and Lorenzo Rimoldini2

1 Departamento de Astronomía, Facultad de Ciencias, Universidad de la República, Iguá 4255, 11000, Montevideo, Uruguay
2 Department of Astronomy, University of Geneva, Chemin d’Ecogia 16, 1290 Versoix, Switzerland
3 Department of Astronomy, University of Geneva, Chemin des Maillettes 51, 1290 Versoix, Switzerland
4 Institute of Astronomy, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

RR Lyrae stars are an important and widely used tracer of the most ancient populations of our Galaxy, mainly due to their standard candle nature. The availability of large scale surveys of variable stars are allowing us to use them to trace the structure of our entire Galaxy, even in previously inaccessible areas like the Galactic disc. In this work we aim to provide an empirical assessment of the completeness of the three largest RRL catalogues available: Gaia DR2, PanSTARRS-1 and ASAS-SN-II. Using a joint probabilistic analysis of the three surveys we compute 2D and 3D completeness maps in each survey’s full magnitude range. At the bright end (\(G < 13\)) we find ASAS-SN-II and Gaia are near 100% complete in RR\(ab\) at high latitude (|\(b| > 20^\circ\)); ASAS-SN-II has the best completeness at low latitude for RR\(ab\) and at all latitudes for RR\(c\). At the faint end (\(G > 13\)), Gaia DR2 is the most complete catalogue for both RRL types, at any latitude, with median completeness rates of 95% (RR\(ab\)) and >85% (RR\(c\)) outside the ecliptic plane (|\(\beta| > 25^\circ\)). We confirm a high and uniform completeness of PanSTARRS-1 RR Lyrae at 91% (\(ab\)) and 82% (\(c\)) down to \(G \sim 18\), and provide the first estimate of its completeness at low galactic latitude (|\(b| \leq 20^\circ\)) at an estimated median 65% (\(ab\)) and 50–60% (\(c\)). Our results are publicly available as 2D and 3D completeness maps, and as functions to evaluate each survey’s completeness versus distance or per line-of-sight.

Key words: catalogues – methods: data analysis – stars: variables: RR Lyrae – Galaxy: stellar content

1 INTRODUCTION

The Gaia mission in its Second Data Release (DR2 Gaia Collaboration et al. 2018) has provided what is currently the largest, deepest, all-sky catalogue of RR Lyrae (RRL) stars (Clementini et al. 2019; Rimoldini et al. 2019), reaching even highly crowded regions in the disc and bulge. The superb astrometry provided by Gaia DR2 combined with the power of using RRL stars as standard candles to measure precise distances beyond the reach of Gaia’s parallaxes, are offering an unprecedented view of the structure and kinematics of the old populations (\(\geq 10\) Gyr) of our Galaxy traced by RRLs. Gaia DR2 RRLs have allowed studies of the shape of the inner halo (Iorio et al. 2018) and it’s velocity ellipsoid and gravitational potential in it’s innermost regions (Wegg et al. 2019); the identification of new and extra-tidal stars associated to ultra-faint dwarfs (Vivas et al. 2020) and globular clusters (Kundu et al. 2019); new tidal tails around known objects Minniti et al. (2018) and the mapping and kinematic characterisation of known tails like the Sagittarius, Orphan and Pal 5 streams (Ramos et al. 2020; Koposov et al. 2019; Price-Whelan et al. 2019).

An aspect shared by these studies is that their results are not sensitive to the selection function and, consequently, do not require a detailed a priori knowledge of the completeness of the catalogue. Other studies do depend critically on this, mainly those that aim at deriving density profiles (e.g. Mateu & Vivas 2018; Hernitschek et al. 2018; Bovy et al. 2016); but also, some techniques aimed to search for stream-like overdensities, which can be prone to yield large rates of false positives if the selection function has sharp variations in completeness (e.g. great circle cell methods, Mateu et al. 2011). For the majority of these studies, a detailed map of the completeness as a function of distance for each line of sight is necessary. However, most variable star surveys typically provide completeness estimates based on simulations.
and can provide only either average estimates, or completeness as a function of distance averaged over the full survey area (e.g. Sesar et al. 2017; Drake et al. 2013; Mateu et al. 2012; Vivas et al. 2004). These strategies have proven successful at high latitude, but at low latitudes they are often not-applicable (Sesar et al. 2017) or simplifications need to be made that lead to discarding valuable information at the faint end (e.g. Mateu & Vivas 2018).

A method to empirically estimate the completeness in a statistical manner can overcome these difficulties. Using two independent catalogues –that can be unequivocally cross-matched– the procedure described by Rybizki & Drimmel (2018) uses the relative fraction of stars in common between the two to provide the completeness estimates for both catalogues; without any prior assumptions on the completeness of either or on their union. To estimate the completeness of Gaia DR2 RRLs we need to combine the other two largest RRL surveys to span Gaia’s full magnitude range ($G < 20.7$) and most of its all-sky coverage: ASAS-SN-II covering the full sky at the bright end, down to $G \sim 16$; and Pan-STARRS-I (PS1), covering 3/4’s of the sky at the faint end from $13 < G < 21$. The joint analysis yields completeness maps for the three catalogues, Gaia, PS1 and ASAS-SN-II, in their full respective magnitude ranges. These will also be the first measurements of completeness at low latitude ($|b| \leq 20^\circ$) for deep large-scale RRL surveys. Our aim in this work is then to provide an empirical assessment of the completeness of the three largest available RRL catalogues: Gaia DR2, Pan-STARRS-1 (PS1) and ASAS-SN-II.

The structure of this paper is as follows. In Section 2 we describe the three RRL catalogues used: Gaia DR2, PS1 and ASAS-SN-II. Gaia DR2 published two partially overlapping catalogues containing stars identified as RRLs based solely on their Gaia photometry: the Specific Objects Study (SOS, Clementini et al. 2019) and VariClassifier (VC, Rimoldini et al. 2019) catalogues, produced by two different pipelines and based on data with different characteristics (Holl et al. 2018). Here we will combine them into a single RRL catalogue VC+SOS. In Section 3 we validate it by comparing it in different aspects to other reference catalogues: we discuss the overall accounting by matching against the PS1 full RRL and bona fide RRL catalogues; the confusion matrix and period recovery statistics comparing against PS1; contamination comparing against the ASAS-SN-II and Catalina Rapid Transient (CRTS) catalogues of variable stars. In Section 4 we compute and discuss the completeness of the three RRL surveys, across the sky and as a function of distance along selected lines of sight. We summarise our conclusions in Section 5.

2 THE RR LYRAE CATALOGUES

2.1 Gaia VC+SOS

The second data release of Gaia (DR2) included 550 737 variables stars (Holl et al. 2018) amongst which 228 904 are RRL stars. These RRLs were published in two partially overlapping outputs: (i) 195 780 sources in the variables classifier (VC, see Rimoldini et al. 2019) which contains a label and classification score for each source, and (ii) 140 784 in the RRL Specific Object Studies (SOS, see Clementini et al. 2019), which contains a large amount of detailed parameters including period, Fourier fitting parameters and photometric metallicity estimates. The overlap between the VC and SOS is 107 660 stars. Because the SOS output is the result of a more detailed analyses we ignore the VC results for these overlapping sources and only discuss the SOS result. This means that in this paper we examine 88 120 VC (hereafter ‘VCnotSOS’) and 140 784 SOS RRLs.

In summary, the notation and Gaia RRL samples1 used in this paper are as follows:

- VC: vari_classifier_result table with results from the variable classifier (Rimoldini et al. 2019), contains 195 780 RRLs (with best_class_name: RRA, RBC, RRD, ARRD).
- SOS: vari_rlyrarc table from the Specific Objects Study pipeline on RRLs, (Clementini et al. 2019) containing 140 784 SOS RRLs.
- VCnotSOS: 88 120 RRLs in VC table but not in SOS.
- VC+SOS: 228 904 RRLs: total of the VC and SOS tables (where SOS results take precedence when there is overlap with the VC).

The first line of Table 1 summarises the initial number of objects in these samples (see also Holl et al. 2018), before the quality cuts described in the following section were applied to filter contaminants out.

2.1.1 Filtering contaminants with Gaia quality flags

There are contaminant objects both in the VC and SOS RRL tables with much redder colours than normal RRLs. They have phot_bp_rp_excess_factor>2 or null, since their $G_B$ and $G_R$ magnitudes are inconsistent with their $G$-band magnitudes. Most, but not all, of these contaminants also have large astrometric_excess_noise (i.e. $\geq 2$). Conversely, a small fraction of stars (6,540; i.e. 3.7%) remain in the clean sample with astrometric_excess_noise> 2. These stars are kept in the clean sample as their $Gaia$ band magnitudes are consistent among themselves and produce distances consistent with their infrared counterparts. They should be culled out, however, in studies aimed at their kinematics.

The 53 740 stars filtered out this way may be either proper contaminants or they might be real RRLs with contaminated photometry: 12% of the contaminants are in the LMC and SMC, 80% are located at low galactic latitude $|b| < 20^\circ$, out of which ~90% are within 50° of the galactic centre. However, only 67 out of the 53 740 filtered out contaminants match stars in PS1. This supports that overall these culled out stars are probably not real RRLs since, as we will see in the next section, there is a large overlap between the Gaia DR2 VC+SOS and PS1 catalogues. As noted by Clementini et al. (2019) and Rimoldini et al. (2019), some of these contaminants (982) turned out to be galaxies mistaken for variable stars. Their source_ids are included in Table C.1 of Clementini et al. (2019).

In what follows we will use the Gaia clean sample, even

\footnote{Data publicly available in the online Gaia archive at \url{http://gea.esac.esa.int/archive/} containing the ‘tables’ and ‘fields’ referred to in the rest of this article.

MNARS 000, 1–18 (2018)
Table 1. Accounting of the contaminant filtering.

|          | VC+SOS | SOS    | VCnotSOS |
|----------|--------|--------|----------|
| Initial  | 228,904 | 140,784 | 88,120   |
| Contaminants | 53,740 | 18,037 | 35,703   |
| Clean    | 175,164 | 122,747 | 52,417   |

Figure 1. Distribution of mean $G$ magnitudes for the (clean) VC+SOS sample and its component SOS and VCnotSOS samples.

at the expense of a possible loss of completeness, since a loss of completeness does not affect the results at all, but a large contamination would (see Sec. 4).

The resulting distribution of mean $G$-band magnitudes, $\text{mean} \_ \text{mag} \_ g \_ \text{fov}$ from the vari\_time\_series\_statistics table, is shown in Figure 1 for the clean VC+SOS catalogue we will use hereafter. The peak at $G \sim 18$, most prominent in VC+SOS and VCnotSOS but almost not visible in SOS, is mostly due to the main body of the Sagittarius dwarf galaxy. The broad peak at $G \sim 19.3$ is due to the Large and Small Magellanic Clouds.

2.2 Distances

We compute distances to RRLs assuming an absolute magnitude $M_G = 0.6$ for all stars. This corresponds to the absolute magnitude of an RRL with a metallicity $[\text{Fe}/\text{H}] = -1.5$, the median halo metallicity, according to the $M_G - [\text{Fe}/\text{H}]$ from Muraveva et al. (2018). We use this simplified approach since only a subset of stars (~35\%) of Clementini et al. (2019) are provided in DR2 with a photometric estimate of $[\text{Fe}/\text{H}]$, and it is the most common way used in the literature to estimate distances for all Gaia DR2 RRL stars and we will only use them to provide an estimate of completeness as a function of distance in Sec. 2.2. We do caution the reader, however, these distances will be overestimated for disc RRL, which have higher metallicity (hence, lower luminosity) on average.

The extinction correction is made taking the $E(B-V)$ from the Schlegel et al. (1998) extinction maps, with the new calibration $A_V = 2.742E(B-V)$ from Schlafly & Finkbeiner (2011), and converted to $G$-band extinction using the coefficients in eq. (A.1) from Ramos et al. (2020) and assuming a dereddened $G_{\text{BP}} - G_{\text{RP}} = 0.7$ for all RRLs (the mean dereddened colour of SOS RRLs).

2.3 PanSTARRS-1

The Sesar et al. (2017) catalogue contains RRL stars identified based on grizy multi-epoch photometry from the PanSTARRS-1 (PS1) photometric survey spanning 3/4 of the sky, i.e. the entire sky observable from Hawaii (DEC $> -30^\circ$), with a limiting magnitude ($r \sim 21-22$) that approximately matches that of Gaia $G \sim 21$ for the typical colours of RRLs. PS1 grizy light curves are sparsely sampled, typically with $\lesssim 12$ asynchronous epochs per band obtained over a time span of four years. It contains a total of 239,044 variable stars and Sesar et al. (2017) provide a ‘classification score’ ($\text{score}_3 = ab/c$) which allows selecting RRL samples of $ab$ and $c$ with varying degrees of purity and completeness. Selecting RRab stars as $\text{score}_3 > 0.8$ and RRc stars $\text{score}_3 > 0.55$, according to Sesar et al. (2017) results in samples with $0.97$ and $0.90$ purity respectively, and $0.92$ and $0.79$ completeness (up to $~ 40$ kpc and at high latitudes) and a total of 44,608 and 17,187 stars for types $ab$ and $c$ respectively. Therefore, based on their suggested thresholds, for our subsequent analysis we define the following sub-samples:

- PS1 bona fide (61,795 stars): $\text{score}_3 > 0.8$ (RRab) or $\text{score}_3 > 0.55$ (RRc)
- PS1 non-bona-fide (177,249 stars): the complement of PS1 bona fide

We also estimate the $G$-band magnitude for the PS1 RRLs based on their (intensity-averaged) $g$ and $r$ magnitudes, in order to assess what is the overlap and depth of the two catalogues in the $G$-band and to be able to make broad $G$-band cuts consistently in the two catalogues when assessing the completeness. For this we use the $G - r$ relation given in Table A2 of Evans et al. (2018). This relation is provided for SDSS filters, we assume they are approximately valid for PS1 filters, for our purposes. Distances are provided for PS1 RRLs by Sesar et al. (2017), computed from a $i$-band Period-Luminosity-Metallicity relation (their eq. 5) and assuming $[\text{Fe}/\text{H}] = -1.5$ for all stars, as we have done for Gaia in Sec. 2.2.

2.4 ASAS-SN-II

ASAS-SN-II (Jayasinghe et al. 2019a,b) is an all-sky catalogue with over half a million classified variable stars2 spanning the magnitude range $10 - 11 < V \lesssim 17$ (corresponding to $G \lesssim 17$ for RRLs, with completeness at $G \sim 16$), including the VSX compilation of variable stars from many different surveys and individual sources. Although the ASAS-SN-II catalogue includes astrometric data from Gaia DR2, it’s variable star identification was done completely independently of Gaia’s, based on it’s own optical observations and classification algorithms (Jayasinghe et al. 2019b).

In total, ASAS-SN-II (Jayasinghe et al. 2019b) contains 44,279 RRLs of types $ab$, $c$ and $d$. Here we have restricted

\footnote{Light curve parameters and data are available at the ASAS-SN-II Variable Stars Database (Shappee et al. 2014) at \url{https://asas-sn.osu.edu/variables}}
the catalogue to the 44 110 RRLs with periods < 0.95 d, removing the clear excess of spuriously identified stars with periods of exactly 1 d (165 stars) and removing also those with suspiciously large amplitudes > 2 mag (4 stars). The G-band magnitude (obtained for all ASAS-SN-II stars from Gaia DR2 against PS1 because of its large area coverage (3π of the sky) and depth comparable to Gaia. However, in Sec. 2.6 we compare against the catalogues of periodic variable stars of the ASAS-SN-II survey to estimate the fraction of contaminants present in the VC+SOS catalogue.

3 VALIDATION OF THE VC+SOS CATALOGUE

In what follows, most of the validation is done against PS1 because of its large area coverage (3π of the sky) and depth comparable to Gaia. However, in Sec. 2.6 we compare against the catalogues of periodic variable stars of the ASAS-SN-II survey to estimate the fraction of contaminants present in the VC+SOS catalogue.

Following (Jayasinghe et al. 2019b).
3.1 Overall validation against PS1

3.1.1 Accounting

Figure 3 shows the G-band magnitude distribution for the PS1, VC+ SOS and different matching subsamples. In order to show the PS1 RRL data in this plot, we compute the expected G-band magnitude for each PS1 star based on its \( g \) and \( r \) intensity-averaged magnitudes (observed, i.e. without extinction correction), using the \( G(g, g−r) \) relation from Evans et al. (2018, Appendix A). Figure 4 summarises the composition in the different sub-samples (SOS, VCnotSOS, PS1 bona fide, PS1 non-bona-fide) the full Gaia VC+SOS catalogue (left); the Gaia VC+SOS sub-catalogue overlapping the PS1 footprint (middle), i.e. restricted to DEC > −30°; and the full PS1 RRL catalogue (right).

Figure 3 shows that, for RRLs, PS1 has a similar limiting magnitude as Gaia. Out of the 239K total variables in PS1, 61 208 (26%) match with Gaia; and the full PS1 RRL catalogue is indeed not RRL stars. Conversely, there are 11K non-bona-fide stars with an RRL counterpart in Gaia; as we will see in our analysis of Period recovery in Sec. 3.1.2, these are probably genuine RRL that simply did not make the bona fide cut in PS1. Hence, it would be useful to analyse these star’s PS1 time series in combination with Gaia’s. Finally, the 50K matching RRL constitute 81% of the total PS1 bona fide sample. So, most of the (VC+SOS)xPS1 matches are bona fide (82%) RRL in PS1 and most (81% also) of the bona fide PS1 RRL are recovered.

The (VC+SOS)xPS1 matching RRL amount to 70% of the objects in the VC+SOS catalogue at DEC > −30°, i.e. in the area overlapping the PS1 footprint where a match is possible. Since roughly two-thirds of the matching objects are bona fide PS1 RRL, this means just over half (57%) the objects in VC+SOS (inside the PS1 footprint) are bona fide PS1 RRL. The question is, in the VC+SOS catalogue, how confident are we in the remaining 45%. These can be classified in two classes: the ones that are in PS1, but are not bona fide; and the ones not in PS1. To check this, we will analyse the period recovery stats for both the bona fide and non-bona-fide samples in Section 3.1.2. The left panel of Fig. 4 also clearly shows that PS1 only represent < 40% of the Gaia RRL, even though PS1 covers 3/4 of the sky at the same limiting magnitude as Gaia and the recall of PS1 RRL by Gaia is large, so roughly > 60% of the RRLs in Gaia VC+SOS are new discoveries. The main reasons for this are: that the LMC and SMC contain thousands of RRLs and are not in PS1’s footprint; and that PS1’s completeness drops of very rapidly at low Galactic latitude due to crowding, where Gaia has a much better performance, as we will see in Section 4.2.

3.1.2 Period Recovery

First, we cross-check against PS1 the classification of VC+SOS RRL into the three types, by computing the confusion matrix, shown in Figure 5. This shows the overall class recovery is good, > 95% of the RRL stars are recovered with the same classification in both surveys. We note that this is particularly good since we are not discriminating by either VCnotSOS/SOS or bona-fide/non-bona-fide and gives us confidence that both the SOS and VC classes are indeed reliable.

The period recovery comparison between PS1 and VC+SOS can only be performed for the SOS subset, since the VCnotSOS RRL do not have reported periods. The top panel of Figure 6 shows the Gaia DR2 versus PS1 frequency, with the bottom panel zooming in on the 1 and 0.5 day aliases, as well as the Gaia aliases around the central part around \( f_{\text{Gaia}} - f_{\text{PS1}} = 0 \). In addition to the identity line, the over-dense horizontal lines where many stars align in this plot correspond to aliased periods, which typically show up when one, or more, external periodicities are embedded in the time sampling. Table 2 summarises the fraction of stars recovered with the same period or an alias, within a tolerance of \( \chi = 3 \) and 10 ‘sigma’ on the normalised frequency difference (see Eq. B4 of Appendix B). These two criteria can be considered as ‘strict’ and ‘loose’ period recovery criteria, respectively. For example, for an RRab star with true period \( P = 0.6 \) d, the 3 and 10 sigma criteria correspond to a 6.0 and 19.8 sec period tolerance, and for an RRe star with true period \( P = 0.3 \) d, the 3 and 10 sigma criteria correspond to a 1.8 and 5.9 sec period tolerance (using Eq. B5). See Figs. B1 and B2 for a visual representation of these tolerances with respect to the data.

Table 2 shows that the vast majority of the RRLs are recovered with the same period; for the full SOSxPS1 sample, more than 77% of the stars are recovered within 3-sigma, and about 83% within 10-sigma. For RRLs it is well known that period aliasing is a relatively common issue since certain common external periodicities, e.g. 1 d, produce period aliases that are also in the typical range of RRL periods (Lafier & Kinman 1965). As Figure 6 and Table 2 show, the 1 d alias and its harmonic the 0.5 d alias are the most common ones and affect mostly RRc stars (blue dots), as is typ-
Figure 4. Accounting: fraction of \emph{Gaia} and PS1 matching RRLs (\emph{Gaia} × PS1), separated as SOS and VCNnotSOS for \emph{Gaia}, and as bona fide and non-bona fide for PS1. Left: for the full \emph{Gaia} VC+SOS sample. Center: for the \emph{Gaia} VC+SOS stars in the 3π PS1 footprint (DEC > 30°). Right: for the full PS1 sample.

4 COMPLETENESS

We estimate the completeness of two independent catalogues following the procedure used by Rybizki & Drimmel (2018). In our case we will take VC+SOS and ASAS-SN-II or VC+SOS and PS1 as the two independent RRL catalogues.

The procedure is as follows. Assuming there is no (or negligible) contamination, the completeness \( C_A \) of survey \( A \) is given by its probability of detection of a star which is, in turn, the number \( N_A \) of stars observed by the survey divided by the true, total, number of stars \( N_{\text{true}} \):

\[
C_A = \frac{P_A}{N_{\text{true}}} = \frac{N_A}{N_{\text{true}}} \quad (1)
\]

Assuming that two surveys are conditionally independent (\( P_{A \cap B} = P_A \cdot P_B \)), the number \( N_{A \cap B} \) of stars in common between the two is given by:

\[
N_{A \cap B} = P_{A \cap B} N_{\text{true}} = C_A C_B N_{\text{true}} \quad (2)
\]

which, substituting Eq. (1) for survey \( A \) and similarly for survey \( B \), and solving for the completeness, gives the two following expressions:

\[
C_A = N_{A \cap B}/N_B \quad (3)
\]
\[
C_B = N_{A \cap B}/N_A \quad (4)
\]

The validity of this procedure to estimate the two survey’s completeness, therefore, hinges on two assumptions: that the surveys are conditionally independent, as already mentioned, and that there is no contamination. In Sec. 2.6 we estimated contamination to be < 10% for the three surveys, we will assume in what follows that the approximation of negligible contamination is valid. The method also makes no assumption on the completeness of either individual survey or on the combination (\( A \cup B \)) of the two.

The method also implicitly assumes the cross-match be-
Figure 6. Gaia DR2 versus PS1 periods for matching RRLs. The identity and main period alias loci are shown with the solid and different dashed lines, respectively. RRL classified as type \(ab\) and \(c\) (SOS) are shown with different colours. The lower left panel is a zoomed-in version of the \(0, \pm 1\) and \(\pm 2\) d\(^{-1}\), and the lower right an expansion around 0, showing a richness of various aliases (mainly from Gaia). The main period aliases were visually identified as the over-dense loci in this plot.

Table 2. Percentage of RRLs (\(ab\) and \(c\)) with periods recovered correctly or as period aliases, using the ‘sigma’ recovery criterion of Eq. (B4).

| X ‘sigma’ recovery | All     | bona fide | non-bona-fide |
|--------------------|---------|-----------|---------------|
| RRab + RRc         | (1)     | (10)      | (3)           |
| Identity           | 77.35   | 83.08     | 82.67         |
| ±0.0055 d\(^{-1}\) (0.5 d) | 0.17    | 0.19      | 0.03          |
| ±1.0027 d\(^{-1}\) (1.0 d) | 6.56    | 7.20      | 4.74          |
| ±0.0613 d\(^{-1}\) (16.3 d) | 0.13    | 0.28      | 0.15          |
| ±0.0590 d\(^{-1}\) (16.9 d) | 0.03    | 0.09      | 0.04          |
| ±0.0397 d\(^{-1}\) (25.2 d) | 0.09    | 0.23      | 0.09          |
| ±0.0372 d\(^{-1}\) (26.9 d) | 0.19    | 0.40      | 0.19          |
| ±0.0317 d\(^{-1}\) (31.5 d) | 0.26    | 0.37      | 0.27          |
| ±0.0053 d\(^{-1}\) (190.0 d) | 0.21    | 0.41      | 0.21          |
| ±0.0027 d\(^{-1}\) (1 yr) | 0.48    | 0.65      | 0.36          |
| Total              | 85.47   | 92.89     | 88.75         |

RRab

| Identity           | 83.78   | 89.71     | 86.67         |
| ±0.0055 d\(^{-1}\) (0.5 d) | 0.00    | 0.01      | 0.00          |
| ±1.0027 d\(^{-1}\) (1.0 d) | 1.25    | 1.49      | 0.43          |
| ±0.0613 d\(^{-1}\) (16.3 d) | 0.18    | 0.38      | 0.20          |
| ±0.0590 d\(^{-1}\) (16.9 d) | 0.05    | 0.13      | 0.05          |
| ±0.0397 d\(^{-1}\) (25.2 d) | 0.12    | 0.31      | 0.13          |
| ±0.0372 d\(^{-1}\) (26.9 d) | 0.25    | 0.53      | 0.24          |
| ±0.0317 d\(^{-1}\) (31.5 d) | 0.35    | 0.51      | 0.36          |
| ±0.0053 d\(^{-1}\) (190.0 d) | 0.29    | 0.53      | 0.29          |
| ±0.0027 d\(^{-1}\) (1 yr) | 0.18    | 0.28      | 0.12          |
| Total              | 86.36   | 93.87     | 89.49         |

RRc

| Identity           | 62.18   | 67.43     | 72.11         |
| ±0.0055 d\(^{-1}\) (0.5 d) | 0.57    | 0.60      | 0.10          |
| ±1.0027 d\(^{-1}\) (1.0 d) | 19.08   | 20.68     | 16.11         |
| ±0.0613 d\(^{-1}\) (16.3 d) | 0.02    | 0.04      | 0.01          |
| ±0.0590 d\(^{-1}\) (16.9 d) | 0.00    | 0.02      | 0.00          |
| ±0.0397 d\(^{-1}\) (25.2 d) | 0.01    | 0.04      | 0.01          |
| ±0.0372 d\(^{-1}\) (26.9 d) | 0.05    | 0.09      | 0.05          |
| ±0.0317 d\(^{-1}\) (31.5 d) | 0.04    | 0.05      | 0.05          |
| ±0.0053 d\(^{-1}\) (190.0 d) | 0.02    | 0.11      | 0.00          |
| ±0.0027 d\(^{-1}\) (1 yr) | 1.17    | 1.53      | 1.00          |
| Total              | 83.13   | 90.58     | 89.44         |

4.1 Completeness at the bright end: Gaia VC+SOS and ASAS

First we take the Gaia VC+SOS and ASAS-SN-II RRL catalogues as surveys A and B in Eqs. (3) and (4) to estimate each catalogue’s completeness. There estimates are valid in the magnitude range 11 \(\leq G \leq 17\), where the two surveys overlap. Figures 7 and 8 show the resulting completeness maps for Gaia VC+SOS and ASAS-SN-II, respectively, sep-
Figure 7. Completeness maps for VC+SOS RR\textsubscript{ab} (top) and RR\textsubscript{c} or RR\textsubscript{d} (bottom) stars, estimated from matches against ASAS-SN-II, in three G-band magnitude ranges. All maps are Mollweide projections in Galactic coordinates, with $l = 0^\circ$ at the centre and longitude increasing to the left. The gray shading denotes pixels with undefined completeness (i.e. zero counts in the comparison survey).

Figure 8. Completeness maps for ASAS-SN-II RR\textsubscript{ab} (top) and RR\textsubscript{c} or RR\textsubscript{d} (bottom) stars, estimated from matches against Gaia VC+SOS, in three G-band magnitude ranges. All maps are Mollweide projections in Galactic coordinates, with $l = 0^\circ$ at the centre and longitude increasing to the left. The gray shading denotes pixels with undefined completeness (i.e. zero counts in the comparison survey).

For Gaia VC+SOS the completeness is remarkably high (> 80%) and homogeneous across the sky for the ab, even close to the Galactic plane. For the c completeness is systematically lower on average than for the ab. Although at the faintest end (right panels) both types start to show the Gaia scanning law pattern, its clear the RR\textsubscript{c} are much more affected by it. These results are to be expected: RR\textsubscript{c} light curves have smaller amplitudes than RR\textsubscript{ab}’s and, because their nearly sinusoidal light curves are more easily confused with those of other types of variables (e.g. eclipsing contact binaries), they are harder to identify with fewer epochs.

For ASAS-SN-II both the RR\textsubscript{ab} and RR\textsubscript{c} completeness maps are remarkably uniform across the sky away from the Galactic plane and show a clear decrease at low Galactic latitudes ($|b| \lesssim 25^\circ$) at the faintest magnitude bin; plus two areas of lower completeness than average around the Magellanic Clouds ($l \sim 300^\circ$, $b \sim -45^\circ$) and at $l \sim 120^\circ$, $b \lesssim 50^\circ$ in the second and third magnitude bins. The lower completeness in at low latitude and the Magellanic Clouds is likely due to crowding, since the ASAS-SN-II telescopes have relatively low angular resolution (pixel size 8”). The other low completeness area at $l \sim 120^\circ$, $b \lesssim 50^\circ$ coincides with the North Celestial Pole (DEC $\lesssim 80^\circ$), the lower completeness could be due to issues in the ASAS-SN-II time sampling. By contrast, its uniformity at intermediate to high latitudes for both types of RRLs is a consequence of the large number of epochs per star (> 200) available in the survey. In these areas, outside the ecliptic plane where Gaia is affected by the scanning law, the mean completeness for Gaia and ASAS-SN-II is similar for the RR\textsubscript{ab} and slightly lower for
the ASAS-SN-II RRc at the faintest two panels and higher for both types in ASAS-SN-II at the brightest panel (left).

4.2 Completeness at the faint end: Gaia VC+SOS and PS1

Following the same procedure as in the previous section, we estimate the completeness of the Gaia VC+SOS and PS1 surveys, valid for the magnitude range 13 ≤ G ≤ 20.7. For this we use only the bona fide subset of PS1 RRLs, to ensure the lowest contamination, so our method’s assumptions remain valid. The possible loss of completeness due to this cut has no effect in our method.

Figures 9 and 10 show the completeness maps for VC+SOS and PS1 respectively, in each figure separated cut has no effect in our method.

4.3 Final Completeness for Gaia VC+SOS

So far we have shown that our separate estimates of the VC+SOS completeness based on two mutually independent catalogues, PS1 and ASAS-SN-II, are consistent with one another in the magnitude range of overlap (13 < G < 16) between the two. Therefore, we can provide a final estimate of the completeness over the full magnitude range using the combined PS1 and ASAS-SN-II catalogues. We compute this taking A as Gaia and B as PS1/ASAS in Eq. (3), making sure objects are not counted twice in the latter.

At the faint end (G > 16) there remains, however, an area of π in the sky (DEC < −30°) where the lack of PS1 coverage prevents us from estimating the completeness. To circumvent this and be able to provide an all-sky map, we choose to report the estimated completeness corresponding to the diametrically opposing field in ecliptic coordinates.

4.4 Completeness as a function of distance and magnitude

Figure 13 and Table 3 summarise the median completeness for VC+SOS, PS1 and ASAS-SN-II as a function of the (intensity-averaged) G-band magnitude, at low (|b| < 20°) and high (|b| > 20°) Galactic latitudes.

At the bright end (G ≤ 13), for RRab stars ASAS-SN-II and Gaia are both near 100% complete at high latitude; at low latitude ASAS-SN-II has a higher completeness. For RRc stars ASAS-SN-II has the best performance, with Gaia’s completeness dropping fast for G < 13. Note, though, that in this magnitude range, for both surveys, the stochastic noise is notoriously high. The higher completeness for ASAS-SN-II is not entirely surprising, at least for the

---

4 https://github.com/cmateu?tab=repositories
RRLs of type c, since identification of variables in ASAS-SN-II is based on several hundreds of epochs, many more than Gaia in its current release, and as we have mentioned RRc stars are more difficult to identify and confidently tell apart from other variables, particularly with few epochs.

At the faint end ($G > 13$) Gaia VC+SOS is more complete than both PS1 and ASAS-SN-II, in all cases. For RRab stars the three surveys are highly complete $> 85\%$ at high latitude, were all surveys are expected to have the best performance. The differences become more noticeable for the RRc stars and at low latitude for both types. Overall, the Gaia VC+SOS catalogue has a remarkably high completeness for RRab, being $> 90\%$ up to $G \sim 19$ both at high and low latitudes; and lower—as expected—for RRc, the difference being more pronounced at low latitudes, where the completeness of the RRc drops to $\sim 70\%$ compared to $> 80\%$ at high latitude. The inter-quartile range is also shown in the figure, to illustrate the combined effect of the stochastic noise and the dispersion of completeness values over the sky for the two types of RRL. This clearly shows the dispersion is more than double for RRc stars ($> 20\%$) than it is for RRab stars ($< 10\%$) over the entire magnitude (or distance) range, as first observed in the mean maps of Figure 9.

In cases where there is zero or constant extinction, Fig. 13 trivially gives the dependence of completeness upon distance. In the Galactic plane, however, extinction varies significantly with distance and with the line of sight. In this case it is useful to conduct the same analysis as we have done so far, but as a function of distance rather than $G$ magnitude. We remind the reader that, although ‘geometric’ distances can be calculated using Gaia parallaxes for the nearest objects, throughout this work we use only photometric distances (see Sec. 2.2). As a use case, Figure 14 illustrates the dependence of the completeness for Gaia VC+SOS as a...
Figure 11. Median completeness versus (intensity-averaged) $G$ stars estimated for Gaia VC+SOS based on ASAS-SN-II (thin) and on PS1 (thick) for RR$ab$ (red) and RR$c$ (blue). The shaded areas represent the completeness’ inter-quartile range for the VC+SOS maps. Top: $|b| > 20^\circ$. Bottom: $|b| \leq 20^\circ$.

function of distance, for lines of sight toward the Galactic Anti-centre ($l = 180^\circ$) with increasing Galactic latitude (left panel) and in the Galactic Plane ($b = 0^\circ$) with increasing Galactic longitude. The bins have variable width in distance, so as to ensure a fixed number of stars per bin and, in turn, a fixed (fractional) Poisson noise. Outside the Galactic plane, for the RR$ab$ there is very little, if any, dependence with Galactic latitude up to $\sim 40$ kpc; for larger distances there is a small dependence, the completeness drops by $\sim 10\%$ for lines-of-sight $b < 10^\circ$ compared to ones at higher latitude. By contrast, for the RR$c$, the completeness depends more strongly on Galactic latitude at all distances. Even at distances < 20 kpc the completeness drops from 80% at $b = 70^\circ$ to 40% at $b \sim 5^\circ$.

The GitHub repository rrl_completeness\textsuperscript{5} provides a convenience function in Python to retrieve completeness as a function of distance for a survey in a given line of sight. Also, for convenience, completeness as a function of distance, per line-of-sight, for each of the three surveys is provided in Table A2, Appendix A.

\textsuperscript{5} https://github.com/cmateu?tab=repositories

4.5 Comparison with the literature

Completeness rates of the RRLs identified in Gaia DR2 were estimated also in other works that validated such candidates. Molnár et al. (2018) compared the RRL candidates in the Gaia DR2 classification (VC) with the ones found in a selection of fields of the Kepler and K2 missions (Borucki et al. 2010; Howell et al. 2014). These fields covered about 110 deg$^2$ each and were distributed along the Ecliptic (except for the original Kepler field); they spanned Galactic latitudes $|b| < 70^\circ$ but excluded the region of the Galactic bulge (as already well resolved in Soszyński et al. 2014). The average completeness rates of the RRL candidates (of all sub-types) in VC were found typically between 70 and

---

**Table 3.** Median Completeness for VC+SOS, PS1 and ASAS-SN-II RRLs at $|b| > 20^\circ$.

| $G$ (mag) | VC+SOS RR$ab$ | VC+SOS RR$c$ | PS1 RR$ab$ | PS1 RR$c$ | ASAS-SN-II RR$ab$ | ASAS-SN-II RR$c$ |
|-----------|----------------|--------------|------------|------------|------------------|------------------|
| 11.0-11.5 | 100.0          | 0.0          | 0.0        | 0.0        | 100.0            | 100.0            |
| 11.5-12.0 | 100.0          | 25.0         | 0.0        | 0.0        | 100.0            | 100.0            |
| 12.0-12.5 | 100.0          | 50.0         | 0.0        | 0.0        | 100.0            | 100.0            |
| 12.5-13.0 | 92.9           | 0.0          | 0.0        | 0.0        | 94.2             | 100.0            |
| 13.0-13.5 | 100.0          | 0.0          | 0.0        | 0.0        | 100.0            | 100.0            |
| 13.5-14.0 | 100.0          | 50.0         | 0.0        | 0.0        | 95.6             | 86.7             |
| 14.0-14.5 | 100.0          | 87.5         | 0.0        | 0.0        | 93.9             | 86.6             |
| 14.5-15.0 | 100.0          | 88.9         | 0.0        | 0.0        | 92.5             | 84.7             |
| 15.0-15.5 | 100.0          | 93.3         | 86.2       | 80.0       | 91.3             | 81.4             |
| 15.5-16.0 | 100.0          | 83.3         | 90.0       | 75.0       | 90.5             | 64.5             |
| 16.0-16.5 | 100.0          | 91.7         | 90.3       | 75.0       | 0.0              | 0.0              |
| 16.5-17.0 | 96.9           | 83.8         | 89.9       | 75.0       | 0.0              | 0.0              |
| 17.0-17.5 | 94.7           | 85.7         | 88.9       | 75.0       | 0.0              | 0.0              |
| 17.5-18.0 | 93.3           | 83.3         | 88.9       | 77.8       | 0.0              | 0.0              |
| 18.0-18.5 | 93.3           | 85.7         | 88.2       | 80.0       | 0.0              | 0.0              |
| 18.5-19.0 | 91.7           | 80.0         | 86.0       | 80.0       | 0.0              | 0.0              |
| 19.0-19.5 | 92.3           | 80.0         | 80.5       | 75.0       | 0.0              | 0.0              |
| 19.5-20.0 | 85.2           | 66.7         | 76.9       | 72.1       | 0.0              | 0.0              |

**Table 4.** Median Completeness for VC+SOS, PS1 and ASAS-SN-II RRLs at $|b| \leq 20^\circ$.

| $G$ (mag) | VC+SOS RR$ab$ | VC+SOS RR$c$ | PS1 RR$ab$ | PS1 RR$c$ | ASAS-SN-II RR$ab$ | ASAS-SN-II RR$c$ |
|-----------|----------------|--------------|------------|------------|------------------|------------------|
| 11.0-11.5 | 87.5           | 0.0          | 0.0        | 0.0        | 100.0            | 50.0             |
| 11.5-12.0 | 50.0           | 0.0          | 0.0        | 0.0        | 100.0            | 100.0            |
| 12.0-12.5 | 62.5           | 0.0          | 0.0        | 0.0        | 100.0            | 75.0             |
| 12.5-13.0 | 68.3           | 0.0          | 0.0        | 0.0        | 100.0            | 50.0             |
| 13.0-13.5 | 73.2           | 13.8         | 0.0        | 0.0        | 87.5             | 100.0            |
| 13.5-14.0 | 100.0          | 62.5         | 50.0       | 100.0      | 85.7             | 60.0             |
| 14.0-14.5 | 100.0          | 65.3         | 80.0       | 80.0       | 80.0             | 65.0             |
| 14.5-15.0 | 100.0          | 75.0         | 67.3       | 80.0       | 76.9             | 48.4             |
| 15.0-15.5 | 100.0          | 62.5         | 66.7       | 66.7       | 78.6             | 25.7             |
| 15.5-16.0 | 100.0          | 52.2         | 66.7       | 57.1       | 62.5             | 13.1             |
| 16.0-16.5 | 96.5           | 74.2         | 67.3       | 56.8       | 0.0              | 0.0              |
| 16.5-17.0 | 94.9           | 69.2         | 60.0       | 50.0       | 0.0              | 0.0              |
| 17.0-17.5 | 96.7           | 74.3         | 62.7       | 49.1       | 0.0              | 0.0              |
| 17.5-18.0 | 94.5           | 66.7         | 62.2       | 50.0       | 0.0              | 0.0              |
| 18.0-18.5 | 94.9           | 66.7         | 53.3       | 38.9       | 0.0              | 0.0              |
| 18.5-19.0 | 100.0          | 52.9         | 52.9       | 40.9       | 0.0              | 0.0              |
| 19.0-19.5 | 91.7           | 66.7         | 33.3       | 33.3       | 0.0              | 0.0              |
| 19.5-20.0 | 81.8           | 56.1         | 27.3       | 42.9       | 0.0              | 0.0              |
Figure 12. Final completeness maps for Gaia VC+SOS RR\textsubscript{ab} (top) and RR\textsubscript{c} (bottom) stars, valid in the full magnitude range of the survey. Maps are Mollweide projections in Galactic coordinates, with $l = 0^\circ$ at the centre and longitude increasing to the left. The gray shading denotes pixels with indefinite completeness (i.e. zero counts in the comparison survey).

Figure 13. Median completeness for RR\textsubscript{ab} (left) and RR\textsubscript{c} (right) stars versus (intensity-averaged) G magnitude for Gaia VC+SOS (solid), PS1 (dashed) and ASAS-SN-II (dotted) RRLs at $|b| \leq 20^\circ$ (top) and $|b| < 20^\circ$ (bottom). The shaded areas represent the completeness' inter-quartile range for each survey, to illustrate the combined effect of the uncertainties due to stochastic noise and line-of-sight variations in completeness.
Figure 14. Completeness of Gaia VC+SOS RRLs as a function of (heliocentric) distance for different lines of sight. **Left:** $l = 180^\circ$ at different Galactic latitudes, with a fixed angular radius of $30^\circ$. **Right:** $b = 0^\circ$ at different Galactic longitudes, with a fixed angular radius of $10^\circ$. The distance bins have a variable width, selected so as to have a minimum number of 100 stars per bin, corresponding to a Poisson noise of $\sim 15\%$.

78\%, depending on the data sets, in agreement with the results as a function of sky location illustrated in Fig. 12.

Further comparisons by Holl et al. (2018) of the RRL identifications (all types combined) in Gaia DR2 versus OGLE-IV (Soszyński et al. 2014, 2016) and Catalina (Drake et al. 2014b) estimated average completeness rates between 49 and 63\%. A higher identification rate is expected from OGLE-IV because of its $I$ band (better suited to the reddened objects in the bulge than the $G$ band) and the significantly higher sampling with respect to Gaia. Besides the effect of extinction in the bulge, the distance to the Magellanic Clouds implied lower completeness rates at the faint end, which are consistent with the range covered by the shaded region in grey in Fig. 13. The 63\% estimated from Catalina is due to the survey’s footprint. Catalina’s footprint ($|b| > 20^\circ$ and $-20^\circ < \text{DEC} < 70^\circ$) covers the ecliptic plane entirely but has holes at high ecliptic latitude; this biases the mean towards lower values since areas were Gaia has higher completeness (ecliptic poles) are under-represented in the survey. Repeating our analysis using Catalina, instead of ASAS-SN-II, as a reference catalogue we find the median completeness of the healpix map as Fig. 7 is 68\%, and a full-sky average of 62\% consistent with the findings from Holl et al. (2018). As a further check, we also computed the median of our Gaia completeness estimated from ASAS-SN-II, but restricting the latter to the Catalina footprint. In this test we find a median completeness of $\sim 75\%$. This seems to be in reasonable agreement with the 68\% found from Catalina when considering the difference in magnitude distributions and completeness of the two surveys.

Rimoldini et al. (2019) reported completeness rates of VC as a function of minimum classification score. The average completeness of RRL candidates (of all sub-types, scores, magnitudes, and locations) was estimated at about 65\%, a lower value than expected but likely biased by the large number of OGLE-IV RRL stars (in the bulge and Magellanic Clouds) used in their comparison set.

We can also compare the completeness we have estimated for PS1 with the estimate Sesar et al. (2017) made based on simulations. Sesar et al. (2017) find, for the bona fide RRL sample at $|b| > 20^\circ$, an expected average completeness of 92\% and 79\% for RR\text{ab} and RR\text{c} stars respectively, up to a distance of 40 kpc ($G \sim 18.5$), in excellent agreement with the averages resulting from Table 3 for $G < 19$, which gives 90.7\% and 82\% for RR\text{ab} and RR\text{c} respectively. At 80 kpc they estimate an expected completeness of 80\% for RR\text{ab}, also in very good agreement with our slightly higher value of $\sim 84\%$. At low latitude, however, Sesar et al. (2017) are not able to provide an estimate for the completeness based on simulations. The empirical approach used here, has allowed us to provide this estimate of PS1’s RRL completeness at low galactic latitude for the first time. We’d also like to highlight that the fact the PS1 RRL search was conducted and published in full and not truncated at low latitude has allowed for our estimate of Gaia’s completeness in its full magnitude range, even close to the Galactic plane, which will be key for upcoming studies of RRLs in the disc.

5 CONCLUSIONS

In this work we have set out to provide a detailed characterisation of the completeness for the three largest RRL surveys available to date: Gaia DR2, PS1 and ASAS-SN-II.

Since Gaia DR2 has two partially-overlapping catalogues –SOS and VC– containing RRLs, we began by validating the combined Gaia DR2 VC+SOS catalogue by comparing it against PS1 and other variable star surveys. The main results of our validation process are the following:

- Most (81\%) of the RRLs in common between Gaia VC+SOS and PS1 are stars classified as bona fide RRL in PS1, a characterisation made by (Sesar et al. 2017) imposing a threshold on the RRL classification scores. The statistics of cross-matches against VC+SOS confirms the
PS1 bona fide sample is indeed highly pure and highly complete. Conversely, only a small fraction of the non-bona fide RRL in PS1 (∼7%) have an RRL counterpart in Gaia, which confirms only a small fraction of these are likely to be real RRLs.

- The period recovery rate of Gaia SOS was estimated by cross-matching against PS1 RRLs. When compared to PS1 bona fide RRL, this was found to be 77% within 3-sigma and about 85% within 10-sigma. The occurrence of period aliases is found to be small (8–9%), with PS1 most likely being more affected than Gaia, as the most common aliases correspond to a 1 d external periodicity.

- The contamination of Gaia VC+SOS RRLs was estimated at ∼6% on average, based on cross-matches against the ASAS-SN-II and CRTS catalogues of periodic variable stars. Separated by RRL type, contamination is found to be 5% and 9% for VC+SOS RRab and RRc respectively.

To estimate the completeness of the RRL surveys we used a probabilistic analysis of stars in common between the different catalogues, following the procedure from Rybizki & Drimmel (2018) and described in Sec. 4. This enabled us to provide completeness estimates for all three catalogues, without assumptions on the completeness of any. The main results of our completeness assessment are summarised as follows:

- Completeness of the Gaia DR2 VC+SOS RRL catalogue is remarkably high with a median ≥95% and >85% for RRab and RRc respectively, even down to the survey limit where it decreases to ∼85% and ∼70% for RRab and RRc respectively. Line-of-sight variations in completeness around the ecliptic plane (β < 25°) due to the Gaia scanning law are ~10% for RRab and 20–30% for RRc. These variations affect only RRab stars fainter than G ~ 15 and RRc stars in all magnitude ranges. Outside the ecliptic plane, Gaia’s completeness is approximately uniform for both RRL types.

- Gaia DR2 VC+SOS’s completeness, remarkably, shows no significant dependence with Galactic latitude even close to the Galactic Plane for RRab. There is a mild dependence with Galactic latitude observed for the RRc, but the main limitation in this estimate is its larger stochastic noise due to the low number of RRls in the comparison catalogue close to the Galactic Plane.

- ASAS-SN-II RRL’s completeness is fairly uniform for both RRL types over the sky and as a function of magnitude outside the Galactic Plane, with an estimated median completeness of 87% and 58% for RRab and RRc respectively. At low Galactic latitude |b| < 20° the completeness rapidly falls off for G > 14. Both RRab and RRc maps show lower completeness than average at the celestial poles at the faint end of the survey 15 ≤ G ≤ 17.

- PS1 RRL’s completeness is remarkably uniform over the sky for both RRL types, and drops sharply to zero when approaching the Galactic Plane at |b| < 20°. At high latitude (|b| > 20°) PS1’s median completeness is estimated at 91% and 82% down to G ~ 18 for RRab and RRc respectively; and drops to 76 and 72% respectively at the faint end (G ~ 20).

- We provide the first estimate of PS1’s RRL completeness at low galactic latitude (|b| ≤ 20°) estimated at a median 65% and 50–60% for RRab and RRc stars respectively, up to G ~ 18 (g ~ 18.5).

- At the bright end (G ≤ 13) ASAS-SN-II has a higher completeness than Gaia VC+SOS, except for RRab at high latitudes (|b| > 20°) where the two catalogues are near 100% complete (Fig. 13). Therefore, when high completeness at the bright end is desired, supplementing VC+SOS with ASAS-SN-II RRLs is the optimal strategy. The combined catalogue will be close to 100% complete all the way down to G ~ 19 even close to the Galactic plane.

- At the faint end (G > 13) Gaia VC+SOS is more complete than both PS1 and ASAS-SN-II, in all cases. ASAS-SN-II is also more complete than PS1 down to G ~ 15, after which completeness rapidly plummets for ASAS-SN-II.

- The completeness maps shown here for Gaia VC+SOS, PS1 and ASAS-SN-II are provided in Appendix A and are publicly available at the rrl_completeness GitHub repository6, which includes convenience functions in Python to retrieve and compute 2D completeness maps and completeness as a function of distance for a given line of sight.

This work has been possible thanks to the unprecedented availability of large independent surveys of RRL stars overlapping in footprint and depth. The approach followed here can, of course, be applied to any type of tracer survey, and will be particularly valuable for other types of variable star or standard candle tracer. It will also be of enormous use to characterise other large surveys already available, e.g. CRTS and VVV, and others soon to come, like the variable star surveys to be made by the Vera Rubin Observatory7.

ACKNOWLEDGEMENTS

CM thanks R. Sanderson for useful discussions. This project was started at the 2018 NYC Gaia Sprint, hosted by the Center for Computational Astrophysics of the FlatIron Institute in New York City. CM thanks the organizers for their encouragement to participate, as well as the Simons Foundation for travel support. JDR thanks the LOC and the SOC for organizing an excellent meeting. Part of this work was performed at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611. The participation of C.M. at the Aspen Center for Physics was supported by the Simons Foundation. The research leading to these results has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement N°760519: MASMIE), from the KU Leuven Research Council (grant C16/18/005: PARADISE), as well as from the Belgian federal Science Policy Office (BELSPO) through PRODEX grants Gaia and PLATO. 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. CM acknowledges support from the DGAPA/UNAM PAPIIT program grant IG100319.

---

6 https://github.com/cmateu?tab=repositories
7 formerly the Large Synoptic Survey Telescope (LSST).
APPENDIX A: COMPLETENESS MAPS FOR RR LYRAE

A1 2D Completeness maps for RR Lyrae

Tables A1 to A4 summarise the 2D completeness maps for RRLs of type ab and c presented in Figures 7, 8, 9, 10 and 12. The last three columns present the completeness map in each survey’s full magnitude range. For PS1 and ASAS-SN-II the completeness map was computed using Gaia DR2 VC+SOS as the reference catalogue. For Gaia it was computed using ASAS-SN-II at the bright end and PS1 at the faint end as described in Sec. 4.1 and 4.2. The completeness indices presented correspond to maps produced with nested ordering. The completeness indices provided correspond to maps produced with nested ordering. The tables’ short versions shown are provided to illustrate their form and content. The full tables can be found in the electronic version of the journal.

Table A1. 2D completeness map for ASAS-SN-II and Gaia DR2 VC+SOS RRab in three G-band magnitude ranges at the bright end.

| Healpix | l (°) | b (°) | ASAS Gaia ASAS Gaia ASAS Gaia |
|---------|------|------|---------------------|
| 0 15.00 78.28 | 1.00 | 1.00 | 0.96 0.69 0.46 |
| 1 135.00 78.28 | 1.00 | 0.83 | 0.95 0.95 0.95 |
| 2 225.00 78.28 | 1.00 | 0.95 | 0.95 0.88 0.82 |
| 3 315.00 78.28 | 1.00 | 0.96 | 0.79 0.79 0.89 |
| 4 405.00 78.28 | 1.00 | 0.90 | 0.86 0.90 0.84 |

... ... ... ... ... ...

Table A2. 2D completeness map for ASAS-SN-II and Gaia DR2 VC+SOS RRc in three G-band magnitude ranges at the bright end.

| Healpix | l (°) | b (°) | ASAS Gaia ASAS Gaia ASAS Gaia |
|---------|------|------|---------------------|
| 0 15.00 78.28 | 0.50 | 0.50 | 0.62 0.24 0.65 |
| 1 135.00 78.28 | 0.89 | 0.89 | 0.80 0.75 |
| 2 225.00 78.28 | 1.00 | 0.25 | 0.82 0.51 0.75 |
| 3 315.00 78.28 | 0.00 | 1.00 | 0.60 0.43 0.78 |
| 4 405.00 78.28 | 1.00 | 0.17 | 0.29 0.71 0.54 |

... ... ... ... ... ...

APPENDIX B: PERIOD RECOVERY CRITERIA

To estimate period recovery for an ensemble of objects, it is desirable to use a statistic that has a fixed variance so that all results can be compared on equal footing and counted using a single threshold. Let us look at two often used statistics:
Table A3. 2D completeness map for ASAS-SN-II, PS1 and Gaia DR2 VC+SOS RR\textsubscript{ab} in three G-band faint magnitude ranges and in the full magnitude range.

| Healpix | $l$ (°) | $b$ (°) | ASAS PS1 | Gaia | ASAS PS1 | Gaia | ASAS PS1 | Gaia | ASAS PS1 | Gaia | ASAS PS1 | Gaia | PS1 | Gaia |
|---------|---------|---------|----------|-------|----------|-------|----------|-------|----------|-------|----------|-------|------|------|
| 0       | 45.00   | +87.08  | 1.00     | 0.50  | 1.00     | 0.50  | 1.00     | 0.67  | 0.00     | 0.50 | 1.00     | 0.60  | 0.89 |
| 1       | 135.00  | +87.08  | 1.00     | 0.83  | 1.00     | 0.83  | 0.00     | 1.00  | 0.00     | 1.00 | 0.40     | 1.00  | 0.92 |
| 2       | 225.00  | +87.08  | 0.50     | 1.00  | 0.67     | 0.40  | 0.80     | 1.00  | 0.00     | 0.67 | 0.67     | 0.75  | 0.92 |
| 3       | 315.00  | +87.08  | 0.67     | 1.00  | 1.00     | 0.62  | 1.00     | 1.00  | 0.00     | 0.67 | 1.00     | 0.86  | 0.93 |
| 4       | 45.00   | +84.15  | 0.67     | 1.00  | 1.00     | 0.40  | 1.00     | 1.00  | 0.00     | 0.67 | 1.00     | 0.83  | 0.91 |

Table A4. 2D completeness map for ASAS-SN-II, PS1 and Gaia DR2 VC+SOS RR\textsubscript{c} in three G-band faint magnitude ranges and in the full magnitude range.

| Healpix | $l$ (°) | $b$ (°) | ASAS PS1 | Gaia | ASAS PS1 | Gaia | ASAS PS1 | Gaia | ASAS PS1 | Gaia | ASAS PS1 | Gaia | PS1 | Gaia |
|---------|---------|---------|----------|-------|----------|-------|----------|-------|----------|-------|----------|-------|------|------|
| 0       | 45.00   | +87.08  | 1.00     | 0.33  | 1.00     | 1.00  | –        | 0.00  | 1.00     | 0.00  | 1.00     | 0.00  | 1.00 |
| 1       | 135.00  | +87.08  | 1.00     | 0.67  | –        | –     | 0.00     | 1.00  | 1.00     | 1.00  | 1.00     | 1.00  | 0.80 |
| 2       | 225.00  | +87.08  | 1.00     | 0.75  | 1.00     | –     | –        | 0.00  | 0.00     | 0.00  | 1.00     | 0.50  | 0.83 |
| 3       | 315.00  | +87.08  | 0.50     | 0.00  | 0.50     | 0.00  | 1.00     | 1.00  | –        | –     | 0.50     | 0.67  | 0.83 |
| 4       | 45.00   | +84.15  | 1.00     | 0.50  | 0.00     | 0.67  | 0.67     | 0.67  | 0.00     | 1.00  | 0.50     | 0.80  | 0.67 |

Table A5. 3D completeness map for Gaia DR2 VC+SOS RR\textsubscript{ab}.

| Healpix | $l$ (°) | $b$ (°) | $D_o$ (kpc) | $D_f$ (kpc) | Completeness | Error |
|---------|---------|---------|-------------|-------------|--------------|-------|
| 0       | 45.00   | +78.28  | 1.3         | 3.9         | 1.00         | 0.26  |
| 0       | 45.00   | +78.28  | 3.9         | 5.5         | 0.97         | 0.25  |
| 0       | 45.00   | +78.28  | 5.5         | 6.8         | 0.93         | 0.25  |
| 0       | 45.00   | +78.28  | 6.8         | 7.9         | 0.97         | 0.25  |
| 0       | 45.00   | +78.28  | 7.9         | 9.2         | 0.93         | 0.25  |

Table A6. 3D completeness map for Gaia DR2 VC+SOS RR\textsubscript{c}.

| Healpix | $l$ (°) | $b$ (°) | $D_o$ (kpc) | $D_f$ (kpc) | Completeness | Error |
|---------|---------|---------|-------------|-------------|--------------|-------|
| 0       | 45.00   | +78.28  | 1.2         | 4.6         | 0.53         | 0.17  |
| 0       | 45.00   | +78.28  | 4.6         | 6.8         | 0.57         | 0.17  |
| 0       | 45.00   | +78.28  | 6.8         | 9.4         | 0.80         | 0.22  |
| 0       | 45.00   | +78.28  | 9.4         | 12.0        | 0.80         | 0.22  |
| 0       | 45.00   | +78.28  | 12.0        | 14.3        | 0.70         | 0.20  |

Table A7. 3D completeness map for PS1 RR\textsubscript{ab}.

| Healpix | $l$ (°) | $b$ (°) | $D_o$ (kpc) | $D_f$ (kpc) | Completeness | Error |
|---------|---------|---------|-------------|-------------|--------------|-------|
| 0       | 45.00   | +78.28  | 0.7         | 3.6         | 0.23         | 0.10  |
| 0       | 45.00   | +78.28  | 3.6         | 5.3         | 0.90         | 0.24  |
| 0       | 45.00   | +78.28  | 5.3         | 6.6         | 0.77         | 0.21  |
| 0       | 45.00   | +78.28  | 6.6         | 7.7         | 0.80         | 0.22  |
| 0       | 45.00   | +78.28  | 7.7         | 9.0         | 0.87         | 0.23  |

the difference $\Delta f$ and ratio $R$ between two independently measured frequencies of the same object:

$$
\Delta f = f_1 - f_2 \quad \sigma^2_{\Delta f} = \sigma^2_R + \sigma^2_f
$$

$$
R = \frac{f_1}{f_2} \quad \sigma^2_R = R^2 \left( \frac{\sigma^2_R}{f_1} + \frac{\sigma^2_f}{f_2^2} \right)
$$

(B1)
with $\sigma_{f_1}^2$ and $\sigma_{f_2}^2$ being the respective uncertainties on the frequencies $f_1$ and $f_2$. The provided variances $\sigma_{f_1}^2$ and $\sigma_{f_2}^2$ explicitly assume that both frequencies are uncorrelated, i.e., estimated from independently measured time series. From Eq. (B1) it is clear that a standard-deviation normalised frequency difference $\Delta f$ will have fixed variance of 1:

$$\Delta f = \frac{f_1 - f_2}{\sigma_{\Delta f}} = \frac{f_1 - f_2}{\sqrt{\sigma_{f_1}^2 + \sigma_{f_2}^2}} = 1 \quad \text{for} \quad \frac{\sigma_{\Delta f}}{\sigma_{f_1}} = 1$$

(B2)

recovery if: $|\Delta f| < X$ (B3)

where recovery can then be defined as being within $X$-sigma. It is however not possible to find such fixed-variance expression for the ratio $R$, and its use is therefore discouraged when used in combination with a single threshold criterion (despite its wide use in literature), unless only a very narrow relative frequency range is examined and when frequency errors are all similar.

Depending on the time-sampling of the examined time series, strong aliases might occur, causing the identified frequency to be offset by a certain value and (generally) a phase shift in the folded light curve. If the time-sampling is known, the exact offset can be computed from the spectral window and, importantly, does not enlarge the uncertainty on the frequency. When the time-sampling is not available, but (very) similar for all objects, it can also be estimated from the ensemble. Using the latter method, we confirm that the prominent 1 day (and 0.5 day harmonic) PanSTARRS-1 alias is at the expected sidereal day, i.e., expressed in frequency: $366.25/365.25 = 1.00273$ (solar) day$^{-1}$.

### B1 Normalised frequency difference from the data

Let us start with the distribution of $f_{\text{Gaia}} - f_{\text{PS1}}$ as shown in Fig. B1. Though the distributions are relatively close to Normal, the heavier tail and more peaked centre suggest that for each type we are dealing with hetrocedastic (non-identical) frequency errors. This is not unexpected as the frequency uncertainty depends on various factors: $\sigma_f \propto \sigma_{\text{noise}}/(AT \sqrt{N})$ with $A$ the signal amplitude, $T$ the time series duration and $N$ the number of observations (Kovacs 1980; Cuypers 1987), and the $Gaia$ SOS peak to peak model amplitude ($A$) can vary already with a factor of 2-3. Note that this theory is also the reason we prefer to work with frequency errors: if all factors stay the same the frequency error is (to first order) independent of the actual frequency (or period) of the signal, while the period error will scale as $\sigma_p = \sigma_f f^{-2}$.

Because we intend to use Eq. (B2) for our recovery criterion, we ideally require individual frequency uncertainty estimates for our sources. For the $Gaia$ DR2 RR Lyrae stars (Clementini et al. 2019) the uncertainty of each individual period was estimated from a 100-fold Monte-Carlo sampling of the time series (see section 2.1, Clementini et al. 2016). The PS1 periods estimates in Sesar et al. (2017), $p_{\text{PS1}}$, unfortunately do not have individual error estimates. They do provide a cumulative distribution of $|p_{\text{PS1}} - p_{\text{SDSS}}|$ where $p_{\text{SDSS}}$ is the period derived in Sesar et al. (2010) from SDSS Stripe 82. In Sesar et al. (2010) it is mentioned that their period errors are probably around a few seconds, though in Sesar et al. (2017) it is assumed to be the ‘true’ period, i.e., (much) better than the PS1 derived period.

The absence of individual PS1 period errors gives us two options: (1) Simply normalise the distributions of the top panel of Fig. B1 using its distribution width. For this we use a sigma derived from a fit to the cumulative distribution function$^8$ between $[-10,10]$ cycles century$^{-1}$. This means that we simply divide the frequency difference of all the stars in the RRab and RRc samples by $\sigma_{\Delta f} = 2.33$ and 2.77 cycles century$^{-1}$, respectively, resulting in the bottom panel of

$^8$ The statistical standard deviations of the data in this range were 2.83 and 3.08 cycles century$^{-1}$, respectively. But the associated normal distributions are (much) too narrow. Hence we opted to fit the regular gridded CDFs in the indicated range to give more weight to any discrepancy of the area under the curve.
Figure B2. Normalized frequency difference between Gaia SOS and PS1 around zero and the main aliases, this corresponds to the option 1 (see text): simply dividing the frequency difference distribution by the best fit standard deviation per type.

Fig. B1. Note that in the figures $\sigma_{\Delta f}$ is denoted as $\sigma_{\text{Gaia-PS1}}$ for clarity. The recovery regarding to this criterion is given in Table 2 and shown for the main aliases in Fig. B2. For ease of use we here provide the practical formula to compute the unit-less $\Delta f$ (using frequency units day$^{-1}$), with $f_{\text{alias}}$ the estimated mean value of $\Delta f$ of the ‘alias’ that is analysed:

$$\Delta f = \frac{f_{\text{Gaia}} - f_{\text{PS1}} - f_{\text{alias}}}{\sigma_{\Delta f}}$$  \hspace{1cm} (B4)

with $f_{\text{alias}}$ = mean $\Delta f$ offset of the ‘alias’

with $\sigma_{\Delta f} = \begin{cases} 6.38 \times 10^{-5} \text{ day}^{-1}, \text{if RRab} \\ 7.58 \times 10^{-5} \text{ day}^{-1}, \text{if RRc} \end{cases}$

and recovery if: $|\Delta f| < X$

The $\Delta \rho$ corresponding to a certain value of threshold $X$ times $|\Delta f|$ for period $p$ (in days) can be computed as (with $\sigma_{\Delta f}$ defined above):

$$|\Delta \rho| \approx X \, p^2 \sigma_{\Delta f} \hspace{1cm} (B5)$$

Option (2) is to try to estimate the error distribution for the PS1 periods (as well as Gaia period errors as they might not all be realistic). We have done so by following the example of Sesar et al. (2017): we cross-matched our SOS sample against the Sesar et al. (2010) sources. For these sources we estimated the width of the distribution in frequency difference of PS1 (and Gaia) against SDSS, like in Fig. B1, and used this to derive independent uncertainty estimates. We also tested the estimated PS1 error distribution in combination with the available Gaia SOS per-source uncertainties. In short: any form of re-estimated error distributions did not provide a significantly different frequency recovery with respect to what was found with option 1. Hence we opted to suppress any further details about this alternative procedure and conclude that the simple approach of option 1 is sufficient for this purpose.

This paper has been typeset from a TeX/LaTeX file prepared by the author.