Gaia Data Release 3. Specific processing and validation of all sky RR Lyrae and Cepheid stars: The Cepheid sample

V. Ripepi, G. Clementini, R. Molinaro, S. Leccia, E. Plachy, L. Molnár, L. Rimoldini, I. Musella, M. Marconi, A. Garofalo, et al.

A&A 674, A17 (2023)
https://doi.org/10.1051/0004-6361/202243990

Astronomy & Astrophysics
Special issue

Gaia Data Release 3

Gaia Data Release 3

Specific processing and validation of all sky RR Lyrae and Cepheid stars: The Cepheid sample*

V. Ripepi[1], G. Clementini[2], R. Molinaro[1], S. Leccia[1], E. Plachy[3,4,5], L. Molnár[3,4,5], L. Rimoldini[6], I. Musella[1], M. Marconi[1], A. Garofalo[2], M. Audard[6,7], B. Holl[6,7], D. W. Evans[8], G. Jevardat de Fombelle[6], I. Lecoeur-Taibi[6], O. Marchal[9], N. Mowlavi[6], T. Muraveva[2], K. Nienartowicz[10], P. Sartoretti[11], L. Szabados[3,4], and L. Eyer[7]

[1] INAF-Osservatorio Astronomico di Capodimonte, Salita Moiariello 16, 80131 Naples, Italy
e-mail: vincenzo.ripepi@inaf.it
[2] INAF-Osservatorio di Astrofisica e Scienza dello Spazio, Via Gobetti 93/3, 40129 Bologna, Italy
[3] Konkoly Observatory, Research Centre for Astronomy and Earth Sciences, Eötvös Loránd Research Network, Konkoly Thege M. út 15-17, 1121 Budapest, Hungary
[4] MTA CSFK Lendület Near-Field Cosmology Research Group, Konkoly Thege Miklós út 15-17, 1121 Budapest, Hungary
[5] ELTE Eötvös Loránd University, Institute of Physics, 1117, Pázmány Péter sétány 1/A, Budapest, Hungary
[6] Department of Astronomy, University of Geneva, Ch. d'Ecogia 16, 1290 Versoix, Switzerland
[7] Department of Astronomy, University of Geneva, Chemin Pegasi 51, 1290 Versoix, Switzerland
[8] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[9] Observatoire Astronomique de Strasbourg, Université de Strasbourg, CNRS, UMR 7550, 11 rue de l'Université, 67000 Strasbourg, France
[10] Sednai Sàrl, Geneva, Switzerland
[11] GEPI, Observatoire de Paris, Université PSL, CNRS, 5 place Jules Janssen, 92190 Meudon, France

Received 10 May 2022 / Accepted 17 June 2022

ABSTRACT

Context. Cepheids are pulsating stars that play a crucial role in several astrophysical contexts. Among the different types, the Classical Cepheids are fundamental tools for the calibration of the extragalactic distance ladder. They are also powerful stellar population tracers in the context of Galactic studies. The *Gaia* Third Data Release (DR3) publishes improved data on Cepheids collected during the initial 34 months of operations.
Aims. We present the *Gaia* DR3 catalogue of Cepheids of all types, obtained through the analysis carried out with the Specific Object Study (SOS) Cep&RRL pipeline.
Methods. We discuss the procedures adopted to clean the Cepheid sample from spurious objects, to validate the results, and to re-classify sources with an incorrect outcome from the SOS Cep&RRL pipeline.
Results. The *Gaia* DR3 includes multi-band time-series photometry and characterisation by the SOS Cep&RRL pipeline for a sample of 15 006 Cepheids of all types. The sample includes 4663, 4616, 321, and 185 pulsators, distributed in the Large and Small Magellanic Cloud, Messier 31, and Messier 33, respectively, as well as 5221 objects in the remaining All Sky subregion which includes stars in the Milky Way field and clusters and in a number of small satellites of our Galaxy. Among this sample, 327 objects were known as variable stars in the literature but with a different classification, while, to the best of our knowledge, 474 stars have not been reported as variable before now and therefore they likely are new Cepheids discovered by *Gaia*.

Key words. stars: distances – stars: variables: Cepheids – Magellanic Clouds – Galaxy: disk – surveys – methods: data analysis

1. Introduction

Cepheids made their appearance on the scene when Edward Pigott discovered their first representative, η Aql, in 1784, opening a field of astrophysical research that is still fully active today. The variable stars that are collectively called Cepheids are actually an ensemble of different types which we now separate into three groups: Classical Cepheids (DCEPs, whose prototype is δ Cep), type II Cepheids (T2CEPs), and anomalous Cepheids (ACEPs).

The crucial role played by DECPs resides in their period–luminosity (PL) and period–Wesenheit (PW) relations, which represent fundamental tools at the basis of the extragalactic distance ladder (e.g. Leavitt & Pickering 1912; Madore 1982; Caputo et al. 2000; Riess et al. 2016). However, DCEPs are also important astrophysical objects for stellar evolution and Galactic studies. Indeed, as their pulsational properties (mainly periods) are linked to the intrinsic stellar parameters (effective temperature, mass, luminosity), DCEPs can be used as an independent test for stellar evolution models. Moreover, given their young age (~50–500 Myr) they are preferentially located in the Milky Way (MW) thin disc, and, thanks to precise distances that can be derived from their PL and PW relations, DCEPs can be used to model the disc and trace their birthplaces in the spiral arms, where star formation is most active (e.g. Skowron et al. 2019; Poggio et al. 2021, and references therein). Furthermore, if the chemical abundance of the DCEPs is available, they can be used to trace the metallicity gradient of the MW (e.g.

* Full Tables 5, 6, 9, 10, and 11 are only available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/674/A17

Genovali et al. 2014; Luck & Lambert 2011; Luck 2018; Ripepi et al. 2022a, and references therein).

While DCEPs are luminous, young, and massive ($M \sim 3-11\,M_{\odot}$) stars, T2CEPs are more evolved objects and are older than 10 Gyr, more luminous, and are slightly less massive than RR Lyrae variables ($M \sim 0.55-0.7\,M_{\odot}$, see e.g., Caputo 1998; Sandage & Tammann 2006, for a more extended description of T2CEPs properties). T2CEPs are preferentially metal-poor objects and, as in the RR Lyrae variables, populate the main Galactic components, that is disc, bulge, and halo. T2CEPs pulsate with periods from ∼1 to ∼24 d and are separated into BL Herculis stars (BLHER; periods between 1 and 4 d) and W Virginis (WVIR; periods between 4 and 24 d) stars. Historically, a third class of variables is considered as an additional subgroup of the T2CEP class, namely the RV Tauri (RVTAU) stars (see e.g. Feast et al. 2008, and references therein), with periods from about 20 to 150 d and often less regular light curves. These latter are post-asymptotic giant branch stars on their way to becoming planetary nebulae. This evolutionary phase corresponds to the latest stage in the evolution of intermediate-mass stars and therefore the link between RVTAU and the low-mass WVIR stars should be considered with caution. T2CEPs follow very tight PL and PW relations, especially in the near-infrared (NIR; see e.g. Matsunaga et al. 2011; Ripepi et al. 2015, and references therein) and are therefore excellent distance indicators.

The third Cepheid-like class of pulsating stars is represented by the anomalous Cepheids (ACEPs). These have periods of between approximately 0.4 d and 2.5 d and brighter absolute magnitudes than RR Lyrae stars by 0.3 mag to 2 mag (Caputo et al. 2004, and references therein). ACEP variables are thought to be in their central He-burning evolutionary phase and to have masses of between approximately 1.3 and 2.1 $M_{\odot}$, as well as metallicities lower than $Z = 0.0004$ (corresponding to an iron abundance lower than ∼−1.6 dex, for $Z_{\odot} = 0.0152$; see Caputo 1998; Marconi et al. 2004, for details). Similarly to RR Lyrae stars, the ACEPs can pulsate in the fundamental or first overtone modes and in both cases show well-defined PL and PW relations, especially in the Large Magellanic Cloud (LMC; Ripepi et al. 2014; Soszyński et al. 2015a).

Great advances in the study of variable sources have been made thanks to the *Gaia* mission (Gaia Collaboration 2016a) and its subsequent data releases (DR1, DR2 and EDR3, Gaia Collaboration 2016b, 2018, 2021a; Riello et al. 2021). Indeed, the multi-epoch nature of *Gaia* observations makes the satellite a very powerful tool to identify, characterise, and classify many different classes of variable stars across the whole Hertzsprung-Russell diagram (HRD; see Gaia Collaboration 2019). In *Gaia* DR1, time-series photometry in the *G*-band and parameters derived from the *G* light curves were released for a small number of objects in and around the LMC, including 599 Cepheids of all types and 2595 RR Lyrae stars (Clementini et al. 2016, hereafter Paper I). In 2018, *Gaia* DR2 released more than 550 000 variable sources belonging to a variety of different classes (see Holl et al. 2018), including about 9500 Cepheids of all types and about 140 000 RR Lyrae stars (Clementini et al. 2019, hereafter Paper II). In December 2020, *Gaia* Early Data Release 3 (EDR3; Gaia Collaboration 2021a) published average photometry, parallaxes, and proper motions but no time-series data. Epoch data are now made available with *Gaia* Data Release 3 (DR3, see Gaia Collaboration 2023a), providing multiband time-series photometry for nearly 12 million variable sources (see Eyer et al. 2023).

The Specific Objects Study (SOS) Cep&RRL pipeline (SOS Cep&RRL pipeline hereafter) was developed to validate and characterise Cepheids and RR Lyrae stars observed by *Gaia*. The pipeline has been described in detail in Papers I and II, to which we refer the interested reader. The general properties of the entire sample of variable objects released in *Gaia* DR3 are discussed in Eyer et al. (2023), which also describes the chain of subsequent steps carried out in the general variability analysis before the SOS Cep&RRL processing of the data (see also Holl et al. 2018).

In this paper, we describe the properties of the Cepheids for which time-series data are released in DR3 – along with their characteristic parameters – and that populate the `vari_cepheid` catalogue, which is part of the data release. More specifically, we (i) illustrate the changes we implemented in the SOS Cep&RRL pipeline to process the DR3 photometric and radial velocity (RV) time series of candidate Cepheids provided by the general variable star classification pipelines (Eyer et al. 2023; Rimoldini et al. 2023); (ii) discuss the procedures adopted to clean the sample of Cepheids that are released in the *Gaia* DR3; (iii) present the ensemble properties of the DR3 Cepheids; and (iv) describe the validation procedures adopted to estimate the completeness and contamination of the sample.

A complementary paper (Clementini et al. 2023) describes the SOS Cep&RRL pipeline and the relative results for the RR Lyrae variables.

2. SOS Cep&RRL pipeline: changes from DR2 to DR3

The main steps of the SOS Cep&RRL pipeline for candidate Cepheids are shown in Figs. 1 and 3 of Papers I and II. The procedures used for DR3 are almost the same as for DR1 and DR2, but with some important changes that we describe in the following sections:

2.1. Pipeline changes

1. *Subregions in the sky.* For the processing of the DR2 data, the SOS pipeline subdivided the sky into three regions: two around the LMC and the Small Magellanic Cloud (SMC), respectively, and a third one, called All Sky, including all the remaining stars, which mainly belonged to the MW. This subdivision was needed because of the different observational properties of Cepheids and RR Lyrae stars in the Magellanic Clouds (MCs) and in the MW. Indeed, while Cepheids (and RR Lyrae) in the LMC and SMC are all more or less at the same distance from us within each galaxy – meaning that we can simply use their apparent magnitudes to define their position in and around the reference PL or PW relations –, for the MW we need absolute magnitudes calculated from *Gaia* parallaxes to place these stars on the PL and PW diagrams. These differences, in turn, required different steps in the SOS pipeline. In DR3, we enlarge the regions around the LMC and SMC and introduce two new subregions encircling the Andromeda (M 31) and Triangulum (M 33) galaxies, whose brightest Cepheids are within reach of the *Gaia* mission. These four subregions are listed in Table 1. The fifth subregion is composed of all the remaining sky after excluding the four subregions defined above; for continuity with Paper II, we refer to this fifth subregion as All Sky. The large majority of the stars contained in this subregion are those of MW, with a small fraction of objects belonging to dwarf galaxies that are satellites of our Galaxy.

Table 1. Sky subregions considered by the SOS Cep&RRL pipeline.

Galaxy	RA_{min} J(2000) (deg)	RA_{max} J(2000) (deg)	Dec_{min} J(2000) (deg)	Dec_{max} J(2000) (deg)
LMC	67.50	97.50	−75.000	−62.000
SMC	0.00	30.00	−76.000	−70.000
M 31	8.75	12.75	39.667	42.833
M 33	22.90	24.00	30.000	31.300

Notes. The fifth subregion, called All Sky, comprises all the sky except for the four subregions listed in the table.

2. *Treatment of multi-mode DCEPs.* To avoid spurious detections of multi-mode DCEPs, we only searched for more than one pulsation mode in the time-series of stars with a number of epochs greater than or equal to 40. In addition, we introduced an analysis of the residuals after the fit of the G light curve with just one pulsation mode, and only retained objects showing a dispersion larger than or equal to 0.025 mag as potential multi-mode (a similar procedure, although with a larger scatter, is adopted for RR Lyrae stars, see Clementini et al. 2023).
3. *RV curves treatment.* As RV time-series are published for a small sample of Cepheids and RR Lyrae stars (see also Sartoretti et al. 2022) as part of DR3, a new module of the pipeline analyses the RV curves, providing average RV values, peak-to-peak amplitudes, and the epoch of minimum RVs (see Clementini et al. 2023, for more details).
4. *Update of the PL and PW relations.* We have updated the PL and PW relations that are used in the pipeline, adding those needed to deal with M 31 and M 33 data. As all these relations are significantly changed with respect to DR2, we describe them in detail in Sect. 2.2.
5. *Errors with bootstrap.* We applied a bootstrap technique to estimate the uncertainties on all Cepheid parameters published in DR3. Specifically, to estimate the uncertainties on the Fourier fit parameters (period, amplitudes and phases) and on all the other quantities characterising the light and RV curves (e.g. mean magnitudes, mean RV, peak-to-peak amplitudes, etc.), the input data were randomly re-sampled (allowing data point repetitions) and all parameters were recalculated on each simulated sample. This procedure was repeated 100 times, and the respective uncertainty was estimated for each parameter by considering the robust standard deviation (1.486·MAD) of the distributions obtained with the bootstrap method. A similar procedure was applied to estimate the uncertainties on all other released quantities, such as the metallicity and the Fourier parameters.
6. *Fine tuning of the ROFABO outlier rejection operator.* The photometric and RV time-series are inserted in the SOS Cep&RRL pipeline after undergoing a chain of routines that elaborate the observations to obtain a standard time, magnitudes, RVs, and relative uncertainties; these constitute the input time-series data. Among these operators, standard outlier rejection techniques are applied to remove as many bad points as possible, without affecting the scientific information contained in the time series. To improve the rejection of outliers from the time series of Cepheids and RR Lyrae stars, the SOS Cep&RRL pipeline adopted a customised configuration set of parameters for the 'Remove Outliers on both FAint and Bright sides Operator' (ROFABO); see (Eyer et al. 2023) routine. To determine the best configuration parameters of ROFABO allowing to maximise the rejection of bad points while preserving good measures (specifically for Cepheids and RR Lyrae stars), we used the SOS pipeline to process a sample of hundreds of time series affected by different kinds of outliers, together with time series not presenting obvious bad measures. A specific ROFABO function for the SOS Cep&RRL pipeline with configuration parameters fine-tuned as described above was then added to the whole operator chain.

2.2. New PL and PW relations employed in the SOS Cep&RRL pipeline

A number of significant changes with respect to DR2 were introduced in the branch of the SOS Cep&RRL pipeline that processes the candidate Cepheids. Specifically, (1) we adopted new PL and PW relations directly derived from the *Gaia* data, while in DR2 we used photometry in the Johnson system transformed into the *Gaia* bands (see Sect. 3.2 of Paper II); and (2) for DR3 we used the new Wesenheit magnitudes defined by Ripepi et al. (2019), that is $W(G, G_{BP} - G_{RP}) = G - 1.90(G_{BP} - G_{RP})$, which replaced the $W(G, G - G_{RP})$ magnitudes used in DR2 (see Eq. (5) in Paper II).

To calculate the PL and PW relations we gathered Cepheids of all types known from the literature and used the SOS pipeline to analyse their light curves in the *Gaia* bands to obtain periods and intensity-averaged magnitudes in the G, G_{BP}, and G_{RP} bands (see Sect. 2.1 of Clementini et al. 2016, for details on how the SOS Cep&RRL pipeline determines these quantities). The calculation of the PL and PW relations required different approaches for the different subregions as specified in the following:

1. *LMC and SMC.* For both galaxies, we adopted 9649 DCEPs and 262 ACEPs from Soszyński et al. (2017) for reference, while the T2CEPs (338 objects) were taken from Soszyński et al. (2018). We retrieved the DR3 time-series photometry of these stars and used the SOS Cep&RRL pipeline to derive periods and intensity-averaged magnitudes in the G, G_{BP}, and G_{RP} bands for the objects with more than 20 epochs (we only wanted good light curves to build the reference PL and PW relations). We discarded all objects for which the SOS and the literature periods did not agree to within 1%. After all these steps, we remained with the number of stars listed in the last column of Table 2. Linear PL and PW relations were derived from them using the `python LtsFit` package (Cappellari et al. 2013), which has a robust outlier-removal procedure.
2. *M 31 and M 33.* Given the faint apparent magnitude of the Cepheids in these distant galaxies, for reference we adopted the PL and PW relations that we calculated for the LMC (see above) – which have the lowest scatter – and simply re-scaled the zero points to take into account the difference in distance moduli between the LMC and M 31/M 33. For the latter, we adopted $\mu_{M31} = 24.40$ mag (the typical value for the M 31 globular clusters, see Perina et al. 2009) and $\mu_{M33} = 24.57$ mag (Conn et al. 2012). However, a different choice for the distance moduli of M 31 and M 33 would not affect our analysis and results, as we used rather large magnitude intervals (up to 0.6−0.8 mag) around the PL and PW relations to select the candidate Cepheids.
3. *All Sky.* The first step consisted in collecting a reliable sample of Cepheids of all types in the MW. To this aim, we adopted the most updated lists of Cepheids available as of October 2020, namely Ripepi et al. (2019, all types);

Table 2. Coefficients and scatter values of the PL and PW relations used for the sky regions including the LMC and SMC.

Type	Mode	α	β	σ	Band	N
			LMC			
DCEP	F	17.333 ± 0.010	-2.793 ± 0.015	0.169	G	2477
DCEP	1O	16.890 ± 0.007	-3.280 ± 0.020	0.197	G	1775
DCEP	F	15.998 ± 0.005	-3.317 ± 0.007	0.075	$W(G, G_{BP} - G_{RP})$	2447
DCEP	1O	15.521 ± 0.003	-3.476 ± 0.009	0.081	$W(G, G_{BP} - G_{RP})$	1745
ACEP	F	18.022 ± 0.026	-2.930 ± 0.17	0.238	G	102
ACEP	1O	17.248 ± 0.062	-3.340 ± 0.290	0.227	G	44
ACEP	F	16.790 ± 0.021	-3.080 ± 0.140	0.180	$W(G, G_{BP} - G_{RP})$	97
ACEP	1O	16.201 ± 0.051	-3.500 ± 0.240	0.184	$W(G, G_{BP} - G_{RP})$	43
T2CEP	–	18.731 ± 0.033	-1.905 ± 0.036	0.275	G	205
T2CEP	–	17.516 ± 0.019	-2.577 ± 0.019	0.138	$W(G, G_{BP} - G_{RP})$	197
			SMC			
DCEP	F ($P < 2.95$)	17.935 ± 0.012	-3.155 ± 0.046	0.226	G	1911
DCEP	F ($P > 2.95$)	17.757 ± 0.026	-2.830 ± 0.033	0.251	G	843
DCEP	1O	17.260 ± 0.007	-3.185 ± 0.029	0.256	G	1790
DCEP	F ($P < 2.95$)	16.711 ± 0.009	-3.627 ± 0.037	0.172	$W(G, G_{BP} - G_{RP})$	1880
DCEP	F ($P > 2.95$)	16.592 ± 0.017	-3.382 ± 0.021	0.156	$W(G, G_{BP} - G_{RP})$	839
DCEP	1O	16.133 ± 0.051	-3.595 ± 0.021	0.177	$W(G, G_{BP} - G_{RP})$	1755
ACEP	F	18.255 ± 0.024	-2.430 ± 0.160	0.171	G	79
ACEP	1O	17.633 ± 0.050	-3.450 ± 0.290	0.186	G	43
ACEP	F	17.161 ± 0.025	-3.020 ± 0.160	0.169	$W(G, G_{BP} - G_{RP})$	77
ACEP	1O	16.712 ± 0.054	-3.540 ± 0.310	0.198	$W(G, G_{BP} - G_{RP})$	40
T2CEP	–	19.105 ± 0.098	-2.140 ± 0.100	0.372	G	42
T2CEP	–	17.843 ± 0.052	-2.505 ± 0.054	0.190	$W(G, G_{BP} - G_{RP})$	42

Notes. All relations are of the form mag $= \alpha + \beta \cdot \log(P)$. The relations for M 31 and M 33 are not shown because they are the same as for the LMC, but scaling the zero points according to the distance moduli and adopting 24.40 mag for M 31, 24.57 mag for M 33, and 18.49 mag for the LMC (see text for details).

Skowron et al. (2019, only DCEPs); Soszyński et al. (2020, including DCEPs, ACEPs, and T2CEPs); and Chen et al. (2020, only DCEPs and T2CEPs, with the former not classified according to the pulsation mode and the latter in the different T2CEPs subtypes). After removing overlaps between catalogues, we filtered the resulting list of objects adopting the *Gaia* EDR3 astrometry. In particular, we retained only objects with relative error on parallax better than 20% and RUWE < 1.4[1]. This choice was driven by the need to clean the sample for contaminants, particularly binaries, which are easily spotted in the PW diagram as they are usually significantly subluminous compared to Cepheids. At the end of this procedure, we were left with a 'clean' sample of All Sky Cepheids, for which numbers divided into various types and/or modes are provided in the last column of Table 3. We note that the T2CEP sample only includes BLHER and WVIR stars, because the physical connection with RVTAU stars is questioned (see Introduction). Table 3 shows that for ACEPs, we have only four stars in each pulsation mode. Therefore, for ACEPs, we adopted the slope of the LMC PW relation and fitted only the zero point. For DCEPs and T2CEP, the number of objects is instead sufficient to obtain good PWs. In fitting the relations to preserve the symmetry of the uncertainties on the parallax as much as possible, we adopted the astrometry-based luminosity (ABL Feast & Catchpole 1997; Arenou & Luri 1999):

$$\mathrm{ABL} = 10^{0.2\,W} = 10^{0.2(\alpha+\beta\log P)} = \varpi 10^{0.2w-2}, \quad (1)$$

where W and w are the absolute and apparent Wesenheit magnitudes and ϖ is the parallax. The fitting procedure is similar to that adopted in Ripepi et al. (2019, 2022a) and is not repeated here. The resulting coefficients for the PW relations of All Sky Cepheids of different type and/or mode are summarised in Table 3.

The PL and PW relations described above represent a fundamental tool of the Cepheid branch in the SOS Cep&RRL pipeline, as we use them for a first classification of the candidate Cepheids of different types and/or modes (see Papers I and II for full details on the pipeline). In practice, we define a band across each PL and PW relation, as $\pm n \times \sigma$, where σ is the dispersion of each relation. For DR3, we used 1σ for the ACEPs, 4σ (10σ for the ABL formalism) for the DCEPs, and 2σ for the T2CEPs. These values were calibrated using the LMC, SMC, and All Sky samples of known Cepheids defined above so as to minimise the overlap between contiguous variable types and modes, and at the same time maximise the number of correct classifications.

3. Application of the SOS Cep&RRL pipeline to the DR3 data: cleaning of the sample

The *Gaia* DR3 data analysed by the SOS Cep&RRL pipeline consist of G and integrated G_{BP} and G_{RP} time-series photometry collected between 25 July 2014 and 28 May 2017, spanning a period of 34 months (for reference, DR2 was based on 22 months of observations). In addition to the time-series photometry, for DR3 we also analysed the RV time series (see Sartoretti et al. 2022, for the general procedures used to measure RV in *Gaia*) for a selected sample of 799 Cepheids

[1] Section 14.1.2 of "*Gaia* Data Release 2 Documentation release 1.2"; https://gea.esac.esa.int/archive/documentation/GDR2/

Table 3. Same as in Table 2, but for All Sky Cepheids.

Type	Mode	α	β	$\sigma_{\rm ABL}$	Band	N
			All sky			
DCEP	F	-2.744 ± 0.045	-3.391 ± 0.052	0.015	$W(G, G_{\rm BP} - G_{\rm RP})$	898
DCEP	1O	-3.224 ± 0.028	-3.588 ± 0.065	0.021	$W(G, G_{\rm BP} - G_{\rm RP})$	416
ACEP	F	-1.717 ± 0.025	-3.080 fixed	0.010	$W(G, G_{\rm BP} - G_{\rm RP})$	4
ACEP	1O	-2.220 ± 0.061	-3.500 fixed	0.013	$W(G, G_{\rm BP} - G_{\rm RP})$	4
T2CEP	–	-1.224 ± 0.039	-2.542 ± 0.088	0.041	$W(G, G_{\rm BP} - G_{\rm RP})$	264

Notes. The reported values of the scatter refer to the residuals around the fit in the ABL formalism. For the ACEPs the slopes are fixed to those of the LMC.

of all types. Among these, 798 are Cepheids present in the `vari_cepheid` catalogue, while one object, previously classified as RR Lyrae, was found to be a DCEP_MULTI variable (source_id = 5861856101075703552) and is present in the `vari_rrlyrae` catalogue (see Clementini et al. 2023, for full details).

The general treatment of the light and RV curves and the processing steps that precede the SOS Cep&RRL pipeline are schematically summarised by Holl et al. (2018) and Eyer et al. (2023). In particular, the SOS Cep&RRL pipeline processed candidate Cepheids (and RR Lyrae stars)[2] identified as such by the supervised classification of the general variability pipeline (see Eyer et al. 2023; Rimoldini et al. 2023, for details) with various probability levels. In order to maximise the number of DCEPs known from the literature that are recovered, we considered classification candidates that also have low probability levels. Among the Cepheid candidates, the SOS Cep&RRL pipeline only retained objects with at least 12 measurements in the G-band for analysis, while the RV time series were only processed for sources with seven RV measurements or more.

At the end of this first processing, we obtained a sample of about 1 million Cepheid candidates of all types. Among them, only about 5000 were in M 31 and M 33. To reduce the huge number of candidate Cepheids in the LMC, SMC, and particularly in the All Sky sample to more manageable numbers, we applied the following series of filters:

1. *Separation of known or suspected Cepheids in the literature.* From the whole sample of Cepheid candidates, we separated sources that are known or suspected Cepheids of all types in the literature. This was done for each of the five subregions defined in Table 1. This first step was necessary to avoid filtering out possible good objects in the following cutting steps. The majority of the literature Cepheids were then validated by visual inspection as described in Sect. 4. For the known Cepheids in the LMC and SMC, we retained those mentioned in Sect. 2.2, while to the All Sky known Cepheids we added all the objects classified as Cepheids as of February 2021 in the SIMBAD database (available at CDS, Centre de Donnés astronomiques de Strasbourg, Wenger et al. 2000). For M 31 and M 33, we adopted the samples by Kodric et al. (2018) and Pellerin & Macri (2011), respectively. After eliminating overlaps, the overall literature sample within the one million candidates contains about 16 000 objects. These literature Cepeheids were elected for visual inspection, with the exclusion of about 9000 Cepheids in the MCs, for which the OGLE classification is already reliable.
2. *Goodness of the light curves.* We filtered the remaining sample based on uncertainties on the light curve parameters. More specifically, we applied the cuts listed in Table 4. This allowed us to filter out about 10% of the sources and, in particular, to reduce the All Sky sample to approximately 667 000 sources.
3. *Probability of the classifiers (LMC and SMC samples).* As the number of candidates remaining from the previous steps was still too large for the LMC and SMC, we reconsidered the probability adopted in selecting Cepheid candidates from the classifiers of the general variability pipeline. Again adopting the highly reliable sample of literature objects in the MCs, for each classifier we calculated the probability that returns 95% of the known Cepheids. This procedure was very effective, leaving us with only about 2500 new Cepheid candidates in the two MCs.
4. *Filtering of aliasing periods (M 31 and M 33 samples):* As discussed in Holl et al. (2023), instrumental effects produce false variable sources with typical periods which are strictly correlated with the position on the sky of the objects. These effects are particularly disturbing in the case of M 31 and M 33, given that for these galaxies we have only the G-band photometry for reference.
   Luckily, as the range in coordinates spanned by the M 31 and M 33 data is rather small, the aliases correlated with the position on the sky produce narrow peaks in period. A histogram of the periods provides five and seven narrow period peaks in M 31 and M 33, respectively. Filtering the stars in those intervals left us with 1923 candidate Cepheids in M 31 and 1332 stars in M 33 for further verification.
5. *Filtering on number of epochs, limiting magnitude, amplitude, and period (All Sky sample).* As the All Sky sample resulting from the previous filtering was still too large, we applied the following further filtering: $G < 19.0$ mag, amp(G) > 0.15 mag, maximum period $P_{\rm max} = 100$ days and number of epochs in the G-band > 30. The selection on the number of epochs was motivated by the need to measure accurate periods, while the limits in magnitude and amplitude allowed us to significantly reduce the number of spurious variability detections caused by instrumental effects (see e.g. Holl et al. 2023) which are more likely among faint sources, whose $G_{\rm BP}$ and $G_{\rm RP}$ magnitudes are also in most cases not accurate. The cut in period is justified because very few Cepheids, that is, both DCEPs and RVTAU, are expected to exceed a period of 100 days. In the end, the above filtering left us with 166k candidates for further analysis.
6. *Machine learning filtering.* While the sample in the MCs was small enough to be checked visually, the All Sky sample was still too large. We therefore applied an additional filtering based on machine learning techniques. We adopted

[2] We recall that the RR Lyrae stars are discussed in a companion paper (Clementini et al. 2023).

a supervised classification method based on a reliable training set. To build the training set, we adopted a sample of Cepheids of all types similar to that described in Sect. 2.2 – but not limited in relative parallax error – to increase the statistics, including about 4100 objects in total. To this sample, we added about 2250 contaminants of different types, including RR Lyrae stars, long-period variables, eclipsing binaries, and so on taken from objects for which the general classification pipeline assigns a very high probability of belonging to the given class. In addition, we verified that the vast majority of the contaminants were also known in the literature with a classification in agreement with that assigned by the classification pipeline.

After establishing the training set, we defined the input attributes for the machine learning algorithm. Based on parameters that are already used by the SOS Cep&RRL pipeline, we adopted: the first periodicity, the second periodicity (if any), the absolute magnitudes in all bands, the absolute Wesenheit magnitudes (in G, $G_{BP}-G_{RP}$), the amplitudes in all bands, the amplitude ratios ($\mathrm{amp}(G_{BP})/\mathrm{amp}(G_{RP})$; $\mathrm{amp}(G_{BP})/\mathrm{amp}(G)$; $\mathrm{amp}(G)/\mathrm{amp}(G_{RP})$), colours ($G_{BP}-G_{RP}$; $G_{BP}-G$; $G-G_{RP}$) and the Fourier parameters (R_{21}; R_{31}; ϕ_{21}; ϕ_{31}). The classes fed to the algorithm were: ACEP_F, ACEP_1O, DCEP_F, DCEP_1O, DCEP_MULTI, BLHER, WVIR, RVTAU, and OTHER, where the last tag included all the non-Cepheid objects. To execute the machine learning procedure we used the `H2O` platform[3]. After ingesting the training set, we divided it into training and validation sets in proportions of 85% and 15%, respectively. We then carried out several tests to find the best model for our case amongst those offered by the `H2O` package. The model that returned the largest percentage of precision in detecting the right classes and modes was found to be the `XGBOOST` algorithm.

We applied this model to the sample of 166k All Sky candidate Cepheids returned by the selection described in point (5) above, obtaining a probability of belonging to one of the classes mentioned above for each candidate. A quick visual examination of samples of light curves for objects with a probability larger than 50% of being Cepheids of any type revealed that there were no reliable candidates with probability <90%. We therefore considered only candidates with probability larger than 90%, giving a total of 10 273 sources. Finally, to further restrict the number of stars for visual inspection, we adopted the peak-to-peak amplitudes, requiring that: $1.3 \leq \mathrm{amp}(G_{BP})/\mathrm{amp}(G_{RP}) \leq 2.0$; $1.1 \leq \mathrm{amp}(G)/\mathrm{amp}(G_{RP}) \leq 1.5$ and $100\times\sigma G/\mathrm{amp}(G) \leq 2.0$. These broad limits include the large majority of bona fine Cepheids according to tests carried out on the training set adopted for the machine learning procedure. After applying this last filtering, we were left with 7349 stars for subsequent visual inspection.

In summary, at the end of the whole filtering procedure described in this section, we were left with about 20 100 Cepheids for subsequent visual inspection in order to validate the classification provided by the SOS Cep&RRL pipeline.

4. Correction of the SOS Cep&RRL pipeline classification

As mentioned in the previous section, a number of sources were selected for further inspection to verify their classification. Different procedures were adopted for the LMC/SMC, M 31/M 33, and All Sky samples because of the different characteristics of the available data. More specifically, for the LMC and SMC, the literature samples have a robust classification and we already knew from DR2 that the SOS Cep&RRL pipeline provides reliable classifications for the Cepheids in these two galaxies. For this reason, we did not visually check the known Cepheids in the MCs, but only the new candidates. On the contrary, the classification of Cepheids in M 31 and M 33 required careful validation because of the low signal-to-noise ratio (S/N) of the *Gaia* data and the much less established literature for Cepheids in these galaxies. Concerning the All Sky sample, the literature sample is likely contaminated by both non-Cepheids and incorrect classifications (i.e. incorrect Cepheid types and pulsation modes) because their classification in these two respects does not rely on solid distances but mainly on the analysis of the light curve shapes. For this reason, we checked all the known Cepheids in addition to the new candidates for the All Sky sample.

Table 4. Constraints on the results from the light-curve fitting.

Parameter
$0.0 < G \leq 22.0$ mag
$0.0 < G_{BP} \leq 22.0$ mag
$0.0 < G_{RP} \leq 22.0$ mag
$0.0 < \sigma G \leq 0.5$ mag
$0.0 < \sigma G_{BP} \leq 0.5$ mag
$0.0 < \sigma G_{RP} \leq 0.5$ mag
$0.0 < \mathrm{amp}(G) \leq 2.5$ mag
$0.0 < \mathrm{amp}(G_{BP}) \leq 2.5$ mag
$0.0 < \mathrm{amp}(G_{RP}) \leq 2.5$ mag
$0.0 < \sigma\mathrm{amp}(G)/\mathrm{amp}(G) \leq 1.0$
$0.0 < \sigma\mathrm{amp}(G_{BP})/\mathrm{amp}(G_{BP}) \leq 1.0$
$0.0 < \sigma\mathrm{amp}(G_{RP})/\mathrm{amp}(G_{RP}) \leq 1.0$
$0.0 < R_{21} \leq 2.0$ mag
$0.0 < R_{31} \leq 2.0$ mag
$0.0 < \sigma R_{21}/R_{21} \leq 1.0$
$0.0 < \sigma R_{31}/R_{31} \leq 1.0$
$0.0 < \sigma\phi_{21}/\phi_{21} \leq 1.0$
$0.0 < \sigma\phi_{31}/\phi_{31} \leq 1.0$

4.1. Visual inspection of the Gaia DR3 light curves

In general, for each star, we evaluated the shape of the light curves in all *Gaia* bands, the position in the period–Fourier parameters diagrams ($P-R_{21}$; $P-R_{31}$; $P-\phi_{21}$; $P-\phi_{31}$), the position on the PL and PW relations, and the amplitude ratios $\mathrm{amp}(G_{BP})/\mathrm{amp}(G_{RP})$ and $\mathrm{amp}(G)/\mathrm{amp}(G_{RP})$. In the case of negative parallax values, the ABL function was used. We adopted the very useful 'OGLE atlas of light curves'[4] as reference for the shapes of the light curves of Cepheids of all types. For Cepheids in M 31 and M 33, we only have the G-band light curves and the position in the PL relation for reference, as for $G \sim 20-21$ mag the G_{RP} and G_{BP} magnitudes are often missing or totally unreliable, hence the Wesenheit relation was not usable in most cases.

In the All Sky sample, the major difficulties were to distinguish DCEP_1O from first overtone RR Lyrae stars with periods smaller than 0.4 days wherever light curves were not very well defined and parallaxes had relative errors of greater than 10%−20%. Similarly, in some cases, it was difficult to separate

[3] h2o.ai

[4] http://ogle.astrouw.edu.pl/atlas/index.html

Table 5. Re-processing of the *Gaia* data for DCEP_MULTI objects not detected as such by the SOS Cep&RRL pipeline.

Source_id	P_L (days)	σP_L (days)	P_S (days)	σP_S (days)	P_S/P_L	Modes
5864639514713019392	0.216830	1.25e–07	0.172544	4.73e–07	0.796	1O/2O
5853820767014992128	0.236592	1.08e–04	0.188584	2.80e–05	0.797	1O/2O
5423800601092727168	0.238321	3.74e–04	0.190348	3.63e–05	0.799	1O/2O
5601418217705666560	0.239664	3.99e–07	0.191709	8.04e–05	0.800	1O/2O
4313476032410287104	0.242839	1.10e–02	0.193298	8.99e–05	0.796	1O/2O
5409512756735301120	0.247992	3.75e–07	0.197926	1.27e–06	0.798	1O/2O
5939019827046790272	0.249715	8.01e–04	0.198971	4.43e–06	0.797	1O/2O
5941658375763435648	0.254389	6.52e–07	0.202699	2.43e–06	0.797	1O/2O
5254261818006768512	0.262113	1.61e–06	0.209564	4.74e–06	0.800	1O/2O
3314887198215151104	0.263087	6.39e–07	0.199706	2.39e–06	0.759	F/1O

Notes. The different columns show: source identification, longest and shortest pulsation periods with relative errors, period ratio, classification (for brevity we use F, 1O, 2O to indicate the fundamental, first and second overtones, respectively). Only the first ten lines are shown to guide the reader about the table content. The entire version of the table will be published at CDS.

DCEP_1O and ACEP_1O with periods ~0.7−0.8 days. Even more challenging was to distinguish ACEP_F from ab-type RR Lyrae for periods of around 0.6 days and from DCEP_F in the period range 1.0−1.4 days. These difficulties arose mainly from the very similar shape of the light curves for these types of variable stars, which can only be distinguished based on fine details of the light curves, such as humps and bumps, which are not always clearly visible. Also, WVIR and DCEP_F can be confused when light curves are noisy and parallaxes inaccurate. In all these cases, the Fourier parameters also provide ambiguous results because they stem directly from the light curve shape.

A main source of contamination is given by contact binary stars, whose light curves mimic those of the overtone Cepheids and, to some extent, also those of the WVIR variables. To mitigate this problem, we always inspected the light curves folded according to once and twice the period provided by the SOS Cep&RRL pipeline. In this way, it was often possible to identify stars for which there was a small but detectable difference between the light curve minima. This check, in conjunction with the amplitude ratios amp(G_{BP})/amp(G_{RP}) and amp(G)/amp(G_{RP}) – which for binaries tend to assume values close to unity (see Sect. 4 in Paper II), while much larger for pulsating stars (see e.g. Table 4 in Ripepi et al. 2019) –, allowed us to detect and reject the large majority of potential contaminants that are contact binaries.

During visual inspection, many objects classified as DCEP_1O variables by the SOS Cep&RRL pipeline were found to show larger scatter than other sources of the same magnitude, leading us to suspect they might be missed multi-mode objects. As discussed in detail in Sect. 4.2, we searched for secondary periodicities in the light curves of the stars in this sample, finding that many of them are actually multi-mode pulsators. A large fraction of them were missed simply because of the overly strict constraint on the number of epochs and scatter in the light curves introduced in the SOS Cep&RRL pipeline (see point 2 of Sect. 2), which allowed us to minimise the number of spurious detections but at the same time also prevented us from detecting many genuine multi-mode pulsators.

4.2. Multi-mode Cepheids

DCEPs in the All Sky sample that, during visual inspection, were suspected to be multi-mode pulsators were further investigated by analysing their light curves with software external to the SOS Cep&RRL pipeline. In particular, we used the `Period04` package (Lenz & Breger 2005) for a first selection of the most promising candidates and to determine their periodicities. We then used a custom program written in `Python` to carry out the non-linear fitting with truncated Fourier series, the prewhitening of the first periodicity, and then the fitting of all periodicities together. In close similarity with the SOS Cep&RRL pipeline, we finally determined the period uncertainties with a bootstrap procedure. The re-processing led to the identification of 109 DCEP_MULTI variables in addition to the 86 DCEP_MULTI for which the SOS Cep&RRL pipeline provided the correct classification. The list of additional DCEP_MULTI variables and their periods are provided in Table 5 with relative errors. The Petersen diagram (period ratios vs longer period) for the DCEP_MULTI in the All Sky sample is shown in Fig. 1, where the loci occupied by the different period ratios are taken from Soszyński et al. (2020).

4.3. Final classification

The processing of the SOS pipeline along with the validation, cleaning, and re-classification procedures described in the previous sections produced a final catalogue of 15 021 Cepheids of all types, which populate the `vari_cepheid` table in the *Gaia* DR3 archive. Despite our efforts to clean the sample from spurious objects, after a deeper analysis, 15 sources turned out to be non-Cepheid variables (these objects are listed with the new classification in Table 6), bringing the total number of bona fide Cepheids of all types in *Gaia* DR3 to 15 006.

In total, we changed the SOS Cep&RRL classification of 1160 stars. This corresponds to about 8% of the total sample. The new classifications are given in Table 6. Taking into account all re-classifications, in Table 7 we report the breakdown of the DR3 Cepheids by type in the different subregions in which we divided our sample.

Comparison with the literature, which is discussed in more detail in Sect. 6.1, along with a cross match with the SIMBAD database[5] (Wenger et al. 2000) allowed us to calculate the number of Cepheids of any type already known in the literature, the number that are classified as variables but of non-Cepheid type, and the number of new discoveries. The result of this exercise is reported in the last line of Table 7. The largest number of new or

[5] `http://simbad.u-strasbg.fr/simbad/`

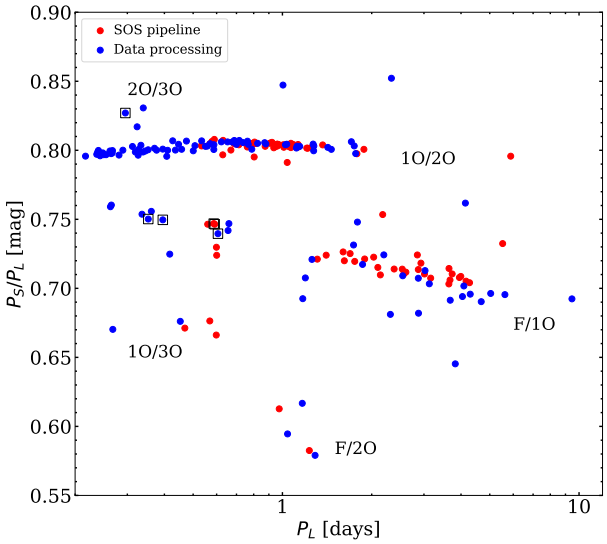

Fig. 1. Petersen diagram for confirmed DCEP_MULTI objects published in the *Gaia* DR3 catalogue (red filled circles) and for additional DCEP_MULTI objects detected in the re-processing of the data (blue filled circles). P_L and P_S represent the longest and shortest pulsation periods of the multi-mode object. Labels show the typical location of the different multi-mode pulsation combinations identified in these sources. Black squares mark six objects known in the literature as ARRDs (see Sect. 6.4).

reclassified objects belongs to the All Sky sample, but we note that many new Cepheids were also discovered in M 31 and M 33.

5. Properties of the Cepheids in the *Gaia* DR3

A summary of the parameters provided by the SOS Cep&RRL pipeline that form the entries of the `vari_cepheid` table is provided in Table 8. In the following subsections we describe the main properties of the Cepheids in *Gaia* DR3.

Examples of light and RV curves for DCEPs of different pulsation modes are shown in Fig. A.1. Similarly, Fig. A.2 displays the *Gaia* time series for the prototypes of the T2CEP classes, namely BL Her, W Vir, and RV Tau. Finally, Fig. A.3 shows the light and RV curves for ACEP_F and ACEP_1O variables.

5.1. Number of epochs

An important quantity affecting the quality of the results is the number of epochs in the light and RV curves. This feature strictly depends on the position of a specific object in the sky, as the *Gaia* scanning law is extremely non-uniform (see Gaia Collaboration 2016a). The more epochs available for the analysis of the time series, the more precise the determination of the periods, amplitudes, and so on. Figure 2 shows histograms with the number of epochs in G band for each subsample (the number of epochs in G_{BP} and G_{RP} provides similar distributions). Restricted regions in the sky such as the SMC, M 31, and M 33 show narrower intervals of epochs than both the All Sky and LMC samples. The latter shows an extended tail with many DCEPs having more than 140 epochs because they are located in the region of the EPSL (Ecliptic Pole Scanning Law) which was covered continuously during the first 28 days of the *Gaia* mission (see Gaia Collaboration 2016b). Unfortunately, for M 31 and M 33, which are the most difficult subregions because of the large distance, the number of epochs is small (less than 40 on average for M 31), making it difficult to study the Cepheids in these systems.

Concerning the RVs, the number of useful epochs for the Cepheids with RV time series published in DR3 is displayed in Fig. 3. There are 15 and 9 DCEPs with RV time series in the LMC and SMC, respectively. The rest of the objects belong to the All Sky sample.

5.2. Spatial distribution

The spatial distribution of Cepheids of different types in the All Sky sample is shown in Fig. 4. The different distributions reflect the progenitor stellar populations of the different types: DCEPs are concentrated in the Galactic disc, as expected for a young population[6]; ACEPs, which are intermediate-age objects, are preferentially located in the Galactic halo; T2CEPs are present in almost all Galactic components, namely disc, thick disc, halo, and bulge, where they are more concentrated. The spatial distributions of the LMC and SMC Cepheids are shown in Fig. 5. Also, in these galaxies, the DCEPs trace the young populations inhabiting the LMC bar and the spiral arms (see e.g. Ripepi et al. 2022b, and references therein) as well as the body and the wing of the SMC (see e.g. Ripepi et al. 2017, and references therein). The spatial distributions of ACEPs and T2CEPs are more sparse and connected with the spheroids describing the intermediate-old populations in both galaxies (see e.g. Gaia Collaboration 2021b).

Figure 6 shows the spatial distribution of the Cepheids in M 31 and M 33. In this case, we mostly find DCEPs, except for two RVTAU stars detected in M 31. The spatial distribution of the M 31 DCEPs closely follows the galaxy spiral arms, where young stars are expected, while the DCEP distribution in M 33 is less ordered because of the different morphology of the galaxy and the different viewing angle from the Sun.

5.3. Fourier parameters

An important product of the SOS Cep&RRL pipeline is the Fourier parameters R_{21}, R_{31}, ϕ_{21}, and ϕ_{31} which represent an important tool to distinguish the different types of variables. The Fourier parameters for the Cepheids in the All Sky sample are shown in Fig. 7, separated in different panels for DCEPs, ACEPs, and T2CEPs in the interest of clarity. The different distributions occupy the expected location for each variable type, confirming the efficacy of our classification. The same kind of considerations are valid for the LMC and SMC as shown in Figs. B.1 and B.2. In the cases of M 31 and M 33 (Figs. B.3 and B.4), the Fourier parameters show a less clear morphology, because light curves are mostly noisy, because we are analysing objects with magnitudes at the limits of *Gaia* capabilities. Nevertheless, it is remarkable that especially for M 31, the morphology of the $P - R_{21}$ and $P - \phi_{21}$ relations is similar to that displayed by the much closer All Sky, LMC, and SMC samples.

5.4. PL and PW diagrams

Figure 8 shows the PW relations for the All Sky sample; shown separately for different Cepheid types and modes. These relationships were adopted by the SOS Cep&RRL pipeline to select and classify the different types of Cepheids, as discussed in

[6] In the figure we have removed from the All Sky sample objects physically bound to the LMC and SMC (see Sect. 5.9).

Table 6. Reclassification of objects incorrectly classified by the SOS Cep&RRL pipeline.

Source_id	RA (deg)	Dec (deg)	Class	Comment	Region
2422853521974230400	0.073909	−10.221463	ACEP_F	WRONG_CLASS	All Sky
565137161224290944	1.836580	80.297101	DCEP_F	WRONG_CLASS	All Sky
419703349473530240	4.678347	54.039237	WVIR	WRONG_CLASS	All Sky
2367033515654862720	4.817180	−18.075243	ACEP_F	WRONG_CLASS	All Sky
430629986799994880	5.630325	63.033041	DCEP_1O	WRONG_MODE	All Sky
431184518613946112	6.407203	64.229891	DCEP_MULTI	WRONG_MODE	All Sky
382372112206462336	7.230740	43.033510	BLHER	WRONG_CLASS	All Sky
4906654274849806592	7.296960	−57.939912	ACEP_1O	WRONG_CLASS	All Sky
4980356188527065472	7.672437	−44.272952	ACEP_F	WRONG_CLASS	All Sky
375318264077848448	11.680155	42.092860	DCEP_1O	WRONG_MODE	M 31

Notes. The column 'class' provides the correct Cepheid type/mode; 'comment' describes whether or not the class or the pulsation mode are incorrect; 'region' shows the sky region to which the particular star belongs. The equatorial coordinates are given at Epoch = 2016.0. Only the first ten lines are shown to guide the reader about the table content. The entire version of the table will be published at CDS.

Table 7. Number and type or mode classification of Cepheids confirmed by the SOS Cep&RRL pipeline and published in *Gaia* DR3.

Type	All sky	LMC	SMC	M 31	M 33
DCEP F	2.008	2357	2487	309	173
DCEP 1O	1101	1931	1803	10	12
DCEP MULTI	195	58	110	–	–
DCEP Total	3304	4346	4400	319	185
ACEP F	150	69	87	–	–
ACEP 1O	132	32	80	–	–
ACEP Total	282	101	167	–	–
T2CEP BLHER	579	66	16	–	–
T2CEP WVIR	795	120	20	–	–
T2CEP RVTAU	261	30	13	2	–
T2CEP Total	1635	216	49	2	–
Cepheid Total	5221	4663	4616	321	185
OTHER	15	–	–	–	–
Reclassified	327	15	1	18	5
New	472	3	11	22	57

Notes. The classification corrections discussed in Sect. 4 have been taken into account in the calculations. The number of objects is provided for each of the five regions in the sky adopted in this work. The last three columns contain the following: OTHER = stars present in the `vari_cepheid` table which after visual inspection resulted in variable stars of type other than Cepheid; Reclassified = objects classified as Cepheids in the `vari_cepheid` table which are known in the literature with different variability types; New = Cepheid variables present in the `vari_cepheid` table which, as far as we know, were not reported before in the literature.

Sect. 2.2[7]. There is a large scatter in Fig. 8 as we also plot objects with very large parallax errors (pulsators with negative parallaxes cannot be shown in the figure). Much better defined PW relationships can be obtained by plotting only objects with relative error in parallax better than 20%, as shown in Fig. 9.

Contrary to the All Sky sample, for the LMC and SMC, we can use the PL relations in the G band in addition to the PW relations, as the reddening in these galaxies is in general rather low and approximately constant over each galaxy. The PL diagrams are shown in Figs. C.1 and C.2 for the LMC and SMC, respectively. Both the PL and the PW diagrams are well defined, especially in the LMC, while the large depth along the line of sight significantly increases the dispersion in the SMC (see Ripepi et al. 2017, and references therein).

[7] We remind the reader that these PW relations are used with the ABL formulation in the SOS Cep&RRL pipeline (see Sect. 2.2).

5.5. Colour–magnitude diagrams

Colour–magnitude diagrams (CMDs) for the Cepheids in all the subregions are shown in Figs. 10, D.1–D.3. The CMDs for the All Sky sample show very large dispersions, as the reddening along the disc and the bulge – where most of the DCEPs and T2CEPs reside – can be of several magnitudes. Not surprisingly, the dispersion of ACEPs is smaller, as the majority of these objects are situated in the halo, where reddening is on average rather low.

The MCs have approximately constant and low reddening, meaning that the CMDs of the Cepheids in these galaxies are more meaningful, with the DCEP_1O clearly bluer than the DCEP_F, as expected. The 'spur' of LMC DCEPs of both modes extending up to $G_{BP} - G_{RP} \sim 1.5$ mag remind us that in the LMC there are regions with high reddening values. The range in colours spanned by ACEPs and different types of T2CEPs reflects their locations in the instability strip. The CMDs of M 31 and M 33 DCEPs are shown only for completeness, as the colours are totally unreliable in most cases.

5.6. Period–amplitude diagrams

Figures 11, E.1–E.3 display the period versus amplitude in the G band (P-Amp(G)) relations for the different subregions and Cepheid types. The morphology of these plots for DCEPs in the All Sky and MC samples is as expected from the literature (see e.g. Ripepi et al. 2017, 2022b, for the SMC and LMC, respectively). The DCEP_1O, as well as most of the DCEP_MULTI objects, have Amp(G) < 0.5 mag, while the DCEP_F objects show the characteristic double peak at periods of 2−3 days and 11−12 days in the All Sky and LMC samples. The P-Amp(G) distribution in the SMC is instead significantly different: the DCEP_1Os show larger amplitudes and the first peak of the DCEP_F pulsators occurs at shorter periods and larger amplitudes than in the All Sky and LMC samples, while the second peak is only barely visible with much smaller amplitudes than the first one, again in contrast with the All Sky and LMC samples. All these differences are most likely due to the much lower

Table 8. Links to *Gaia* archive table to retrieve the pulsation characteristics: period(s), epochs of maximum light and minimum radial velocity (E), peak-to-peak amplitudes, intensity-averaged mean magnitudes, mean radial velocity, ϕ_{21}, R_{21}, ϕ_{31}, R_{31} Fourier parameters with related uncertainties and metallicity computed by the SOS Cep&RRL pipeline for the 15 021 objects (15 006 Cepheids and 15 stars of different type) released in *Gaia* DR3.

Table URL	http://archives.esac.esa.int/gaia/
Cepheids main parameters computed by the SOS Cep&RRL pipeline	
Table name	`gaiadr3.vari_cepheid`
Source ID	`source_id`
Type	`type_best_classification` (one of `T2CEP`, `DCEP` or `ACEP`)
Type2	`type2_best_classification` (for type-II Cepheids, one of `BL_HER`, `W_WVIR` or `RV_TAU`)
Mode	`mode_best_classification` (one of `FUNDAMENTAL`, `FIRST_OVERTONE`, `SECOND_OVERTONE` `MULTI`, `UNDEFINED`, or `NOT_APPLICABLE`)
Multi-mode	`multi_mode_best_classification` (for multi-mode δ Cepheids, one of `F/10`, `F/20`, `10/20`, `10/30`, `20/30`, `F/10/20`, or `10/20/30`)
$Pf, P1O, P2O, P3O$	`p_f, p1_o, p2_o, p3_o`
$\sigma(Pf, P1O, P2O, P3O)$	`pf_error, p1_o_error, p2_o_error, p3_o_error`
E [(a)](G, G_{BP}, G_{RP}, RV)	`epoch_g, epoch_bp, epoch_rp, epoch_rv`
$\sigma E(G, G_{BP}, G_{RP}, RV)$	`epoch_g_error, epoch_bp_error, epoch_rp_error, epoch_rv_error`
$\langle G\rangle, \langle G_{BP}\rangle, \langle G_{RP}\rangle, \langle RV\rangle$	`int_average_g, int_average_bp, int_average_rp, average_rv`
$\sigma\langle G\rangle, \sigma\langle G_{BP}\rangle, \sigma\langle G_{RP}\rangle, \sigma\langle RV\rangle$	`int_average_g_error, int_average_bp_error, int_average_rp_error, average_rv_error`
$Amp(G, G_{BP}, G_{RP}, RV)$	`peak_to_peak_g, peak_to_peak_bp, peak_to_peak_rp, peak_to_peak_rv`
$\sigma[Amp(G)], \sigma[Amp(G_{BP}], \sigma[Amp(G_{RP}], \sigma[Amp(RV)]$	`peak_to_peak_g_error, peak_to_peak_bp_error, peak_to_peak_rp_error, peak_to_peak_rv_error`
$\phi_{21}(G)$	`phi21_g`
$\sigma[\phi_{21}(G)]$	`phi21_g_error`
$R_{21}(G)$	`r21_g`
$\sigma[R_{21}(G)]$	`r21_g_error`
$\phi_{31}(G)$	`phi31_g`
$\sigma[\phi_{31}(G)]$	`phi31_g_error`
R_{31}	`r31_g`
$\sigma[R_{31}(G)]$	`r31_g_error`
[Fe/H] [(b)]	`metallicity`
σ([Fe/H])	`metallicity_error`
N_{obs}(G band)	`num_clean_epochs_g`
N_{obs}(G_{BP} band)	`num_clean_epochs_bp`
N_{obs}(G_{RP} band)	`num_clean_epochs_rp`
N_{obs}(RV)	`num_clean_epochs_rv`

Notes. To ease table access, we also provide the correspondence between parameter [period(s), E, etc.] and the name of the parameter in the *Gaia* archive table. [(a)]E corresponds to the time of maximum in the light curve and the time of minimum in the RV curve. The BJD of all epochs is offset by JD 2455197.5 d (=J2010.0). [(b)]Photometric metallicity based on the Fourier parameters (see Sect. 5.8)

metallicity of the SMC DCEPs with respect to the MW and LMC samples (see e.g. De Somma et al. 2023). The P-Amp(G) diagrams for the DCEPs in the M 31 and M 33 galaxies appear rather different from the other samples. This is mainly because only a handful of stars with period shorter than 10 days were detected in these galaxies, which means that the first amplitude peak for DCEP_F is completely missed. Instead, we observe the second peak, at least in M 31, but shifted to about $P \sim 30$ days. However, this feature requires confirmation, as in M 31, *Gaia* is operating at the extreme limits of its capabilities.

The P-Amp(G) distributions of ACEPs and T2CEPs are also very interesting: (i) as expected, ACEP_1O objects have smaller amplitudes than those of ACEP_F; (ii) at periods in the range 1−2 days, ACEP_F can reach significantly higher amplitudes than both DCEP_F and BLHER, providing us with an additional tool to distinguish them from the different Cepheid types; and (iii) the period separation between different T2CEP types also corresponds to a difference in amplitude, meaning that the WVIR stars have a minimum and a maximum at the extreme periods characterising this class. These features are clearly visible in the data of the All Sky survey because of the large sample size, but are also clearly discernible in the LMC, while in the SMC the paucity of T2CEPs prevents any conclusions.

5.7. Radial velocities

One of the new products of *Gaia* DR3 is the publication of time-series RV data. The final catalogue of Cepheids of all types includes 799 objects for which RV time series are released. The SOS Cep&RRL pipeline only obtained average RV and peak-to-peak amplitude values for 786 objects, as for 13 objects the number of epochs is smaller than seven, which is the minimum required for the RV curve fitting. In total, the time-series are released for 582 DCEP_F, 133 DCEP_1O, 14 DCEP_MULTI, 12 BLHER, 35 WVIR, 17 RVTAU, 3 ACEP_F, and 2 ACEP_1O pulsators. Among the DCEP_Fs, 15 and 9 objects belong to the LMC and SMC, respectively. In addition to the time series, median RV values calculated by the general RV data processing in *Gaia* (Sartoretti et al. 2022) are published for 3190 Cepheids of all types in the `gaia_source` table. As shown in Fig. 12, there is excellent agreement between the two estimates for the 736 stars in common between the two samples (see Clementini et al. 2023, for further details). Indeed, the median and mean difference between the two average values are of 0.43 and 0.33 km s^{-1}, respectively, with a standard deviation of 6.40 km s^{-1}.

The spatial distributions of Cepheids with average RV values from both the general and the SOS Cep&RRL pipelines are shown in Fig. 13 and are colour coded according to the RV

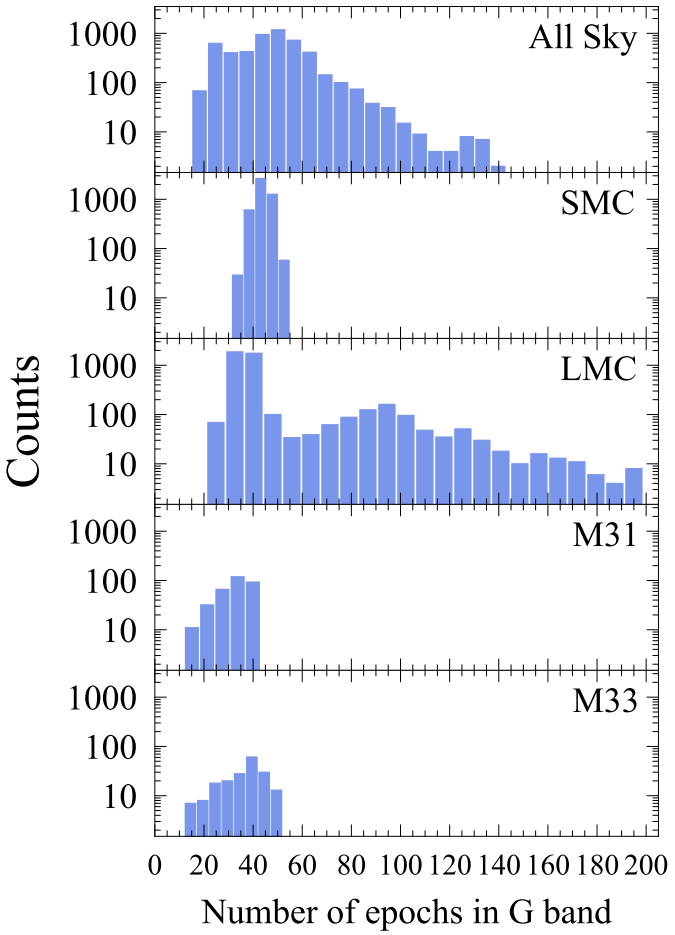

Fig. 2. Number of epochs in the *G*-band time series. *From top to bottom*, the different panels show the data for the different subsamples corresponding to the five regions of the sky defined in Sect. 2.

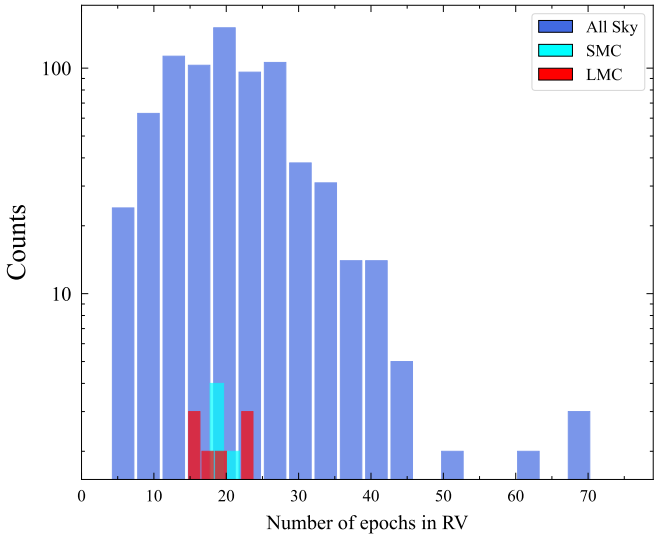

Fig. 3. Number of epochs in the RV time series for the labelled subsamples.

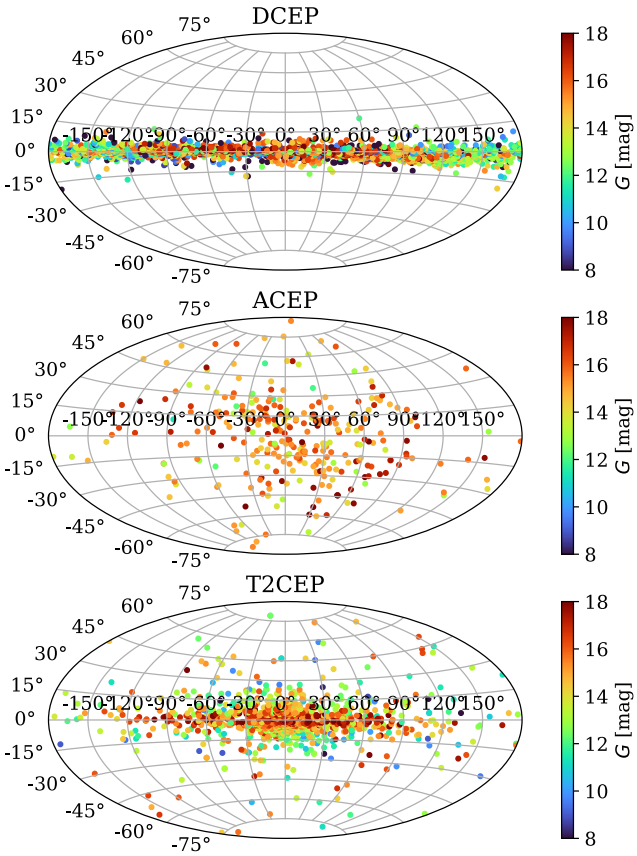

Fig. 4. Map in Galactic coordinates of the different Cepheid types in the MW. The objects are colour coded according to their apparent *G* magnitude.

values. As expected, the objects lying in the disc (mainly DCEPs, see Gaia Collaboration 2023b, for the an example of exploitation of these data) show low values of RV, while the halo Cepheids show both highly positive and negative RV values. The LMC and SMC are clearly identified by the RV values shared by all stars belonging to the two galaxies.

The uncertainties measured by the SOS Cep&RRL pipeline on the average RV ($\langle RV \rangle$) and on the RV peak-to-peak amplitude (Amp(RV)) are shown in Fig. 14. The typical uncertainties on $\langle RV \rangle$ are on the order of 1−1.5 km s^{-1}, as expected (see Clementini et al. 2023). However there are a few objects showing large errors as measured by the bootstrap procedure. These cases are often correlated with the low number of RV epochs available for these Cepheids (see Fig. 3). Similarly, the typical uncertainty is ∼3−4 km s^{-1} for the Amp(RV), but there are a few objects with uncertainties larger than 30−40 km s^{-1} which can be an indication of unreliable Amp(RV) values. This is verified in Fig. 15, where, in analogy to the photometry, we show the relation between amplitude in RV and period. The general trend closely follows that shown from photometry, with DCEP_F objects having larger amplitudes than DCEP_1O or DCEP_MULTI objects and showing the typical bell shape starting from a minimum amplitude at a period of ∼9 days and a maximum at ∼20 days. The figure shows that despite the large uncertainties in the Amp(RV) of some objects, only a few Cepheids appear out of their expected position in this plot. We conclude that the RV amplitudes calculated by the SOS Cep&RRL pipeline are generally reliable.

5.8. Metallicities

An additional product of the SOS Cep&RRL pipeline are the photometric iron abundances inferred from the Fourier

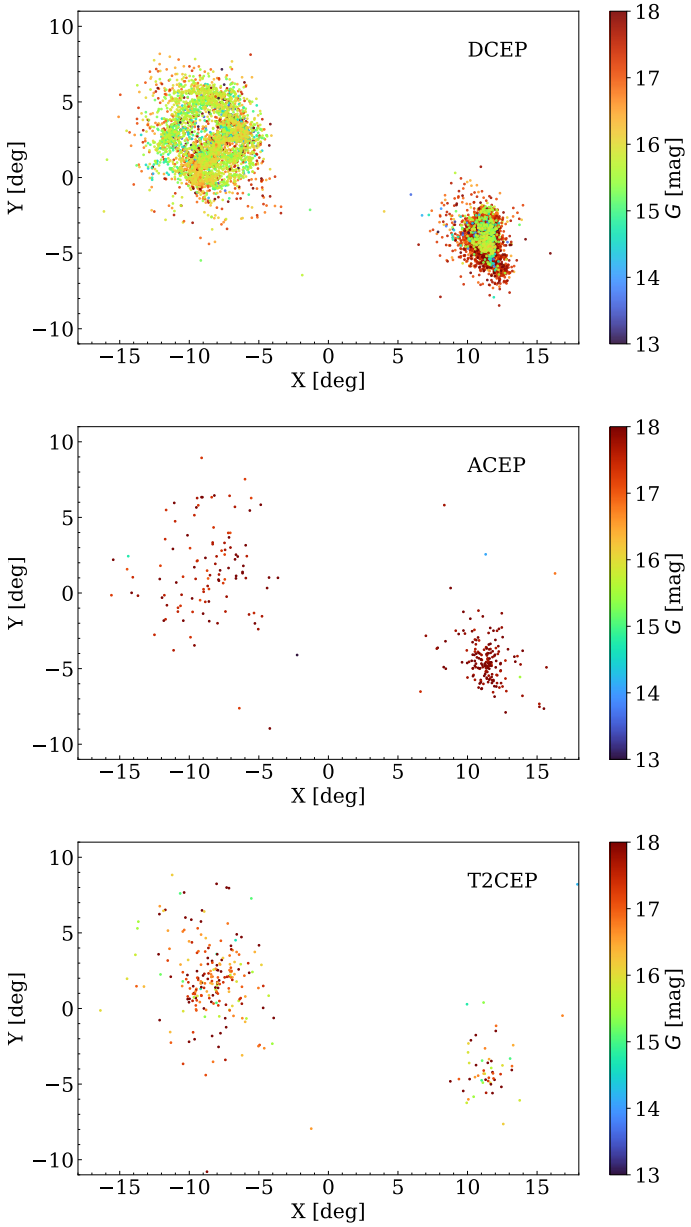

Fig. 5. Map of the different Cepheid types in the MCs. The objects are colour coded according to their apparent G magnitude. The map is a zenithal equidistant projection centred at equatorial coordinates RA, Dec = 56.0, −73.0 deg (J2000).

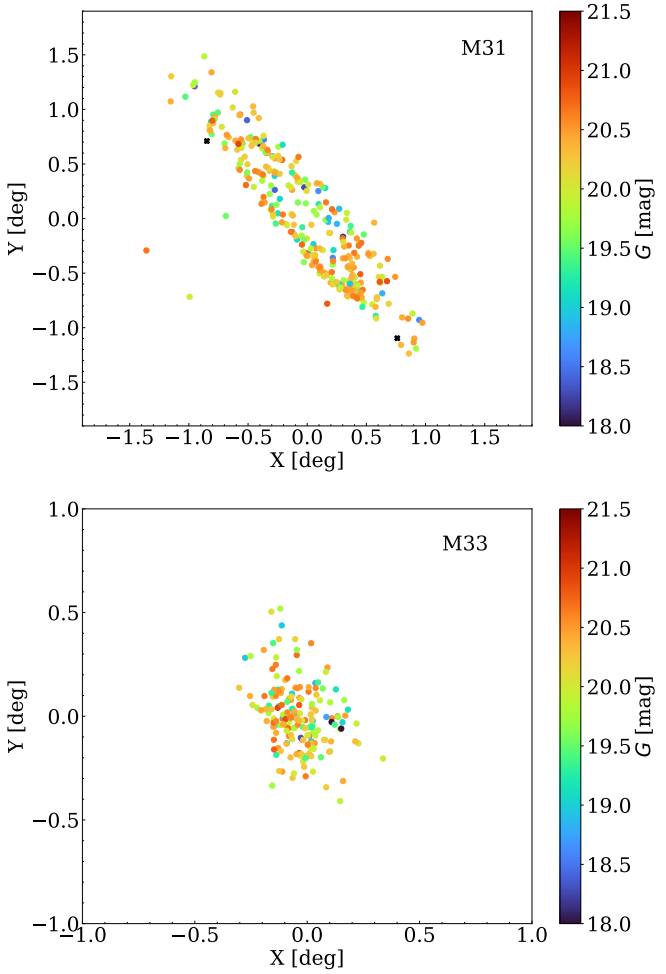

Fig. 6. Map of the DCEPs in M 31 (*top panel*) and M 33 (*bottom panel*). The symbols are colour coded based on the apparent G magnitude of the DCEPs. The two black crosses identify two RVTAU stars in M 31. The maps are in zenithal equidistant projection centred at equatorial coordinates $(\mathrm{RA, Dec})_{\mathrm{M\,31}} = 10.6, 41.2$ deg (J2000) and $(\mathrm{RA, Dec})_{\mathrm{M\,33}} = 23.5, 30.65$ deg (J2000).

parameters R_{21} and R_{31} according to the calibration by Klagyivik et al. (2013), which is valid for DCEP_Fs with periods shorter than 6.3 days and for an interval of metallicity reaching the average [Fe/H] values of the LMC and SMC DCEPs (see Clementini et al. 2019, for details). As the metallicity estimates rely on the R_{21} and R_{31} Fourier parameters, which sometimes have large errors calculated with the bootstrap technique, we suggest using the [Fe/H] values with uncertainties larger than ∼0.5 dex with care. The catalogue includes a total of 5265 DCEP_Fs with [Fe/H] estimates. However, as we have changed the classification (see Sect. 4) for the 142 objects reported in Table 6, some of these objects are no longer DCEP_Fs, and therefore their metallicity estimates are incorrect and should not be used. The DCEP_Fs with an [Fe/H] estimate are 1053, 1882, 2174, 7, and 7 in the All Sky, LMC, SMC, M 31, and M 33 samples, respectively. The distribution of the metallicities in the SMC, LMC, and All Sky samples is shown in Fig. 16. The figure shows that, as expected, the DCEPs in the All Sky sample (exclusively MW objects) are, on average, more metal rich than the LMC ones, which in turn are more metal rich than those in the SMC. From a quantitative point of view, we can see that the peak of the All Sky distribution is [Fe/H] ∼ +0.05 dex, which is in general agreement with the literature (see e.g. Ripepi et al. 2019). On the contrary, for the LMC and SMC, we have peaks of approximately −0.2 dex and −0.3 dex for the LMC and SMC, respectively. These values are significantly larger than those found in the literature, namely $[\mathrm{Fe/H}]_{\mathrm{LMC}} = -0.41$ dex ($\sigma = 0.08$ dex Romaniello et al. 2022) and $[\mathrm{Fe/H}]_{\mathrm{SMC}} = -0.75$ dex ($\sigma = 0.08$ dex Romaniello et al. 2008). Therefore, the photometric metallicities are not particularly reliable for metallicity values lower than [Fe/H] ∼ −0.3 dex, which is not unexpected as the work by Klagyivik et al. (2013) relies on very few calibrators in this metallicity range.

For M 31 and M 33, the PL relations are more accurate than the PW relations because the magnitudes in the G_{BP} and G_{RP} bands, if any, are less accurate that that in the G band,

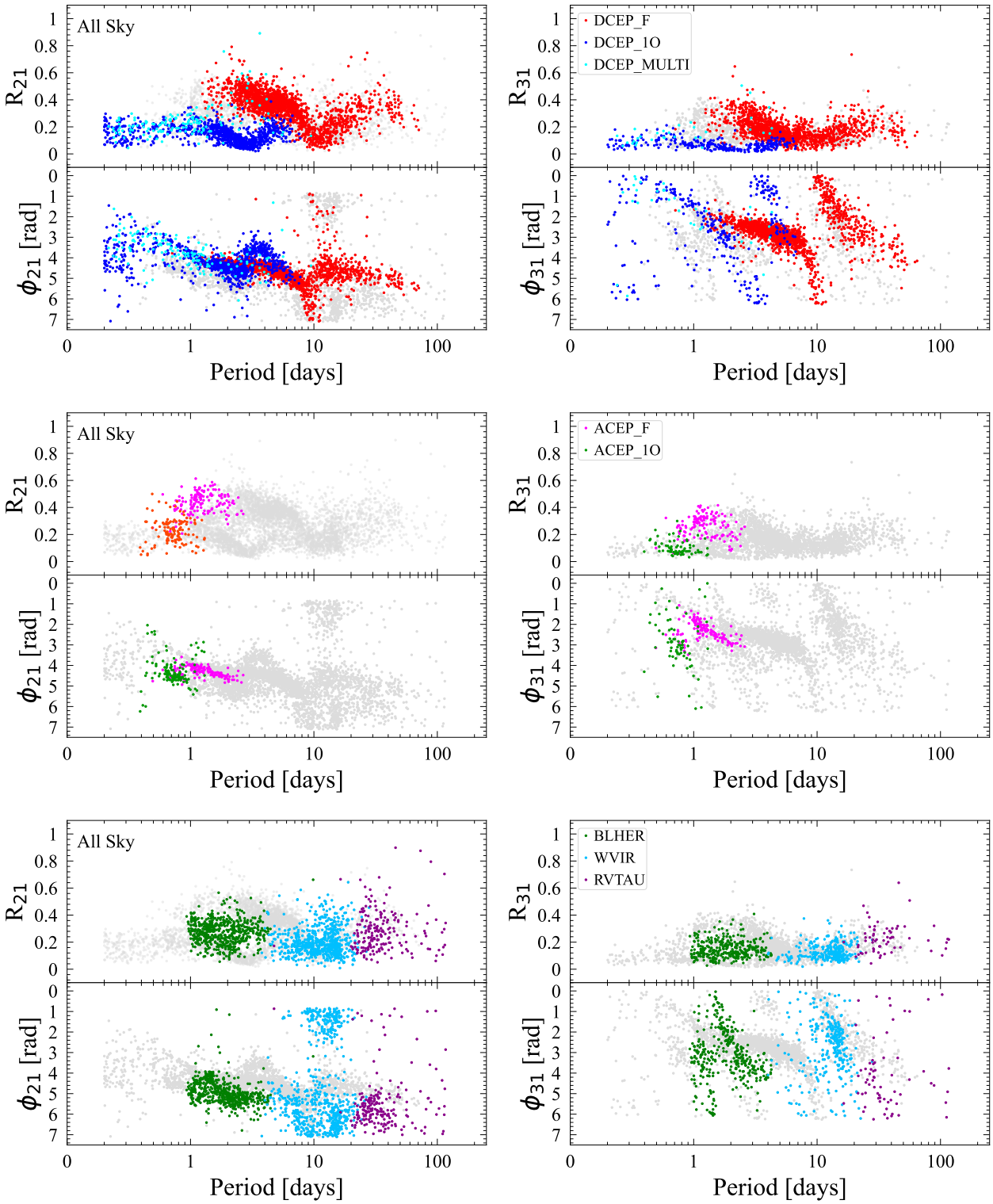

Fig. 7. Fourier parameters for the All Sky sample. *From top to bottom* the different panels show the results for DCEPs, ACEPs, and T2CEPs, respectively.

which leads to much greater dispersion in the PW relations (see Fig. C.3). The PL relations for both M 31 and M 33 show a remarkable linearity up to about $G \sim 21$ mag.

We can perform a more detailed comparison between the photometric metallicities from the SOS Cep&RRL pipeline and the literature by cross-matching the All Sky sample with the list of DCEPs that have metallicities measured from high-resolution spectroscopy recently published by Ripepi et al. (2022a)[8]. The metallicity estimates for the 185 DCEPs in common between the two samples are displayed in Fig. 17. The photometric [Fe/H]

[8] In this and many other phases of this work, we made use of the TOPCAT package (Tool for OPerations on Catalogues And Tables Taylor 2005).

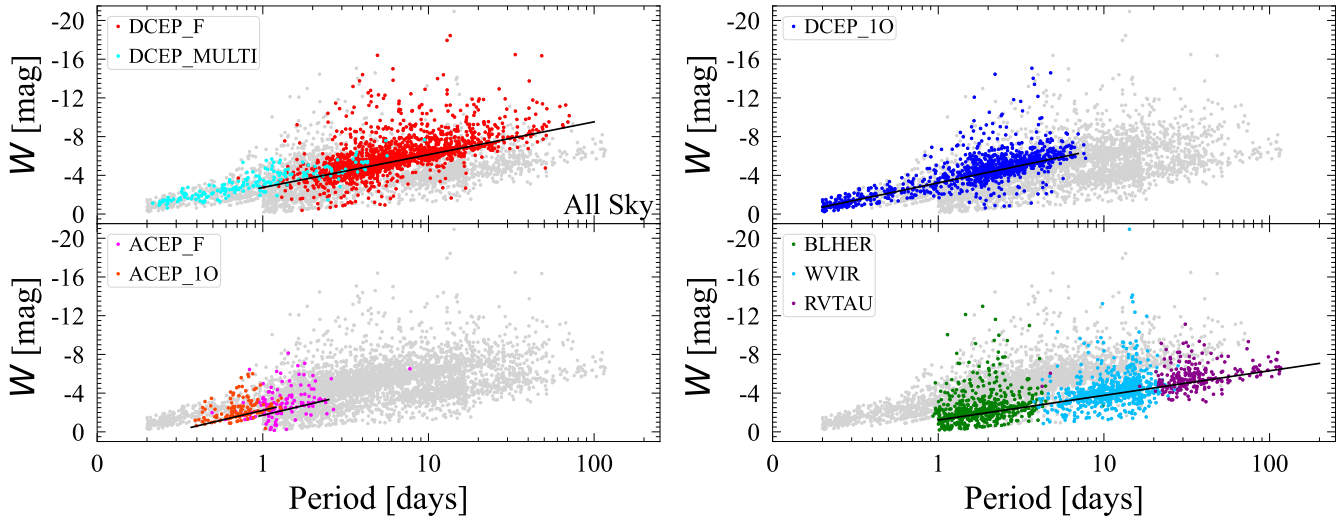

Fig. 8. PW relation for the All Sky sample. The different types and modes of the Cepheids displayed in the figures are labelled in each panel.

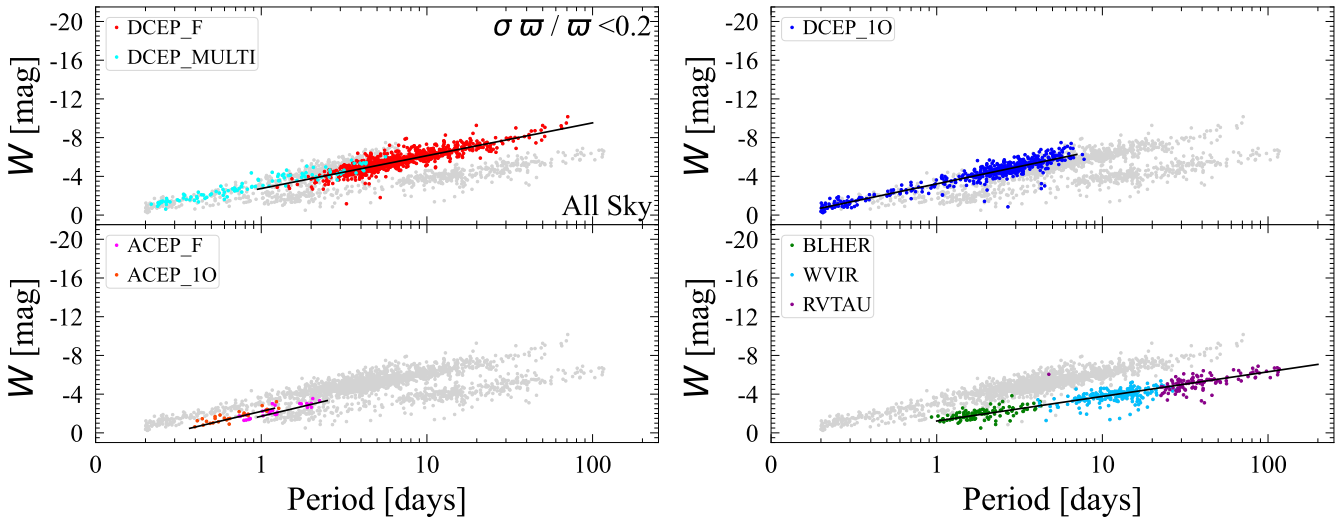

Fig. 9. Same as in Fig. 8 but restricting the sample to objects with $\sigma\varpi/\varpi < 0.2$.

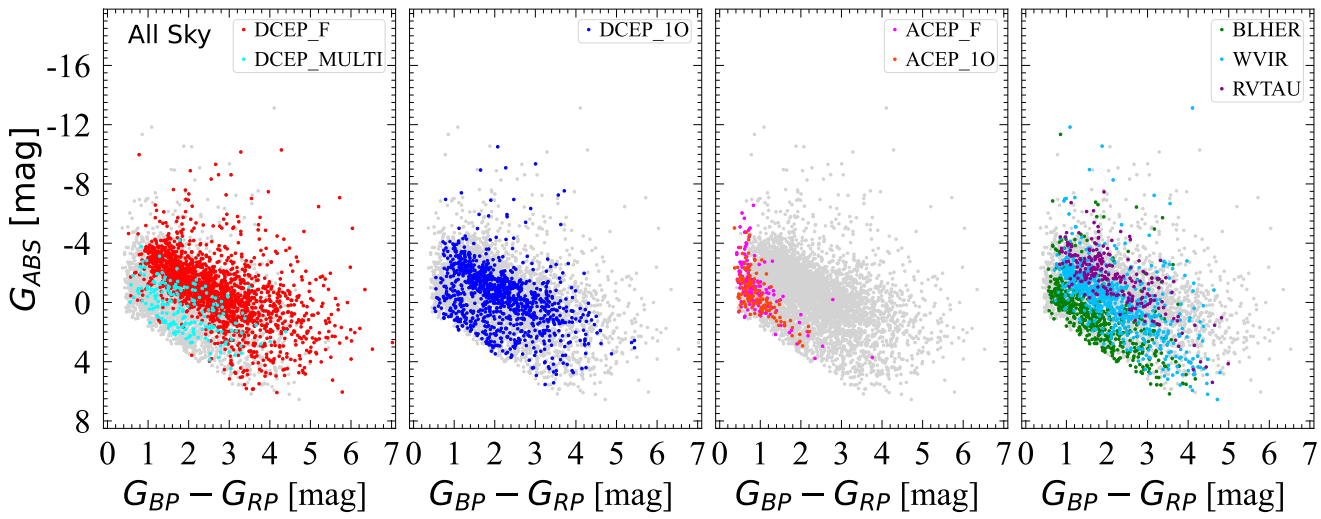

Fig. 10. CMD of the All Sky Cepheid sample.

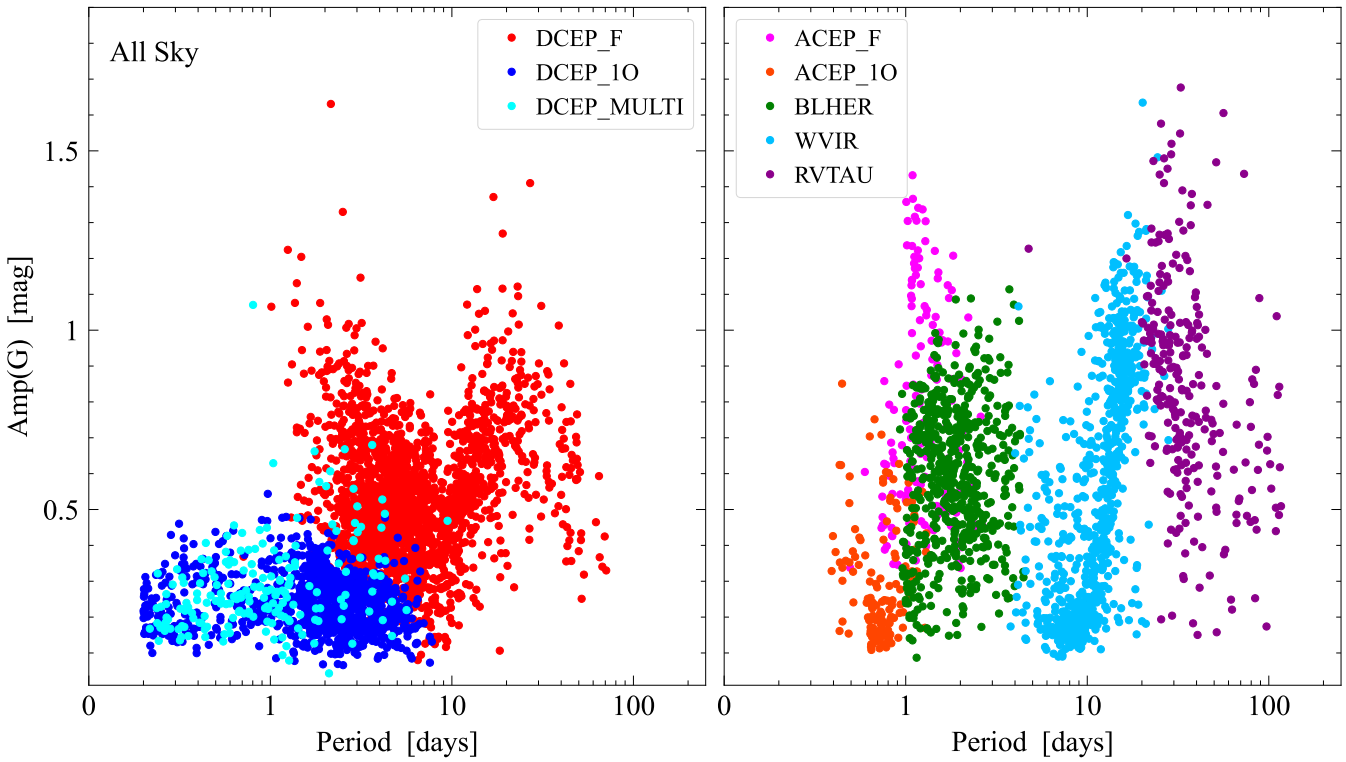

Fig. 11. Period–amplitude (G) diagram for the All Sky sample.

values appear to be systematically higher than the spectroscopic abundances. The average difference is $[Fe/H]_{Lit}$–$[Fe/H]_{SOS}$ = −0.08 dex, with $\sigma = 0.16$ dex and no apparent trend with the $[Fe/H]_{Lit}$ value. The mean shift and relative dispersion are modest, meaning that as far as the All Sky sample is concerned, or at least in the metallicity range $-0.3 < [Fe/H] < +0.4$ dex, the photometric metallicites can be used. We speculate that for lower values, the metallicity sensitivity of the R_{21} and R_{31} parameters may vanish. This could explain the poor performance of the method for the LMC and SMC DCEP samples (see Table 9).

5.9. Cepheids hosted by stellar clusters and satellite dwarf galaxies of the MW

We searched for any association of Cepheids in the All Sky sample with stellar clusters hosted by the MW or with dwarf galaxies orbiting our Galaxy. For the open clusters (OCs), we adopted the list of likely member stars by Cantat-Gaudin et al. (2020) supplemented with new data provided by Castro-Ginard et al. (2022) and Tarricq et al. (2022); for the globular clusters (GCs) we used the list by Clement et al. (2001) (continuously updated); for the dwarf galaxies we used a variety of literature sources including (Soszyński et al. 2017, 2018). Results are shown in Table 10. An additional 35 objects from the All Sky sample can be associated with the MCs, 45 with Galactic GCs, 24 with OCs, and one with the Draco dwarf spheroidal galaxy (variable data for Draco by Kinemuchi et al. 2008).

6. Validation

In the following sections we discuss the many different procedures adopted to validate the catalogue of Cepheids of all types published in *Gaia* DR3. Also, we discuss its completeness and contamination level.

6.1. Literature adopted for the validation

To validate the results of the SOS Cep&RRL pipeline classification, we adopted different literature sources according to the different subregions of reference. Starting with the All Sky, for the DCEPs we adopted the recent compilation by Pietrukowicz et al. (2021, hereafter, P21) – including 3352 reliable bona fide DCEPs – which is mainly based on results from the OGLE survey (Udalski et al. 2018; Soszyński et al. 2020). For ACEPs and T2CEPs, we adopted the results of the OGLE survey (Soszyński et al. 2020, and references therein) complemented by entries in Chen et al. (2020), which is based on the ZTF (Zwicky Transient Factory) survey, and by Drake et al. (2014), Torrealba et al. (2015), which are based on the Catalina sky survey (CSS). As the classification of the latter papers does not distinguish the mode or type of pulsation, we assigned the fundamental mode to the ACEP detected by CSS[9] and separated BLHER from WVIR and WVIR from RVTAU using period thresholds of 4 and 24 days, respectively (in analogy with the SOS Cep&RRL pipeline). The total sample of sources with a positive cross-match with the *Gaia* DR3 catalogue includes 3917 Cepheids. We note that we have intentionally not included results from *Gaia* DR2 re-classifications by Ripepi et al. (2019) to preserve the independence of the counterpart. We have also not included Cepheids by ASAS-SN (All Sky Automated Survey for Supernovae Shappee et al. 2014; Jayasinghe et al. 2019) or ATLAS (Asteroid Terrestrial-impact Last Alert System Heinze et al. 2018), who adopt automatic classification procedures and perform no careful visual inspection of the light curves. However, many stars originally detected by these surveys were analysed by Pietrukowicz et al. (2021) and are included in their catalogue.

[9] For analogy with their studies on RR Lyrae stars, for which they only consider fundamental mode pulsators.

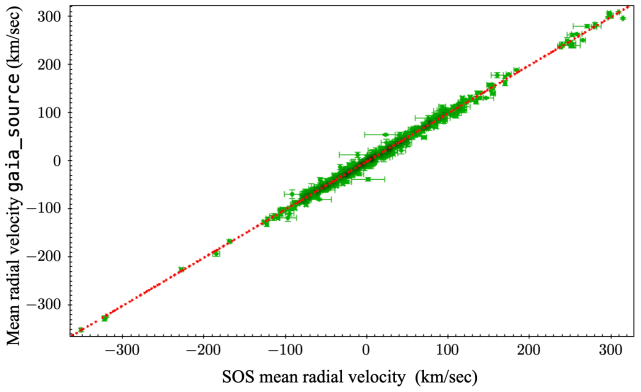

Fig. 12. Comparison between the average RV calculated by the SOS Cep&RRL pipeline from fitting the RV curves and the mean values published in the `gaia_source` table (see Sartoretti et al. 2022, for details).

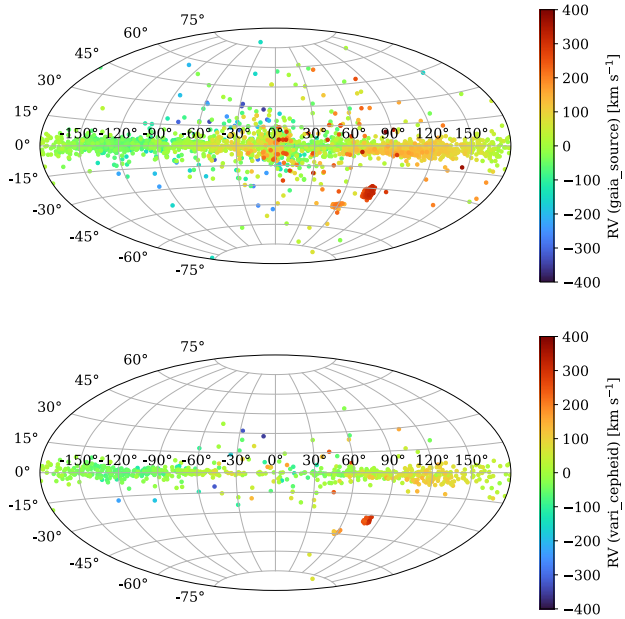

Fig. 13. RV maps defined by the 3190 Cepheids in the DR3 `gaia_source` table (*top panel*) and 786/799 Cepheids in the DR3 `vari_cepheid` table (*bottom panel*).

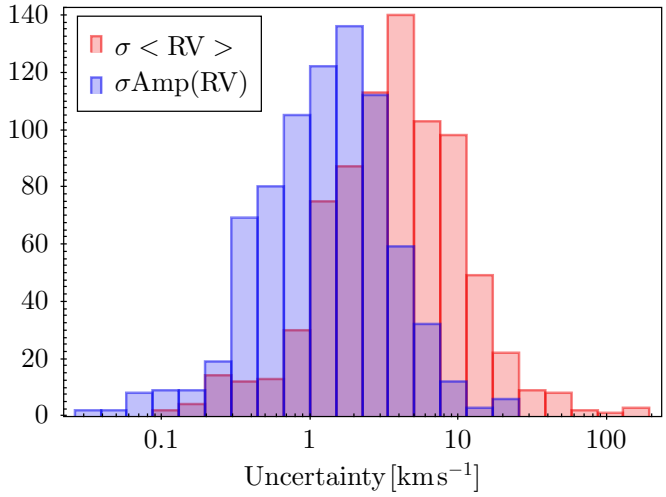

Fig. 14. Uncertainties on the average and peak-to-peak RV values measured by the SOS Cep&RRL pipeline for a sample of 786 Cepheids.

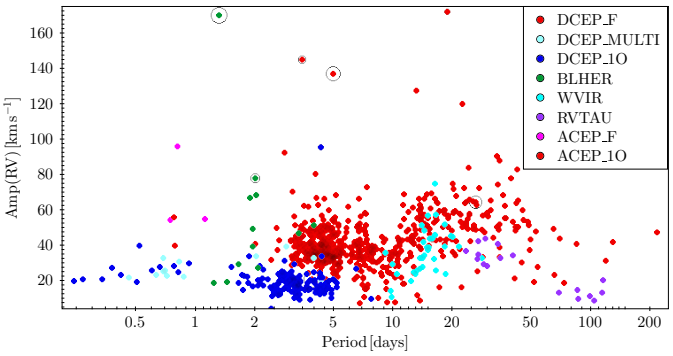

Fig. 15. Period–amplitude (RV) for the 786 Cepheids whose RV curves were analysed by the SOS Cep&RRL pipeline. The different Cepheid types are labelled. The size of the circles surrounding the symbols is proportional to the uncertainty in Amp(RV) (see also Fig. 14).

As for the MCs, we adopted the OGLE catalogue by Soszyński et al. (2019a), including 9650 DCEPs, 343 T2CEPs, and 278 ACEPs. A cross-match with *Gaia* DR3 results provides 4638 and 4608 matches for the LMC and SMC, respectively.

For M 31 we used the work by Kodric et al. (2018) who provide the classification for 2247 Cepheids, including DCEP_F, DCEP_1O, and RVTAU stars. We have 262 stars in common with this work. As for M 33, 112 of the 185 objects classified as Cepheid from the SOS Cep&RRL pipeline are present in the work by Pellerin & Macri (2011). However, these latter authors do not provide a classification in DCEPs or T2CEPs, and therefore we refrained from any comparison.

6.2. Accuracy of the classification, completeness, and contamination

On the basis of the literature data discussed in the previous section, we produced confusion matrices for the LMC, SMC, and All Sky samples. There are 2739 stars in common with P21, corresponding to 82% of the sample. A further 130 objects are published in the general classification (Rimoldini et al. 2023) as the SOS Cep&RRL pipeline found an incorrect period for these objects. Therefore, taking the latter objects into account, the completeness of the *Gaia* catalogue for the All Sky DCEP sample is of 85.6% at least. However, the catalogue by Pietrukowicz et al. (2021) is not free of contamination, especially for the DCEP_1Os, which can be easily confused with binaries if the distance is not used in the classification. This is shown in Fig. 18, which shows the PW relation for a selected sample of DCEPs with parallax relative errors of better than 20% and good astrometric solution (RUWE ≤ 1.4). The vast majority of the objects shown in the figure are common to the *Gaia* DR3 catalogue and Pietrukowicz et al. (2021), and the figure nicely depicts the expected linear relations for both DCEP_F and DCEP_1O pulsators. The second sample includes objects present only in the Pietrukowicz et al. (2021) list. Most of the DCEP_1O are clearly too faint to be DCEPs or any other type of Cepheid, and are likely binaries contaminating the DCEPs sample. Although the numbers of objects with a good parallax is too small to obtain statistical significance, it is plausible that the completeness of the *Gaia* DR3 catalogue for DCEPs is larger than 85.6% once the purity of the comparison samples is taken into account.

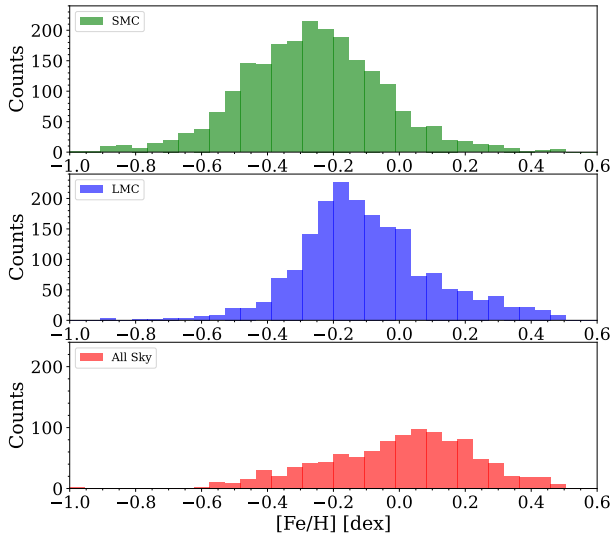

Fig. 16. Photometric metallicities in the LMC, SMC, and All Sky samples.

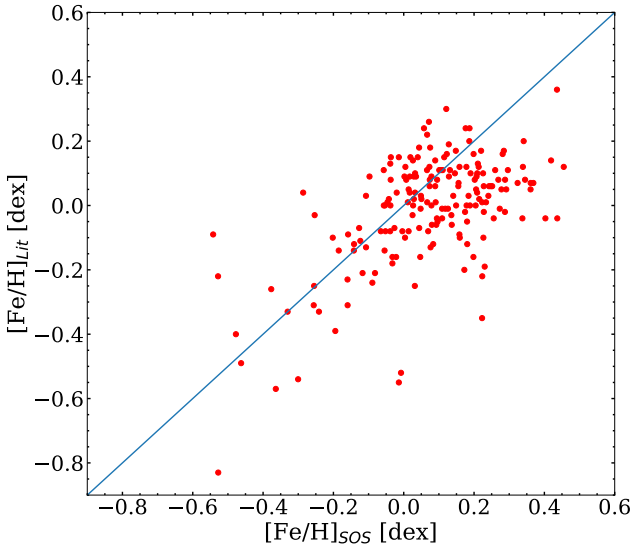

Fig. 17. Comparison between photometric metallicities computed by the SOS Cep&RRL pipeline ([Fe/H]$_{SOS}$) and metal abundances from high-resolution spectroscopy available in the literature ([Fe/H]$_{Lit}$).

The completeness for ACEPs and T2CEPs is more difficult to establish as there are no homogeneous catalogues for these Cepheid types, except for regions of the sky covered by the OGLE survey. Therefore, we restricted our estimates to the bulge and a portion of the disc (see e.g. Soszyński et al. 2020), and calculated the ratio of the number of ACEPs and T2CEPs in DR3 and the OGLE catalogues. Given the small numbers involved compared with DCEPs, we summed ACEPs and T2CEPs, obtaining an overall completeness of about 25%. Such a low completeness compared to the DCEPs is due to the fact that the large majority of the OGLE ACEPs and T2CEPs are in the bulge, a region where *Gaia* has still a low number of epochs on average. In addition, the bulge is also almost devoid of DCEPs, meaning that the *Gaia* low detection efficiency in this region does not impact the DCEP completeness.

The confusion matrix of the All Sky sample is shown in Fig. F.1. The apparent accuracy of our DCEPs classification

Table 9. *Gaia* source_id of sources for which the SOS Cep&RRL pipeline provides a metallicity estimate which should not be used as these stars are not DCEP_F pulsators.

Source_id
375318264077848448
375435873166554752
431184518613946112
513074186146353536
543759459725179136
1208200864738741376
1248397910338129664
1374376207437762688
1400474455952839168
1682922431734385152

Notes. Only the first ten lines are shown to guide the reader as to the content of the table. The table in its entirety is published at CDS.

(‘Recall’ column) is satisfactorily high, being 96%, 92%, and 95% for DCEP_F, DCEP_1O, and DCEP_MULTI, respectively. A similar result is obtained for T2CEP variables, namely >94% for all Cepheid types. The percentages are less good for the ACEPs which are much more difficult to classify, given the similarities in light curve shape with DCEP and BLHER variables. We therefore tend to classify more ACEPs than the literature, where the classification is usually only based on the light curve shape. Precision is again very high for T2CEPs and DCEPs with the exception of DCEP_MULTI, of which we appear to have missed about 30%. This is not surprising, as for many pulsators we just do not have enough epochs to resolve more than one pulsating mode. For ACEPs, the precision is about 70%, which means that we are able to detect a large fraction of the literature ACEPs.

The same kind of comparison is shown in Figs. F.2 and F.3 for the LMC and SMC, respectively. The results are very good in the LMC for both accuracy and precision for all types, with the exception of the DCEP_MULTI, which we massively missed and classified as DCEP_1O because the low number of epochs prevented the detection of the second (or third) periodicity. The results are slightly worse in the SMC, where the elongation along the line of sight produces far less separated PL and PW relations. This especially impacts the ACEPs, which were confused with DCEPs, introducing a 2% contamination among the latter. In the SMC, we missed a smaller percentage of DCEP_MULTI sources.

Concerning the overall completeness (e.g. ignoring the subclassification in types or modes), in both the LMC and SMC the *Gaia* DR3 catalogue includes 90% of the known Cepheids of all types. As for M 31, we do not show the confusion matrix as the agreement between our classification and the literature is 100%. The completeness is much less, because we were only able to detect reasonable light curves for the brightest Cepheids in M 31, which is due to the *Gaia* limiting magnitude. This corresponded to only 12.1% of the known Cepheids of all types. We do not have an accurate literature control sample for the M 33 Cepheids, and therefore we only mention that we detected about 23% of the known Cepheids in this galaxy.

6.3. Contamination by variables other than Cepheids

In the previous section, we established the reliability of the Cepheid classification in the *Gaia* DR3 catalogue by comparison

Table 10. Association of Cepheids in the All Sky sample with open and globular clusters and with dwarf galaxies that are satellites of the MW.

Source_id	RA (deg)	Dec (deg)	Class	System	Other ID
429385923752386944	0.246798	60.959002	DCEP_F	UBC 406	CG Cas
4707044742055169152	9.219742	−66.593232	BLHER	SMC	OGLE-SMC-CEP-4693
4684386345732125696	11.086946	−76.195508	DCEP_F	SMC	OGLE-SMC-CEP-4710
4702506576531479424	12.344029	−69.50825	DCEP_F	SMC	OGLE-SMC-CEP-4723
4635176637678468096	14.144298	−77.920315	DCEP_F	SMC	OGLE-SMC-CEP-4967
4691023998645738368	17.496078	−69.937937	DCEP_F	SMC	OGLE-SMC-CEP-4816
4691296265212110720	22.594176	−69.427211	RVTAU	SMC	–
4636490008613940992	23.448788	−77.656158	ACEP_F	SMC	OGLE-SMC-ACEP-091
4691302823627312640	23.541042	−69.256368	DCEP_F	SMC	OGLE-SMC-ACEP-092
4698416118397803776	25.204992	−67.494995	ACEP_F	SMC	–

Notes. The equatorial coordinates are given at Epoch = 2016.0. Only the first ten lines are shown to guide the reader as to the content of the table. The table is published in its entirety at the CDS.

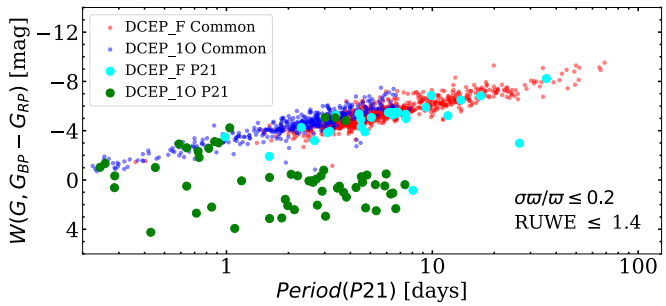

Fig. 18. PW relation for a selected sample of DCEPs. Red and blue small filled circles show the DCEP_Fs and DCEP_1Os in common between *Gaia* DR3 and Pietrukowicz et al. (2021, abbreviated as P21 in the labels), respectively. Cyan and green large filled circles show DCEP_Fs and DCEP_1Os present in the P21 catalogue only. For all objects, we applied a selection in parallax, requiring that the relative precision be better than 20%. We also required the RUWE parameter to be lower than 1.4, so as to ensure a good astrometric solution (see text).

with high-quality Cepheid catalogues in the literature. For the All Sky sample, we use the same literature catalogues –namely OGLE (Soszyński et al. 2019b), ZTF (Chen et al. 2020), and CSS (Drake et al. 2014), which also list variability types other than Cepheids– to assess the possible contamination of the *Gaia* DR3 catalogue by non-Cepheids. As a result, we found 93 objects which are listed in Table 11. The main source of possible contamination is from RR Lyrae stars, eclipsing binaries, and eruptive variables. Even if we restrict our comparison to the aforementioned surveys, we can nevertheless conclude that contamination of the *Gaia* DR3 Cepheid catalogue is on the order of 1%−2%.

6.4. The case of ARRD stars

Anomalous double-mode RR Lyrae stars (ARRDs) differ from normal RRDs because of the smaller ratio between the 1O and F pulsation modes (see Soszyński et al. 2016a,b). The ARRDs were originally discovered in the LMC, but Soszyński et al. (2019b) reported the presence of many ARRDs also among the OGLE bulge and disc collection of RR Lyrae stars. Six of these ARRDs are in the All Sky sample with classification as DCEP_MULTI. The position of these stars in the Petersen diagram is highlighted in Fig. 1. Five objects lie in the region where DCEPs pulsate in the F/1O multi-mode, while one (*Gaia* EDR3 4091104989668551936) is placed in the locus of 2O/3O pulsators. However, the two periods of the latter differ from those found by OGLE and could be incorrect, as we have only 23 epochs in *Gaia*. Adoption of the OGLE periods would also place this sixth source close to the F/1O DCEP multi-mode pulsators.

The location of the six objects in the PW plane is shown in Fig. 19. The uncertainty of the W values for three objects is rather high because of the large uncertainty in their parallaxes. Nevertheless, the location on the PW relation of all six objects seems compatible with them being short-period DCEPs. We conclude that at least some of the objects classified as ARRDs in the MW are actually DCEPs and not RR Lyrae variables. This is due to the difficulty in determining the distances in the MW compared with the LMC, where all the objects are at approximately the same distance from us.

6.5. Validation with TESS photometry

For validation we used photometric data collected by the Transiting Exoplanet Survey Satellite (TESS, Ricker et al. 2015), which is collecting continuous photometry over a large (24° × 96°) area with four cameras with adjacent fields of view over segments of 27 days in length, called sectors. In mission years 1, 2, and 3, the field of view was rotated around the centre of camera 4, positioned towards the southern and then the northern and then again the southern ecliptic pole, while avoiding a 12-degree band along the Ecliptic. In year 4, five sectors were rotated so that all cameras were pointing towards the Ecliptic and observations cover a roughly 230° segment of it. We searched the full-frame image data up to Sector 43, which was the fourth sector in year 4 and the second along the Ecliptic. Sampling cadence of the full-frame images was initially 30 min in years 1–2, and was lowered to 10 min in the first extended mission (years 3–4).

The spatial resolution of TESS is limited to 21″ px^{-1}. Therefore, although it is capable of reaching the brighter Cepheids in the LMC and SMC, the images suffer from severe crowding and blending (Plachy et al. 2021). To avoid that, we only looked at Galactic Cepheids in this study. We cross-matched the *Gaia* coordinates with the sector coverage using the Web TESS Viewing Tool[10] and then queried the TESS Quick Look Pipeline (QLP) database for light curves (Kunimoto et al. 2021;

[10] https://heasarc.gsfc.nasa.gov/cgi-bin/tess/webtess/wtv.py

Table 11. Potential contaminants of type other than Cepheids.

Source_id	RA (deg)	Dec (deg)	P_SOS (days)	Class_DR3	Class_Lit	Lit_source
526109377526041344	12.70783	66.30146	0.34443	DCEP_1O	RRC	ZTF
523287961970713728	16.24640	63.19386	0.51796	DCEP_1O	RRAB	ZTF
525971972931888256	16.61012	66.18293	0.33474	DCEP_1O	RRC	ZTF
2454747674236106240	18.90559	−16.24635	7.54712	WVIR	EW	CATALINA
514299866732307584	37.15235	63.32395	0.25799	DCEP_1O	EW	ZTF
249876593076412032	55.50083	50.70071	0.65550	DCEP_1O	RRAB	ZTF
4858120560289808256	60.61599	−35.48554	0.95495	BLHER	RRAB	CATALINA
3286936002024112896	66.40184	7.48227	0.73534	ACEP_F	RRAB	CATALINA
258545623786523904	69.12216	49.27242	26.17633	RVTAU	SR	ZTF
3239597250445418624	73.44271	4.91058	0.71103	ACEP_1O	RRAB	CATALINA

Notes. The equatorial coordinates are given at Epoch = 2016.0. The meaning of the variability types listed in column 'Class_Lit' can be found at the following address: http://cdsarc.u-strasbg.fr/viz-bin/getCatFile_Redirect/?-plus=-%2b&B/gcvs/./vartype.txt. The 'Lit_source' acronyms are: CATALINA = Drake et al. (2014, 2017), Torrealba et al. (2015); OGLE = Soszyński et al. (2019b); ZTF = Chen et al. (2020). The entire version of the table will be published at CDS. Only the first ten lines are shown to guide the reader as to the content of the table.

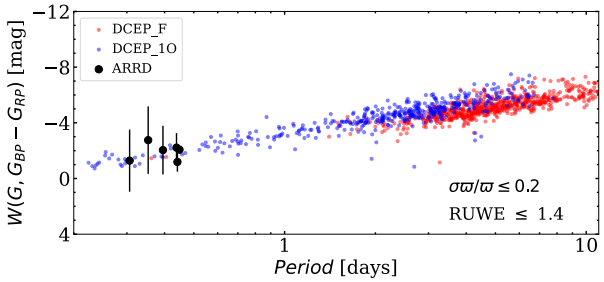

Fig. 19. Position on the PW diagram of the six stars that are known in the literature as ARRD stars but are classified here as DCEP_MULTI (black filled circles). For reference, red and blue dots show the PW for the same DCEP_F and DCEP_1O samples displayed in Fig. 18.

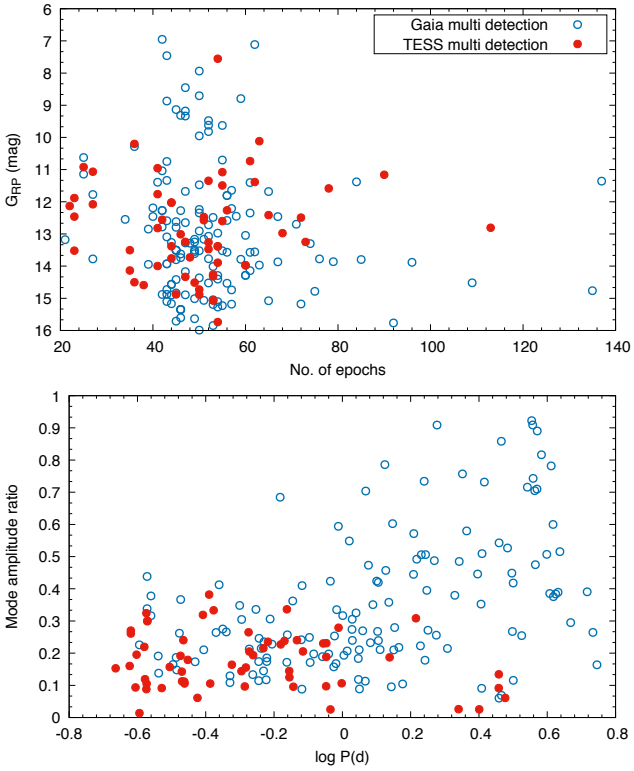

Fig. 20. Comparison of the parameters of the multi-mode stars detected in *Gaia* (blue circles) or from the TESS light curves (red dots). The upper plot compares the G_{RP} brightness (which is close to the TESS passband) and the number of photometric epochs available in DR3. The lower plot compares the amplitude ratio of the modes and the logarithm of the longer pulsation period.

Huang et al. 2020a,b). The pre-processed QLP light curves have a faint limit of $T = 15$ mag, which is equivalent to the same G_{RP} magnitude, and are produced primarily for searches of exoplanet transits. As a consequence, not all Cepheid candidates have good QLP light curves. Therefore, we also extracted photometry from the full-frame images with the `eleanor` software, which is capable of both pixel aperture and PSF photometry and post-processing of the light curves via regression against a systematic error model or via principal component analysis (Feinstein et al. 2019). We then selected the best light curves from the QLP and the four `eleanor` results (raw, corrected, PCA-corrected, and PSF photometry), and applied further corrections: sigma clipping to remove outliers and detrending to remove residual slow variations. For the trend removal, we used the method described by Bódi et al. (2022). Briefly, the algorithm searches for the dominant periodicity in the light curve, computes the phase dispersion of the folded data, and then fits a polynomial to the data by minimising against the phase dispersion. This way even high-order polynomials can be fitted that still follow the changes in average brightness and are much less affected by the effects of incomplete pulsation cycles at the edges.

We then calculated the pulsation periods and A_{i1} and ϕ_{i0} relative Fourier coefficients of the first few harmonics from the processed light curves and compared them to that of the OGLE *I*-band measurements (Soszyński et al. 2015a,b, 2018, 2019b, 2021). This validation only focused on the periods, light curve shapes, and Fourier coefficients and we did not use positions on the PL or PW relations for classification here. If the software failed to calculate the Fourier coefficients, we only classified the star if we deemed the light curve shape conclusive enough through visual inspection. For the DCEP_MULTI candidates, we fitted all possible pulsation frequencies and calculated the frequency ratios. We also checked for the presence of significant secondary periodicities in the single-mode stars and calculated

Table 12. Data for the validation of 14 DCEPs with high accuracy RV curves available in the literature.

Source_id	Name	Mode	Period (days)	$\langle G \rangle$ (mag)	N_{RVS}	Amp(RV) (km s^{-1})	γ_{RVS} (km s^{-1})	$\gamma_{Lit.}$ (km s^{-1})	[Fe/H]$_{Lit.}$ (dex)	$\langle T_{eff} \rangle_{Lit.}$ (K)	$\langle \log g \rangle_{Lit.}$ (km s^{-1})	V_t	Bin.	P_{orbit} (days)	Ref1	Ref2	Ref3	Ins.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
5519380081746387328	AH Vel	1O	4.227132	5.53	16	16.4 ± 0.4	26.34 ± 0.10	24.4 ± 0.5	0.09	6040	2.2	4.3	B	>1000	1, 2	8	12	a, b
5597379741549105280	AQ Pup	F	30.18194	8.12	24	56.7 ± 3.6	65.45 ± 1.47	61.0 ± 0.8	−0.14	4940	0.6	5.6	B	~1400	16, 17	11	–	b
5848500161483878400	AV Cir	1O	3.06526	7.10	31	14.3 ± 0.2	5.11 ± 0.04	4.8 ± 0.4	0.14	6170	2.1	3.3	V	–	2	8	–	a
5873984023533350400	AX Cir	F	5.27337	5.63	74	30.2 ± 0.9	−15.78 ± 0.28	−14.7 ± 0.7	−0.04	5760	1.8	3.7	OV	6532 ± 35	3	8	13	b
6058439910929477120	BG Cru	1O	3.3425241	3.34	23	9.9 ± 0.1	−19.42 ± 0.07	−19.7 ± 0.5	−0.11	6250	2.1	4.3	B	4050, 4950, 6650	1, 2	9	14	a, b
5877460679352962048	BP Cir	1O	2.398106	7.29	38	16.7 ± 0.2	−18.18 ± 0.05	−18.0 ± 0.5	−0.01	6530	2.4	3.7	B	–	1, 3	8	–	b
2011892703004353792	CF Cas	F	4.875122	10.73	12	36.8 ± 10.6	−71.87 ± 3.64	−78.4 ± 1.1	−0.01	5510	1.7	4.0	–	–	4	10	–	a
1853025642297186688	DT Cyg	1O	2.498763	5.61	57	13.8 ± 0.1	−1.04 ± 0.03	−1.9 ± 0.6	0.16	6270	2.4	3.6	–	–	4	10	–	a
5546476927338700416	RS Pup	F	41.49233	6.46	25	50.8 ± 8.3	25.80 ± 1.58	25.0 ± 0.8	0.21	5070	1.0	5.0	–	–	5, 6	11	–	b
3409635486731094400	SZ Tau	1O	3.148786	6.23	20	20.7 ± 0.4	0.27 ± 0.12	−0.61 ± 0.5	0.15	5990	2.2	3.6	B	1244	4	10	15	a
6060173364074372352	S Cru	F	4.689765	6.36	31	44.9 ± 3.0	−6.58 ± 0.40	−5.1 ± 0.5	0.08	6460	2.1	4.1	–	–	7	11	–	a
6054829806275577216	T Cru	F	6.73324	6.38	23	28.8 ± 0.8	−5.94 ± 0.44	−9.6 ± 0.5	0.11	5590	1.7	4.3	B	–	7	9	–	a
5932569709575669504	V340 Nor	F	11.28895	7.98	37	18.8 ± 0.3	−38.89 ± 0.07	−30.3 ± 0.6	0.16	5730	2.0	5.3	–	–	4	8	–	a
2027263738133623168	X Vul	F	6.3196418	8.23	35	42.3 ± 5.2	−14.94 ± 0.94	−15.7 ± 0.6	0.13	5650	1.7	3.9	–	–	4	10	–	a

Notes. Meaning of the different columns is as follows: (1) *Gaia* DR3 source id; (2) Literature Name; (3) Mode of pulsation (for brevity F = Fundamental; 1O = First Overtone); (4) Pulsation period (P), as re-evaluated in the present analysis; (5) Intensity-averaged G-band mean magnitude, as derived from the SOS Cep&RRL pipeline; (6) Number of valid RVS RV measurements; (7) Peak-to-peak amplitude of the RVS RV curve and relative uncertainty; (8) Center of mass RV (γ) as estimated by SOS Cep&RRL pipeline and uncertainty; (9) As for the previous column but for the literature; (10) Iron abundance; (11) Mean effective temperature; (12) Mean gravity; (13) Microturbulent velocity; (14) Binary type; (15) Orbital period (P_{orbit}); (16) References for the literature RVs; (17) References for the stellar parameters; (18) References for P_{orbit}; (19) Instrument type: a = CORAVEL; b = Other spectrograph. Metallicities in Col. (10) are taken from Genovali et al. (2014). The meaning of the numbers in Cols. (16)–(18) is as follows: 1 = Gallenne et al. (2019); 2 = Kienzle et al. (1999); 3 = Petterson et al. (2005); 4 = Bersier et al. (1994); 5 = Anderson (2014); 6 = Storm et al. (2004); 7 = Bersier (2002); 8 = Luck & Lambert (2011); 9 = Usenko et al. (2014); 10 = Andrievsky et al. (2002); 11 = Andrievsky et al. (2013); 12 = Gieren (1977); 13 = Petterson et al. (2004); 14 = Szabados (1989); 15 = Gorynya et al. (1996); 16 = Anderson et al. (2016); 17 = Storm et al. (2011). The binary types listed in Col. (14) are taken from the website `https://konkoly.hu/CEP/nagytab3.html` maintained by L. Szabados. The different symbols mean: B = spectroscopic binary; O = spectroscopic binary with known orbit; V = visual binary.

period ratios for any potential DCEP_MULTI stars. As TESS sectors are 27 d in duration, we were effectively limited to <20 d periods. For some long-period stars, we were able to stitch data from consecutive sectors but this was limited to high and low ecliptic latitudes and was prone to brightness differences and other systematic errors.

Overall we searched for light curves for 4690 stars and were able to classify 2378 (51%) of those. The validation results show strong agreement between the *Gaia* and TESS classifications. The largest discrepancy occurs among the 1O/2O DCEP_MULTI stars, where we identified a significant number of further stars classified as single-mode DCEP_1O in DR3. We also identified six stars as 1O/2O/3O DCEP_MULTI pulsators. This subclass is not included in DR3 but is known among the OGLE Cepheids.

Finally, we investigated the possible reasons for missing a significant amount of DM Cepheids in the DR3 classification. Figure 20 displays four diagnostic quantities: the upper panel shows the brightness of the stars (in G_{RP} band) against the number of epochs in the light curves; the lower panel shows the amplitude ratio of the modes (calculated from the Fourier amplitudes of the pulsation frequencies) against the logarithm of the periods. The plots indicate that the number of epochs and brightness had little effect on the detection, with the brightest and most well-sampled stars having the highest positive detection rate in *Gaia*. The main driver for detection success appears to be the mode amplitude ratio, with all stars above 40% identified from the *Gaia* data. Longer period DCEP_MULTI stars also seem to be easier to discover from the sparse photometry collected by the mission.

6.6. Validation of RV data

It is important to validate the RV curves of Cepheids published in *Gaia* DR3, as they have important applications, especially in the case of DCEPs. For example, the Baade-Wesselink method is widely used to estimate the radius and the distance of radial pulsators of different types and Cepheids in particular from the combination of light and RV curves (e.g. Wesselink 1946; Gautschy 1987; Ripepi et al. 1997; Gieren et al. 2018, and references therein). We searched the literature for Cepheids with complete and reliable RV curves. As a result, we considered 14 DCEPs, complete properties of which are listed in Table 12. The comparison between the centre-of-mass velocity estimated from the *Gaia* RVs and those from the literature shows good agreement within 1−2σ. The only discrepant object is V340 Nor for which the SOS Cep&RRL pipeline uncertainty is perhaps underestimated. In summary, the Cepheid RV time series released with *Gaia* DR3 are reliable and can be used to derive the intrinsic parameters of the stars. Examples of the comparison between *Gaia* and literature RVs are shown in Fig. 21.

7. Conclusions

In this paper, we present the *Gaia* DR3 catalogue of Cepheids of all types. We discuss the changes in the SOS Cep&RRL pipeline with respect to DR2, including the derivation of a full set of *PL* and *PW* relations adopted in the pipeline. The major novelties in DR3 compared to the previous release are the analysis of DCEPs in the distant galaxies M 31 and M 33, and the analysis of the RV data for a subsample of 799 Cepheids of all types, including 24 objects belonging to LMC and SMC.

We describe the techniques adopted to carry out a first gross cleaning of the sample for the large number of spurious objects retrieved from the general classification catalogue. In this process we also made use of machine learning techniques which significantly helped to single out the most promising candidates for further analysis.

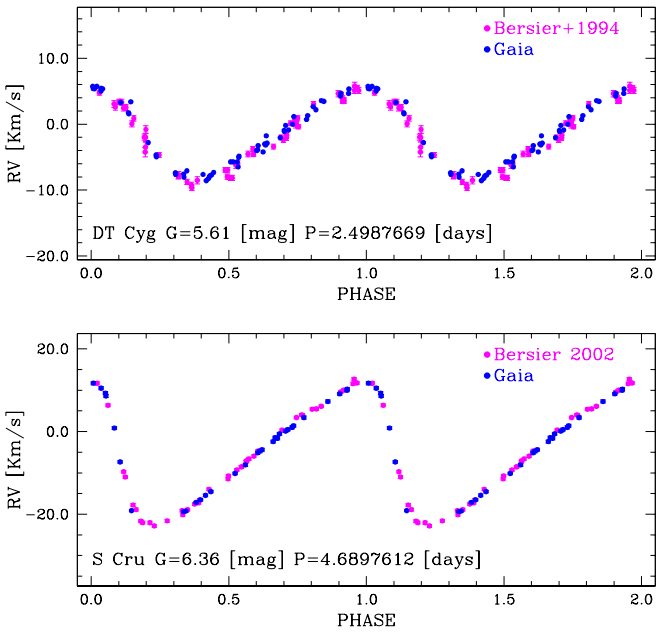

Fig. 21. Examples of the comparison between the *Gaia* and the literature RV curves for a DCEP_1O (DT Cyg) and a DCEP_F (S Cru).

To obtain maximum purity in the sample, we visually analysed almost all the candidates and corrected the classification provided from the *Gaia* SOS Cep&RRL pipeline when it was incorrect. In this context, the *G* time series of a number of suspect multi-mode pulsators was re-processed to determine correct pulsation periods.

In total, the *Gaia* DR3 catalogue counts 15 006 Cepheids of all types, among which 327 objects were known variable stars with a different classification in the literature, while, to our knowledge, 474 stars either had not been reported previously or had non-Cepheid type classification in the literature, and are therefore likely new Cepheid discoveries by *Gaia*.

The validation of the DR3 catalogue was carried out via comparison with literature results and through analysis of a consistent sample of light curves from TESS. The overall purity of the sample is very high and certainly larger than 90%−95%. The completeness varies significantly from one region in the sky to another and also as a function of Cepheid type. Completeness is larger than 90% in LMC and SMC overall, and is on the order of 10%−20% in M 31 and M 33. Concerning the All Sky sample, which is largely dominated by MW objects, the completeness for DCEPs is likely between 85% and 90%, with contamination of a few percent. The completeness is lower for ACEPs and especially T2CEPs, which are located in large numbers in the MW bulge, a region for which *Gaia* has not yet collected a sufficient amount of epoch data. Validation of the RV curves with literature data showed that the *Gaia* RV curves for Cepheids are generally accurate and usable for astrophysical purposes.

Compared to DR2, the Cepheids in DR3 represent a huge improvement both quantitatively, given the addition of about 5000 Cepheids of all types, and qualitatively, as the DR3 Cepheid catalogue has a much improved purity, especially for the All Sky sample. In addition, a significant benefit of DR3 is the release of RV time series for 799 Cepheids of all types.

The following release (DR4) will present further improvements compared to DR3, mainly due to the additional 24 months of data, which in turn will lead to more accurate period determinations. For the next release, we plan to thoroughly use the machine learning technique that was implemented to clean the DR3 sample. In this respect, the present *Gaia* DR3 Cepheid sample, with its high purity, will represent an excellent training set.

Acknowledgements. We wish to thank our anonymous Referee whose comments helped us to improve the manuscript. 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. The Italian participation in DPAC has been supported by Istituto Nazionale di Astrofisica (INAF) and the Agenzia Spaziale Italiana (ASI) through grants I/037/08/0, I/058/10/0, 2014-025-R.0, 2014-025-R.1.2015 and 2018-24-HH.0 to INAF (PI M.G. Lattanzi). The Swiss participation by the Swiss State Secretariat for Education, Research and Innovation through the "Activités Nationales Complémentaires". UK community participation in this work has been supported by funding from the UK Space Agency, and from the UK Science and Technology Research Council. This work was supported in part by the French Centre National de la Recherche Scientifique (CNRS), the Centre National d'Etudes Spatiales (CNES), the Institut des Sciences de l'Univers (INSU) through the Service National d'Observation (SNO) *Gaia*. This research was supported by the 'SeismoLab' KKP-137523 Élvonal grant of the Hungarian Research, Development and Innovation Office (NKFIH), and by the LP2018-7 Lendület grant of the Hungarian Academy of Sciences. This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France. It is a pleasure to thank M.B. Taylor for developing the TOPCAT software, which was very useful in carrying out this work.

References

Anderson, R. I. 2014, A&A, 566, L10
Anderson, R. I., Casertano, S., Riess, A. G., et al. 2016, ApJS, 226, 18
Andrievsky, S. M., Kovtyukh, V. V., Luck, R. E., et al. 2002, A&A, 392, 491
Andrievsky, S. M., Lépine, J. R. D., Korotin, S. A., et al. 2013, MNRAS, 428, 3252
Arenou, F., & Luri, X. 1999, in Harmonizing Cosmic Distance Scales in a Post-Hipparcos Era, eds. D. Egret, & A. Heck, ASP Conf. Ser., 167, 13
Bersier, D. 2002, ApJS, 140, 465
Bersier, D., Burki, G., Mayor, M., & Duquennoy, A. 1994, A&AS, 108, 25
Bódi, A., Szabó, P., Plachy, E., Molnár, L., & Szabó, R. 2022, PASP, 134, 014503
Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, A&A, 640, A1
Cappellari, M., Scott, N., Alatalo, K., et al. 2013, MNRAS, 432, 1709
Caputo, F. 1998, A&ARv, 9, 33
Caputo, F., Marconi, M., Musella, I., & Santolamazza, P. 2000, A&A, 359, 1059
Caputo, F., Castellani, V., Degl'Innocenti, S., Fiorentino, G., & Marconi, M. 2004, A&A, 424, 927
Castro-Ginard, A., Jordi, C., Luri, X., et al. 2022, A&A, 661, A118
Chen, X., Wang, S., Deng, L., et al. 2020, ApJS, 249, 18
Clement, C. M., Muzzin, A., Dufton, Q., et al. 2001, AJ, 122, 2587
Clementini, G., Ripepi, V., Leccia, S., et al. 2016, A&A, 595, A133
Clementini, G., Ripepi, V., Molinaro, R., et al. 2019, A&A, 622, A60
Clementini, G., Ripepi, V., Molinaro, R., et al. 2023, A&A, 674, A18 (*Gaia* DR3 SI)
Conn, A. R., Ibata, R. A., Lewis, G. F., et al. 2012, ApJ, 758, 11
De Somma, G., Marconi, M., Molinaro, R., et al. 2023, ApJS, submitted
Drake, A. J., Graham, M. J., Djorgovski, S. G., et al. 2014, ApJS, 213, 9
Drake, A. J., Djorgovski, S. G., Catelan, M., et al. 2017, MNRAS, 469, 3688
Eyer, L., Audard, M., Holl, B., et al. 2023, A&A, 674, A13 (*Gaia* DR3 SI)
Feast, M. W., & Catchpole, R. M. 1997, MNRAS, 286, L1
Feast, M. W., Laney, C. D., Kinman, T. D., van Leeuwen, F., & Whitelock, P. A. 2008, MNRAS, 386, 2115
Feinstein, A. D., Montet, B. T., Foreman-Mackey, D., et al. 2019, PASP, 131, 094502
Gaia Collaboration (Prusti, T., et al.) 2016a, A&A, 595, A1
Gaia Collaboration (Brown, A. G. A., et al.) 2016b, A&A, 595, A2
Gaia Collaboration (Brown, A. G. A., et al.) 2018, A&A, 616, A1
Gaia Collaboration (Eyer, L., et al.) 2019, A&A, 623, A110
Gaia Collaboration (Brown, A. G. A., et al.) 2021a, A&A, 649, A1
Gaia Collaboration (Luri, X., et al.) 2021b, A&A, 649, A7
Gaia Collaboration (Vallenari, A., et al.) 2023a, A&A, 674, A1 (*Gaia* DR3 SI)
Gaia Collaboration (Drimmel, R., et al.) 2023b, A&A, 674, A37 (*Gaia* DR3 SI)
Gallenne, A., Kervella, P., Borgniet, S., et al. 2019, A&A, 622, A164
Gautschy, A. 1987, Vistas Astron., 30, 197
Genovali, K., Lemasle, B., Bono, G., et al. 2014, A&A, 566, A37
Gieren, W. 1977, A&AS, 28, 193

Gieren, W., Storm, J., Konorski, P., et al. 2018, A&A, 620, A99
Gorynya, N. A., Samus', N. N., Rastorguev, A. S., & Sachkov, M. E. 1996, Astron. Lett., 22, 175
Heinze, A. N., Tonry, J. L., Denneau, L., et al. 2018, AJ, 156, 241
Holl, B., Audard, M., Nienartowicz, K., et al. 2018, A&A, 618, A30
Holl, B., Fabricius, C., Portell, J., et al. 2023, A&A, 674, A25 (*Gaia* DR3 SI)
Huang, C. X., Vanderburg, A., Pál, A., et al. 2020a, Res. Notes Am. Astron. Soc., 4, 204
Huang, C. X., Vanderburg, A., Pál, A., et al. 2020b, Res. Notes Am. Astron. Soc., 4, 206
Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019, MNRAS, 486, 1907
Kienzle, F., Moskalik, P., Bersier, D., & Pont, F. 1999, A&A, 341, 818
Kinemuchi, K., Harris, H. C., Smith, H. A., et al. 2008, AJ, 136, 1921
Klagyivik, P., Szabados, L., Szing, A., Leccia, S., & Mowlavi, N. 2013, MNRAS, 434, 2418
Kodric, M., Riffeser, A., Hopp, U., et al. 2018, AJ, 156, 130
Kunimoto, M., Huang, C., Tey, E., et al. 2021, Res. Notes Am. Astron. Soc., 5, 234
Leavitt, H. S., & Pickering, E. C. 1912, Harv. Coll. Obs. Circ., 173, 1
Lenz, P., & Breger, M. 2005, Commun. Asteroseismol., 146, 53
Luck, R. E. 2018, AJ, 156, 171
Luck, R. E., & Lambert, D. L. 2011, AJ, 142, 136
Madore, B. F. 1982, ApJ, 253, 575
Marconi, M., Fiorentino, G., & Caputo, F. 2004, A&A, 417, 1101
Matsunaga, N., Feast, M. W., & Soszyński, I. 2011, MNRAS, 413, 223
Pellerin, A., & Macri, L. M. 2011, ApJS, 193, 26
Perina, S., Federici, L., Bellazzini, M., et al. 2009, A&A, 507, 1375
Petterson, O. K. L., Cottrell, P. L., & Albrow, M. D. 2004, MNRAS, 350, 95
Petterson, O. K. L., Cottrell, P. L., Albrow, M. D., & Fokin, A. 2005, MNRAS, 362, 1167
Pietrukowicz, P., Soszyński, I., & Udalski, A. 2021, Acta Astron., 71, 205
Plachy, E., Pál, A., Bódi, A., et al. 2021, ApJS, 253, 11
Poggio, E., Drimmel, R., Cantat-Gaudin, T., et al. 2021, A&A, 651, A104
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, J. Astron. Telesc. Instrum. Syst., 1, 014003
Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3
Riess, A. G., Macri, L. M., Hoffmann, S. L., et al. 2016, ApJ, 826, 56
Rimoldini, L., Holl, B., Gavras, P., et al. 2023, A&A, 674, A14 (*Gaia* DR3 SI)
Ripepi, V., Barone, F., Milano, L., & Russo, G. 1997, A&A, 318, 797
Ripepi, V., Marconi, M., Moretti, M. I., et al. 2014, MNRAS, 437, 2307
Ripepi, V., Moretti, M. I., Marconi, M., et al. 2015, MNRAS, 446, 3034
Ripepi, V., Cioni, M.-R. L., Moretti, M. I., et al. 2017, MNRAS, 472, 808
Ripepi, V., Molinaro, R., Musella, I., et al. 2019, A&A, 625, A14
Ripepi, V., Catanzaro, G., Clementini, G., et al. 2022a, A&A, 659, A167
Ripepi, V., Chemin, L., Molinaro, R., et al. 2022b, MNRAS, 512, 563
Romaniello, M., Primas, F., Mottini, M., et al. 2008, A&A, 488, 731
Romaniello, M., Riess, A., Mancino, S., et al. 2022, A&A, 658, A29
Sandage, A., & Tammann, G. A. 2006, ARA&A, 44, 93
Sartoretti, P., Blomme, R., David, M., et al. 2022, Gaia DR2 Documentation, European Space Agency; Gaia Data Processing and Analysis Consortium, 6
Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, ApJ, 788, 48
Skowron, D. M., Skowron, J., Mróz, P., et al. 2019, Science, 365, 478
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2015a, Acta Astron., 65, 233
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2015b, Acta Astron., 65, 297
Soszyński, I., Pawlak, M., Pietrukowicz, P., et al. 2016a, Acta Astron., 66, 405
Soszyński, I., Smolec, R., Dziembowski, W. A., et al. 2016b, MNRAS, 463, 1332
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2017, Acta Astron., 67, 103
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2018, Acta Astron., 68, 89
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2019a, Acta Astron., 69, 87
Soszyński, I., Udalski, A., Wrona, M., et al. 2019b, Acta Astron., 69, 321
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2020, Acta Astron., 70, 101
Soszyński, I., Pietrukowicz, P., Skowron, J., et al. 2021, Acta Astron., 71, 189
Storm, J., Carney, B. W., Gieren, W. P., et al. 2004, A&A, 415, 521
Storm, J., Gieren, W., Fouqué, P., et al. 2011, A&A, 534, A94
Szabados, L. 1989, Commun. Konkoly Obs. Hung., 94, 1
Tarricq, Y., Soubiran, C., Casamiquela, L., et al. 2022, A&A, 659, A59
Taylor, M. B. 2005, in Astronomical Data Analysis Software and Systems XIV, eds. P. Shopbell, M. Britton, & R. Ebert, ASP Conf. Ser., 347, 29
Torrealba, G., Catelan, M., Drake, A. J., et al. 2015, MNRAS, 446, 2251
Udalski, A., Soszyński, I., Pietrukowicz, P., et al. 2018, Acta Astron., 68, 315
Usenko, I. A., Kniazev, A. Y., Berdnikov, L. N., & Kravtsov, V. V. 2014, Astron. Lett., 40, 800
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
Wesselink, A. J. 1946, Bull. Astron. Inst. Neth., 10, 91

Appendix A: Light curve examples

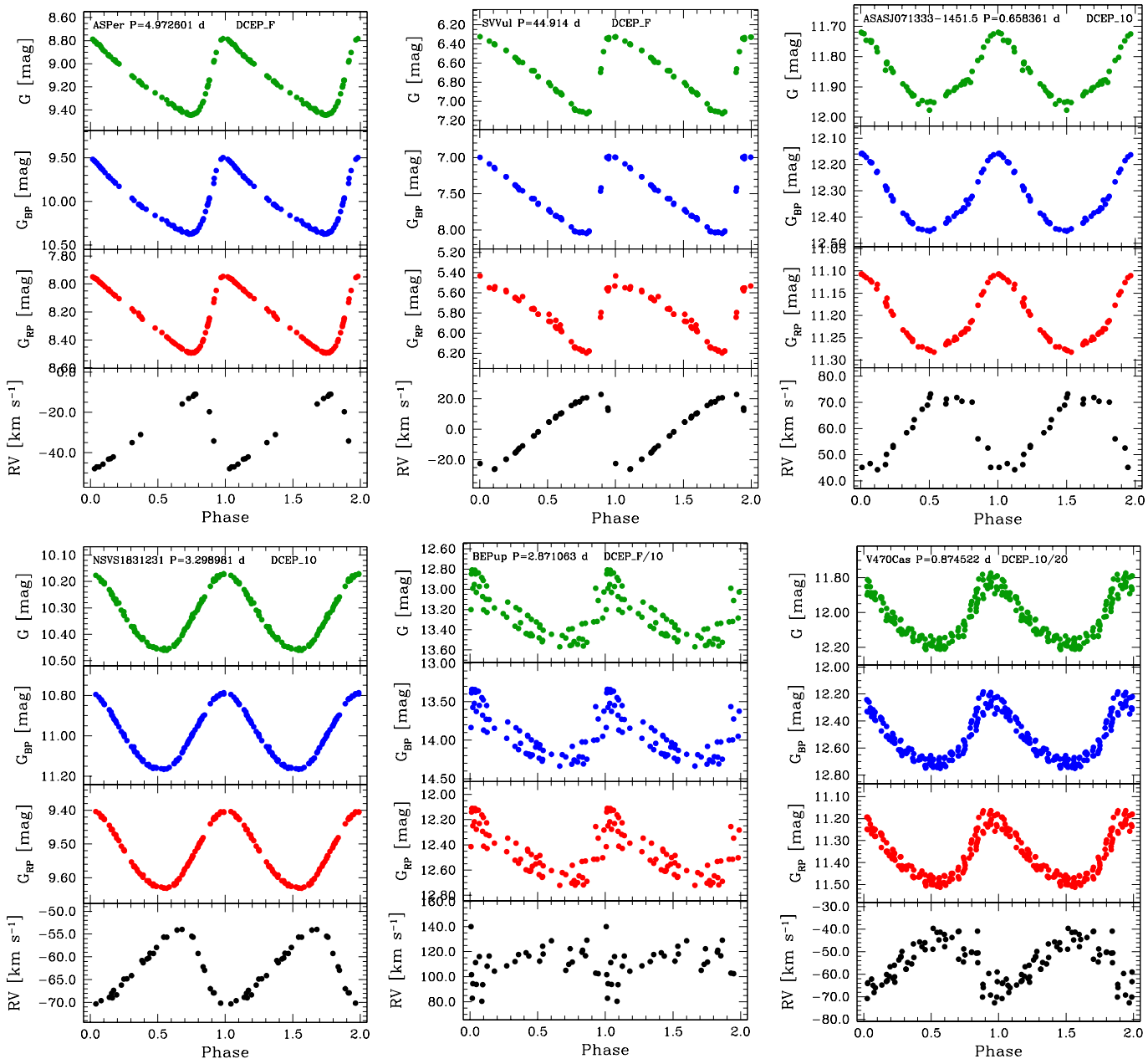

Fig. A.1. Light and RV curves for a selected sample of DCEPs of different modes.

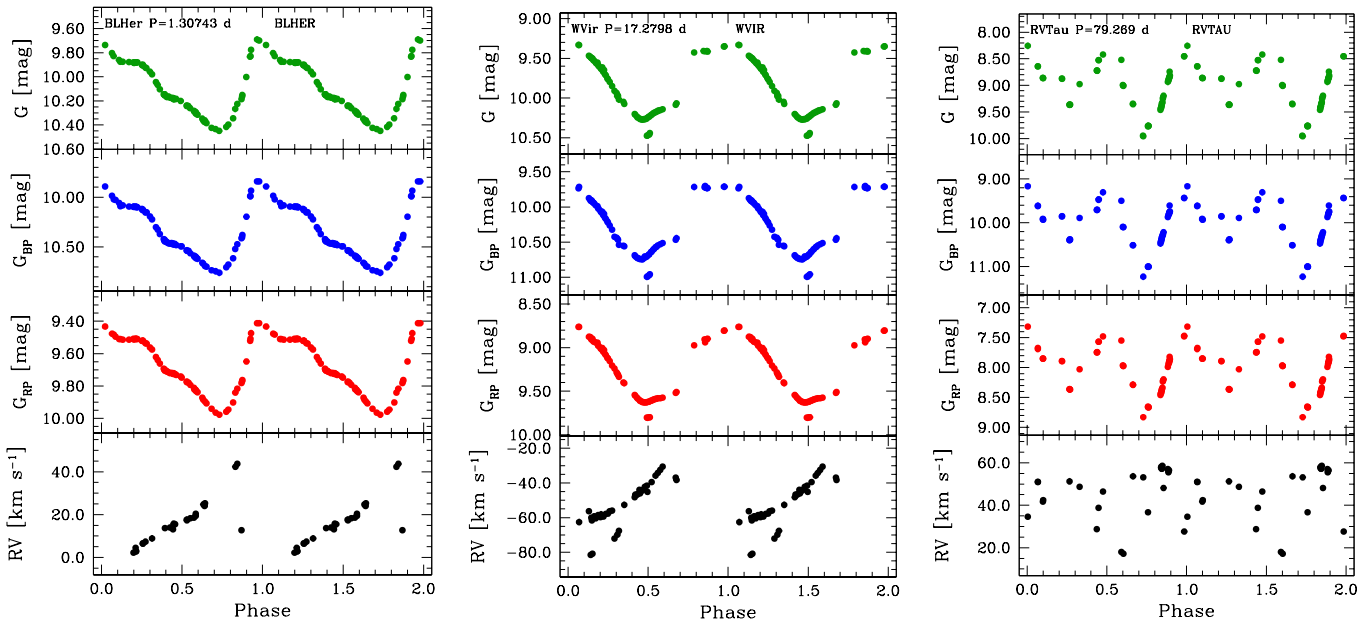

Fig. A.2. Light and RV curves for the prototypes of the BLHER (left), WVIR (centre), and RVTAU (right) classes.

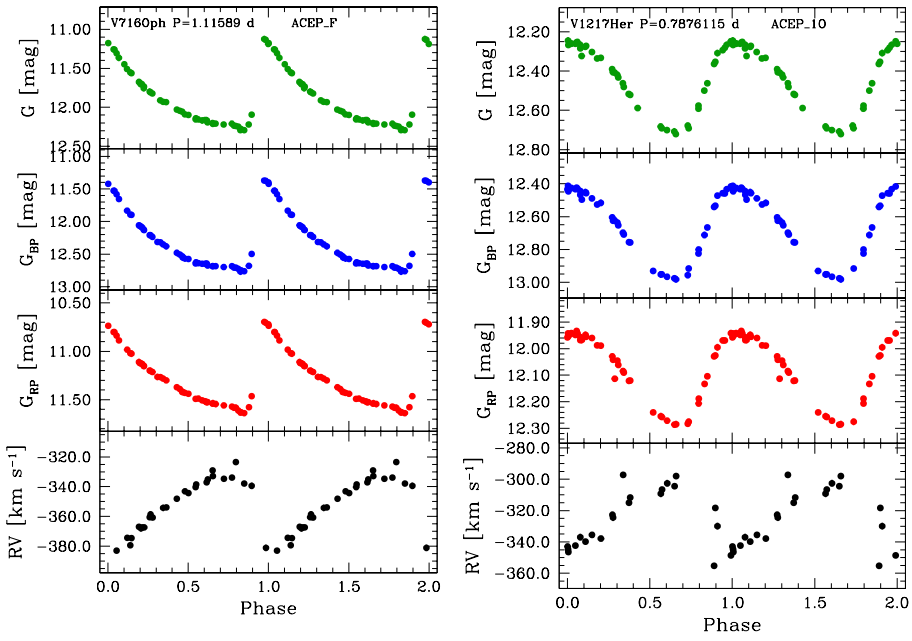

Fig. A.3. Light and RV curves for ACEP_F (left) and ACEP_1O (right) variables.

Appendix B: Fourier parameters for the LMC, SMC, M31, and M33 Cepheid samples

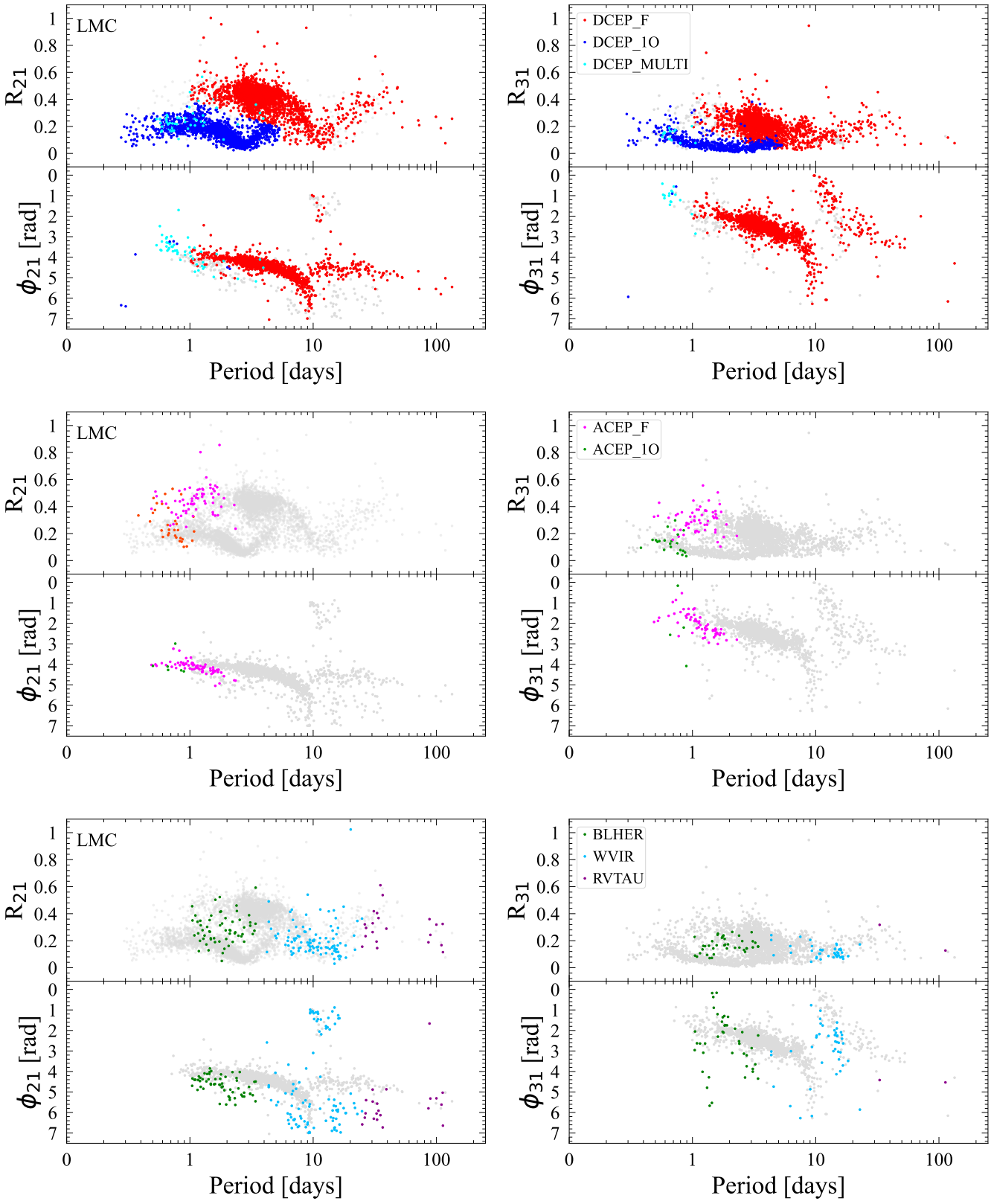

Fig. B.1. Same as in Fig. 7 but for the LMC.

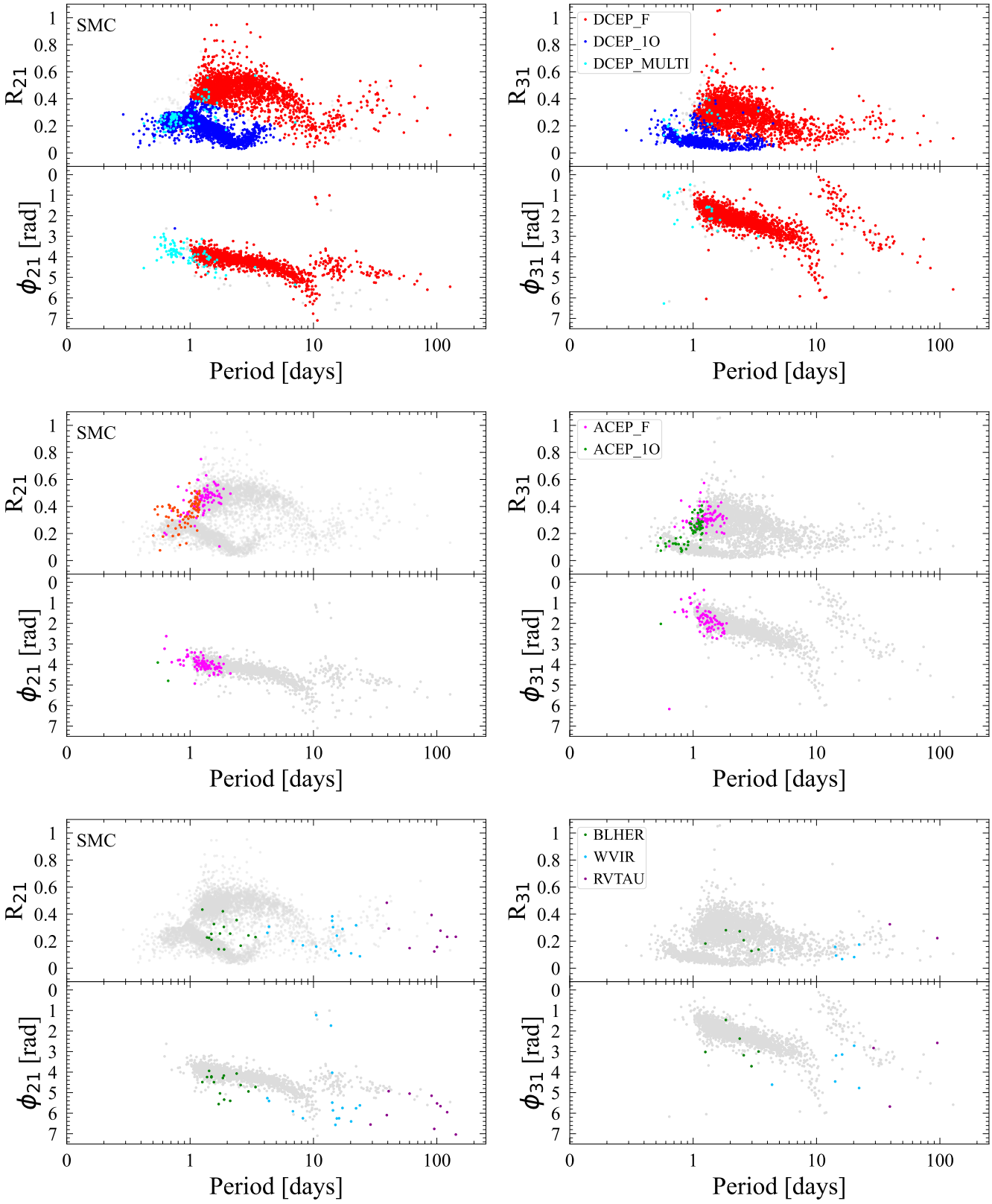

Fig. B.2. Same as in Fig. 7 but for the SMC.

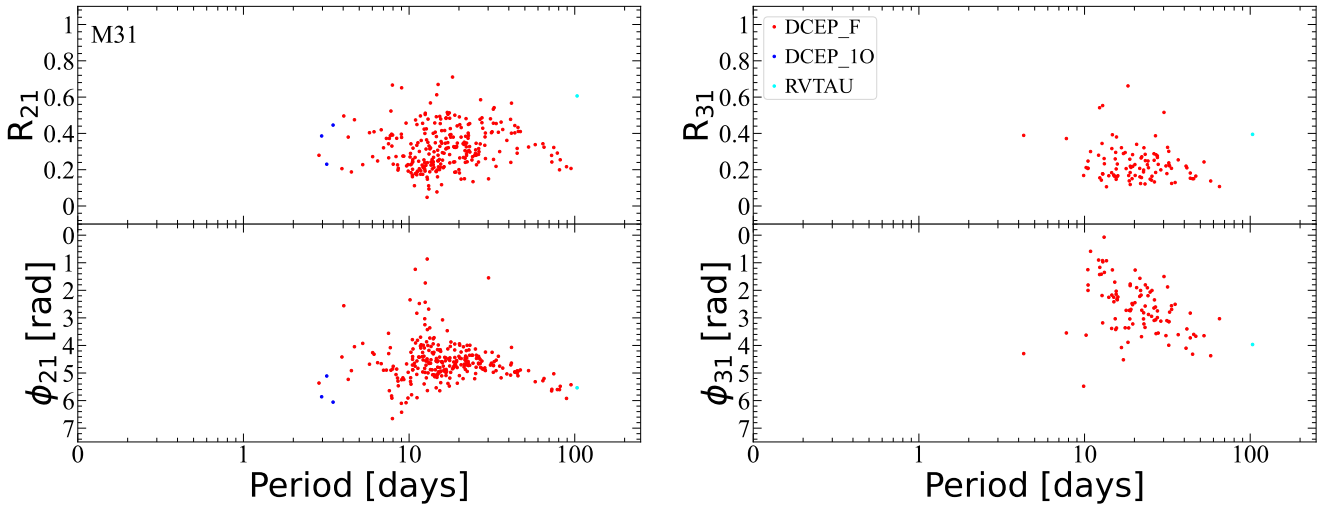

Fig. B.3. Fourier parameters for the M31 DCEPs.

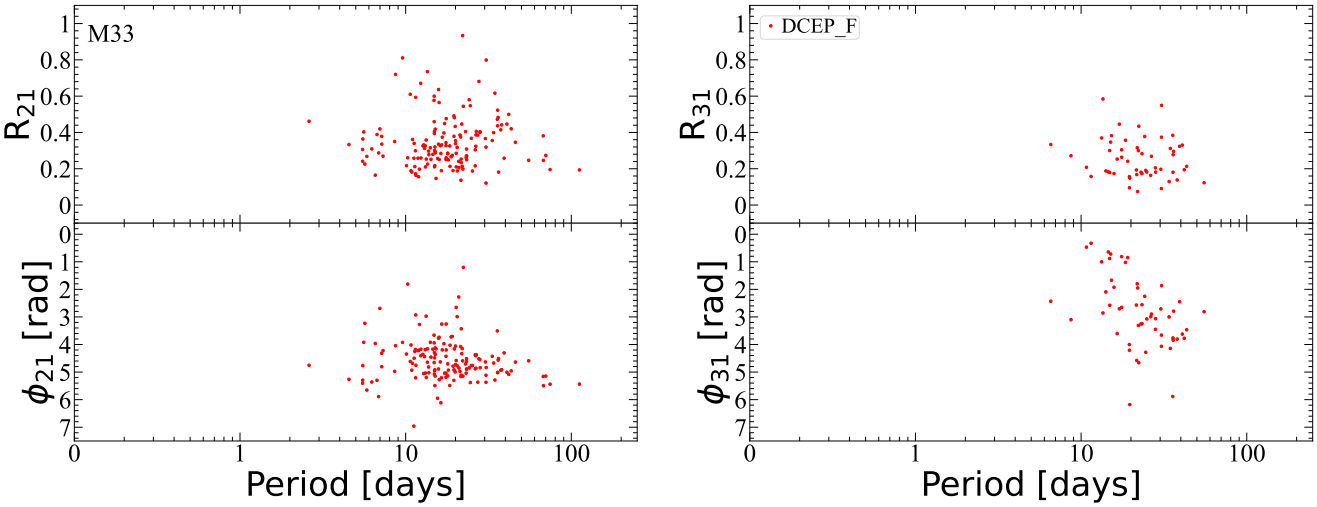

Fig. B.4. Fourier parameters for the M33 DCEPs.

Appendix C: *PL* and *PW* relations for the LMC, SMC, M31, and M33 Cepheid samples

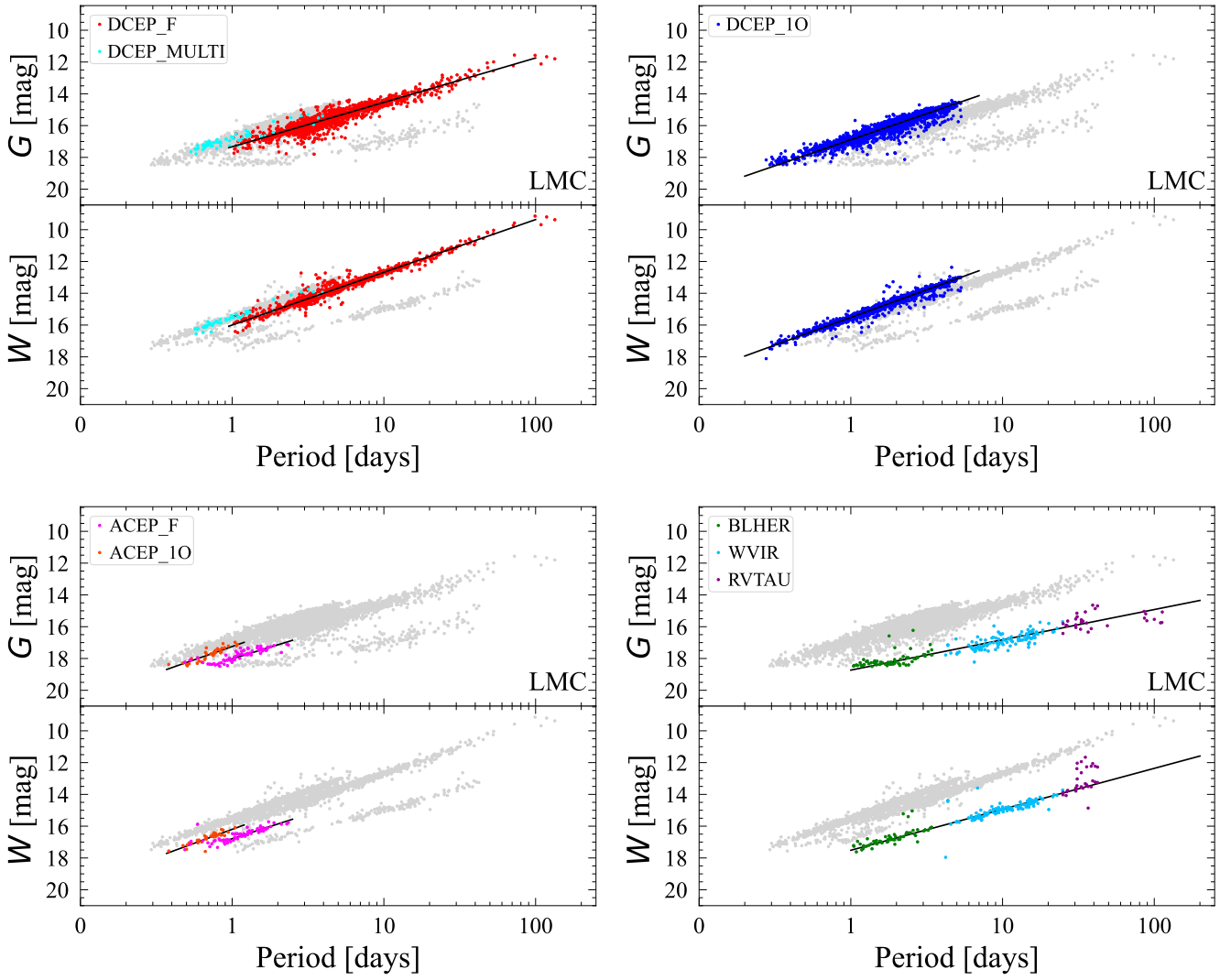

Fig. C.1. *PL* in the *G*-band and *PW* relations for the LMC Cepheids. The top panels show results for the DCEPs, while the bottom panels display ACEPs and T2CEPs.

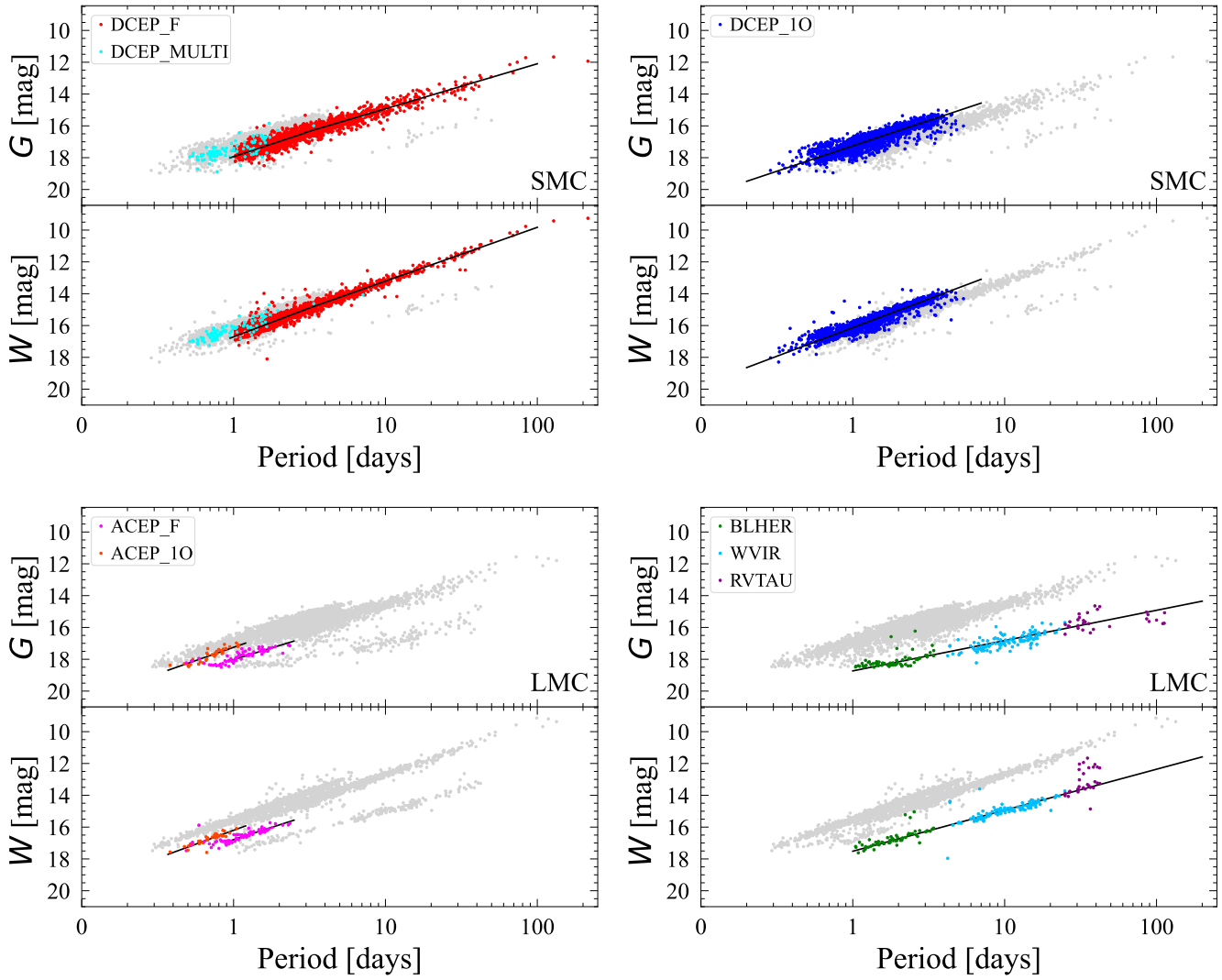

Fig. C.2. Same as in Fig. C.1 but for the SMC.

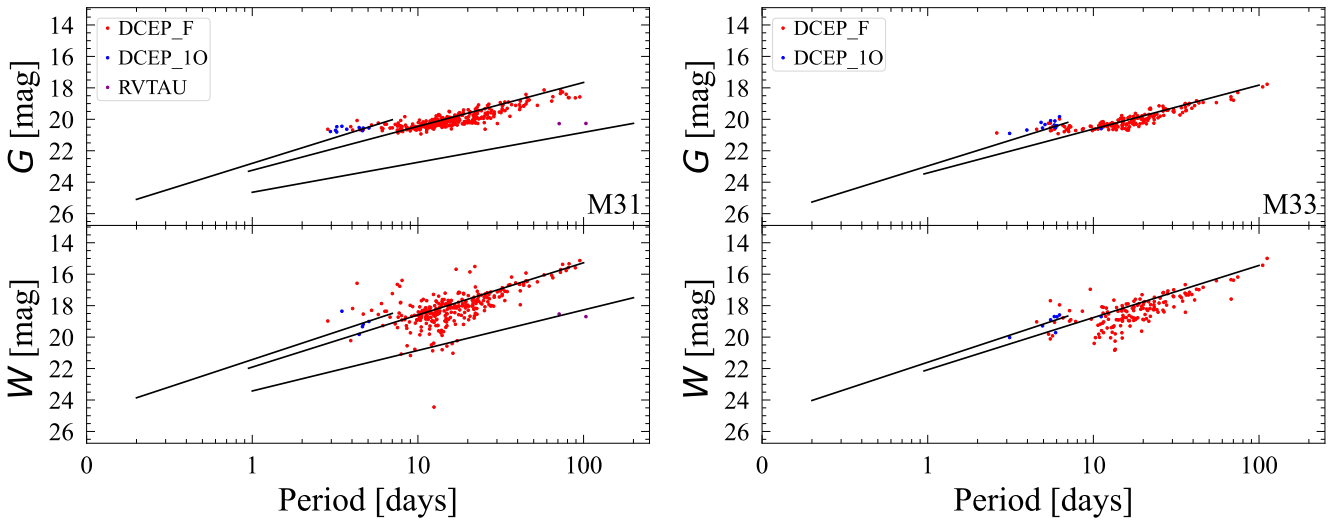

Fig. C.3. *PL* in the *G*-band and *PW* relations for the Cepheids in M31 (left panel) and M33 (right panel).

Appendix D: CMDs for the LMC, SMC, M31, and M33 Cepheid samples

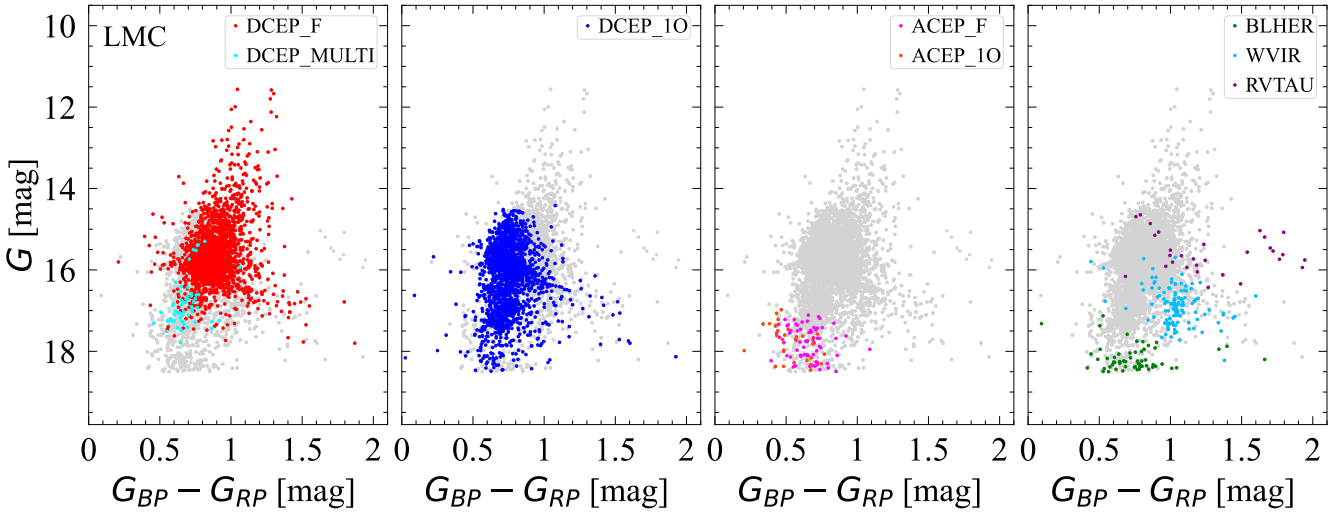

Fig. D.1. CMD in apparent G magnitude of the LMC Cepheid sample.

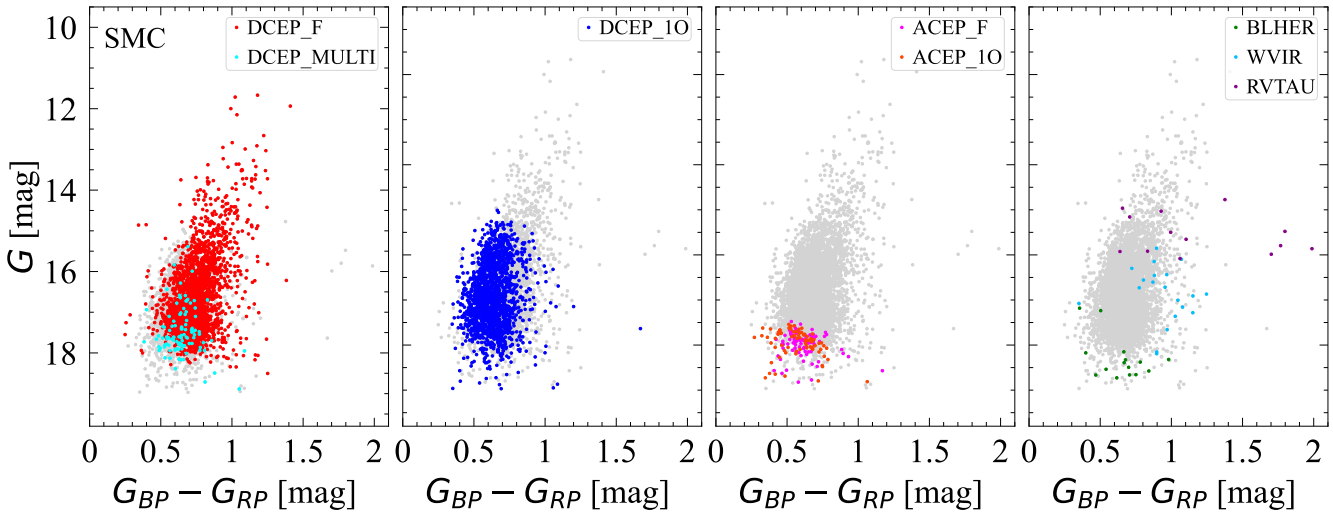

Fig. D.2. CMD in apparent G magnitude of the SMC Cepheid sample.

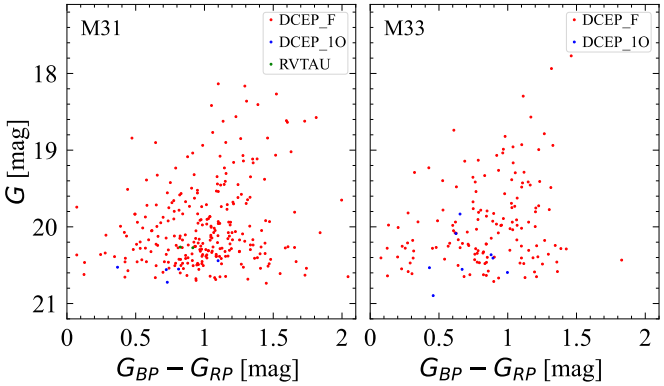

Fig. D.3. CMD in apparent G magnitude of the M31 (left pane)l and M33 (right panel) Cepheid samples.

Appendix E: Period–amplitude diagram for the LMC, SMC, M31, and M33 Cepheid samples

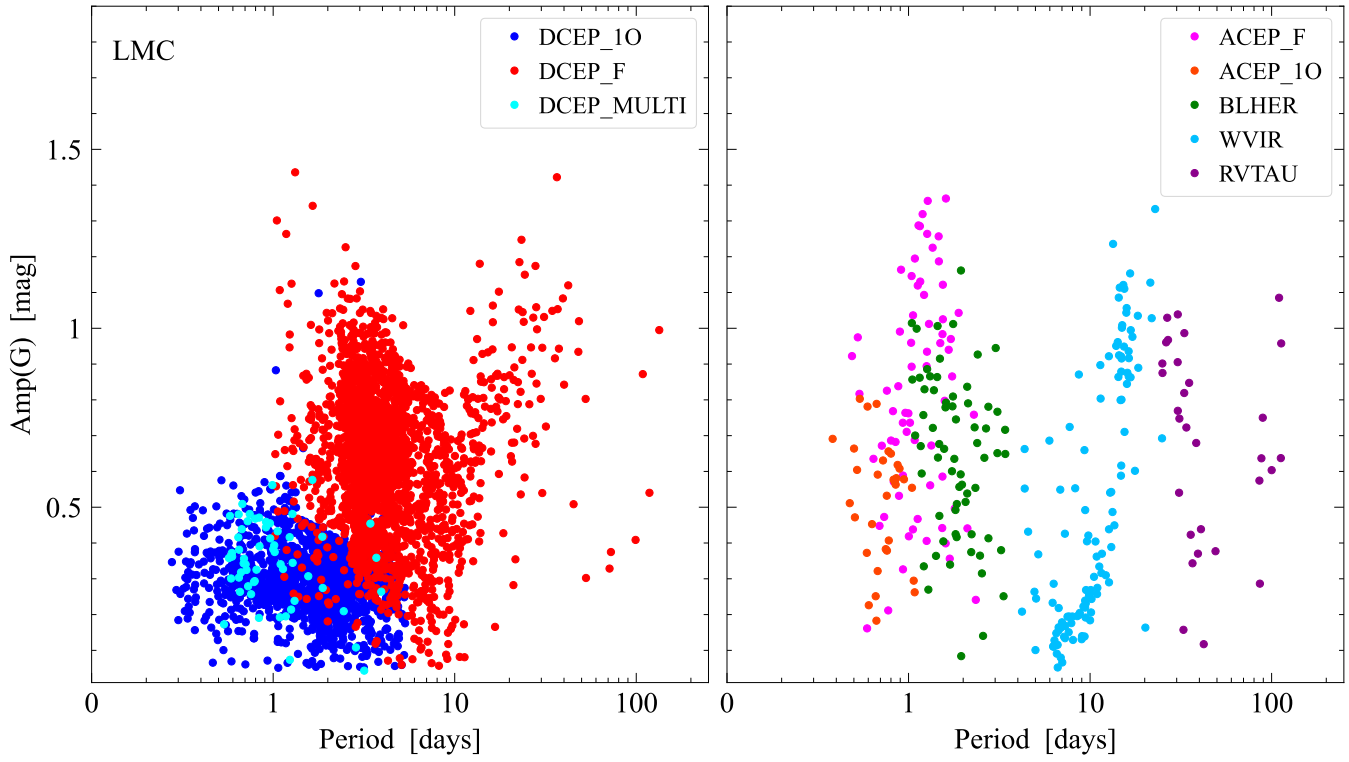

Fig. E.1. Period–amplitude(G) diagram for the LMC sample.

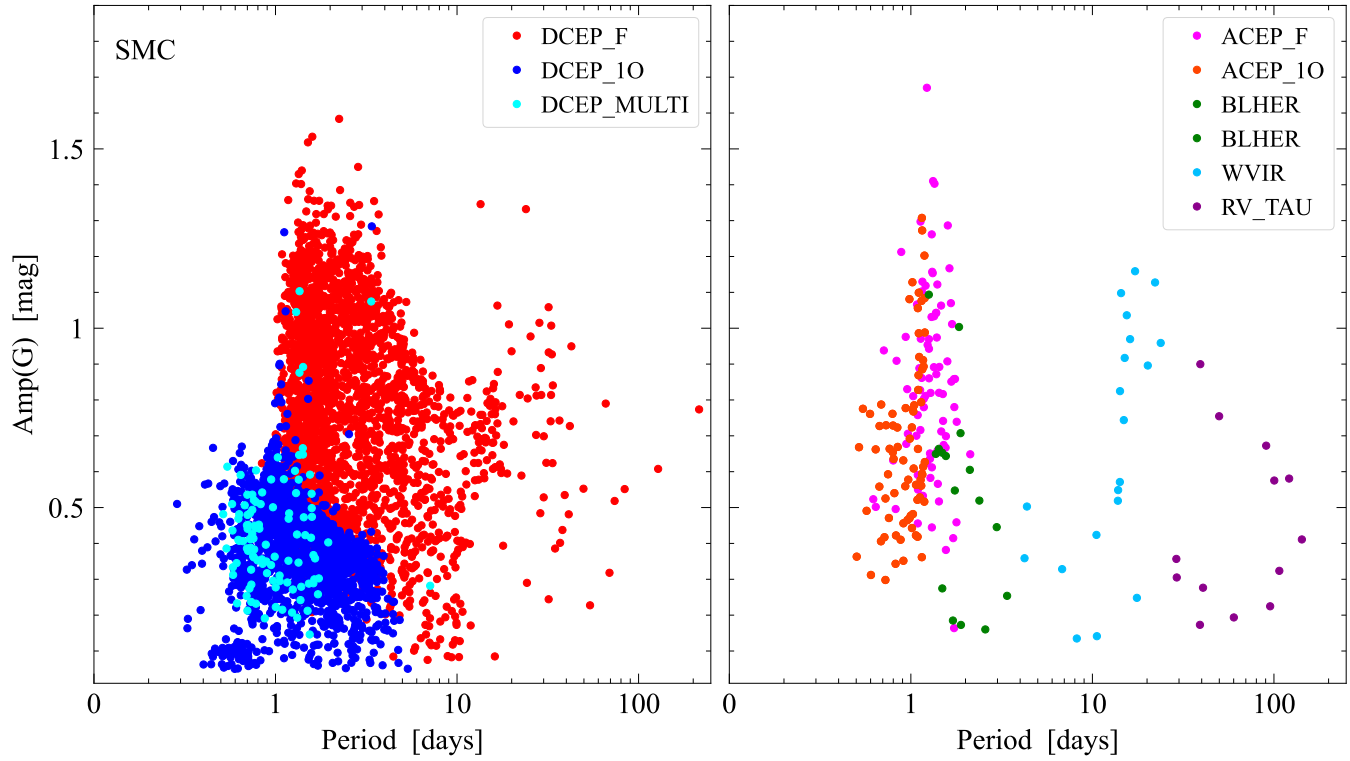

Fig. E.2. Period–amplitude(G) diagram for the SMC sample.

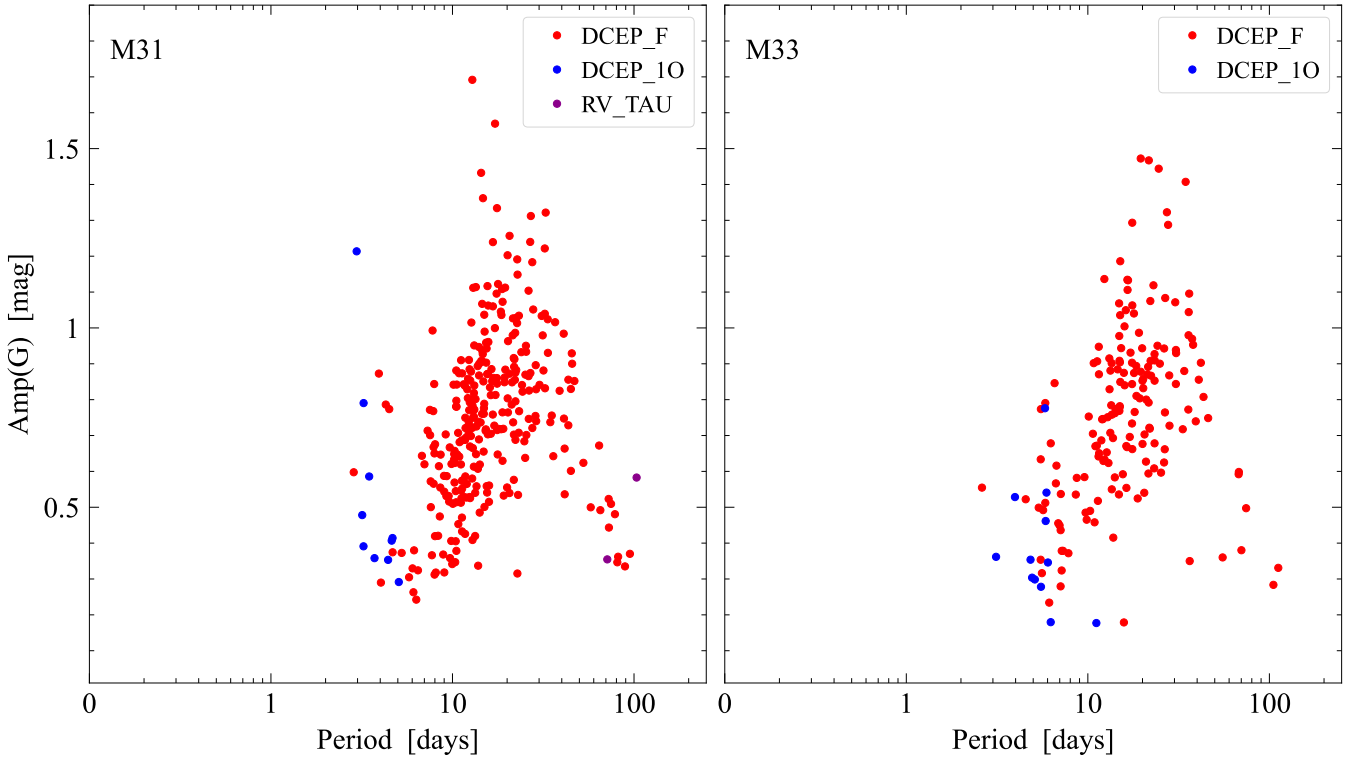

Fig. E.3. Period–amplitude(G) diagram for the M31 (left panel) and M33 (right panel) samples, respectively.

Appendix F: Confusion matrices

Literature SOS	ACEP_F (38)	ACEP_1O (11)	DCEP_F (1837)	DCEP_1O (771)	DCEP_M (188)	BLHER (420)	WVIR (496)	RVTAU (156)	Recall (%)
ACEP_F (46)	**29 (76.3%)**	0 (0.0%)	5 (0.2%)	0 (0.0%)	0 (0.0%)	12 (2.8%)	0 (0.0%)	0 (0.0%)	63.0
ACEP_1O (20)	1 (2.6%)	**8 (72.7%)**	0 (0.0%)	4 (0.5%)	0 (0.0%)	7 (1.6%)	0 (0.0%)	0 (0.0%)	40.0
DCEP_F (1860)	6 (15.7%)	0 (0.0%)	**1799 (97.9%)**	20 (2.5%)	13 (6.9%)	2 (0.4%)	19 (3.8%)	1 (0.6%)	96.7
DCEP_1O (790)	0 (0.0%)	3 (27.2%)	10 (0.5%)	**727 (94.2%)**	42 (22.3%)	7 (1.6%)	1 (0.2%)	0 (0.0%)	92.0
DCEP_M (139)	0 (0.0%)	0 (0.0%)	2 (0.1%)	4 (0.5%)	**133 (70.7%)**	0 (0.0%)	0 (0.0%)	0 (0.0%)	95.6
BLHER (409)	2 (5.2%)	0 (0.0%)	7 (0.3%)	12 (1.5%)	0 (0.0%)	**388 (92.3%)**	0 (0.0%)	0 (0.0%)	94.8
WVIR (505)	0 (0.0%)	0 (0.0%)	13 (0.7%)	4 (0.5%)	0 (0.0%)	4 (0.9%)	**475 (95.7%)**	9 (5.7%)	94.0
RVTAU (148)	0 (0.0%)	0 (0.0%)	1 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.2%)	**146 (93.5%)**	98
Precision (%)	76.3	72.7	97.9	94.2	70.7	92.3	95.7	93.5	

Fig. F.1. Confusion matrix for the All Sky sample. The percentages between parenthesis are calculated with respect to the literature.

Literature SOS	ACEP_F (68)	ACEP_1O (35)	DCEP_F (2343)	DCEP_1O (1667)	DCEP_M (318)	BLHER (63)	WVIR (113)	RVTAU (25)	Recall (%)
ACEP_F (64)	**61 (89.7%)**	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (3.1%)	0 (0.0%)	0 (0.0%)	95.3
ACEP_1O (30)	0 (0.0%)	**27 (77.1%)**	0 (0.0%)	2 (0.1%)	1 (0.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	90.0
DCEP_F (2356)	6 (8.8%)	0 (0.0%)	**2284 (97.4%)**	33 (1.9%)	30 (9.4%)	0 (0.0%)	3 (2.6%)	0 (0.0%)	96.9
DCEP_1O (1926)	0 (0.0%)	8 (22.8%)	50 (2.1%)	**1631 (97.8%)**	231 (72.6%)	1 (1.5%)	0 (0.0%)	0 (0.0%)	84.6
DCEP_M (58)	0 (0.0%)	0 (0.0%)	1 (0.0%)	1 (0.0%)	**56 (17.6%)**	0 (0.0%)	0 (0.0%)	0 (0.0%)	96.5
BLHER (66)	1 (1.4%)	0 (0.0%)	5 (0.2%)	0 (0.0%)	0 (0.0%)	**60 (95.2%)**	0 (0.0%)	0 (0.0%)	90.9
WVIR (118)	0 (0.0%)	0 (0.0%)	3 (0.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	**110 (97.3%)**	5 (20.0%)	93.2
RVTAU (20)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	**20 (80.0%)**	100
Precision (%)	89.7	77.1	97.4	97.8	17.6	95.2	97.3	80.0	

Fig. F.2. As in Fig. F.1 but for the LMC.

Literature SOS	ACEP_F (46)	ACEP_1O (23)	DCEP_F (2674)	DCEP_1O (1548)	DCEP_M (210)	BLHER (15)	WVIR (18)	RVTAU (10)	Recall (%)
ACEP_F (87)	**30 (65.2%)**	2 (8.6%)	54 (2.0%)	0 (0.0%)	0 (0.0%)	1 (6.6%)	0 (0.0%)	0 (0.0%)	34.4
ACEP_1O (80)	4 (8.6%)	**18 (78.2%)**	41 (1.5%)	12 (0.7%)	5 (2.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	22.5
DCEP_F (2488)	8 (17.3%)	0 (0.0%)	**2460 (91.9%)**	11 (0.7%)	5 (2.3%)	0 (0.0%)	2 (11.1%)	2 (20.0%)	98.8
DCEP_1O (1800)	1 (2.1%)	3 (13.0%)	100 (3.7%)	**1521 (98.2%)**	110 (52.3%)	0 (0.0%)	1 (5.5%)	0 (0.0%)	84.5
DCEP_M (110)	0 (0.0%)	0 (0.0%)	16 (0.5%)	4 (0.2%)	**90 (42.8%)**	0 (0.0%)	0 (0.0%)	0 (0.0%)	81.8
BLHER (20)	3 (6.5%)	0 (0.0%)	3 (0.1%)	0 (0.0%)	0 (0.0%)	**14 (93.3%)**	0 (0.0%)	0 (0.0%)	70.0
WVIR (18)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	**15 (83.3%)**	3 (30.0%)	83.3
RVTAU (5)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	**5 (50.0%)**	100
Precision (%)	65.2	78.2	91.9	98.2	42.8	93.3	83.3	50.0	

Fig. F.3. As in Fig. F.1 but for the SMC.