A&A 622, A60 (2019)
https://doi.org/10.1051/0004-6361/201833374

Astronomy & Astrophysics

Special issue

Gaia Data Release 2

Gaia Data Release 2

Specific characterisation and validation of all-sky Cepheids and RR Lyrae stars⋆

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

[1] INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Gobetti 93/3, 40129 Bologna, Italy
e-mail: gisella.clementini@inaf.it
[2] INAF – Osservatorio Astronomico di Capodimonte, Salita Moiariello 16, 80131 Napoli, Italy
[3] Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy
[4] Department of Astronomy, University of Geneva, Ch. d'Ecogia 16, 1290 Versoix, Switzerland
[5] SixSq, Rue du Bois-du-Lan 8, 1217 Geneva, Switzerland
[6] GEPI, Observatoire de Paris, Université PSL, CNRS, 5 Place Jules Janssen, 92190 Meudon, France
[7] Institute of Astronomy, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium
[8] Departamento Inteligencia Artificial, UNED, c/ Juan del Rosal 16, 28040 Madrid, Spain
[9] Konkoly Observatory, Research Centre for Astronomy and Earth Sciences, Hungarian Academy of Sciences, 1121 Budapest, Konkoly Thege M. ut 15-17, Hungary
[10] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

Received 6 May 2018 / Accepted 29 October 2018

ABSTRACT

Context. The *Gaia* second Data Release (DR2) presents a first mapping of full-sky RR Lyrae stars and Cepheids observed by the spacecraft during the initial 22 months of science operations.
Aims. The Specific Objects Study (SOS) pipeline, developed to validate and fully characterise Cepheids and RR Lyrae stars (SOS Cep&RRL) observed by *Gaia*, has been presented in the documentation and papers accompanying the *Gaia* first Data Release. Here we describe how the SOS pipeline was modified to allow for processing the *Gaia* multi-band (G, G_{BP}, and G_{RP}) time-series photometry of all-sky candidate variables and produce specific results for confirmed RR Lyrae stars and Cepheids that are published in the DR2 catalogue.
Methods. The SOS Cep&RRL processing uses tools such as the period–amplitude and the period–luminosity relations in the G band. For the analysis of the *Gaia* DR2 candidates we also used tools based on the G_{BP} and G_{RP} photometry, such as the period–Wesenheit relation in (G, G_{RP}).
Results. Multi-band time-series photometry and characterisation by the SOS Cep&RRL pipeline are published in *Gaia* DR2 for 150 359 such variables (9575 classified as Cepheids and 140 784 as RR Lyrae stars) distributed throughout the sky. The sample includes variables in 87 globular clusters and 14 dwarf galaxies (the Magellanic Clouds, 5 classical and 7 ultra-faint dwarfs). To the best of our knowledge, as of 25 April 2018, the variability of 50 570 of these sources (350 Cepheids and 50 220 RR Lyrae stars) has not been reported before in the literature, therefore they are likely new discoveries by *Gaia*. An estimate of the interstellar absorption is published for 54 272 fundamental-mode RR Lyrae stars from a relation based on the G-band amplitude and the pulsation period. Metallicities derived from the Fourier parameters of the light curves are also released for 64 932 RR Lyrae stars and 3738 fundamental-mode classical Cepheids with periods shorter than 6.3 days.

Key words. stars: general – stars: oscillations – stars: variables: Cepheids – stars: variables: RR Lyrae – methods: data analysis – Galaxy: structure

1. Introduction

The *Gaia* mission is repeatedly monitoring the celestial sphere since the start of scientific operation on 25 July 2014. The spacecraft collects multi-band photometry and astrometric measurements of sources crossing its field of view (FoV) down to a limiting magnitude $G \sim 21$ mag, along with low-resolution spectroscopy for sources brighter than $G \sim 16$ mag. A description of the *Gaia* mission (spacecraft, instruments, survey, and measurement principles) as well as the structure and activities of the *Gaia* Data Processing and Analysis Consortium (DPAC) can be found in Gaia Collaboration (2016a).

On 25 April 2018, *Gaia* Data Release 2 (DR2) has published photometry in three passbands (*Gaia* G, G_{BP}, and G_{RP}), five-parameter astrometry and radial velocities collected over the initial 22 months of observations. A summary of the *Gaia* DR2

⋆ Full Table C.1 is 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/qcat?J/A+A/622/A60

contents and survey properties is provided in Gaia Collaboration (2018). The photometric dataset and the processing of the G, G_{BP}, and G_{RP}-band (time series) photometry used for the analysis in this paper are described in Riello et al. (2018) and Evans et al. (2018), and a detailed description of the astrometric dataset and processing is provided in Lindegren et al. (2018).

The multi-epoch operating procedure makes *Gaia* a very powerful tool to identify and characterise stellar variability phenomena across the whole HR diagram (e.g. Gaia Collaboration 2019). In *Gaia* Data Relase 1 (DR1; Gaia Collaboration 2016b), G-band time-series photometry and characteristic parameters were released for only a small number of pulsating variables, comprising 599 Cepheids and 2595 RR Lyrae stars (Clementini et al. 2016, hereafter Paper 1) in a region of the Large Magellanic Cloud (LMC) that the spacecraft observed at high cadence during the first 28 days of scientific operation in ecliptic poles scanning law. The catalogue of variables released in *Gaia* DR2 (Holl et al. 2018) contains thousands of Cepheids and hundred thousands RR Lyrae stars in the Milky Way (MW) and its nearest neighbours. They represent a first census of full-sky RR Lyrae stars and Cepheids and provide an indication of *Gaia* capabilities by recovering most of the known MW Cepheids, identifying a few bona fide new ones, and increasing the number of known Galactic RR Lyrae stars well beyond the current value of more than one hundred thousand.

The general approach of variability analysis and classification developed within the *Gaia* DPAC was presented in Eyer et al. (2017b). For DR2, an additional fully statistical approach was developed to classify all-sky high-amplitude pulsating stars. This approach is extensively described in Rimoldini et al. (2018). The Specific Objects Study (SOS) pipeline (hereafter referred to as SOS Cep&RRL pipeline), which is specifically designed to validate and fully characterise Cepheids and RR Lyrae stars observed by *Gaia*, is described in detail in Paper 1. The general properties of the whole sample of variable sources released in *Gaia* DR2 are described in Holl et al. (2018), which also briefly summarises steps of the general variability analysis prior the SOS Cep&RRL processing.

In this paper we describe how the SOS Cep&RRL pipeline was modified and further developed to process the DR2 multi-band time-series photometry of candidate Cepheids and RR Lyrae stars identified by the general variable star classification pipelines (Eyer et al. 2017b; Rimoldini et al. 2018). We describe our validation procedures and briefly present results from the SOS Cep&RRL processing of Cepheids and RR Lyrae stars that were confirmed by the pipeline, which are released in the *Gaia* DR2 variability catalogue. We recall that according to the strict DPAC policy, only a very limited interpretative overview and no scientific exploitation of the data is presented in this paper. We also recall that the characterisation and classification of any RR Lyrae stars and Cepheids released in DR2 is purely and exclusively based on the *Gaia* data of the sources. That is to say, we do not complement the *Gaia* time series with external non-*Gaia* data. Literature data are only used as training sets for the classification tools and for the final validation of the results.

The paper is organised as follows. Section 2 provides a brief summary of the SOS Cep&RRL pipeline that specifically highlights changes and improvements with respect to the DR1 processing. Section 3 presents the datasets and selections of all-sky sources on which the SOS Cep&RRL pipeline was run. Sections 4 and 5 present the SOS Cep&RRL analysis and extensively describe the procedures we used to validate the SOS results that we publish in *Gaia* DR2 along with a comparison with the literature and the DR1 results. In these sections we specifically acknowledge the limitation of the current analysis and results and caution about oversimplifications and possible biases of the SOS Cep&RRL processing. We also discuss why well-known sources are missing in DR2 as well as a few misclassifications of the SOS Cep&RRL processing for DR1 (Sect. 5.2). After DR2, we became aware of some misclassifications in the released DR2 Cepheid and RR Lyrae samples. A partial listing of these misclassifications is provided in Appendix C. Finally, the main results and future developments of the SOS Cep&RRL pipeline are summarised in Sect. 6.

2. SOS Cep&RRL pipeline applied to the DR2 data: general overview

The main purposes, tools, and steps of the SOS Cep&RRL pipeline are extensively described in Sect. 2 of Paper 1, to which we refer for details. Here we only summarise the main differences with respect to the DR1 processing. An overview of the different parts of the SOS Cep&RRL processing is presented in Figs. 1–3, which were updated from Figs. 1–3 of Paper 1 to summarise changes made in the pipeline, such as parts that were activated or tools that were newly developed to allow for processing the multi-band (G, G_{BP} and G_{RP}) time-series photometry of all-sky sources that are released in DR2.

As schematically summarised in Fig. 2 of Holl et al. (2018), the SOS Cep&RRL pipeline received as input for the DR2 processing the calibrated G and integrated G_{BP}, G_{RP} time-series photometry (Evans et al. 2018; Riello et al. 2018) collected by *Gaia* in the 22 months between 25 July 2014 and 23 May 2016 of the sources that were pre-classified as candidate Cepheids and RR Lyrae stars by the classifiers of the general variability pipeline (Eyer et al. 2017b, Rimoldini et al. 2018). Specifically, SOS Cep&RRL processed candidate Cepheids and RR Lyrae stars with ≥20 (hereafter geq20) G-band measurements (or FoV transits) identified by the supervised classification of the general variability pipeline (Eyer et al. 2017b) and candidates with ≥12 measurements (hereafter geq12) identified by the semi-supervised classification described in Rimoldini et al. (2018) for a total of about 640 000 candidate RR Lyrae stars and more than 72 000 candidate Cepheids, after removing overlaps between geq20 and geq12 samples. A preliminary version of the DR2 astrometry was also used in order to test our classification of the All-Sky Cepheids (Sect. 3.2). However, due to the tight DR2 data-processing schedule, we were unable to use the final astrometric solution for the processing (Lindegren et al. 2018), nor could we access the astrophysical parameters (e.g. reddening values and temperatures, Andrae et al. 2018) and radial velocity measurements (Sartoretti et al. 2018) that are published in DR2.

Main outliers and measurements of insufficient quality are removed from the G, G_{BP}, and G_{RP} time series by specific operators of the general variability pipeline (see Sect. 4.1 in Holl et al. 2018). However, we additionally cleaned the G_{BP} and G_{RP} time series by setting more stringent limits that were specifically tailored for Cepheids and RR Lyrae stars to the cleaning operators before the source time series were entered in the main trunk of the SOS pipeline that is common to Cepheids and RR Lyrae stars (Fig. 1).

Determining the main periodicity of the source is the initial step in the SOS Cep&RRL pipeline (see Fig. 1), and the subsequent classification and characterisation steps significantly rely on the accuracy and precision of the SOS period, which

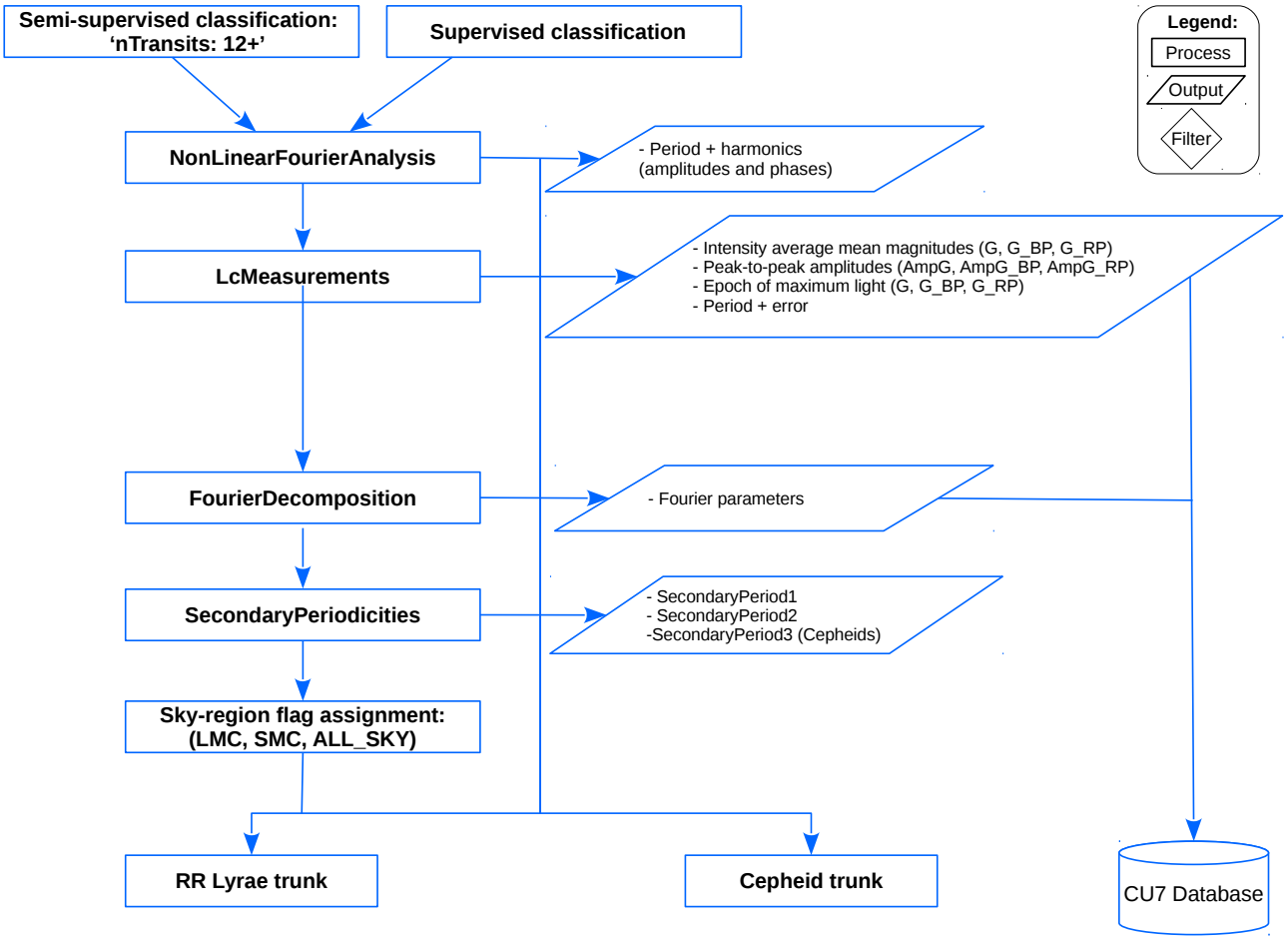

Fig. 1. Overview of the SOS Cep&RRL processing that is common to Cepheids and RR Lyrae stars. Rectangles show different processing modules of the general trunk, with names highlighted in boldface. Their outputs are indicated within parallelograms. The figure is updated from Fig. 1 of Paper 1 to include the processing of the G_{BP} and G_{RP} time series and the search for a third periodicity for Cepheids. A further addition for DR2 is the selection of the region of the sky where the sources are located, whether LMC, SMC, or All-Sky (see text for details).

ultimately depends upon the number of epoch-data that is available for the sources. A comparison between periods derived by the SOS pipeline and periods published in the literature for a sample of 37 941 RR Lyrae stars in common between *Gaia* and the OGLE catalogues for the Large Magellanic Cloud (LMC), Small Magellanic Cloud (SMC), and Galactic bulge shows that agreement can already be satisfactory if at least 20 epochs are available, and the results definitely improve for sources with 30 or more epoch data. Unless stated otherwise, all periods in this paper have been derived with the SOS Cep&RRL pipeline (P_{SOS}).

In the SOS Cep&RRL main trunk, the main periodicity of the source is determined with the Lomb–Scargle algorithm (Lomb 1976, Scargle 1982, see Sect. 2.1 in Paper 1 for details) that searches for a periodicity in the range $0.2 \leq P < 1$ day (frequency between 1 and 5) for RR Lyrae stars and $0.2 \leq P < 333$ days (frequency between 0.003 and 5) for Cepheids[1].

The light curves are modelled with truncated Fourier series and the pulsation characteristics (period, peak-to-peak amplitude, epoch of maximum light, and intensity-averaged mean magnitude in each of the three passbands) are determined from the modelled light curves along with the related errors. The Fourier parameters (ϕ_{21}, ϕ_{31}, R_{21}, and R_{31}) of the modelled G-band light curve are also determined. Secondary periodicities, if any, are identified in the G-band data by searching for one additional period in the case of RR Lyrae stars and up to two periods for Cepheids (see Fig. 1).

A significant change with respect to DR1 that was introduced in the main trunk of the SOS Cep&RRL pipeline to process all-sky sources was that before the RR Lyrae (Fig. 2) and Cepheid (Fig. 3) branches were entered, the region of the sky was selected where the sources are located, whether LMC, SMC, or All-Sky. The sky region corresponding to the LMC is defined as a box with a centre at RA = 82.5°, Dec = −68.25° that extends from 67.5° to 97.5° in RA and from −73.0° to −63.5° in Dec. The SMC is defined as a box region with a centre at RA = 16°, Dec = −73° that extends from 2° to 30.0° in RA and from −75.0° to −71.0° in Dec. The All-Sky region is defined by what is left after subtracting the LMC and SMC selections. Different reference relations are used to classify the Cepheids and RR Lyrae stars in these three distinct regions, as detailed in Sects. 3.1 and 3.2. After the region of the sky is selected, the source enters the RR Lyrae (Fig. 2) or the Cepheid (Fig. 3) branches according to the following schema. If the source belongs to the All-Sky region, it is first ingested into the RR Lyrae branch (Fig. 2), and if it is not classified as an RR Lyrae star, it is then sent for analysis to the Cepheid branch (Fig. 3), where the star is processed if the parallax value conforms to specific quality

[1] The period range for Cepheids was later reduced to $0.2 \leq P \lesssim 160$ days during validation to account for the actual time interval spanned by the DR2 time-series data.

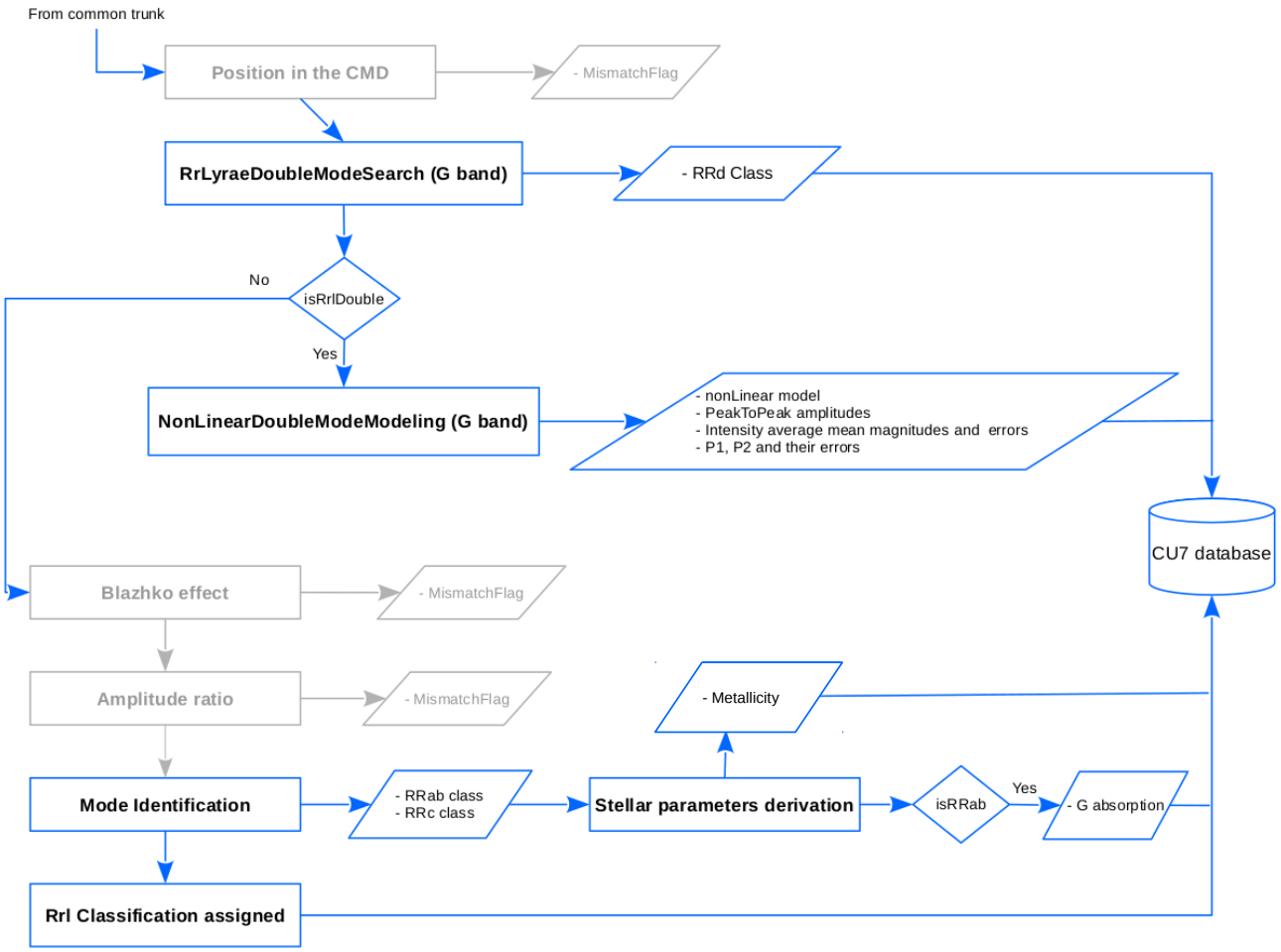

Fig. 2. Flow chart of the RR Lyrae branch in the SOS Cep&RRL pipeline. As in Fig. 1, rectangles show the different processing modules (with names highlighted in boldface) of this branch. Their outputs are indicated within parallelograms. Rhombi indicate filters leading to different processing options. We mark in light grey modules that were not operational for the *Gaia* DR2 processing. The figure is updated from Fig. 2 of Paper 1 to include modifications that were implemented to process the DR2 photometry.

assurance conditions defined in the astrometric processing (see Sect. 3.2 and Lindegren et al. 2018, for details). If this is not the case, the source is rejected. If the source belongs to the LMC or SMC regions, the branch that is entered first is chosen based on a number of checks on the intensity-averaged mean G magnitude[2], the period, and the amplitude of the G-band variation.

In the RR Lyrae branch (Fig. 2), the SOS Cep&RRL pipeline can use specific features and diagnostic tools such as (a) the (apparent and/or absolute) colour–magnitude diagram (CMD; G vs. $G_{\rm BP} - G_{\rm RP}$, G vs. $G - G_{\rm BP}$, and G vs. $G - G_{\rm RP}$), (b) the parameters of the G-band light curve Fourier decomposition (R_{21}, and ϕ_{21} vs. P and R_{31}, and ϕ_{31} vs. P plots), (c) the *Gaia* period–amplitude (PA) diagram, also known as Bailey diagram (Bailey 1902) in the G band, and (d) the amplitude ratios in the different passbands (specifically, Amp($G_{\rm BP}$)/Amp($G_{\rm RP}$)) to classify the sources and infer their pulsation mode(s) and type(s)[3]. We refer to Sects. 2.3.2, 2.3.3, 2.4.1, and 2.4.2 in Paper 1 for a general description of the procedures and provide in the following sections here new relations that replace or complement those used for DR1. We note that of the various tools described above, tool (c) is most frequently applicable as it only requires knowledge of the source period and G-band amplitude, but it is also the least constraining tool. In contrarst, tools (a) and (d) are increasingly less applicable because they require availability of well-modelled $G_{\rm BP}$, $G_{\rm RP}$ light curves from which to derive mean colours to build the CMDs needed for tool (a) and derive reliable amplitudes to compute the amplitude ratios needed for tool (d) and increasingly sampled light curves modelled with at least two or three harmonics to correctly infer the Fourier parameters used by tool (b). This also shapes the limitations of our analysis, which we discuss in the following sections. Furthermore, for the SOS processing of RR Lyrae stars we did not use the parallax information because we were able to access only a preliminary version of the DR2 parallax values and also because RR Lyrae stars, which are at least 3 magnitudes fainter than Cepheids, are more likely to have parallaxes with large errors that may even scatter the parallaxes to negative values if they are very close to zero, as in the case of distant systems such as the Magellanic Clouds, especially in this still initial astrometric solution.

To classify the sources processed in the Cepheid branch (Fig. 3) and identify their pulsation mode and the multi-mode pulsators, we used the Fourier parameters and the

[2] For the sake of clarity, we recall that the intensity-averaged magnitude is the mean magnitude of a variable star obtained by transforming into intensity each individual value of the Fourier model that best fits the light curve and then averaging all these intensities and converting the mean intensity back into a magnitude.

[3] All quantities (mean magnitudes, colours, amplitudes, and Fourier parameters) used in these tools are inferred from the non-linear model of the light curves computed by the SOS Cep&RRL pipeline.

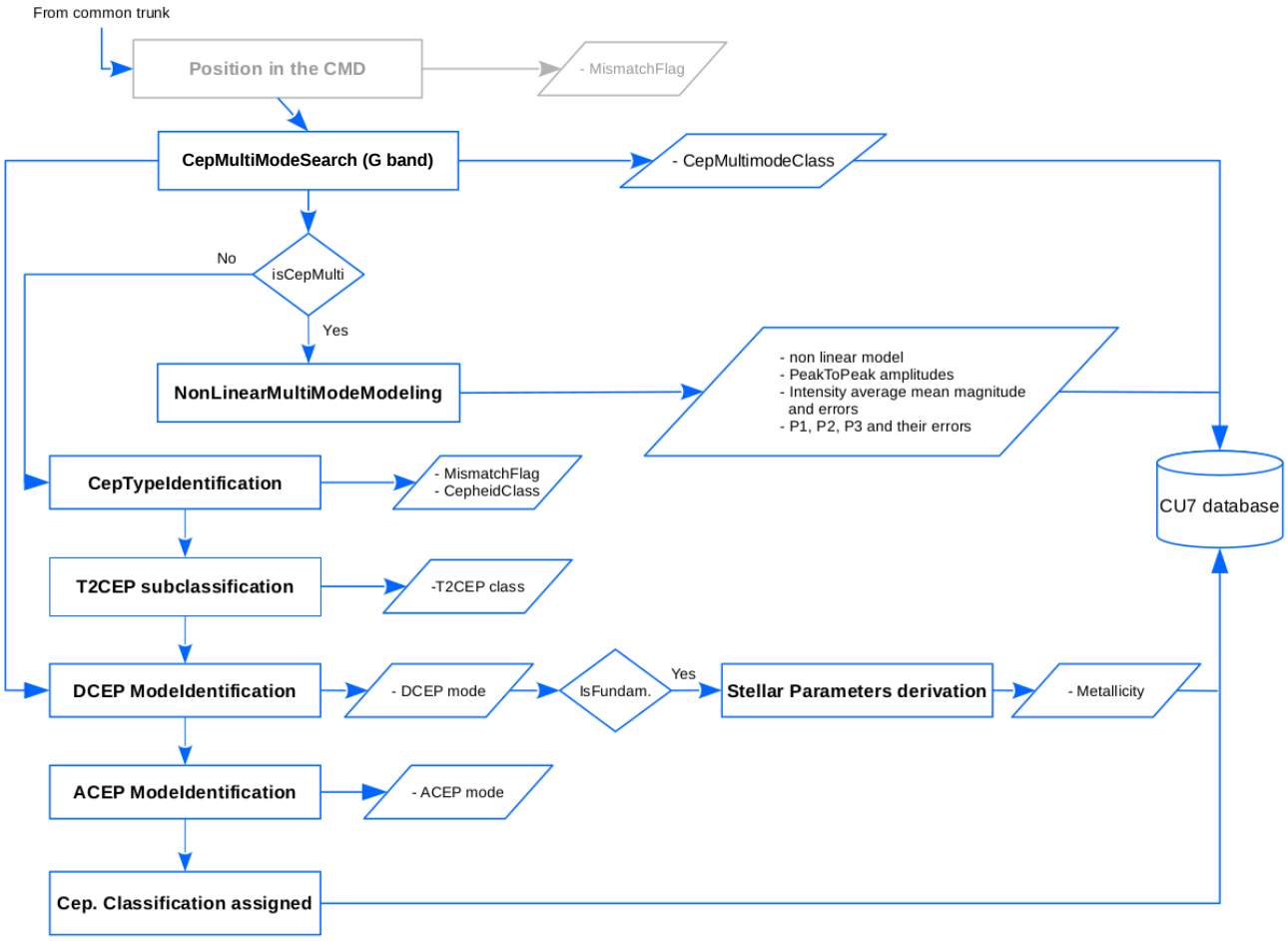

Fig. 3. Flow chart of the Cepheid branch in the SOS Cep&RRL pipeline as properly modified from Fig. 3 in Paper 1 to process the DR2 photometry. Layout and colour-coding are the same as in Fig. 2.

period–luminosity (PL) and period–Wesenheit (PW) relations. For the LMC and SMC, the zero-points of the PL and PW relations (in the *Gaia* passbands) were calibrated adopting 18.5 and 19.0 mag for the distance modulus of the LMC and SMC, respectively. To classify the MW All-Sky Cepheids, we primarily relied on the Fourier parameter planes, and as a test only, we also considered the PW distributions, for which we had to rely on parallaxes (see Sect. 3.2).

In Figs. 2 and 3 we indicate in light grey several modules that were not fully operational during the SOS DR2 processing. This includes the module that checks the position of the sources on the CMD and the check of amplitude ratios to reject binaries that contaminate the samples. Although not yet automatically activated during the pipeline processing, these two tools were later extensively used during the validation of the sources that were confirmed as Cepheids and RR Lyrae stars by the SOS Cep&RRL pipeline (see Sect. 4).

The first working step in the RR Lyrae (Fig. 2) and in the Cepheid (Fig. 3) branches is to verify whether any detected secondary periodicity is consistent with the RR Lyrae star being a double-mode pulsator (RRd; Fig. 2) and the Cepheid being a multi-mode pulsator (Fig. 3). If this is confirmed, a non-linear multi-mode modelling of the light curve follows that takes all excited periodicities into account. We refer to Sects. 2.3.1 and 2.4.1 of Paper 1 for a general and more detailed presentation of the algorithms used in the SOS Cep&RRL pipeline to identify and characterise double- and multi-mode sources.

The subsequent step in the Cepheid branch (Fig. 3) is the identification of the Cepheid type: DCEP, ACEP, or T2CEP (see Table B.1 for the meaning of the acronyms), and pulsation mode (only for DCEP, ACEP) for which the path was as described in Sect. 2.4.2 of Paper 1, with the additional use of PW relations to distinguish the different types and differentiation of the PL and PW relations according to the location of the sources on the sky. The reference PL and PW relations used in DR2 are described in detail in Sect. 3.2. A further minor change with respect to DR1 was to use a stricter limit in period of 0.234 d$< P <$ 6 days when $R_{21} < 0.214$ to identify DCEPs that pulsate in the first overtone (1O), using the R_{21} versus P diagram. Since the SOS Cep&RRL modules for detection of the Blazhko effect and calculation of the amplitude ratios (see Sect.2.3.2 in Paper 1) were not activated for DR2, the next active step in the RR Lyrae branch (Fig. 2) is the RR Lyrae mode identification, which follows the path described in Sect. 2.3.3 of Paper 1 with only the following change: a new line to separate RRc and RRab types was adopted. This line is described by the equation $\mathrm{Amp}(G) = -3.5 \times P + 2.08$.

2.1. Deriving the stellar parameters

A major addition in the SOS Cep&RRL processing of DR2 sources was that a module for the derivation of stellar parameters for the confirmed Cepheids and RR Lyrae stars was activated. This occurs in the *StellarParametersDerivation* module of the Cepheid and RR Lyrae branches, where stellar intrinsic

Table 1. Metallicity and absorption in the G band for a sample of GCs and dSphs, obtained by averaging individual values derived for the RR Lyrae stars by the SOS Cep&RRL pipeline.

Name	$[Fe/H]_{C09}$ dex	$[Fe/H]_{Gaia}$ dex	St. dev.($[Fe/H]_{Gaia}$) dex	N_1^a	$A(G)$ mag	St. dev.$[A(G)]$ mag	$A(V)$ mag	N_2^a	Ref.
NGC 1261	−1.27	−1.33	±0.55	14	1.07	±1.33	0.03	10	(1)
NGC 1851	−1.18	−1.17	±0.48	9	0.46	±0.34	0.06	8	(2)
NGC 288	−1.32	−1.33	±0.22	1	0.17		0.09	1	(3)
NGC 3201	−1.59	−1.30	±0.04	82	0.67	±0.21	0.74	73	(4)
NGC 5024	−2.10	−1.62	±0.48	8	0.73	±0.75	0.06	20	(5)
NGC 5139 - ω Cen	−1.53	−1.17	±0.79	8	0.32	±0.14	0.37	9	(6)
NGC 5272 (M3)	−1.50	−1.32	±0.43	111	0.49	±0.87	0.03	92	(7)
NGC 6266 (M62)	−1.18	−0.71	±0.39	81	2.18	±0.78	1.46	52	(8)
IC 4499	−1.53	−1.47	±0.53	56	0.81	±0.35	0.71	54	(9)
NGC 7078 (M15)	−2.37	−1.26	±0.87	5	0.22		0.31	1	(10)
Sculptor	−1.68	−1.52	±0.53	102	0.68	±0.43	0.06	105	(11)
Draco	−1.93	−1.70	±0.74	30	0.54	±0.32	0.09	13	(12)
Ursa Major I	−2.18	−1.92	±0.18	2	0.53	±0.05	0.02	2	(13)
Ursa Major II	−2.47	−2.28	±0.23	1	0.33		0.30	1	(14)

Notes. $^{(a)}N_1$ and N_2 are the number of RR Lyrae stars on which the mean [Fe/H] and $A(G)$ were computed, respectively.
References. (1) Salinas et al. (2016); (2) Walker (1998); (3) Kaluzny et al. (1997), Arellano Ferro et al. (2013); (4) Layden & Sarajedini (2003), Arellano Ferro et al. (2014); (5) Cuffey (1966), Goranskij (1976), Arellano Ferro et al. (2011); (6) Braga et al. (2016); (7) Benkő et al. (2006); (8) Contreras et al. (2010); (9) Walker & Nemec (1996); (10) Corwin et al. (2008); (11) Kaluzny et al. (1995), Clementini et al. (2005); (12) Kinemuchi et al. (2008); (13) Garofalo et al. (2013); (14) Dall'Ora et al. (2012).

parameters are derived through a variety of methods appropriate for RR Lyrae stars or Cepheids. For DR2, we specifically implemented the estimate of the photometric metal abundance ([Fe/H]) from the ϕ_{31} parameter of the Fourier light-curve decomposition of fundamental-mode (RRab) and first-overtone (RRc) RR Lyrae stars, and from the R_{21} and R_{31} parameters for fundamental-mode classical Cepheids with periods shorter than 6.3 days. For RRab stars we also activated a tool to estimate the absorption in the G band, $A(G)$, from the light and colour curves.

2.1.1. Metallicity

Jurcsik & Kovács (1996) were the first to devise a method for inferring a photometric metal abundance ([Fe/H]) from the ϕ_{31} parameter of the visual light-curve Fourier decomposition of fundamental-mode RR Lyrae stars. The method was later extended to RRc stars as well by Morgan et al. (2007). Nemec et al. (2013) revised and recalibrated the ϕ_{31}–[Fe/H] relations based on very accurate light curves of RRab and RRc stars observed by the *Kepler* spacecraft, along with metallicities derived from an abundance analysis of high-resolution spectroscopy.

In the *StellarParametersDerivation* module of the RR Lyrae branch, photometric metal abundance of RRab and RRc stars observed by *Gaia* are derived from the ϕ_{31} parameter of the Fourier G-band light curve decomposition using the relations derived by Nemec et al. (2013) separately for RRc and RRab stars. Specifically, Nemec et al. (2013) estimated the values of [Fe/H] using the ϕ_{31} parameters that were calculated by fitting the observed time series in the *Kepler* photometric system through a Fourier series of sine functions for RRab stars and cosine functions for RRc sources. In order to use Eqs. (2) and (4) of Nemec et al. (2013), which are valid for RRab and RRc stars, respectively, we first transformed the ϕ_{31} parameters from the *Gaia* G band into the *Kepler* photometric system with the following steps: (i) the ϕ_{31} parameter in the G band was first transformed into the V band using the relation $\phi_{31}(V) = \phi_{31}(G) - 0.104$; and (ii) the ϕ_{31} parameter in the *Kepler* system was then obtained using the following relations: $\phi_{31}^s = \phi_{31}(V) + \pi + 0.151$ and $\phi_{31}^c = \phi_{31}(V) + 0.151$ (Nemec et al. 2011) for RRab and RRc stars, respectively, where the superscript s stands for the sine function, while c indicates the use of the cosine function. The uncertainties of the estimated [Fe/H] values were derived via Monte Carlo simulations: (i) the G-band ϕ_{31} parameter was varied using 100 random shifts extracted from a normal distribution with a standard deviation equal to the error on the Fourier parameter itself; and (ii) the metallicity was recalculated for each simulated ϕ_{31} value and the standard deviation, $\sigma_{\rm sim}$, of these 100 values was estimated. The final uncertainty of the derived [Fe/H] values was obtained as the sum in quadrature of the $\sigma_{\rm sim}$ error (derived above) and of a conservative systematic error assumed to be of 0.2 dex to account for systematics in the various calibrations and passband transformations[4]. We also recall that according to Cacciari et al. (2005), photometric metallicities inferred with this method are better suited to describe the average metal abundance of a population of RR Lyrae stars than individual metallicities. As a check, we list in Table 1 the mean photometric metallicity derived by the SOS Cep&RRL pipeline for RR Lyrae stars in a number of Galactic globular clusters (GCs) and dwarf spheroidal galaxies (Sphs) observed by *Gaia* and released in DR2 (see also Figs. 21, 22, and 24).

The metallicity of fundamental-mode classical Cepheids was calculated using the relations derived by Klagyivik et al. (2013). These authors calculated equations to estimate the metallicity ([Fe/H]) for classical Cepheids with period $\log P < 0.8$ ($P = 6.3$ days) using the Fourier parameters R_{21} and/or R_{31}. To use the Klagyivik et al. (2013) equations, we first transformed the G-band R_{21} and R_{31} parameters into the corresponding

[4] Gratton et al. (2004) and Di Fabrizio et al. (2005) found average differences between photometric (from the ϕ_{31} parameter) and spectroscopic metallicities of about 0.30 ± 0.07 dex from a sample of RR Lyrae stars in the LMC.

V-band values through the equations $R_{21}(V) = 0.985 \times R_{21}(G)$ and $R_{31}(V) = 0.982 \times R_{31}(G) + 0.0098$, which were obtained by inverting Eqs. (9) and (10) in Paper 1. The errors in metallicity were estimated through a Monte Carlo simulation applied to the R_{21} and R_{31} Fourier parameters, and, as for the RR Lyrae stars, adding in quadrature a systematic error of 0.2 dex.

The reliability of individual metallicities inferred with these methods also significantly depends on the reliability and accuracy of the G-band Fourier parameters (ϕ_{31} for RR Lyrae stars and R_{21} and R_{31} for Cepheids). We thus advise users to check errors of the Fourier parameters and inspect the light curves before blindly trusting the published photometric metallicities, which are inferred by automatically processing all sources for which these Fourier parameters are available through the *StellarParametersDerivation* module of the SOS Cep&RRL pipeline.

2.1.2. Absorption in the G band

The absorption in the G band, $A(G)$, of fundamental-mode RR Lyrae stars was determined from the following empirical relation derived by Piersimoni et al. (2002):

$$(V-I)_0 = (0.65 \pm 0.02) - (0.07 \pm 0.01) \times [\mathrm{Amp}(V)] + (0.36 \pm 0.06) \times \log(P) \quad (1)$$
$$(\sigma = 0.02).$$

However, to use this relation, we first had to transform amplitudes and colours from the Johnson to the *Gaia* passbands. To this purpose, we used the following transformation equation for the amplitude in the V band (inverting Eq. (3) of Paper 1):

$$\mathrm{Amp}(V) = (1.081 \pm 0.003) \times \mathrm{Amp}(G) + (0.013 \pm 0.003) \quad (2)$$
$$(\sigma = 0.012).$$

Then, we calculated the $(V-I)$ colours from those of $(G - G_{\rm RP})$,

$$(V-I) = (0.027 \pm 0.003) + (1.13 \pm 0.02) \times (G - G_{\rm RP}) + (0.55 \pm 0.03) \times (G - G_{\rm RP})^2 \quad (3)$$
$$(\sigma = 0.013),$$

and the conversion between the G absorption, $A(G)$, and the reddening $E(V-I) = (V-I) - (V-I)_0$

$$A(G) = [(2.3116 \pm 0.006) - (0.3097 \pm 0.0011) \times (V-I)_0] \times E(V-I) \quad (4)$$
$$(\sigma = 0.013),$$

which are both based on Jordi et al. (2010) passband transformations[5].

In practice, for each ab-type RR Lyrae, the observed amplitude in the G band, $\mathrm{Amp}(G)$, and the star period give the expected intrinsic value of $(V-I)_0$ through Eqs. (1) and (2). The "observed" Johnson $(V-I)$ colour is calculated from the observed *Gaia* $(G - G_{\rm RP})$ colour by means of Eq. (3). Finally, the calculated $(V-I)_0$ and $(V-I)$ colours are inserted into Eq. (4) to obtain the source $A(G)$ absorption.

[5] The Jordi et al. (2010) transformations are superseded by new transformations published in Evans et al. (2018). These new transformations became available when the whole variability processing for DR2 had already been completed. We are currently updating the SOS Cep&RRL pipeline to the new transformations, in preparation for the next *Gaia* releases.

3. Application of the SOS Cep&RRL pipeline to the *Gaia* DR2 dataset: source selections and processing simplifications

The dataset processed by the SOS Cep&RRL pipeline to produce the results that are published in *Gaia* DR2 consisted of G, $G_{\rm BP}$, and $G_{\rm RP}$ time-series photometry[6] of candidate Cepheids and RR Lyrae stars, observed by *Gaia* in the 22 months between July 2014 and May 2016. The time-series data provided in units of flux by the photometry pipeline were converted into magnitudes by the variability processing prior SOS, using the magnitude zero-points defined in Evans et al. (2018). For DR2 the minimum number of data points in the G-band time series of sources to be fed into the SOS Cep&RRL pipeline was reduced to 12, as this limit was deemed to be sufficient for a reliable estimate of the period and other (pulsation) characteristics of confirmed Cepheids and RR Lyrae stars, based upon experience of the DR1 analysis and results (see Paper 1).

Sources with more than 12 epoch-data in the G band, pre-classified as candidate Cepheids and RR Lyrae stars by the classifiers of the general variability pipeline following the two separate paths described in Sect. 2 (Eyer et al. 2017b, Rimoldini et al. 2018), were ingested into the SOS Cep&RRL pipeline (see Fig. 2 in Holl et al. 2018) for a total of 639 828 univocally defined candidate RR Lyrae stars and 72 717 candidate Cepheids. These rather large numbers of candidates included even low probability levels and also candidates flagged as class outliers in order to maintain a high level of completeness and avoid losing potentially valid candidates. However, as also done in DR1 (see discussion in Sect. 3.1 and Fig. 15 of Paper 1), we dropped sources with $\mathrm{Amp}(G) \leq 0.1$ mag. Further cuts were applied during validation to conform to the quality assurance limits set by *Gaia* photometric and astrometric processing teams. That is, we dropped candidate RR Lyrae stars and Cepheids with an excess flux above the limit recommended in the photometry processing (Evans et al. 2018), and we also rejected All-Sky candidate Cepheids with excess noise above the limits recommended in the astrometry processing (Lindegren et al. 2018).

In order to facilitate an early release, the very tight schedule of the DR2 data-processing did not allow a full iteration between the pipelines that process the different *Gaia* data. In particular, since the photometry and variability pipelines run in parallel, the SOS Cep&RRL pipeline could be tested only on a preliminary version of the multi-band time-series photometry of the candidate Cepheids and RR Lyrae stars, and it was not possible to update the SOS Cep&RRL pipeline and compute its diagnostic relations (see, e.g. Sects. 2.1.2 and 3.2) using the *Gaia* G, $G_{\rm BP}$ and $G_{\rm RP}$ photometry of the sources directly.

In addition, the SOS Cep&RRL pipeline was only able to access a preliminary version of the astrometric solution, which differs from the final astrometry released in DR2. We did not implement the use of the PW relations to process the RR Lyrae candidates because only limited time was available, because the parallaxes (and related errors) that were available for the SOS Cep&RRL initial processing were preliminary, and because the vast majority of RR Lyrae stars that are released with *Gaia* DR2

[6] Each point in the G-band time series is the mean of the nine measurements taken in the Astrometric Field (AF) CCDs and collected during one observation of a source by *Gaia*, while each $G_{\rm BP}$, $G_{\rm RP}$ measurement is integrated over the low-resolution spectra collected in the blue and red photometer (BP and RP) CCDs.

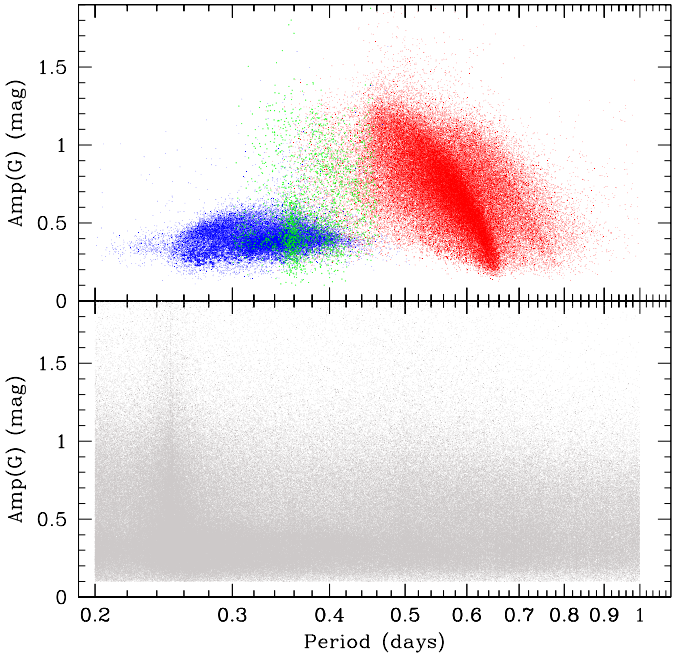

Fig. 4. *Upper panel*: G-band PA diagram of the 140 784 RR Lyrae stars confirmed by the SOS Cep&RRL pipeline that are released in DR2 (in blue the first-overtone pulsators: RRc. In green the double-mode pulsators: RRd. In red the fundamental-mode pulsators: RRab). *Lower panel*: PA diagram of RR Lyrae candidates that were rejected from the SOS pipeline: 499 044 sources in total (see text for details).

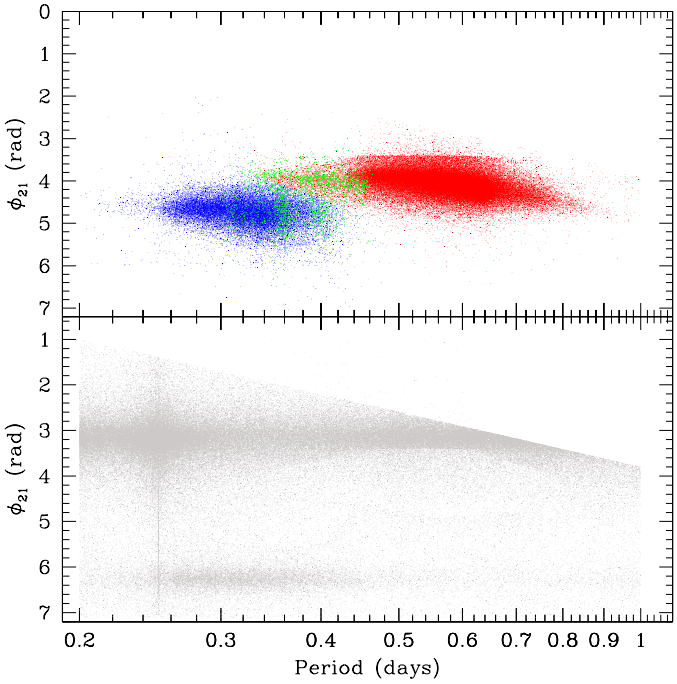

Fig. 5. *Upper panel*: G-band ϕ_{21} vs. period diagram for RR Lyrae stars confirmed by the SOS Cep&RRL pipeline. The colour-coding is the same as in Fig. 4. *Lower panel*: G-band ϕ_{21} vs. period diagram of RR Lyrae candidates that were rejected from the SOS pipeline. The significant peak around $P = 0.25$ d (also shown in the *lower panel* of Fig. 4) is an alias that is due to the rotation period of *Gaia* around its axis.

are typically in the Galactic halo and the Magellanic Clouds[7], where reddening is a minor issue. For similar reasons and because the Magellanic Cloud Cepheids have small true parallaxes that are close to zero and have still rather large uncertainties in DR2, we did not use them. To classify the SMC and LMC Cepheids, we preferred to use the PL and PW relations in apparent magnitude taken from the OGLE studies, transformed to the *Gaia* G, $G_{\rm RP}$ passbands and with zero-points set by adopting literature values for the LMC and SMC distance moduli. We did not have access either to any astrophysical parameters (reddening values in particular) or to radial velocity measurements of the candidates.

In summary, we used the same pipeline as for DR1 (see Paper 1) to process the DR2 sources. The basic relations in the pipeline were obtained by transforming to the *Gaia* passbands through Jordi et al. (2010) passband transformations the quantities and relations that were originally defined in the Johnson-Cousins passbands, according to the procedures that are extensively described in Sect. 2.2., Eqs. (2)–(19), Sect. 2.4.2, Eqs. (20)–(24), and Appendix A of Paper 1.

3.1. SOS Cep&RRL processing of all-sky RR Lyrae stars released in DR2

Although partially reduced by cutting in Amp(G), excess flux, and excess noise, the number of sources to be processed through the SOS Cep&RRL pipeline still consisted of several hundred thousand. This along with the very tight schedule of *Gaia* DR2 made it mandatory to run the SOS pipeline in a fully automatic way, which limited the quality controls and verification of derived products to a detailed analysis of only small randomly selected samples of specific subtypes, whose folded light curves were visually inspected during the validation of the results (see Sect. 4). The following diagnostic tools were specifically used and their results combined to extract the bona fide RR Lyrae stars that are published in DR2 from the candidates : the PA, ϕ_{21} versus P, R_{21} versus P, ϕ_{31} versus P, and R_{31} versus P diagrams. All these tools rely on parameters that are derived only from the G-band light curves, where the time-series data were folded according to period and epoch of maximum light determined by the SOS Cep&RRL pipeline and modelled by the non-linear fitting algorithm of module *NonLinearFourierAnalysis* (see Fig. 1). Tools and diagnostics of the SOS pipeline were run in fully automatic way, producing validation plots from which the bona fide RR Lyrae stars and Cepheids were selected. The validation plots of the PA and ϕ_{21} versus P diagrams are shown in Figs. 4 and 5. The PA tool is the most applicable, but it is also the least constraining and most contaminated tool (see the lower panel of Fig. 4), as it requires knowledge of only the source period (P) and amplitude in the G band [Amp(G)]. These quantities were available for all sources as a result of the SOS Cep&RRL processing. The upper panel of Fig. 4 shows the G-band PA diagram of the 140 784 RR Lyrae stars that were confirmed by the SOS Cep&RRL pipeline that are released in DR2. The lower panel of the figure shows RR Lyrae candidates that were rejected from the SOS pipeline: 499 044 sources in total. The plot shows the high contamination of the RR Lyrae candidate sample that in large part is due to binaries and other types of variable sources. However, we note that a significant fraction of the sources in the lower PA diagram may be bona fide RR Lyrae stars that are not confirmed by the SOS pipeline for a number of different reasons, among them the lack of a good period determination. During the validation process (Sect. 4), the

[7] The Galactic bulge region, where reddening is very high, is only partially covered by the *Gaia* DR2 observations and still with a limited number of epochs.

cross-match with RR Lyrae catalogues in the literature revealed that about 27 000–30 000 sources in the lower panel of Fig. 4 are known RR Lyrae stars that the SOS pipeline rejected due to an incorrect period determination. We note that although not listed in the variability tables, mean G, $G_{\rm BP}$, $G_{\rm RP}$ magnitudes are released in DR2 for all these rejected candidate RR Lyrae stars.

Only for ~220 000 of the candidate RR Lyrae was it possible to model the light curve with at least two harmonics, hence measure the ϕ_{21}, R_{21} Fourier parameters, and construct the ϕ_{21} versus P diagram shown in Fig. 5[8]. The light curves of all remaining candidates, having a sinusoidal shape, were modelled with only one harmonic and were classified only by means of the PA diagram. The upper panel of Fig. 5 shows the ϕ_{21} versus P diagram for RR Lyrae stars that are confirmed by the SOS Cep&RRL pipeline processing and validation, and the lower panel shows RR Lyrae candidates that were rejected. The sharp diagonal cut in the figure is due to the limits set in the SOS Cep&RRL pipeline to separate RR Lyrae stars from Cepheids in the ϕ_{21} versus P plane (see Fig. 21 in Paper 1 for details). The significant peak around $P = 0.25$ d is an alias that is due to the rotation period of *Gaia* around its axis. The two stripes running parallel to the abscissa axis are populated by binaries (that at ϕ_{21} ~6.28 rad) and mainly by other types of variables sources (the one at ϕ_{21} ~3.14 rad) whose removal caused a quite sharp cut to appear in the RRab star distribution in the upper panel of the figure.

Finally, about half of the sources shown in the upper panels of Figs. 4 and 5 were modelled with three or more harmonics. For these sources we were therefore also able to measure the metal abundance from the ϕ_{31} Fourier parameter (see Sect. 4.1).

In summary, after running the SOS Cep&RRL on the 639 828 RR Lyrae candidates and final validation of the results, 140 784 were confirmed as bona fide RR Lyrae stars, and 499 044 were rejected. P and amplitude in the G band [Amp(G)] are available for all 140 784 confirmed RR Lyrae stars, the ϕ_{21}, R_{21} Fourier parameters are available for 121 234 of them, and the ϕ_{31}, R_{31} only for 67 681. This also affects the quality and reliability of the SOS Cep&RRL classifications into type, subtype, and pulsation mode(s). A complete discussion of the validation process and final results for the RR Lyrae stars are provided in Sects. 4 and 4.1.

3.2. SOS Cep&RRL processing of all-sky Cepheids released in DR2

As anticipated at the end of Sect. 3, because of the scheduling constraints of the DR2 processing, it was not possible to update the diagnostic relations of the SOS Cep&RRL pipeline using the *Gaia* light curves directly, and we used the same version of the pipeline and relations as defined for DR1 to process the DR2 sources (see Paper 1). In particular, for the analysis of the DR2 candidate Cepheids and the classification of the confirmed ones into different types (DCEP, ACEP, and T2CEP subtypes) and pulsation modes, the SOS Cep&RRL pipeline relied on (i) the Fourier parameters ϕ_{21} and R_{21} of the G-band light curve decomposition and the source position on the ϕ_{21} versus P, R_{21} versus P planes as defined in Sect. 4 and Figs. 21 and 22 of Paper 1, and (ii) the source position with respect to the G-band PL and the PW($G, G_{\rm RP}$) relations, using different values for the slope and zero-point depending on the region of the sky where the sources were located.

Specifically, to process candidate Cepheids within the LMC region (Sect. 2), we used the G-band PL relations described by Eqs. (20)–(24) in Paper 1 (see Sect. 2.4.2. in that paper) and the (G, $G_{\rm RP}$) PW relations defined by Eqs. (6)–(10) in this paper. These relations were obtained by transforming into the *Gaia* passbands, according to Jordi et al. (2010), the relations derived from LMC Cepheids whose light curves in the Johnson-Cousins passbands and pulsation characteristics have been published by the OGLE team. We adopted an absolute de-reddened distance modulus for the LMC of $\mathrm{DM_{LMC}} = 18.49$ mag from Pietrzynski et al. (2013) and the definition of the Wesenheit function in the *Gaia* G, $G_{\rm RP}$ passbands provided by Eq. (5), which was derived for this study assuming for the ratio of total-to-selective absorption the value $R_\lambda = 3.1$,

$$\begin{aligned} W_{(G,G_{\rm RP})} = {} & G - 0.08193 - 2.98056 \times (G - G_{\rm RP}) \\ & - 0.21906 \times (G - G_{\rm RP})^2 \\ & - 0.6378 \times (G - G_{\rm RP})^3, \end{aligned} \tag{5}$$

$$\begin{aligned} & \mathrm{DCEP_F} : W_{(G,G_{\rm RP})} = 15.861 - \mathrm{DM_{LMC}} - 3.317 \times \log P, \\ & \sigma = 0.069 \text{ mag}, \end{aligned} \tag{6}$$

$$\begin{aligned} & \mathrm{DCEP_{1O}} : W_{(G,G_{\rm RP})} = 15.365 - \mathrm{DM_{LMC}} - 3.456 \times \log P, \\ & \sigma = 0.067 \text{ mag}, \end{aligned} \tag{7}$$

$$\begin{aligned} & \mathrm{T2CEP} : W_{(G,G_{\rm RP})} = 17.321 - \mathrm{DM_{LMC}} - 2.527 \times \log P, \\ & \sigma = 0.088 \text{ mag}, \end{aligned} \tag{8}$$

$$\begin{aligned} & \mathrm{ACEP_F} : W_{(G,G_{\rm RP})} = 16.567 - \mathrm{DM_{LMC}} - 3.19 \times \log P, \\ & \sigma = 0.15 \text{ mag}, \end{aligned} \tag{9}$$

$$\begin{aligned} & \mathrm{ACEP_{1O}} : W_{(G,G_{\rm RP})} = 15.995 - \mathrm{DM_{LMC}} - 3.26 \times \log P, \\ & \sigma = 0.14 \text{ mag}. \end{aligned} \tag{10}$$

Similarly, for candidate DCEPs within the SMC region (Sect. 2), we used the G-band PL and the PW($G, G_{\rm RP}$) relations described by the following set of equations:

$$\begin{aligned} & \mathrm{DCEP_F} : M_G = 17.984 - \mathrm{DM_{SMC}} - 2.898 \times \log P, \\ & \sigma = 0.266 \text{ mag}, \end{aligned} \tag{11}$$

$$\begin{aligned} & \mathrm{DCEP_{1O}} : M_G = 17.368 - \mathrm{DM_{SMC}} - 3.155 \times \log P, \\ & \sigma = 0.271 \text{ mag}, \end{aligned} \tag{12}$$

which were derived by transforming to the *Gaia* passbands the relations of Soszyński et al. (2015a) and adopting an absolute de-reddened distance modulus of $\mathrm{DM_{SMC}} = 19.00$ mag obtained by adding +0.51 mag to the LMC distance modulus of Pietrzynski et al. (2013),

$$\begin{aligned} & \mathrm{T2CEP} : M_G = 19.31 - \mathrm{DM_{SMC}} - 1.96 \times \log P, \\ & \sigma = 0.188 \text{ mag}, \end{aligned} \tag{13}$$

which was derived from Soszyński et al. (2015a).

$$\begin{aligned} & \mathrm{ACEP_F} : M_G = 18.33 - \mathrm{DM_{SMC}} - 2.63 \times \log P, \\ & \sigma = 0.22 \text{ mag}, \end{aligned} \tag{14}$$

$$\begin{aligned} & \mathrm{ACEP_{1O}} : M_G = 17.81 - \mathrm{DM_{SMC}} - 3.78 \times \log P, \\ & \sigma = 0.20 \text{ mag}, \end{aligned} \tag{15}$$

[8] In DR2 we did not publish the number of harmonics used to model the light curves. This parameter will be added for DR3.

which were defined by adopting the ACEP F and ACEP FO slopes of the I-band PL relations from Soszyński et al. (2015b) and a $(V-I)$ average colour for the ACEPs of 0.8 mag.

$$\mathrm{DCEP_F}: W_{(G,G_{\rm RP})} = 16.493 - \mathrm{DM_{SMC}} - 3.46 \times \log P, \quad \sigma = 0.155 \text{ mag}, \tag{16}$$

$$\mathrm{DCEP_{1O}}: W_{(G,G_{\rm RP})} = 15.961 - \mathrm{DM_{SMC}} - 3.548 \times \log P, \quad \sigma = 0.169 \text{ mag}, \tag{17}$$

$$\mathrm{T2CEP}: W_{(G,G_{\rm RP})} = 17.64 - \mathrm{DM_{SMC}} - 2.32 \times \log P, \quad \sigma = 0.23 \text{ mag}, \tag{18}$$

$$\mathrm{ACEP_F}: W_{(G,G_{\rm RP})} = 17.01 - \mathrm{DM_{SMC}} - 2.85 \times \log P, \quad \sigma = 0.15 \text{ mag}, \tag{19}$$

$$\mathrm{ACEP_{1O}}: W_{(G,G_{\rm RP})} = 16.64 - \mathrm{DM_{SMC}} - 3.69 \times \log P, \quad \sigma = 0.14 \text{ mag}, \tag{20}$$

which were derived from Soszyński et al. (2015a) for DCEPs and Soszyński et al. (2015b) for ACEPs and T2CEPs.

Candidate Cepheids located in the LMC and SMC regions that fall within 4σ of any of the above PL, PW relations were assigned the Cepheid type and pulsation mode of the closest PL or PW relation. The classification was then refined using the Fourier decomposition parameters R_{21} and ϕ_{21}. Conversely, candidate Cepheids falling beyond 4σ were rejected as misclassified sources. These large intervals were adopted in order to achieve a higher completeness. However, they introduced a higher contamination that we have tried to mitigate during the validation process (see Sect. 4). As in Paper 1, the T2CEPs were further subclassified into the BL Her, W Vir, and the RV Tau type classes, depending on the pulsation period. Following Soszyński et al. (2008), in the *T2CEPSubclassification* module, T2CEPs with periods in the range $1 \le P < 4$ d were classified as BL Her, the T2CEPs with periods in the range $4 \le P < 20$ d as W Vir, and those with periods equal to or longer than 20 days as RV Tau.

Similarly, ACEPs are known to pulsate in the F and 1O modes. The *ACEPModeIdentification* module assigns the pulsation mode to an ACEP by combining results from the classification into types based on the PL relations (*CepTypeIdentification* module) and the source period, as 1O ACEPs have periods in the range $0.35 < P \le 1.20$ d, but F ACEPs have periods in the range $1.20 < P \le 2.5$ d. These limits were inferred from the PL relations of ACEPs based on OGLE-III data.

The selection of bona fide Cepheids from the All-Sky candidate sample (Sect. 2) was made based on the star position in the ϕ_{21} versus P and R_{21} versus P Fourier parameter planes and retaining only candidates that are located in the regions populated by Cepheids known in the literature, as defined in Sect. 4 and Figs. 21 and 22 of Paper 1. As a test, we then compared the All-Sky Cepheids selected from the Fourier parameter planes with the PW$(G, G_{\rm RP})$ relations in Eqs. (21)–(23), which for DCEPs were derived from the TGAS DR1 sample using the Astrometry-Based Luminosity (ABL, Arenou & Luri 1999; see e.g. Gaia Collaboration 2017, and references therein), but for T2CEPs we used the Soszyński et al. (2015b) relation, transformed into the *Gaia* passbands and with an LMC distance modulus of 18.49 mag subtracted,

$$\mathrm{DCEP_F}: W_{(G,G_{\rm RP})} = -3.21 - 2.93 \times \log P, \quad \sigma = 0.37 \text{ mag}, \tag{21}$$

$$\mathrm{DCEP_{1O}}: W_{(G,G_{\rm RP})} = -4.31 - 2.98 \times \log P, \quad \sigma = 0.69 \text{ mag}, \tag{22}$$

$$\mathrm{T2CEP}: W_{(G,G_{\rm RP})} = -1.15 - 2.53 \times \log P, \quad \sigma = 0.11 \text{ mag}. \tag{23}$$

We considered only the PW$(G, G_{\rm RP})$ relation, because it is reddening free and no information on the reddening of the All-Sky Cepheids, which are highly reddened since they mainly reside in the Galactic disc, was available at the time of the processing[9].

For the comparison with the PW relations, we used the parallaxes and worked directly in parallax space with the ABL. The deviation (Δ) from the reference PW relations was computed according to the equation

$$\Delta = \left| \mathrm{ABL} - 10^{0.2(a \log P + b)} \right|, \tag{24}$$

where ABL is defined as

$$\mathrm{ABL} = \varpi 10^{0.2 W_{(G,G_{\rm RP})} - 2.0} \tag{25}$$

and ϖ is the stellar parallax.

The slope of the PW relation for the All-Sky fundamental-mode DCEPs (Eq. (21)) appears to be shallower than the corresponding OGLE slopes for the LMC and SMC samples, and more importantly, the scatter in the PW relations for the All-Sky DCEPs (Eqs. (21) and (22)) is much larger than the LMC values: it extends from ± 0.069 mag for fundamental-mode and 0.067 mag for first-overtone DCEPs in the LMC to ±0.37 and 0.69 mag for the corresponding All-Sky samples. This is due to the large uncertainty of the DR1 TGAS parallaxes we used to derive the DCEP reference PW relations in Eqs. (21) and (22), and likely also to the adoption of $R_\lambda = 3.1$ in the definition of the PW functions.

In total, the SOS pipeline analysed 72 455 candidate Cepheids provided by the classifiers. A total of 62 880 objects were rejected. Of the remaining 9575 objects, 3767, 3692 and 2116 are distributed in the LMC, SMC, and All-Sky regions, respectively. The vast majority of the Cepheids of all types in the Magellanic Clouds are known from the OGLE survey, with the exception of 118 new objects. In the All-Sky sample, 998 objects are classified as Cepheids of any type in the literature, 419 have an uncertain or other than Cepheid classification, and for the remaining 699 targets, no cross-match with known sources was found (Ripepi et al. 2018).

Sources classified as Cepheids of any type by the SOS pipeline are displayed in Figs. 6 and 7 for the PL and PW relations, respectively. In both figures, the upper, middle, and lower panels display LMC, SMC, and All-Sky sources, respectively. Clearly, only objects having a positive parallax value can be shown in the All-Sky panels of these figures. An intriguing feature of Figs. 6 and 7 is the sequence of stars that runs almost parallel to the abscissa axis with absolute G magnitudes fainter than 2 mag. This sequence is also clearly visible in Fig. 8, which shows the PL and PW relations for the 998 All-Sky sources that have a classification as Cepheids in the literature. We have

[9] The Wesenheit function W (Madore 1982) is reddening free by construction, but it depends on the assumed value of the ratio of total-to-selective absorption R_λ. We adopted $R_\lambda = 3.1$, which perhaps is too low for classical Cepheids in dusty regions of the Galactic disc, where a higher value might be more appropriate. For DR3 we foresee the adoption of a varying R_λ value depending on the source location on the sky.

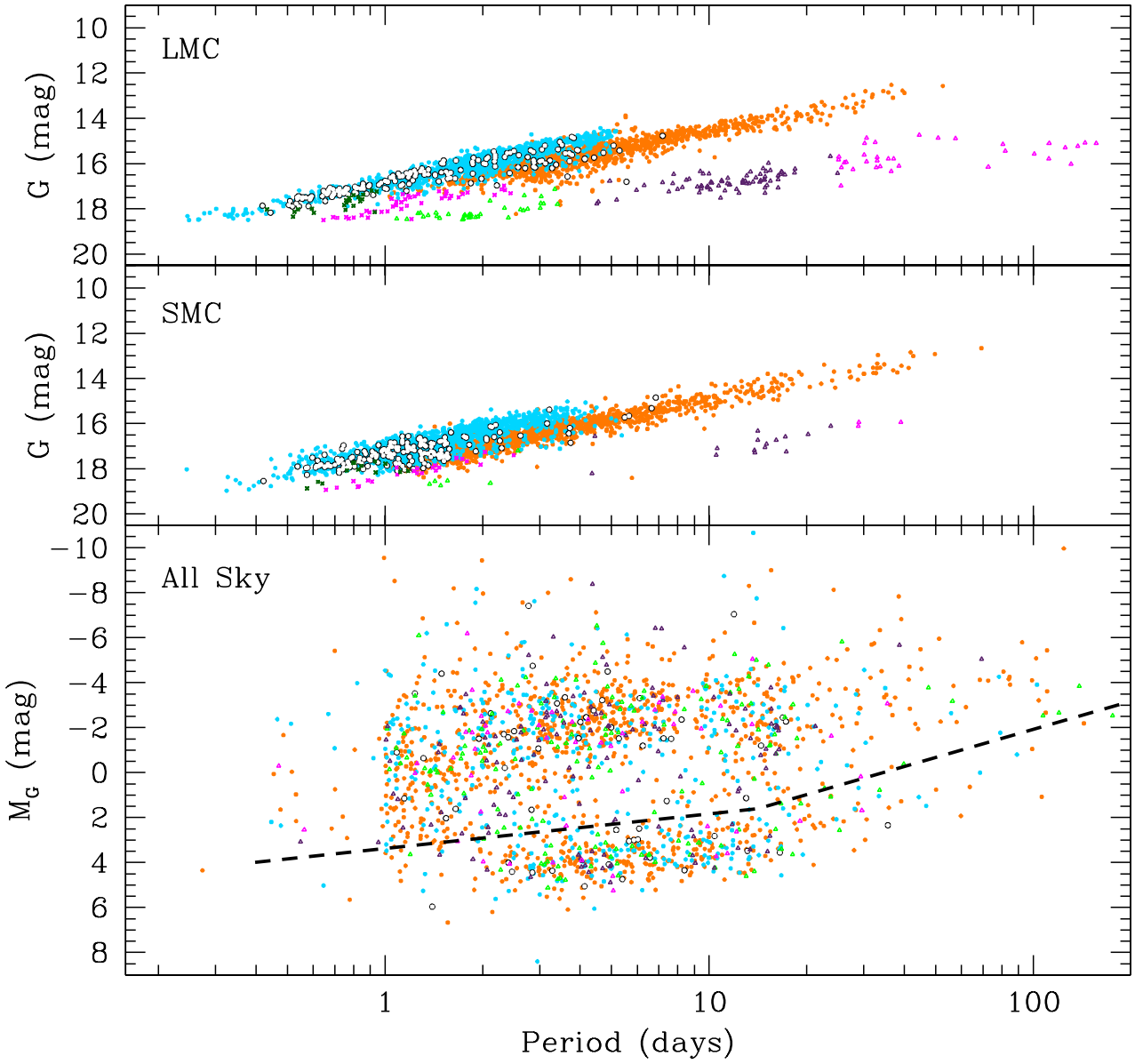

Fig. 6. *Upper and middle panels:* G-band PL distribution in apparent magnitude not corrected for reddening of DCEPs, ACEPs, and T2CEPs in the LMC and SMC, respectively. *Lower panel:* PL distribution in absolute G magnitude (M_G) not corrected for extinction of All-Sky Cepheids of the different types. Orange filled circles: DCEPs F; cyan filled circles: DECPs 1O; blank filled circles: multi-mode DCEPs; magenta four-stars: ACEPs F; dark green four-stars: ACEPS 1O; green open triangles: BL Her; violet open triangles: W Vir; magenta open triangles: RV Tau. The much larger scatter of the All-Sky Cepheid PL distribution is clearly visible: the Y-axis in the lower panel of the figure spans a magnitude range of 20.0 mag, to compare with the 11.5 mag range of the two upper panels. Several All-Sky sources lie below the dashed line in the lower panel of the figure. They are a mix of misclassifications (spurious sources), sources with very high reddening, and Cepheids with an incorrect parallax value that is due to the still simplified astrometric processing applied for DR2, among which in particular the lack of a proper treatment of binary or multiple sources (see Sect. 3.2 and Ripepi et al. 2018, for more details).

drawn dashed lines in Figs. 6–8 to better highlight the regions that are populated by these faint sources. The true nature of the sources populating these faint sequences is unclear. If they are bona fide Cepheids, their *Gaia* DR2 parallax must be incorrect. Conversely, if their parallax is correctly measured, they cannot be Cepheids, and the SOS Cep&RRL classification is in error, as must be some of the literature classifications as well. The All-Sky Cepheid sample may be significantly contaminated by spurious sources, and a large fraction of the sources below the dashed lines in the lower panels of Figs. 6 and 7 and in Fig. 8 might be misclassifications. In the Fourier parameters versus period planes (Figs. 37 and 38), these spurious sources share the same loci, hence are indistinguishable from bona fide Cepheids. Therefore, they were retained although they were more than 4σ away from the PW relations in Eqs. (21)–(23), but they are a different type of variables that is much fainter than Cepheids. This is indeed the case of the star with *Gaia* sourceid 2077108036182676224. The SOS Cep&RRL pipeline classifies this source as a multi-mode classical Cepheid with fundamental-mode period $P = 1.045$ day. However, the parallax places the star on the lower sequences in Figs. 6 and 7. The source was also observed by *Kepler* (*Kepler* sourceid: KIC 6619830) and has a full *Kepler* light curve showing a hump in the phase interval 0.7–0.9. This indicates that it is a spotted rotating star, which

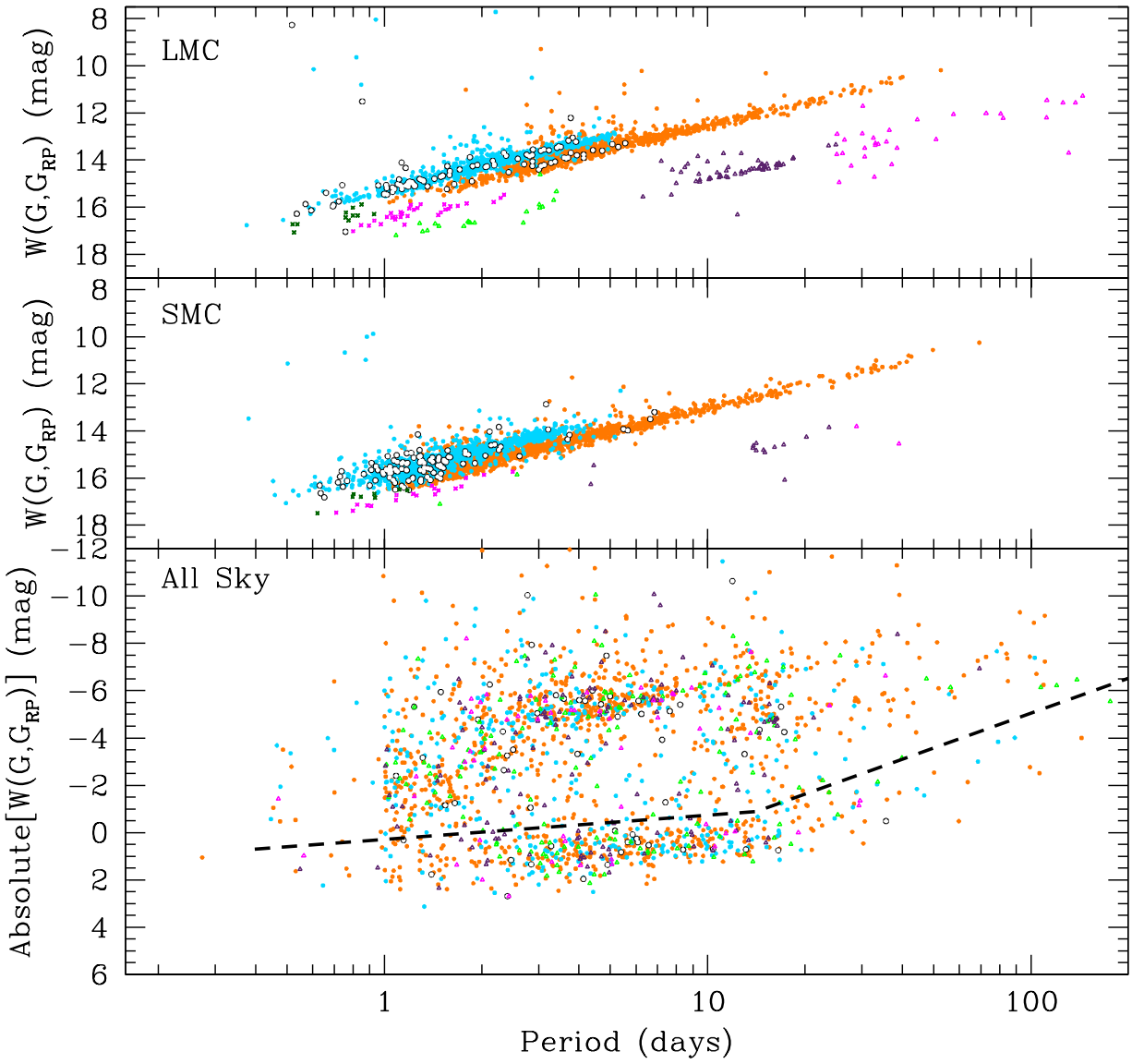

Fig. 7. *Upper and middle panels:* PW(G, G_{RP}) distribution in apparent magnitude of DCEPs, ACEPs, and T2CEPs in the LMC and SMC, respectively. *Lower panel:* period–absolute ($W(G, G_{RP})$) distribution of All-Sky Cepheids of the different types. Symbols and colour-coding are the same as in Fig. 6. The much larger scatter of the All-Sky Cepheid PW distribution is clearly visible: the Y-axis in the lower panel of the figure spans a magnitude range of 18.0 mag, to be compared with the 11.5 mag range of the two upper panels. Several All-Sky sources lie below the dashed line in the lower panel of the figure. They are a mix of misclassifications (spurious sources) and Cepheids with an incorrect parallax value that is due to the still simplified astrometric processing applied for DR2, among which in particular the lack of a proper treatment of binary or multiple sources (see Sect. 3.2 and Ripepi et al. 2018, for more details).

is consistent with its absolute magnitude (T. Bedding & D. Hey, priv. comm.). The phase coverage of the *Gaia* light curve is poor and the hump at phase 0.7–0.9 is not sampled, so it could easily be mistaken for a Cepheid. Perhaps some of the sources on the lower sequence of Figs. 6 and 7 are rotators with a few observations and poor coverage of the light curve that were misclassified by the SOS Cep&RRL pipeline. The large scatter in Figs. 6 and 7 could also be due in part to the DR2 parallax for some of the bona fide Cepheids being in error due to an incorrect determination of the stellar mean G magnitude and because the colour variation during the pulsation cycle was not taken into account in the derivation of the DR2 astrometric solution[10]. An additional source of scatter in the PL distribution in Fig. 6 is that the reddening that affects the sources was not known, therefore we were unable to apply any correction for reddening. All these effects may have combined to inflate the dispersions observed in the PL and PW distributions of the All-Sky MW samples in Figs. 6 and 7, as is suggested by the large scatter seen also for the known Cepheids in Fig. 8.

We refer to Ripepi et al. (2018), where these issues are discussed in more detail and a catalogue of bona fide DR2 new All-Sky Cepheids is presented after excising other types of variable sources and misclassifications from the sample.

4. Validations and results

The SOS Cep&RRL pipeline produced lists of confirmed Cepheids and RR Lyrae stars. However, as discussed in the

[10] See also https://www.cosmos.esa.int/web/gaia/dr2-known-issues and the technical note: GAIA-C3-TN-LU-LL-124-01.

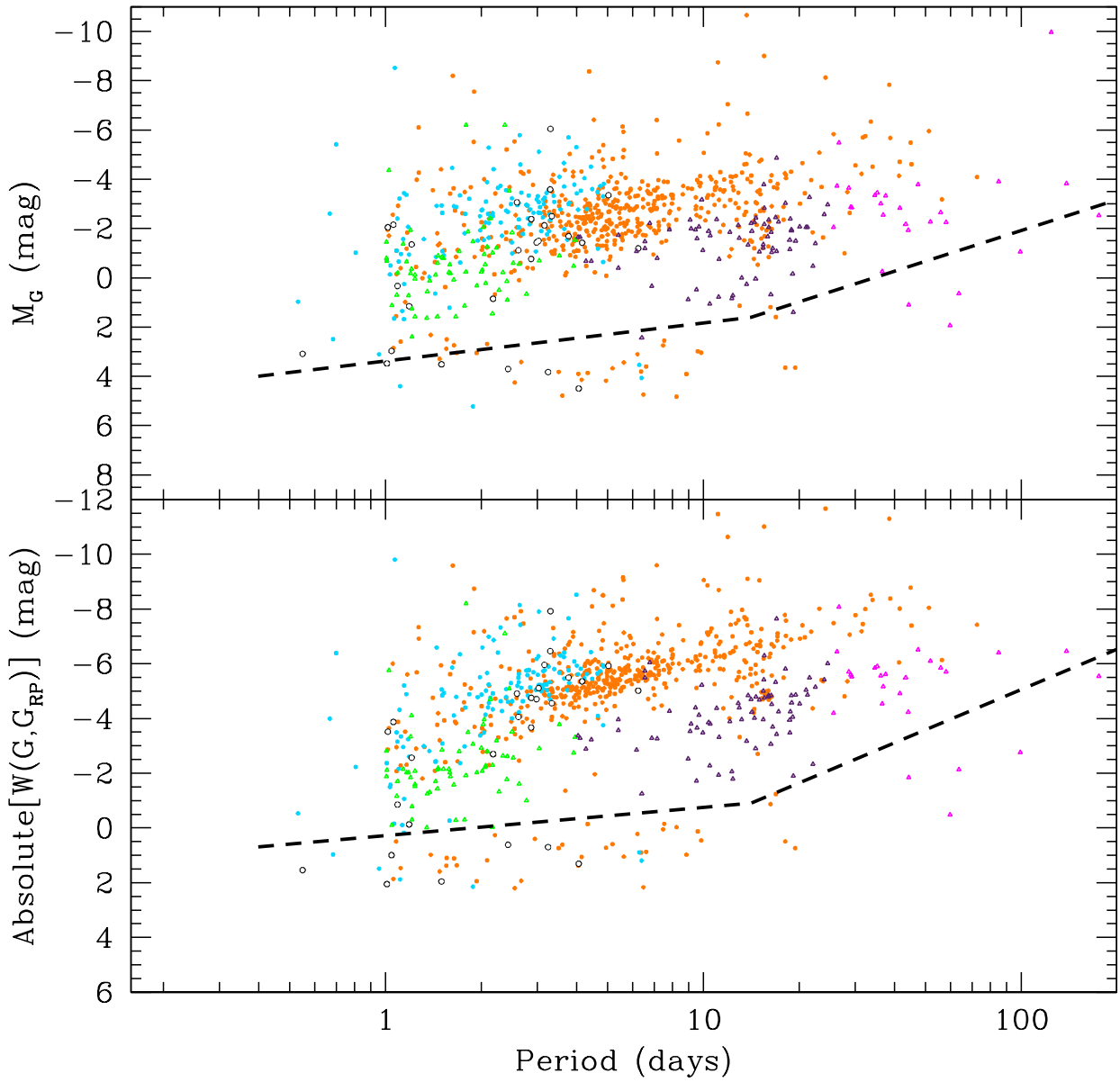

Fig. 8. *Upper panel:* G-band PL distribution in absolute G magnitude not corrected for extinction of All-Sky DCEPs and T2CEPs known in the literature. *Lower panel:* period–absolute($W(G, G_{RP})$) distribution of All-Sky DCEPs and T2CEPs known in the literature. Symbols and colour-coding are the same as in Fig. 6. The dashed lines in each panel indicate the regions below which the sources may be misclassifications (spurious sources) or may be bona fide Cepheids with an incorrect parallax value that is due to the still simplified astrometric processing applied for DR2, among which in particular the lack of a proper treatment of binary or multiple sources (see Sect. 3.2 and Ripepi et al. 2018, for more details).

previous sections, not all sources are correctly classified by the SOS processing. An additional significant source of contamination for RR Lyrae stars in particular is due to contact binary systems, because contact binaries and RRc stars populate the same regions of the PA and ϕ_{21} versus P diagrams in Figs. 4 and 5 (specifically, the region around ϕ_{21} ~6.28 rad). The *AmplitudeRatios* module in the RR Lyrae branch of the SOS Cep&RRL pipeline is designed to clear the RR Lyrae sample from contact binaries that mimic RRc-like light curves, based on the principle that amplitude ratios in different photometric bands are close to unity for binaries. This module was not activated during the DR2 processing, but during validation, we still computed the G_{BP}, G_{RP} peak-to-peak amplitude ratio and removed sources with $\mathrm{Amp}(G_{BP})/\mathrm{Amp}(G_{RP}) = 1 \pm 0.2$ as binaries. A visual inspection of random samples of the removed sources confirmed that they were indeed binaries.

In order to gauge the contamination of the SOS Cep&RRL results by other types of variables, we also cross-matched them against catalogues of variables available in the literature. Main references were the OGLE catalogues of variable stars in the Magellanic Clouds and Galactic bulge, which we complemented with entries from the ASAS, LINEAR, and CTRS catalogues of variables in the MW halo and with further Galactic Cepheids taken from SIMBAD and VSX.

The final step of the validation process was a cleaning procedure that exploits a training set of sources with well-established classifications in the literature. We specifically considered the fiducial regions occupied by the training sources

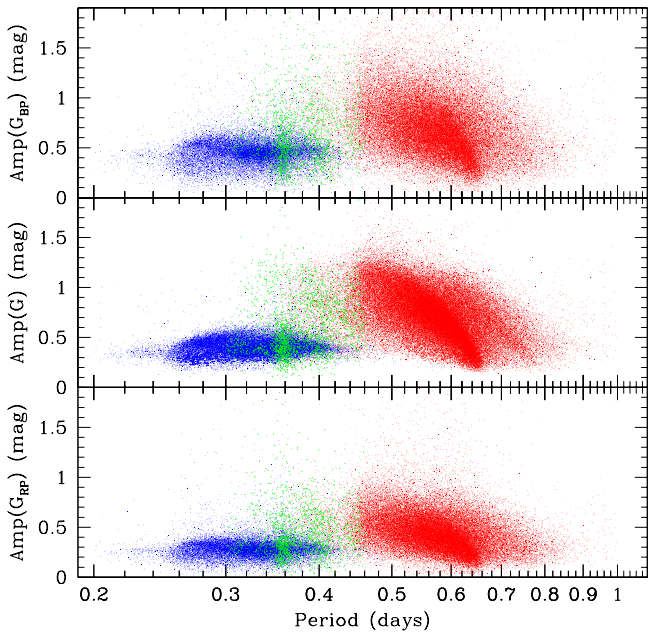

Fig. 9. PA diagrams in the G (*middle panel*), $G_{\rm BP}$ (*upper panel*), and $G_{\rm RP}$ (*lower panel*) bands of the RR Lyrae stars (140 784 in the G-band panel) confirmed by the SOS Cep&RRL pipeline that are released in DR2. The sources are colour-coded as in Fig. 4.

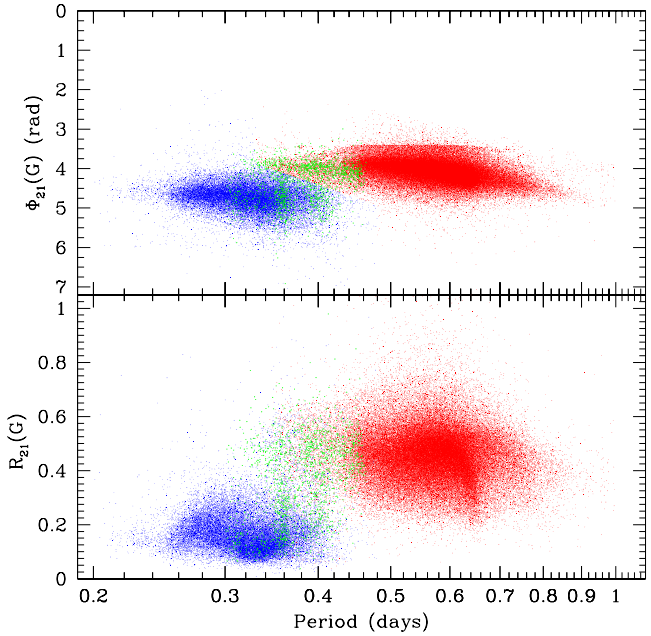

Fig. 10. G-band ϕ_{21} vs. period (*upper panel*) and R_{21} vs. period (*lower panel*) diagrams for the RR Lyrae stars that are confirmed by the SOS Cep&RRL pipeline. The colour-coding is the same as in Fig. 4.

in the planes of parameters used by the SOS code for classification, that is, the Amp(G), R_{21}, ϕ_{21} and the ϕ_{31} versus period planes. We expect sources that are correctly classified to populate the same regions as the training sources. To achieve a quantitative mathematical definition of these "fiducial" regions, we subdivided each parameter plane using a rectangular grid, assigning an occupation frequency value to every rectangular bin. The frequency of occupation of each bin was defined as $f_i = n_i/\mathrm{MAX}(n_i)_{(i=1,n_{\rm bins})}$, where n_i is the number of sources that are present in the ith bin. A smoothing algorithm, given by the *matrixSmooth* routine (R Core Team 2018), was applied to the resulting frequency matrix to avoid sharp edges for the "fiducial" regions. The occupation frequency matrices were also calculated for the complete SOS output sample using the same rectangular grids, but without the smoothing step. In order to select only sources in the defined "good" regions, we multiplied the occupation matrix of the complete sample by that of the training set, and we performed this operation for the four parameter planes described above. Finally, we retained only the sources that were located in bins with a combined occupation frequency value higher than 0 in all parameter planes.

The final result of the cleaning procedures described above is a validated sample comprising 140 784 RR Lyrae stars and 9575 Cepheids. These form the final sample of SOS Cep& RRL confirmed sources released in *Gaia* DR2.

4.1. Results for RR Lyrae stars

Figure 9 shows the PA diagrams in the G (middle panel), $G_{\rm BP}$ (upper panel), and $G_{\rm RP}$ (lower panel) bands of the RR Lyrae stars that are confirmed by the SOS Cep&RRL pipeline and released in DR2. The two panels of Fig. 10 show the G-band ϕ_{21} versus period (upper panel) and R_{21} versus period (lower panel) diagrams of the subsample of 121 234 RR Lyrae stars whose light curves could be modelled with at least two harmonics. Finally,

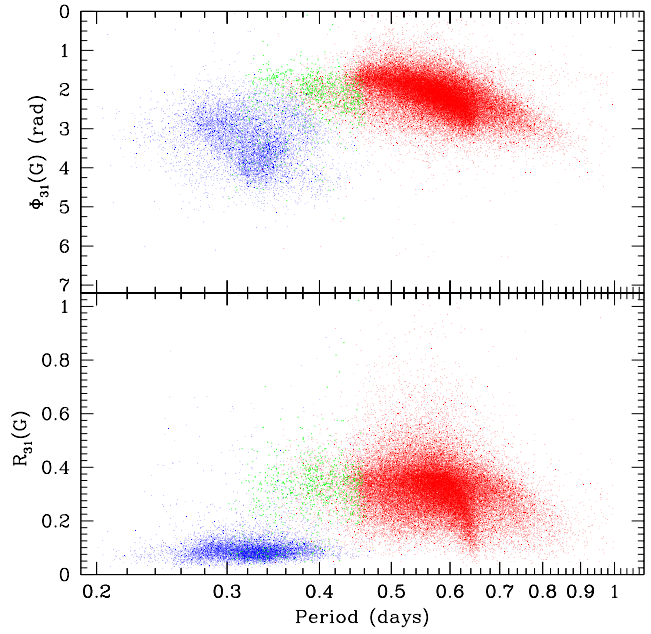

Fig. 11. Same as in Fig. 10 for the G-band ϕ_{31} vs. period (*upper panel*) and R_{31} vs. period (*lower panel*) diagrams.

the two panels of Fig. 11 show the G-band ϕ_{31} versus period (upper panel) and R_{31} versus period (lower panel) diagrams for 67 681 RR Lyrae stars whose light curves were modelled with at least three harmonics. The sample of confirmed RR Lyrae stars released in *Gaia* DR2 includes variables in the MW (disc, bulge, and halo), in the two Magellanic Clouds, in 5 dSph galaxies, 7 ultra-faint dwarfs, and in 87 globular clusters (GCs). We present some tests that we performed on the RR Lyrae stars in GCs and one of the dSphs (Sculptor) in Sect. 4.1.2. Examples of light curves for RR Lyrae stars in these various systems are presented in Figs. 12–14. In all plots the light curves are

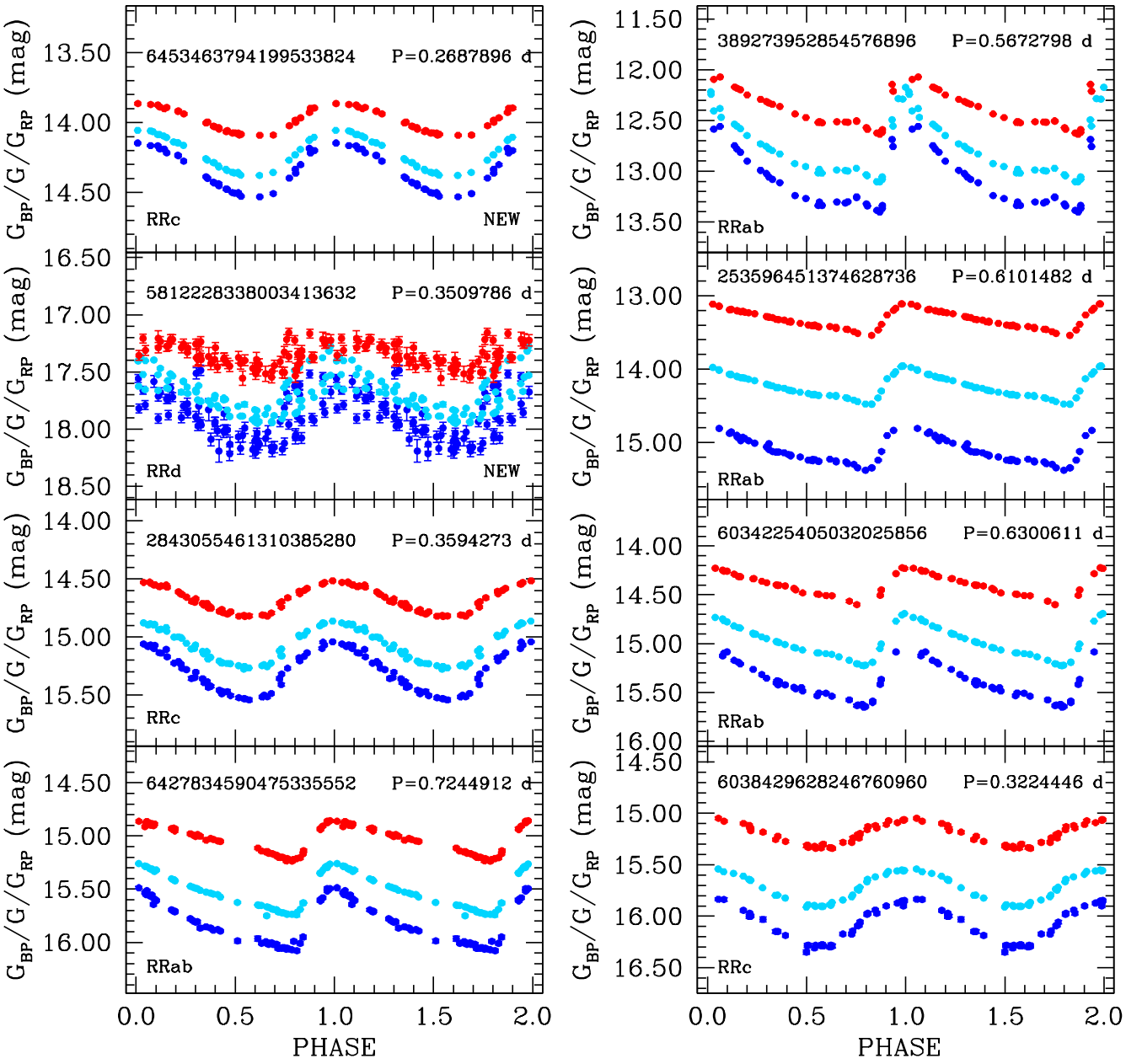

Fig. 12. G (cyan), G_{BP} (blue), and G_{RP} (red) light curves of All-Sky RR Lyrae stars of different pulsation mode released in *Gaia* DR2. The multi-band time-series data are folded according to the period and epoch of maximum light derived by the SOS Cep&RRL pipeline. Error bars are comparable to or smaller than the symbol size. New discoveries by *Gaia* are flagged.

folded according to the period and epochs of maximum light in the G, G_{BP} , and G_{RP} bands determined by the SOS Cep&RRL pipeline. The G_{BP} light curve of the LMC RR Lyrae star in the top right panel of Fig. 13 is brighter and has a lower amplitude than the G-band light curve, likely because the star is blended with a companion source that affects its G_{BP} photometry but not the other bands. This may happen in crowded fields such as the internal regions of the Magellanic Clouds, where this specific RR Lyrae star is located, or the core of a globular cluster, because of the longer extraction windows of the G_{BP}, G_{RP} spectrophotometric data compared to the G band (see Evans et al. 2018).

Figure 15 shows the CMDs defined by all the confirmed RR Lyrae stars for which G_{BP} and G_{RP} photometry is available (83 097 sources in total). The figure updates and improves Fig. 3 in Eyer et al. (2017a). A different colour-coding has been used for RRab (red), RRc (blue), and double-mode (green) pulsators. As expected, RRc and RRd pulsators are slightly bluer than RRab stars. The high concentrations of sources at $G \sim 19$ and 19.7 mag are the LMC and SMC variables, respectively, whereas the overdensities fainter than $G \sim 20$ mag are due to RR Lyrae stars in the Draco and Sculptor dSphs. The arm extending towards redder colours is produced by reddened variables in the MW disc and bulge. The figure confirms that sources with extreme and unphysical red colours were efficiently removed by cutting in excess flux. CMDs in the *Gaia* passbands showing such a large number of All-Sky RR Lyrae stars with different pulsation type and intensity-averaged mean magnitudes and colours computed over the full pulsation cycle have never been published before. Figure 16 instead shows the CMDs defined by

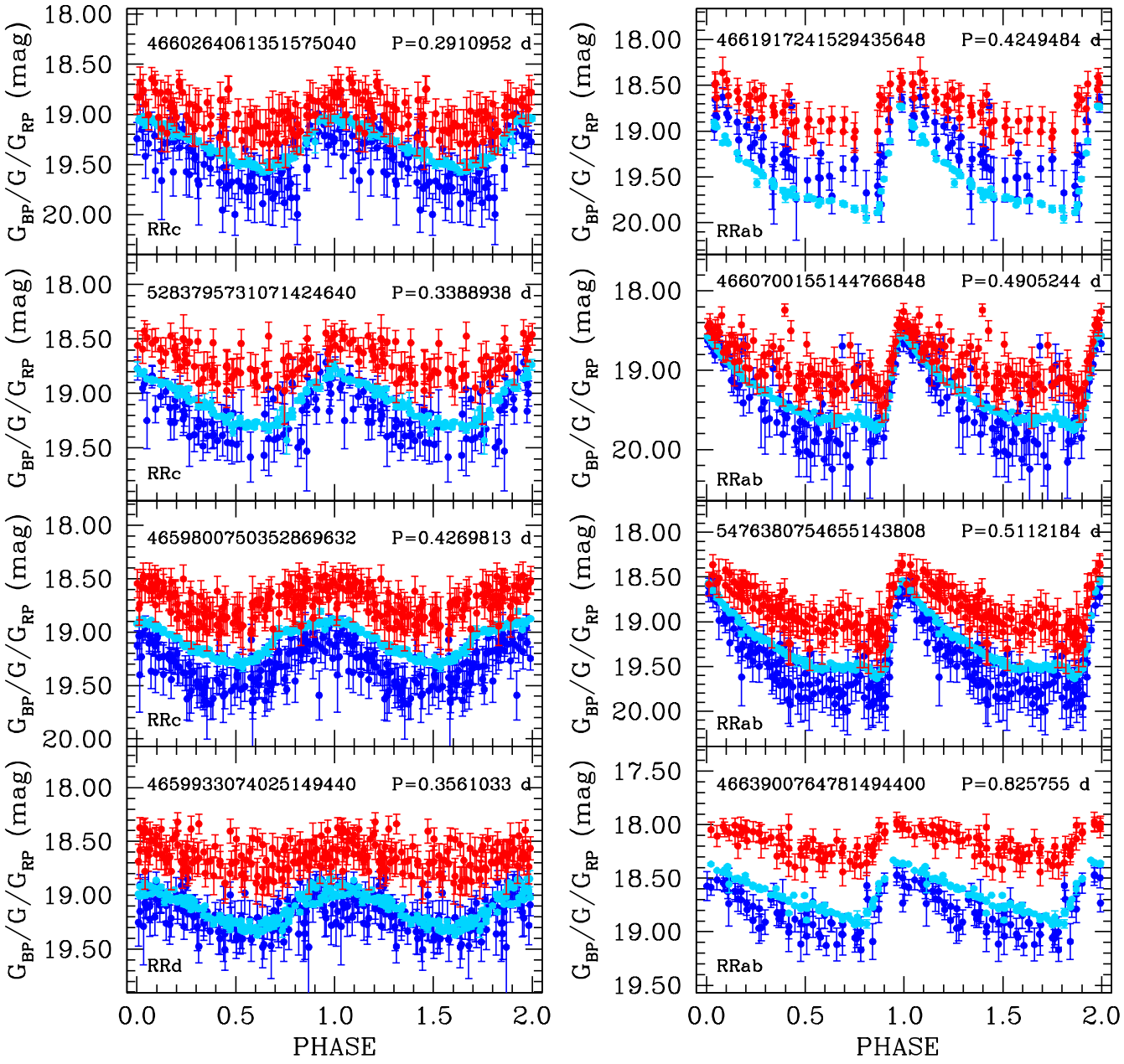

Fig. 13. G (cyan), $G_{\rm BP}$ (blue), and $G_{\rm RP}$ (red) light curves of RR Lyrae stars of different pulsation mode in the LMC and SMC, released in *Gaia* DR2. The multi-band time-series data are folded according to the period and epoch of maximum light derived by the SOS Cep&RRL pipeline. Although the errors are larger than in the G band, the colour light curves are well defined. The $G_{\rm BP}$ light curve of the RR Lyrae star in the *top right panel* is brighter and has a lower amplitude than the G-band curve, likely because the star is blended with a companion that contaminates the $G_{\rm BP}$ photometry (see Sect. 4.1 for details).

the confirmed RR Lyrae stars in GCs (red) and dSphs (blue) for which $G_{\rm BP}$ and $G_{\rm RP}$ photometry is available (1167 sources in total). Each concentration of red points in this figure corresponds to a different GC. Although not used for the DR2 processing, CMDs like those in Figs. 15 and 16 will be the first tool used by the SOS Cep&RRL pipeline (see Fig. 2) for the classification of RR Lyrae stars during the processing for the next *Gaia* release (DR3), which is currently foreseen for the first half of 2021.

Individual photometric metallicities ([Fe/H]) were derived from the ϕ_{31} parameter of the light-curve Fourier decomposition for 64 957 of the confirmed RR Lyrae stars. The corresponding metallicity distributions are shown in Fig. 17, where the variables are divided according to the three separate regions (LMC, SMC, and All-Sky) defined in Sect. 2. The three distributions peak at mean values of [Fe/H] approximately -1.15 ± 0.6, -1.3 ± 0.7, and -1.6 ± 0.7 dex for the MW, LMC, and SMC RR Lyrae stars, respectively.

4.1.1. Results for double-mode RR Lyrae stars published in DR2

The SOS Cep&RRL pipeline detected a secondary periodicity and classified as double-mode pulsators 2378 of the 140 784 confirmed RR Lyrae stars. According to the comparison with the literature, this rather large number of RRd stars comprises 558 cross-matches with known RR Lyrae variables that are classified

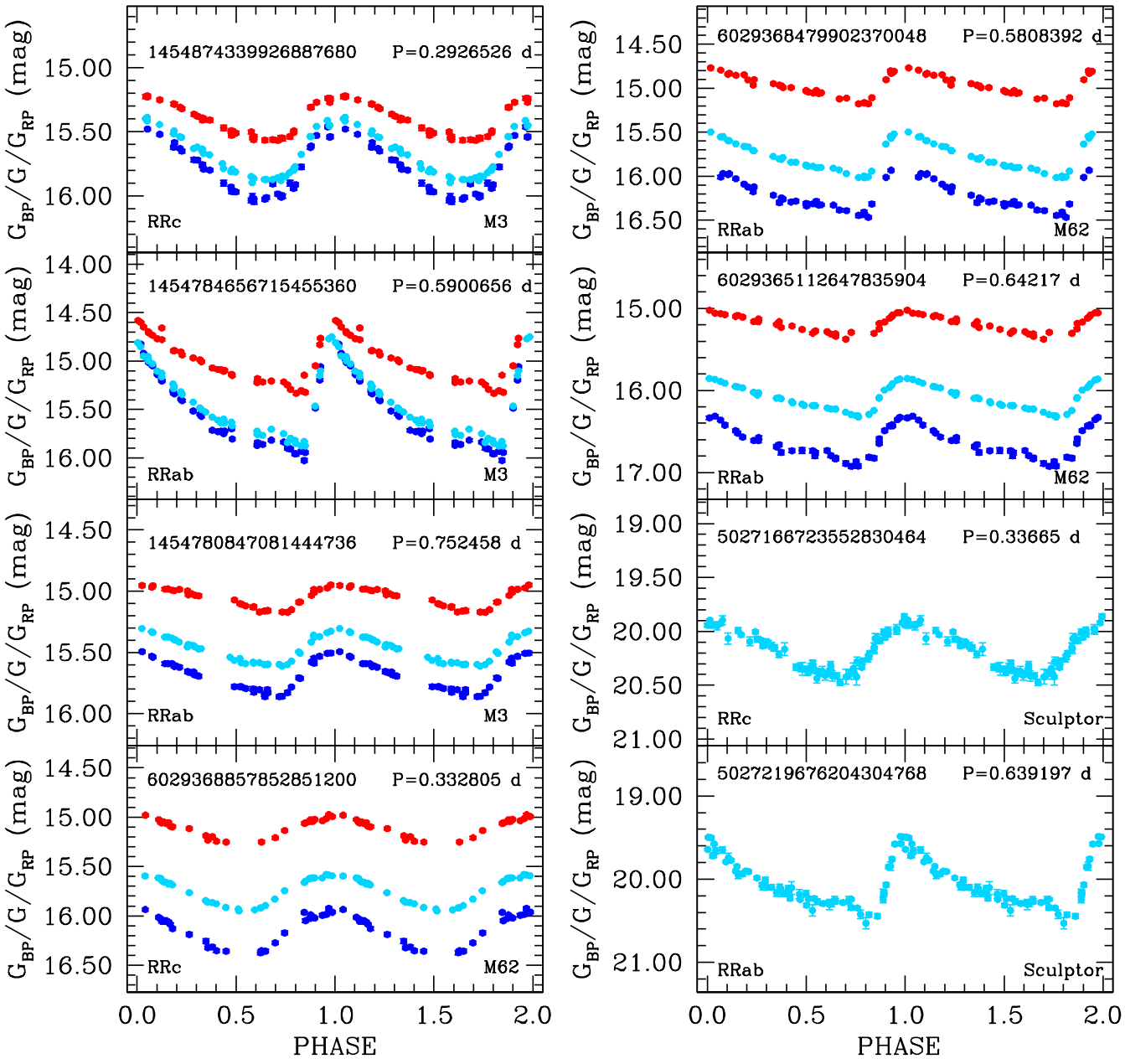

Fig. 14. *G* (cyan), G_{BP} (blue), and G_{RP} (red) light curves of RR Lyrae stars of different pulsation mode in the Galactic globular clusters M3 and M62 and in the Sculptor dSph galaxy (*G* band only), released in DR2. *Gaia* DR2 1454784656715455360 is located within the M3 half-light radius ($r_h = 2.31$ arcmin, Harris 1996), star 6029368479902370048 is within twice the r_h of M62 ($r_h = 0.92$ arcmin, Harris 1996). *Gaia* DR2 5027166723552830464 in Sculptor is a new discovery by *Gaia*. Only the *G*-band light curves are shown for Sculptor because of the low S/N of the G_{BP} and G_{RP} time-series data at the faint magnitudes of the RR Lyrae stars in this dSph galaxy. The multi-band time-series data are folded according to the period and epoch of maximum light derived by the SOS Cep&RRL pipeline. Error bars are comparable to or smaller than symbol size.

as double-mode pulsators in the literature (a *Gaia* light curve for one of them is shown in the bottom left panel of Fig. 13), 1067 cross-matches with known RR Lyrae stars that are classified as single-mode pulsators in the literature, either of RRab or RRc types, and the remaining 753 sources without a counterpart in the literature. We specifically note that the SOS Cep&RRL classification as an RRd pulsator relies only on the detection of two periodicities in the proper period ratio in the time-series data (see Sect. 2.3.1 of Paper 1), but currently does not take into account whether the source exhibits additional scatter or noise in the light curve folded according to the primary, dominant periodicity, as is commonly observed among RRd stars. Hence, while those 2378 sources all are found by the SOS Cep&RRL to have two periodicities in the proper ratio excited, a clear confirmation of their actual nature as double-mode pulsators will require further analysis and additional data. A refinement of the SOS Cep&RRL algorithm for the detection of double-mode pulsators is foreseen for *Gaia* DR3.

4.1.2. RR Lyrae stars in globular clusters and dwarf spheroidal galaxies

Of the 140 784 SOS-confirmed RR Lyrae stars released in *Gaia* DR2, 1986 reside in GCs and dSphs (classical and ultra-faint)

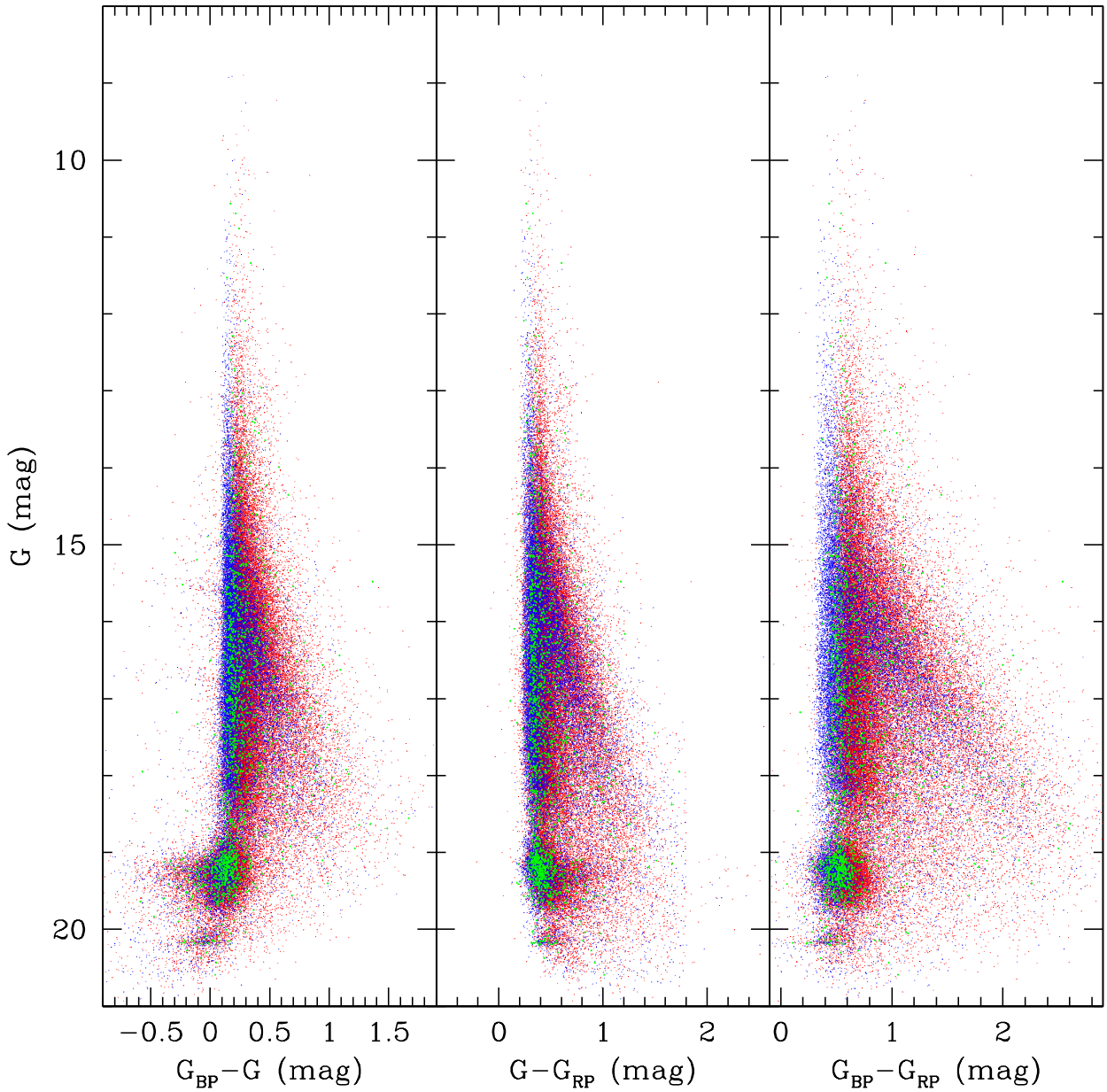

Fig. 15. G, G_{BP}–G; G, G–G_{RP} ; and G, G_{BP}–G_{RP} CMDs of the RR Lyrae stars confirmed by the SOS Cep&RRL pipeline for which the colour information is available (83 097 sources out of 140 784). Blue shows RRc stars, green represents RRd stars, and red shows RRab stars. The high concentrations of sources at $G \sim 19$ and 19.7 mag are the LMC and SMC variables, respectively, and the overdensities below $G \sim 20$ mag are due to RR Lyrae stars in the Draco and Sculptor dSphs. The arm extending towards redder colours is produced by reddened variables in the MW disc and bulge.

within reach of the *Gaia* limiting magnitude ($G \lesssim 20.7$ mag). Specifically, 1569 are distributed over 87 GCs and 417 over 12 dSphs, with the largest numbers being in M3 (159), NGC 3201 (83), Sculptor (176), and Draco (176). Figure 16 shows the CMDs defined by RR Lyrae stars in these systems. Examples of light curves in the Galactic GCs M3 and M62 and in the Sculptor dSph are shown in Fig. 14. We specifically tested the SOS Cep&RRL pipeline on the RR Lyrae stars in GCs, in order to establish its performance in crowded fields and to verify the reliability of the derived pulsation characteristics (e.g. periods and amplitudes) and stellar parameters (metallicity and G-band absorption) on systems such as GCs for which metallicity and reddening are generally well known in the literature. In M3, a relatively low central concentration cluster (c = 1.89, Harris 1996), *Gaia* recovered 159 of the 222 known RR Lyrae stars. They are plotted as large black filled circles in the map in Fig. 18, where different colours are used for stars at different distance from the cluster centre. *Gaia* was capable to identify RR Lyrae stars even within the core radius of M3 (r_c = 0.37 arcmin). Centre and right panels of Fig. 18 show the *Gaia* CMD of M3 using the same colour-coding as in the left panel. RR Lyrae stars that are brighter than the cluster horizontal branch (HB) level in the right panel of Fig. 18 are located closer to the cluster centre, where their G_{BP} photometry is more likely to be contaminated by companions (a similar effect was observed in the light curve shown in the upper right panel of Fig. 13 and discussed in Sect. 4.1). The effect of contamination by companions in the M3 central regions is better visible in the two panels

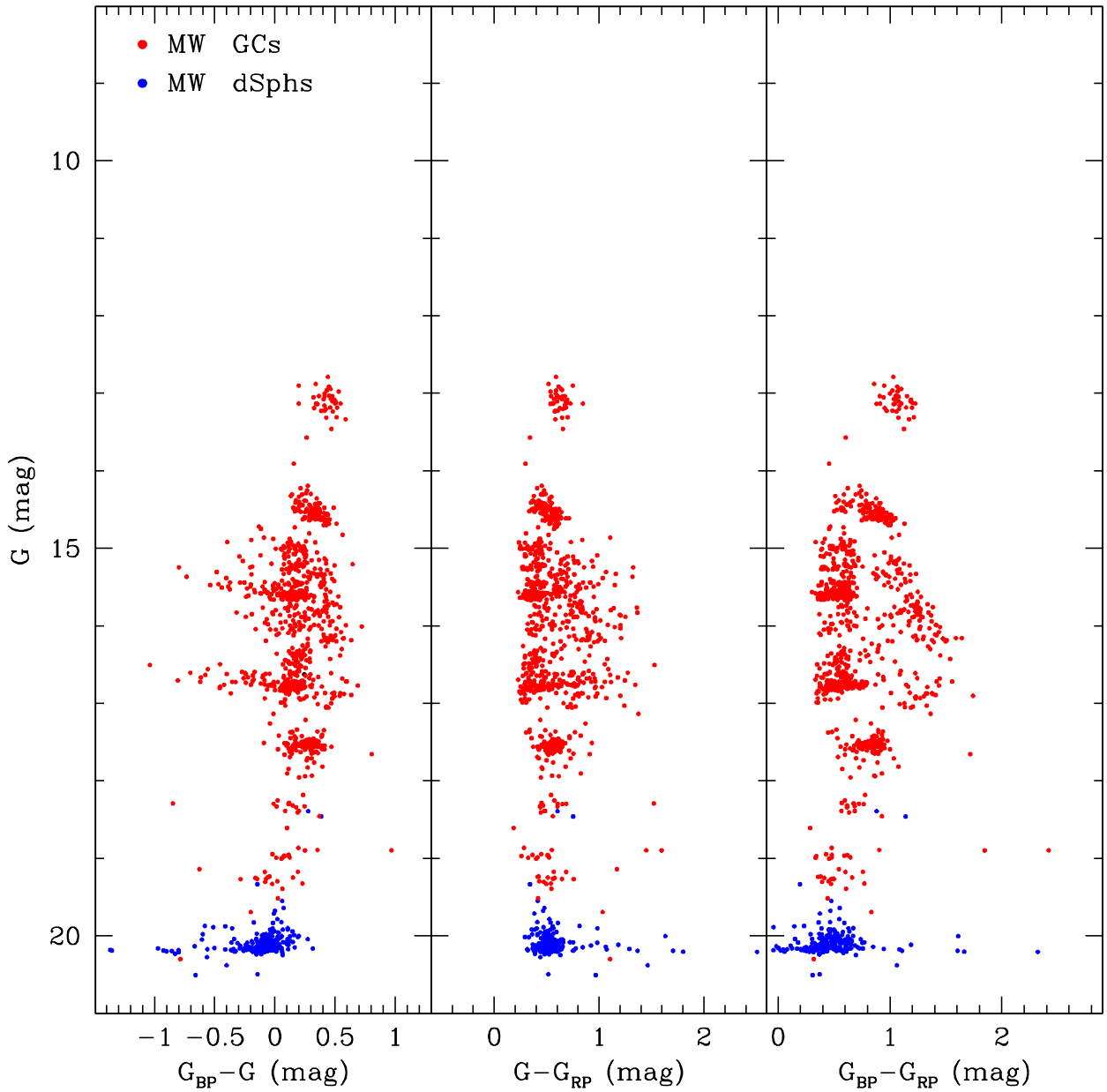

Fig. 16. Same as in Fig. 15, but for the RR Lyrae stars confirmed by the SOS Cep&RRL pipeline for which the colour information is available (1167 sources out of 1986) that are in globular clusters (red points) and dwarf spheroidal galaxies (blue points). Each concentration of red points corresponds to a different GC, the blue overdensity below $G \sim 20$ mag is due to RR Lyrae stars in the Draco and Sculptor dSphs.

of Fig. 19, which show an enlargement of the HB region in the *Gaia* CMD of the cluster. Although the *Gaia* photometry might suffer from blending in crowded regions like the centre of a GC, the periods measured by the SOS Cep&RRL pipeline and the literature periods for the 159 RR Lyrae stars observed in M3 by *Gaia* are in excellent agreement, as shown in Fig. 20. The three deviating objects in the figure are variable stars known to be affected by the Blazhko effect (Blazhko 1907), a modulation of shape and amplitude of the light variation that may occur on time spans ranging from a few tens to hundreds of days. Finally, metal abundances were measured from the ϕ_{31} Fourier parameter for 111 RR Lyrae stars in the cluster. The corresponding metallicity distribution is shown in Fig. 21.

The same comparison was also made for M62, a suspected core-collapsed GC with $c = 1.71$, $r_c = 0.22$ arcmin and half-light radius $r_h = 0.92$ arcmin (Harris 1996). Because of the higher concentration, only 91 out of 214 (~42.5%) RR Lyrae stars known in M62 were recovered, to be compared with 71.6% in M3, which hosts the same RR Lyrae population. Their metallicity distribution is shown in Fig. 22.

More than 200 RR Lyrae stars are known in the Sculptor dSph from the study of Kaluzny et al. (1995). We recovered 176 of them by cross-matching the SOS confirmed RR Lyrae stars against the catalogue of Kaluzny et al. (1995), which so far remains the only study where identifications and coordinates for the Sculptor variable stars were published in the literature. Figure 23 shows the G-band PA diagram of the Sculptor RR Lyrae stars, where filled symbols are known variables that are cross-matched with Kaluzny et al. (1995), while open symbols are new RR Lyrae stars observed by *Gaia*. Figure 24 shows that a good agreement is found between individual metallicities measured by the SOS Cep&RRL pipeline for RR Lyrae stars in

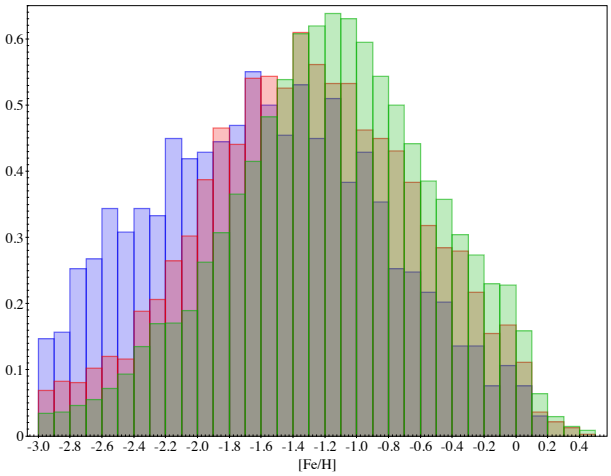

Fig. 17. Normalised metallicity distribution of 64 957 RR Lyrae stars in the sample of 140 784 confirmed sources for which a photometric [Fe/H] value was inferred from the ϕ_{31} Fourier parameter of the G-band light curve. The sources were divided according to the three separate regions (All-Sky, LMC, and SMC) defined in Sect. 2. Green, pink, and blue histograms represent All-Sky, LMC, and SMC variables, respectively. The three distributions peak at mean values of [Fe/H] approximately -1.15 ± 0.6, -1.3 ± 0.7, and -1.6 ± 0.7 dex for the MW, LMC, and SMC RR Lyrae stars, respectively.

Sculptor (blue histogram) and the corresponding spectroscopic metallicities from Clementini et al. (2005; red histogram).

For a more direct comparison in Table 1, we report the mean metallicity and G-band absorption (with the related standard deviations) for a few GCs and dSphs as derived by averaging individual [Fe/H] and $A(G)$ values inferred by the SOS Cep&RRL pipeline from the RR Lyrae stars (Cols. 3, 4, 6, and 7, respectively) and the corresponding literature reference values (Carretta et al. 2009 for the GC metal abundances, Clementini et al. 2005 and Kirby et al. 2011 for the RR Lyrae stars in Sculptor, Draco, UMa I and II, respectively). In the table, N_1 and N_2 are the number of variable stars on which the mean quantities were computed. The overall agreement is quite satisfactory. On the other hand, the rather high value of $A(G)$ derived for Sculptor is likely due to the poor signal-to-noise ratio (S/N) of the G_{BP}, G_{RP} photometry at the faint magnitudes of the RR Lyrae stars in this dSph.

4.1.3. RR Lyrae stars: validation and comparison with the literature

In order to establish the completeness and purity of the RR Lyrae stars confirmed by the SOS Cep&RRL pipeline and to estimate the number of new discoveries by *Gaia,* we performed a deep and careful comparison with the literature. As a first step, the catalogue of 140 784 confirmed sources was cross-matched against all major catalogues of known RR Lyrae stars that are available. We primarily used the OGLE catalogues for RR Lyrae stars (version IV of the survey, Soszyński et al. 2014, 2016), but we also used RR Lyrae stars by CTRS (Drake et al. 2013a,b, 2014, 2017; Torrealba et al. 2015), ASAS (Pojmanski 1997; Richards et al. 2012), ASAS-SN (Jayasinghe et al. 2018), ATLAS (Tonry et al. 2018), IOMC (Alfonso-Garzón et al. 2012), LINEAR (Palaversa et al. 2013), NSVS (Kinemuchi et al. 2006), Pann-Stars (PS1 Sesar et al. 2017), and from the works based on *Kepler*/K2 (Debosscher et al. 2011; Nemec et al. 2011; Molnár et al. 2015a,b, 2016) and on the Simbad database (Wenger et al. 2000). These cross-matches returned a list of 88 578 known RR Lyrae stars in our sample of 140 784 stars. The SOS Cep&RRL confirmed RR Lyrae stars were also cross-matched against catalogues of candidate RR Lyrae stars discovered by the VVV survey (Gran et al. 2016; Minniti et al. 2017; D. Minniti, priv. comm.) in the MW disc and bulge. This returned 319 VVV cross-identified sources in the MW disc and 222 in the MW bulge. We thus confirm these VVV candidates. For known RR Lyrae stars in GCs, the main reference was the catalogue of Clement et al. (2001), which was updated to the latest literature as described in Garofalo et al. (in prep.). For variables in dSphs, we used the following references: Kaluzny et al. (1995), Clementini et al. (2005), Kinemuchi et al. (2008), Dall'Ora et al. (2012) and Garofalo et al. (2013). These latter cross-matches returned a list of 1986 further known RR Lyrae stars. At the end of this cross-match procedure, of the 140 784 RR Lyrae stars that are confirmed by the SOS Cep&RRL pipeline, 90 564 were shown to be known previously, and 50 220 are new discoveries by *Gaia*.

A detailed confusion matrix was derived only for the RR Lyrae stars in the Magellanic Clouds and the MW bulge because only for these variables do we have a catalogue available as reference that is as complete and pure as that of OGLE in the LMC, SMC, and MW bulge. The confusion matrix is shown in Fig. 25. For the Magellanic Cloud RR Lyrae stars we have <0.1% contamination. We achieved this high purity because binaries were removed during the validation process either using the amplitude ratios or by cross-matching against binaries and ellipsoidals from the OGLE catalogues.

A more general assessment of the contamination that affects the SOS-confirmed RR Lyrae sample was achieved through the procedure described in Holl et al. (2018), which consisted of the visual inspection of random and sky-uniform samples of SOS-confirmed RR Lyrae stars, each composed of 500 sources without overlap between the two groups. Based on this procedure, we estimate a contamination of about 9%, of which 4% is due to faint stars ($20.0 < G < 20.7$ mag; see Holl et al. 2018, for details). We note that of the 140 784 RR Lyrae stars that are confirmed by the SOS pipeline, 8306 are fainter than $G \sim 20.0$ mag. The colours for these sources are less reliable, hence misclassifications and contamination by other types of variable sources are definitely possible. On the other hand, of these 8306 faint sources, 16 are bona fide RR Lyrae stars in Sculptor and 637 are OGLE-confirmed RR Lyrae in the two Magellanic Clouds. This reduces the number of possible faint misclassifications to 7653.

The number of epochs available in the G-band light curves of the 140 784 RR Lyrae stars confirmed by the SOS Cep&RRL pipeline is shown in Fig. 26. The distribution of the sources on sky is shown in Figs. 27 and 28, and Figs. 29 and 30 present their distribution on sky according to the metallicity and G-band absorption derived by the SOS pipeline.

4.2. Results for Cepheids

Examples of G, G_{BP}, and G_{RP} light curves for known and new Cepheids confirmed by the SOS Cep&RRL pipeline and released in *Gaia* DR2 are presented in Figs. 31 (known All-Sky Cepheids), 32 (known LMC and SMC Cepheids), and 33 (new All-Sky Cepheids). Light curves are folded according to period and epoch of maximum light determined by the SOS Cep&RRL pipeline.

Figures 34 and 35 show the CMDs in apparent G magnitude not corrected for reddening of confirmed Cepheids in the LMC and SMC, and Fig. 36 shows the CMD in absolute G magnitude (M_G) not corrected for extinction of the confirmed All-Sky Cepheids (only objects with positive parallax can be plotted). An

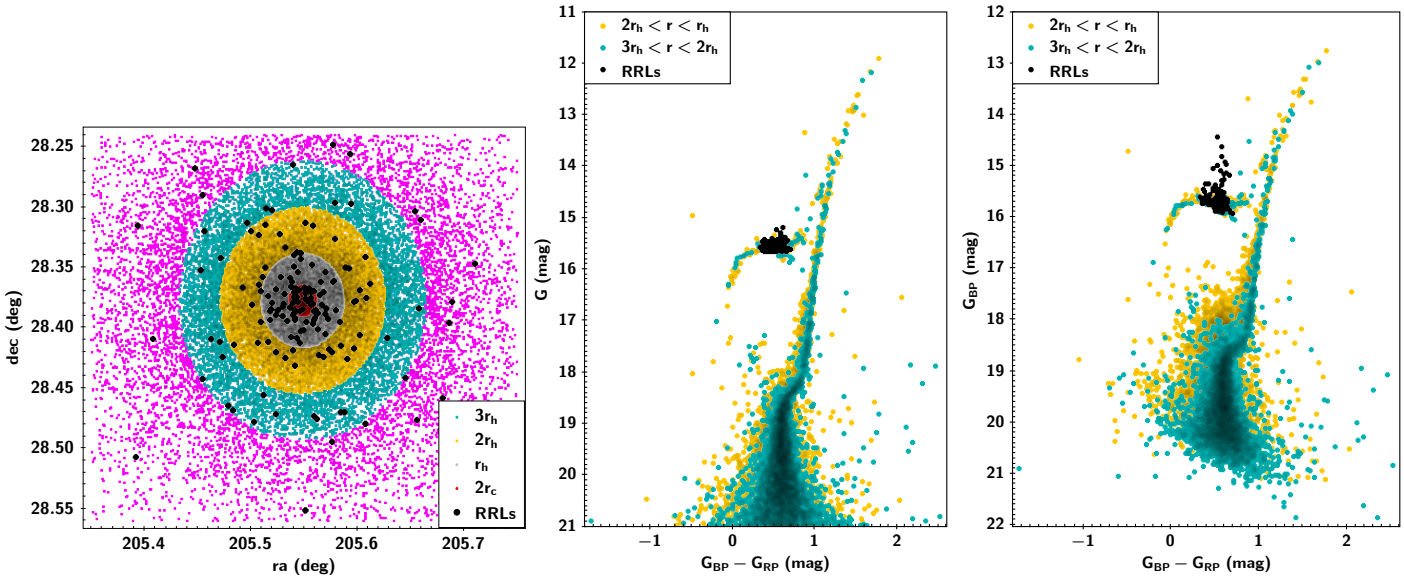

Fig. 18. *Left panel:* map of *Gaia* DR2 stars in an area of 0.31 × 0.30 degree^2 centred on the Galactic GC M3 (NGC 5272). Different colours highlight stars within twice the cluster r_c (red filled circles), once (grey filled circles), twice (yellow filled cicles), three times (cyan filled circles) the cluster half-light radius ($r_h = 2.31$ arcmin), and beyond three times r_h (magenta filled circles). The RR Lyrae stars observed by *Gaia* in the cluster (159 of the 222 RR Lyrae stars known in M3) are marked as large black filled circles. The M3 centre coordinates and r_c, r_h values are taken from Harris (1996). *Centre panel:* G vs. G_{BP}–G_{RP} CMD for stars within twice the M3 r_h. The colour-coding is the same as in the left panel. *Right panel:* same as in the centre panel, but for the cluster G_{BP} vs. G_{BP}–G_{RP} CMD.

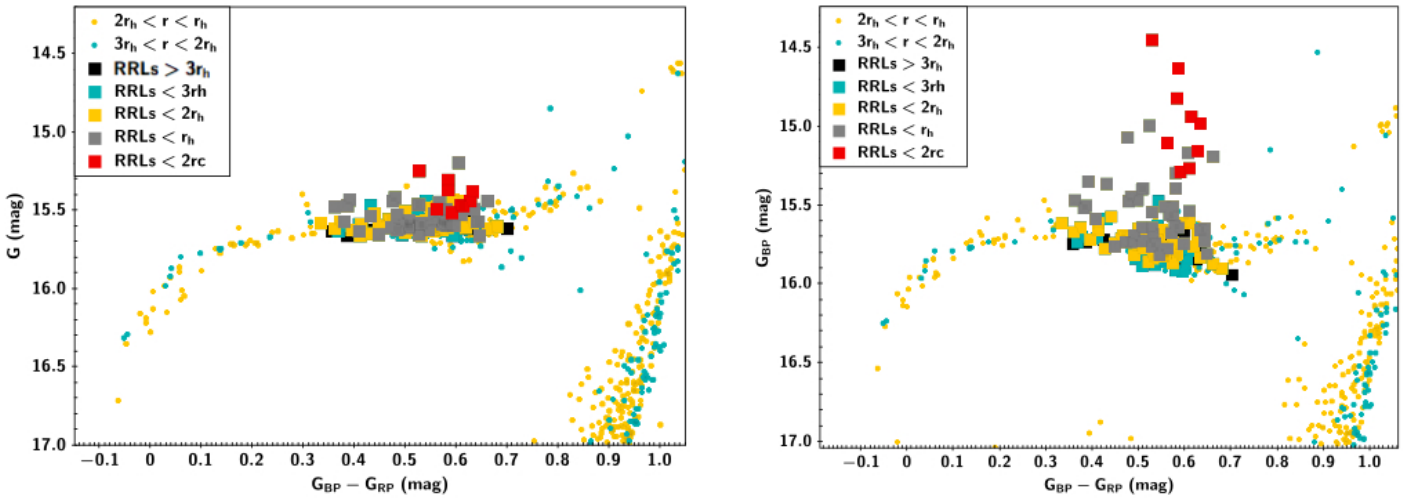

Fig. 19. *Left panel:* zoom of the HB region in the G vs G_{BP}–G_{RP} CMD of M3. RR Lyrae stars are plotted as filled squares and with different colours according to their position with respect to the cluster centre. The same colour-coding is adopted as in the left panel of Fig.18, and black filled squares show the variables beyond three times r_h. *Right panel:* same as in the left panel, but for the G_{BP} vs. G_{BP}–G_{RP} CMD.

inspection of Figs. 34 and 35 shows the expected behaviour of DCEP F and 1O variables (the first are redder, and the latter are hotter), as well as of ACEPs and T2CEPs, in order of increasing (apparent) luminosity passing from BLHER to RVTAU objects. This neat behaviour is less well visible in the All-Sky sample displayed in Fig. 36. This is due to the likely contamination by other types of variables especially for absolute G magnitudes $\gtrsim$2 mag.

Figures 37 and 38 show the distributions of the confirmed Cepheids in the ϕ_{21}, R_{21}, ϕ_{31}, and R_{31} versus period planes. These figures display a very good separation between DCEP F and 1O variables, especially in the R_{21}-period plane, where the two modes have significantly different locations. As with the CMDs, the different diagrams appear more mixed in the case of All-Sky Cepheids.

Figure 39 presents the spatial distribution of the bona fide Cepheids released in *Gaia* DR2 (about 8900 sources in total) after cleaning the sample of All-Sky Cepheids from other types of variable objects not following the PL and PW relations (see Sect. 3.2). Finally, Fig. 40 presents the metallicity distributions of 3738 fundamental-mode classical Cepheids with periods shorter than 6.3 days published in DR2. Individual metallicities for these stars were estimated from the R_{21} and R_{31} Fourier parameters of the light curves according to the procedure described in Sect. 2.1.1. The sources are divided into LMC (red histogram in the top panel), SMC (green histogram in the middle panel), and All-Sky (blue histogram in the bottom panel). Their distributions have median values of [Fe/H] approximately −0.2, −0.1, and 0.0 dex for the SMC, LMC, and All-Sky samples,

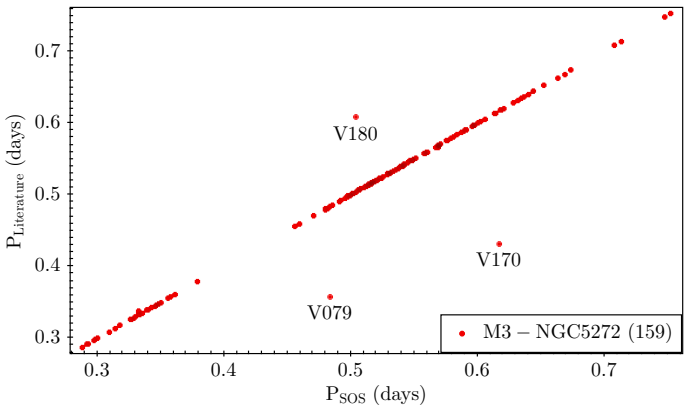

Fig. 20. Comparison between periods measured by the SOS Cep&RRL pipeline and literature values for 159 RR Lyrae stars observed in M3 by *Gaia*. The three deviating objects are variable stars affected by the Blazhko effect (Blazhko 1907).

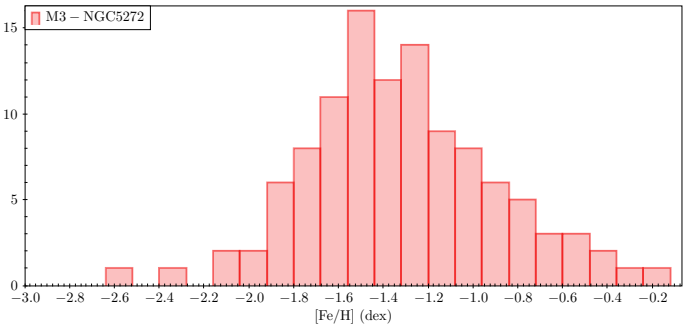

Fig. 21. Metallicity distribution of the RR Lyrae stars observed by *Gaia* in M3, based on the individual measurements from the SOS Cep&RRL pipeline.

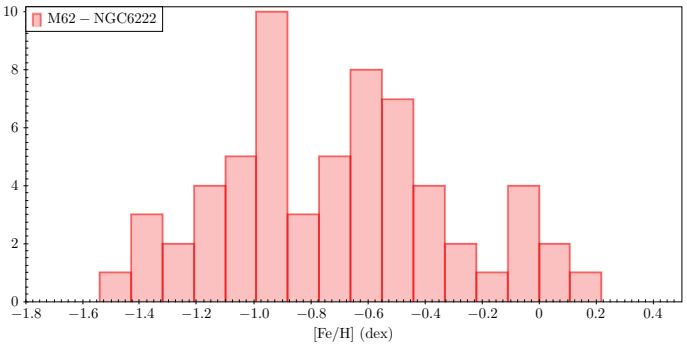

Fig. 22. Same as in Fig. 21, but for RR Lyrae stars in M62.

respectively. The yellow histogram in the bottom panel highlights 235 fundamental-mode classical Cepheids in the All-Sky sample that are known in the literature. They are indistinguishable from the total sample (blue histogram). We caution potential users of these metallicity values that while the median metallicity of the All-Sky sample is consistent with the literature values, there seems to be a shift of approximately +0.2 dex in the LMC and SMC distributions, which might be a hint of calibration issues in the equations of Klagyivik et al. (2013). These equations are mainly based on Galactic Cepheids that do not cover the metal-poorer regime typical of the Magellanic Cloud Cepheids.

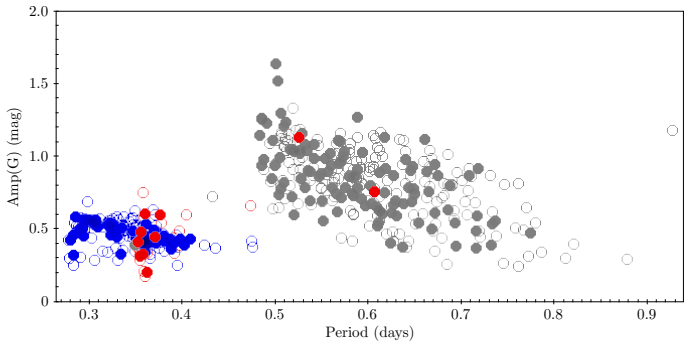

Fig. 23. PA diagram in the *G* band of known (filled circles; Kaluzny et al. 1995) and new RR Lyrae stars observed by *Gaia* in the Sculptor dSph. Blue, red, and grey symbols are RRc, RRd, and RRab stars, respectively.

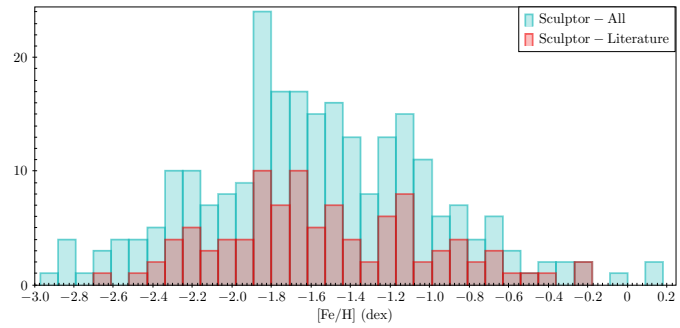

Fig. 24. Metallicity distribution of RR Lyrae stars in Sculptor. The blue histogram is obtained from metallicities measured by the SOS Cep&RRL pipeline. The red histogram shows the metallicity distribution of the RR Lyrae stars that were studied spectroscopically by Clementini et al. (2005).

Gaia SOS

OGLE	RRab	RRc	RRd	Sum
RRab	25779 98.3%	232 0.9%	206 0.8%	26217 100%
RRc	163 1.6%	9725 97.3%	105 1.0%	9993 100%
RRd	159 9.1%	1040 60.0%	532 30.7%	1731 100%

Fig. 25. Confusion matrix for the RR Lyrae stars. As control sample we used all the variable stars classified as RR Lyrae by the OGLE survey in the LMC, SMC, and MW bulge that have a cross-match within a radius of 3 arcsec with the SOS-confirmed RR Lyrae stars published in *Gaia* DR2 for a total of 37 941 objects. Rows refer to literature results and columns to results of the SOS Cep&RRL pipeline. The corresponding success percentage is shown in the diagonal cells.

4.2.1. Cepheids: validation and comparison with the literature

As with the RR Lyrae stars, a confusion matrix was derived only for the Cepheids in the Magellanic Clouds (Fig. 41) because only in the Magellanic Clouds do we have complete and homogeneous reference catalogues for ACEPs, DCEPs, and T2CEPs published by the OGLE group (Soszyński et al. 2008, 2010, 2015b,a). An inspection of Fig. 41 reveals a very low (almost 0%)

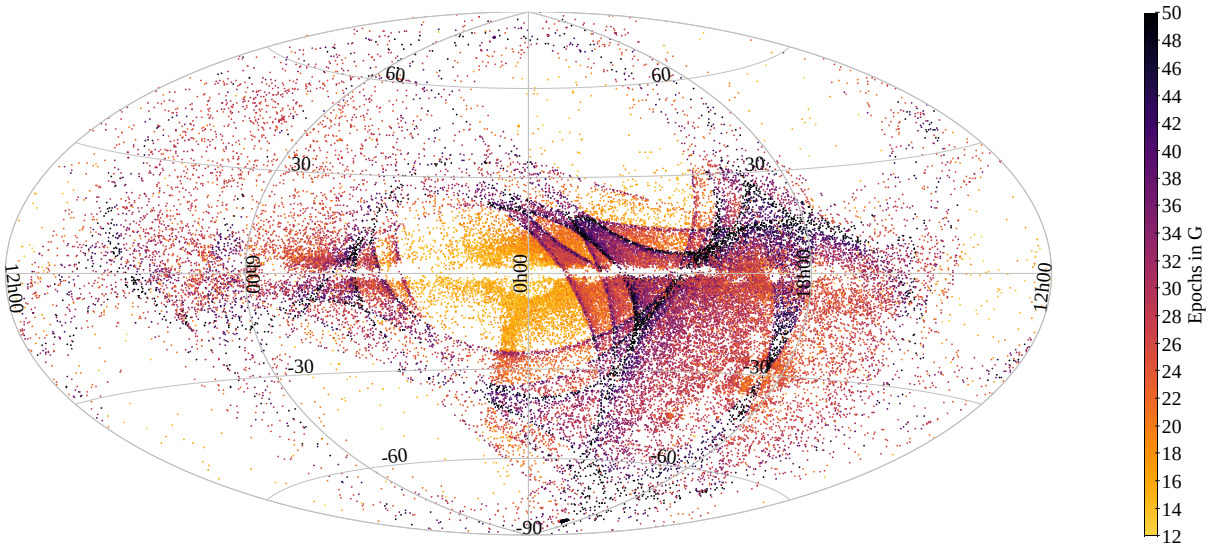

Fig. 26. Map in galactic coordinates showing RR Lyrae stars that are new discoveries by *Gaia* and that are confirmed by the SOS Cep&RRL pipeline, colour-coded according to the number of G-band epochs available for each source.

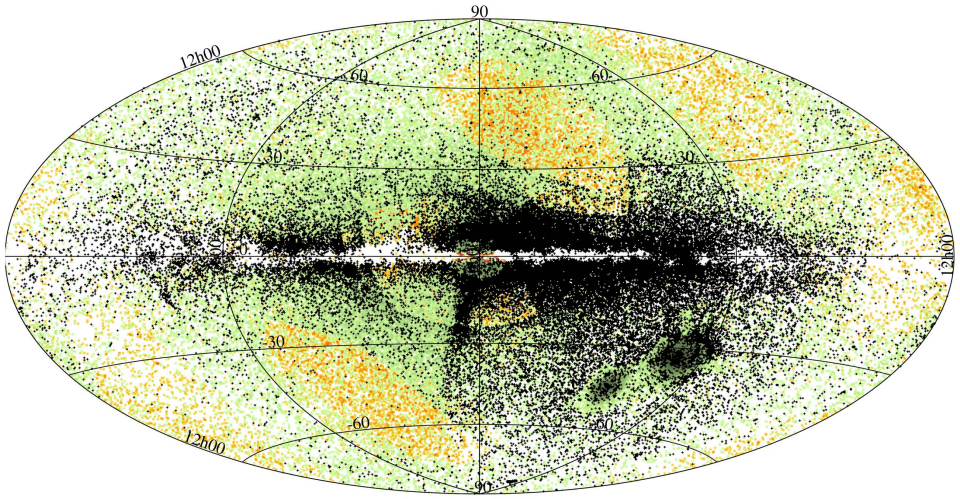

Fig. 27. Sky distribution in galactic coordinates of previously known and RR Lyrae stars that are new discoveries by *Gaia* and that are confirmed by the SOS Cep&RRL pipeline. Orange points: known RR Lyrae stars that do not have a counterpart among the SOS-confirmed RR Lyrae stars published in DR2. Green points: known RR Lyrae stars that are cross-matched with SOS-confirmed RR Lyrae stars. Black points: new RR Lyrae detected by *Gaia* and confirmed by the SOS Cep&RRL pipeline. The green and black points clearly reflect the pattern of the *Gaia* scanning law. More than 220 000 RR Lyrae stars are shown in the figure, of which 46 443 are in the Magellanic Clouds, 2860 are in GCs, 984 in classical dSphs (885) and ultra-faint dwarfs (99; Garofalo et al., in prep.), and 50 220 are new discoveries by *Gaia*. To avoid further overcrowding, we did not highlight GCs and dSphs, but refer to Figs. 45 and 46 for the most complete post-*Gaia* DR2 view of All-Sky RR Lyrae stars down to the *Gaia* faint-magnitude limit of $G \sim 20.7$ mag.

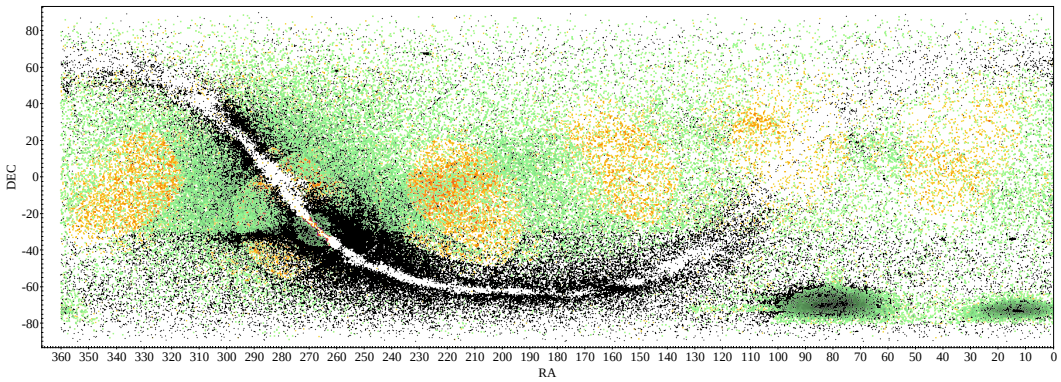

Fig. 28. Same as in Fig. 27, but in equatorial coordinates.

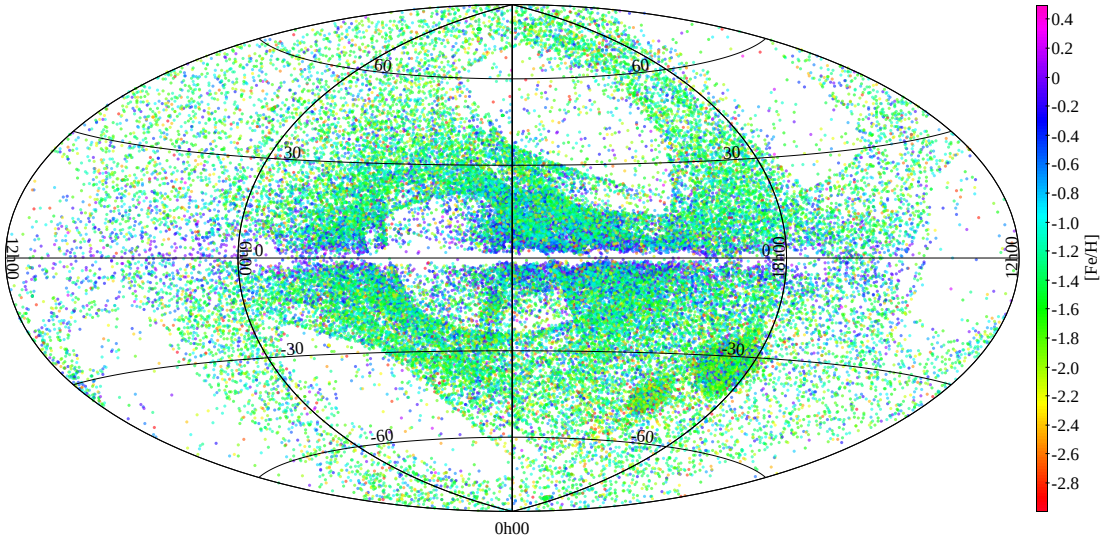

Fig. 29. Map in galactic coordinates of 64 932 RR Lyrae stars for which a photometric [Fe/H] metallicity was inferred using the ϕ_{31} parameter in the Fourier decomposition of the G-band light curve. The map is colour-coded according to the source metallicity.

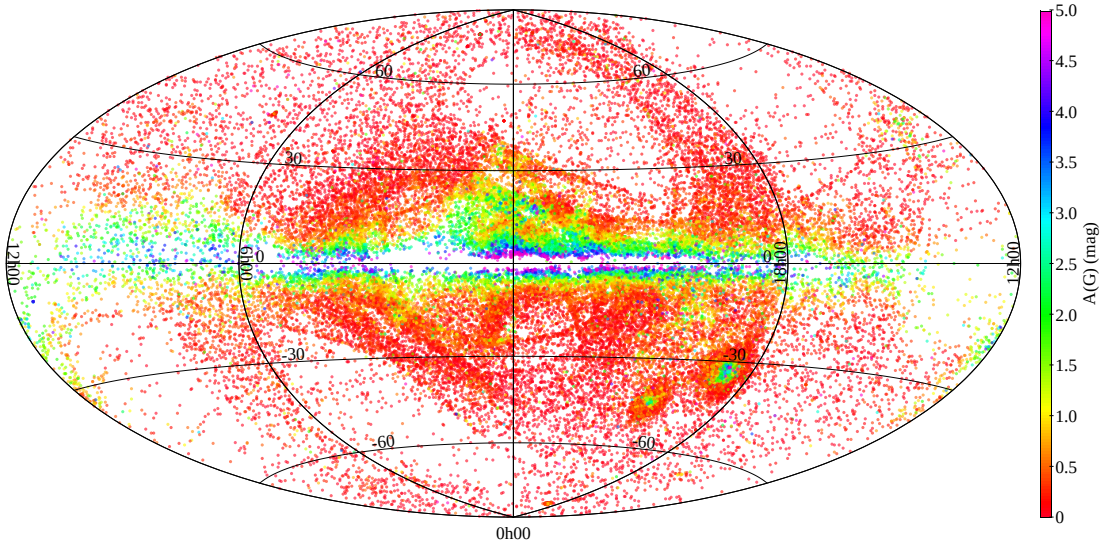

Fig. 30. Map in galactic coordinates of 54 272 fundamental-mode RR Lyrae stars with absorption in the G-band [$A(G)$] derived from an empirical relation that connects the amplitude of the light variation in the G-band [Amp(G)] and the stellar pulsation period (see text for details). The map is colour-coded according to the stellar $A(G)$ value.

contamination level for all Cepheid types and modes, with the exception of the multi-mode Cepheids, for which we have more than 50% false positives, mostly among ordinary DCEP 1O. This is a common problem for double-mode RR Lyrae stars as well. It will be mitigated in next *Gaia* releases by improvements of the SOS pipeline and a natural increase of the number of epochs in the light curves, which is a fundamental ingredient for a successful recovery of multi-mode pulsators. By applying the same procedure to Cepheids as for the RR Lyrae stars (see Sect. 4.1.3 and Holl et al. 2018), we estimate a contamination of about 5% in regions of the sky that extend partially beyond OGLE-IV footprint.

Finally, we note that a small number of new Cepheids, 118 in total, and 1640 new RR Lyrae stars were identified in the LMC and SMC areas that are extensively monitored by the OGLE survey. While we are quite confident that they are bona fide new Cepheids and RR Lyrae stars as they did pass all diagnostics and a very careful visual inspection, we would be indebted to our colleagues of the OGLE team if they could verify the reliability of these sources with their OGLE photometry.

5. Results and final accounting

The run of the SOS Cep&RRL pipeline on the candidate RR Lyrae stars and Cepheids, combined with extensive validation procedures and random visual inspection of some of the resulting light curves, produced final samples of 140 784 confirmed RR Lyrae stars and 9575 Cepheids for a total of 150 359 sources. Position, G, $G_{\rm BP}$, and $G_{\rm RP}$ time-series photometry, and final results of the SOS Cep&RRL processing are published in *Gaia* DR2 for all these 150 359 sources.

The subdivision of the 150 359 sources according to type, subtype, and pulsation mode is summarised in Table 2, while

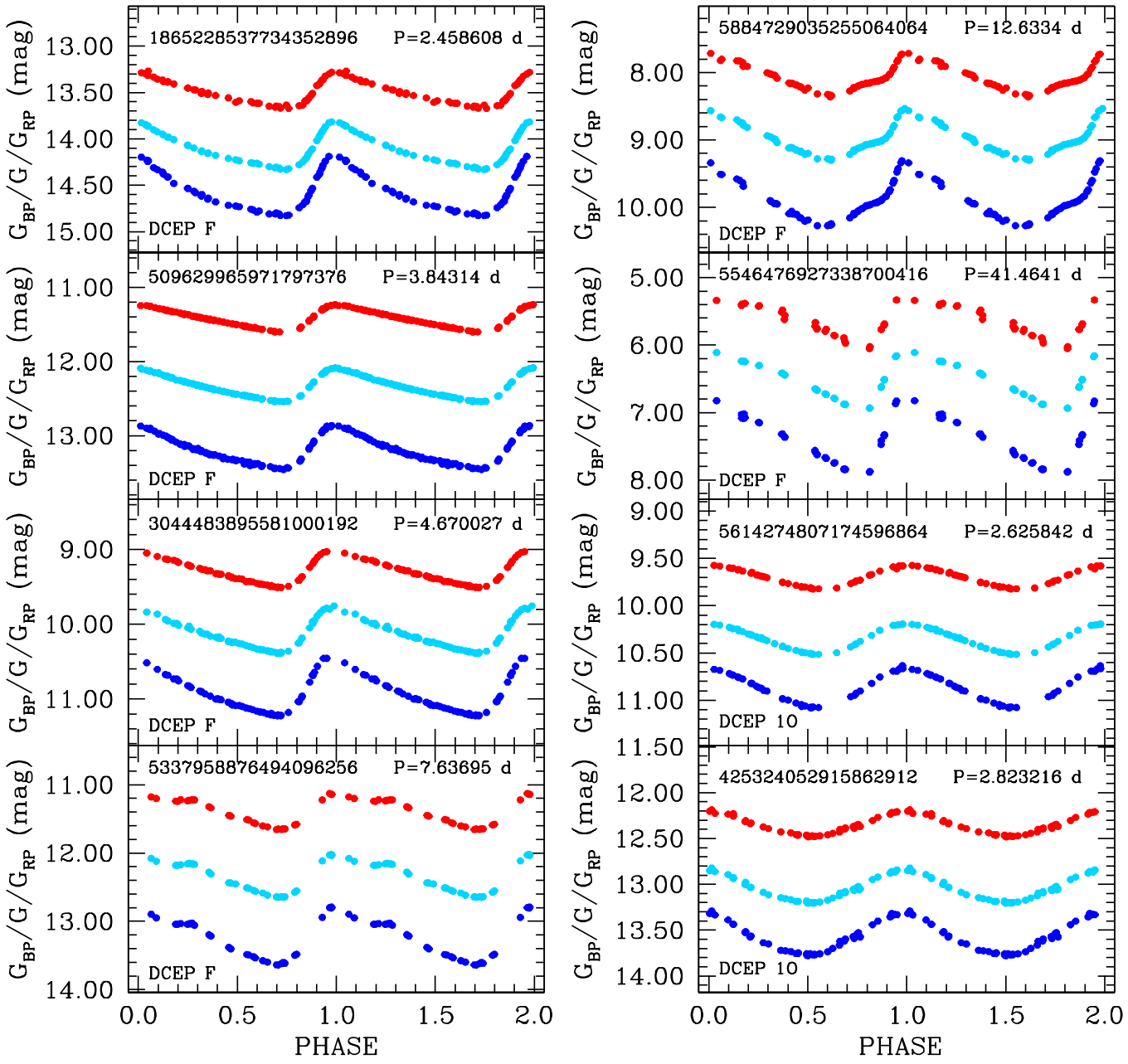

Fig. 31. *G* (cyan), G_{BP} (blue), and G_{RP} (red) light curves of All-Sky known classical Cepheids of different pulsation modes released in *Gaia* DR2. The multi-band time-series data are folded according to the period and epoch of maximum light derived by the SOS Cep&RRL pipeline. Error bars are comparable to or smaller than the symbol size.

Table 3 provides the subdivision between known and new discoveries by *Gaia*.

The following parameters, computed by the SOS Cep&RRL pipeline, have been released in *Gaia* DR2 for these 140 784 RR Lyrae stars and 9575 Cepheids along with the related errors:

- source pulsation periods (main and secondary periodicities, if any);
- intensity-averaged mean G, G_{BP}, G_{RP} magnitudes;
- epochs of maximum light in the three passbands;
- ϕ_{21} and R_{21} Fourier parameters;
- ϕ_{31} and R_{31} Fourier parameters;
- peak-to-peak G, G_{BP}, and G_{RP} amplitudes [Amp(G)], [Amp(G_{BP})], and [Amp(G_{RP})];
- RR Lyrae star subclassification into RRab, RRc, and RRd types;
- Cepheid classification into DCEP, ACEP, and T2CEP types;
- DCEPs and ACEPs pulsation mode (F, 1O, DCEP MULTI);
- T2CEPs subclassification into BLHER, WVIR, and RVTAU types;
- absorption in the G band, $A(G)$, of RRab stars;
- photometric metallicity, [Fe/H], for RRab and RRc stars;
- photometric metallicity, [Fe/H], for fundamental-mode DCEPs with $P < 6.3$ days.

The *Gaia* sourceids, coordinates, values of the above quantities, and associated statistics, along with the G, G_{BP}, and G_{RP} time-series photometry for each of the 140 784 RR Lyrae stars and 9575 Cepheids that are confirmed and characterised by the

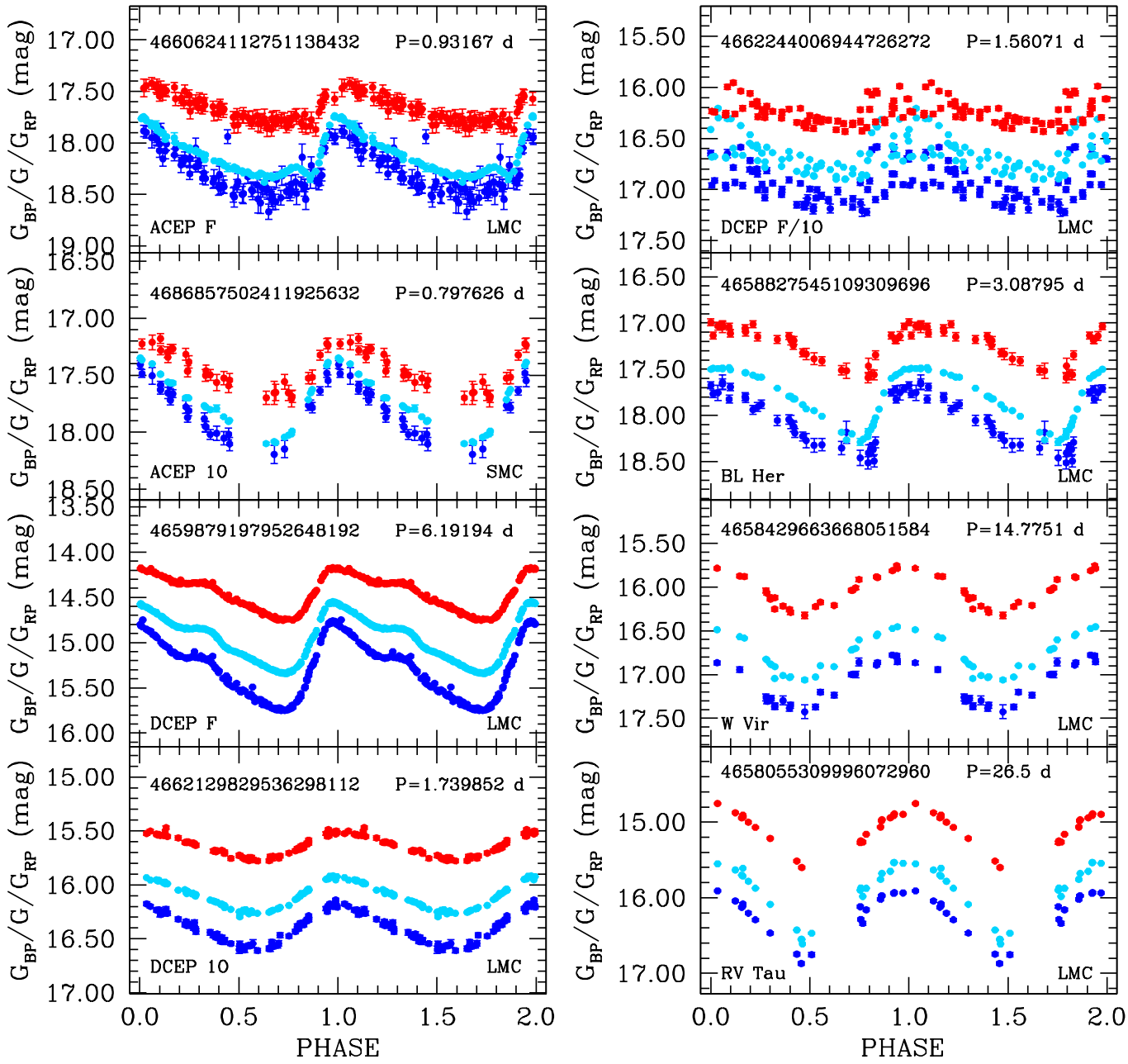

Fig. 32. G (cyan), G_{BP} (blue), and G_{RP} (red) light curves of known Cepheids of different types and pulsation modes released in *Gaia* DR2. All of them are in the LMC, except for the ACEP 1O, which is in the SMC, as labelled. The multi-band time-series data are folded according to the period and epoch of maximum light derived by the SOS Cep&RRL pipeline. Error bars are comparable to or smaller than the symbol size, except for the G_{BP} and G_{RP} light curves of the two ACEPs in the *upper left panels*, as expected, which is due to the faintness of these two Magellanic Cloud variable stars.

SOS Cep&RRL pipeline can be retrieved from the *Gaia* data release archive[11] and other distribution nodes. The archives also provide tools for queries and for cross-matching *Gaia* data with other catalogues in the literature.

We provide in Tables 4 and 5 specific links to the archive tables, and summarise the names of the parameters computed by SOS Cep&RRL that can be retrieved from the archive tables. In Appendix A we provide examples of queries to retrieve some of the quantities and parameters listed in Tables 4 and 5.

[11] http://archives.esac.esa.int/gaia/

5.1. Sky maps

Figures 42 and 43 show sky maps of the SOS-confirmed RR Lyrae stars released in DR2 in the region of the Magellanic Clouds. In the latter map, variables are colour-coded according to their apparent magnitude. Figure 44 shows the same region of the sky as mapped by the SOS-confirmed Cepheids. The rather smooth distribution of RR Lyrae stars around the two Clouds (Fig. 43) is remarkable. It very nicely reproduces the shape of the far-extended halo that surrounds the LMC.

Figures 45 and 46 show the distribution on sky in galactic and equatorial coordinates of RR Lyrae stars within the limiting magnitude of *Gaia* (orange points). The map combines

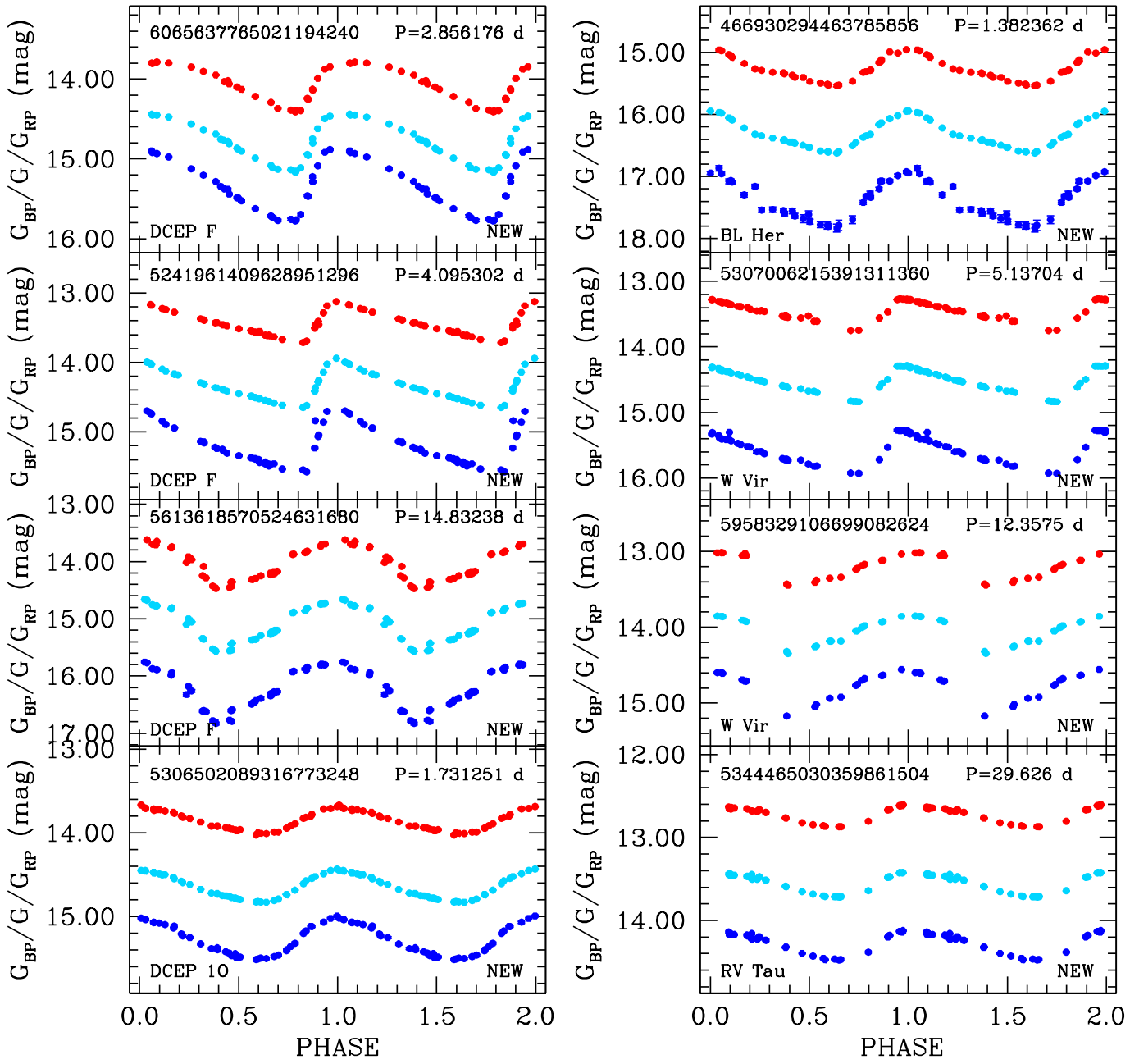

Fig. 33. *Left panels:* G (cyan), G_{BP} (blue), and G_{RP} (red) light curves of All-Sky new classical Cepheids of different pulsation modes released in *Gaia* DR2. *Right panels:* same as in the left panels, but for new Type II Cepheids of different types. The multi-band time-series data are folded according to the period and epoch of maximum light derived by the SOS Cep&RRL pipeline. Error bars are comparable to or smaller than the symbol size.

known literature (with and without a *Gaia* counterpart) and new RR Lyrae stars discovered by *Gaia* and confirmed by the SOS Cep&RRL pipeline for more than 223 000 RR Lyrae stars in total. This number favourably compares with estimates of the total number RR Lyrae stars in Holl et al. (2018) and Rimoldini et al. (2018). Blue filled dots and magenta filled squares indicate 87 GCs and 12 dSphs (classical and ultra-faint) in which *Gaia* observed SOS-confirmed RR Lyrae stars that are published in DR2. This figure presents a post-DR2 update of Fig. 4 in Clementini et al. (2018). It displays the largest-ever census of RR Lyrae stars in our Galaxy and its close companions.

5.2. Limitations of the SOS Cep&RRL results for DR2

The catalogues of SOS Cep&RRL confirmed Cepheids and RR Lyrae stars released in *Gaia* DR2 have some limitations. The main problem is incompleteness, which is mainly due to the still small number of epochs available for a significant fraction of the sky (see e.g. Fig. 26) according to the *Gaia* scanning law. Furthermore, we recall that (i) objects with fewer than 12 epoch data were not processed by the SOS Cep&RRL pipeline and (ii) a sizeable sample of true pulsators with fewer than ~20 epochs were rejected because the small number of observations and the uneven sampling of the time-series data

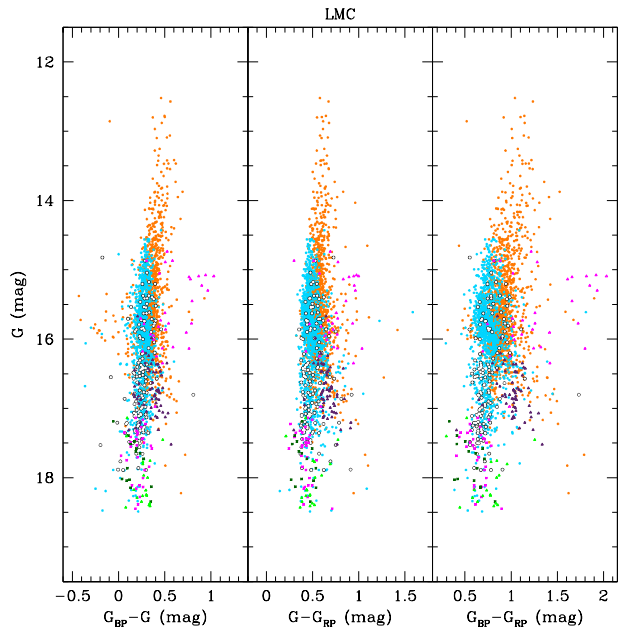

Fig. 34. G, G_{BP}–G; G, G–G_{RP} ; and G, G_{BP}–G_{RP} CMDs in apparent magnitude of Cepheids in the LMC published in *Gaia* DR2. The symbols and colour-coding are the same as in Fig. 6.

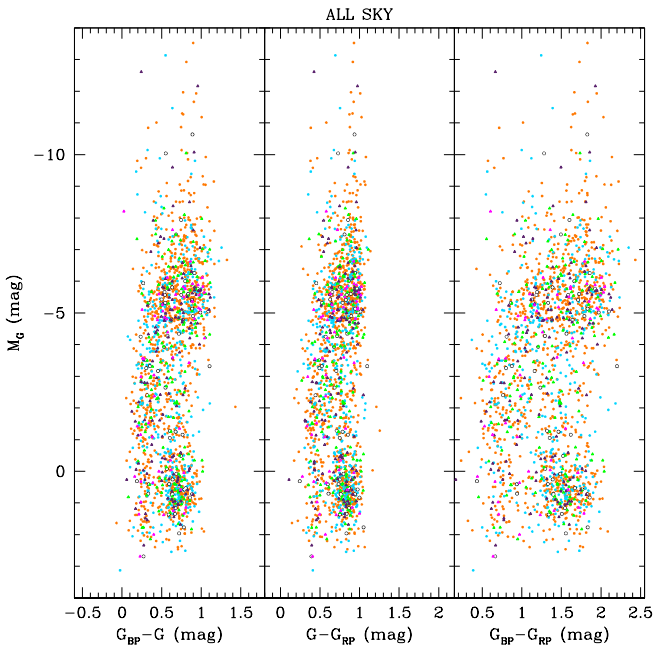

Fig. 36. Same as in Fig. 34, but in absolute G magnitude (M_G) for All-Sky Cepheids.

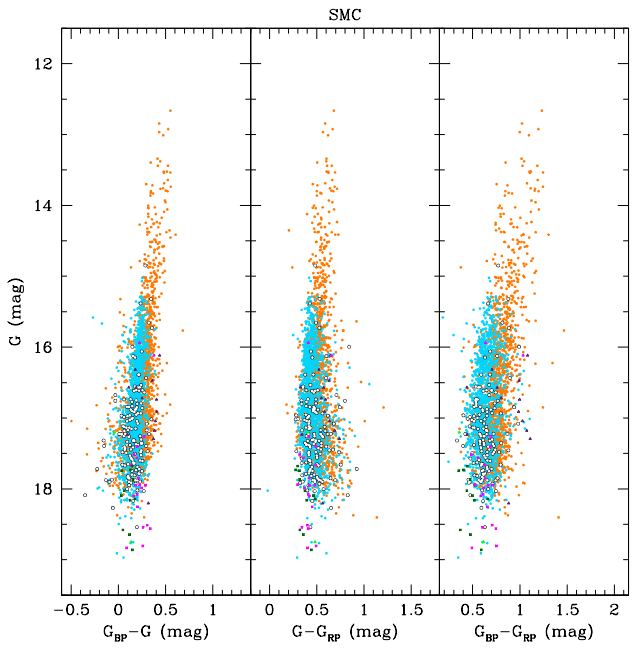

Fig. 35. Same as in Fig. 34, but for Cepheids in the SMC.

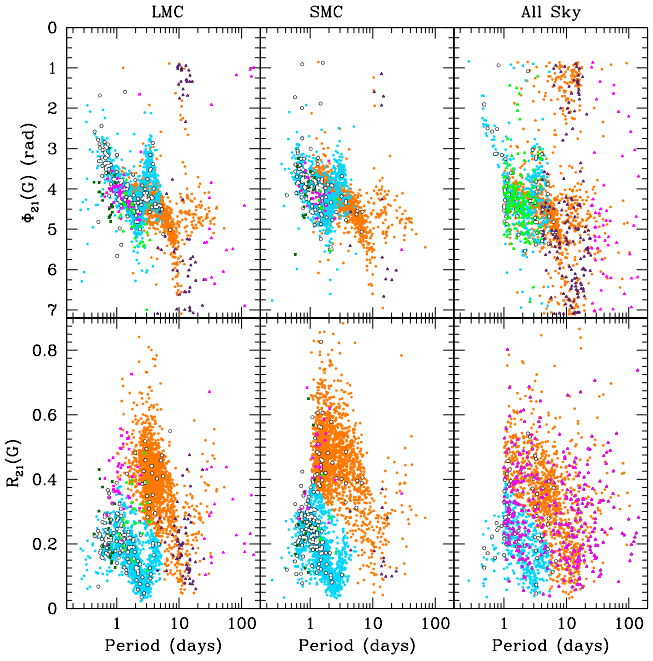

Fig. 37. *Upper panels*: G-band ϕ_{21} vs. period diagram for Cepheids in the LMC (upper left), SMC (middle), and All-Sky (upper right). *Lower panels*: same as in the upper panels, but for the G-band R_{21} vs. period diagram. The symbols and colour-coding are the same as in Fig. 6.

caused by the scanning law resulted in an incorrect period determination. Hence, completeness is a strong function of the source position on the sky. This can be easily checked by comparing the DR2 SOS Cep&RRL RR Lyrae catalogue to OGLE data for the LMC, SMC, and bulge. In these systems our recovery percentages are 67, 82, and 15%, respectively. Taking into account that the OGLE data are complete at more or less the same magnitude as the *Gaia* data, that the typical magnitudes of the RR Lyrae variables in the Magellanic Clouds are generally fainter than in the bulge, and that the level of contamination is similar, it is clear that the very low recovery rate in the bulge is almost entirely due to the lack of enough epoch data for the bulge region due to the *Gaia* scanning law. Similar considerations can be drawn for the rest of the sky (mainly the MW halo), even though for the Galactic halo sky zones rich and poor in number of epochs are more or less the same, hence giving in the end an indication of the average completeness of the DR2 catalogue. For example, the comparison with the ATLAS and ASAS-SN surveys returned a recovery rate of 66 and 64%, respectively. A

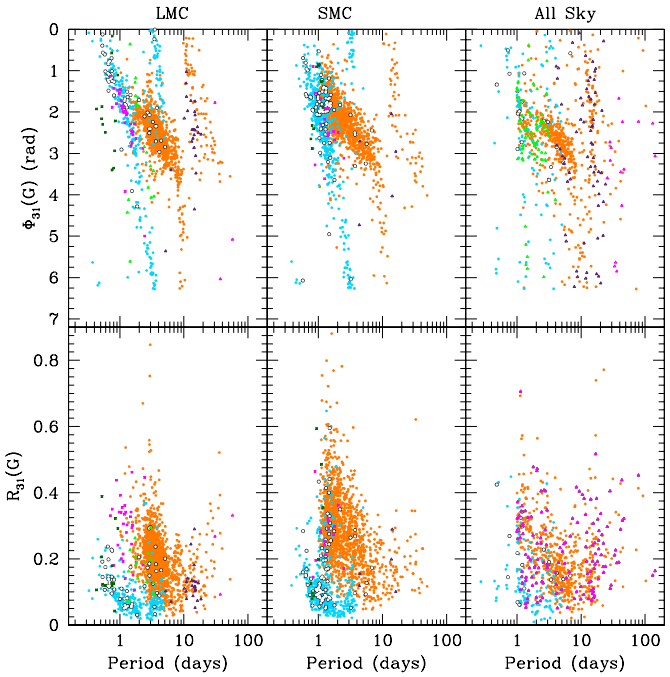

Fig. 38. *Upper panels*: *G*-band ϕ_{31} vs. period diagram for Cepheids in the LMC (*left*), SMC (*middle*), and All-Sky (*right*). *Lower panels*: same as in the upper panels, but for the *G*-band R_{31} vs. period diagram. The symbols and colour-coding are the same as in Fig. 6.

Table 2. Number and type/mode classification of RR Lyrae stars and Cepheids confirmed by the SOS Cep&RRL pipeline that are published in *Gaia* DR2, broken down into LMC, SMC, and All-Sky regions of the sky.

Type	LMC	SMC	All-Sky	Total
RRab	20 264	4 629	73 133	98 026
RRc	8 395	1 213	30 772	40 380
RRd	693	164	1 521	2 378
RR Lyrae Total	29 352	6 006	105 426	140 784
DCEP F	1 925	1 915	1 158	4 998
DCEP 1O	1 461	1 540	471	3 472
DCEP MULTI	175	177	68	420
DCEP Total	3 561	3 632	1697	8 890
ACEP F	43	23	–	66
ACEP 1O	21	13	–	34
ACEP Total	64	36		100
T2CEP BLHER	35	6	182	223
T2CEP WVIR	73	15	165	253
T2CEP RVTAU	34	3	72	109
T2CEP Total	142	24	419	585
Cepheid Total	3 767	3 692	2 116	9 575

Notes. Counts for Cepheids correspond to the number of sources in the LMC, SMC, and All-Sky regions defined in Sect. 2. Counts for the RR Lyrae correspond to the number of sources in regions defined by taking into account where the number density of RR Lyrae stars in the LMC and SMC drops and becomes comparable to the counts in the field.

Table 3. Total numbers of SOS confirmed RR Lyrae stars and Cepheids published in *Gaia* DR2 subdivided into new and known sources.

Type	Grand total	New	Known
RR Lyrae stars	140 784	50 220	90 564
Cepheids	9 575	350	9 225

Notes. "New" sources mean "new to the best of our knowledge".

general comparison with all our literature collection, but excluding OGLE, returns a completeness of about 60%. This is of course only an indicative number, as the literature data are a mixture of different surveys with very different characteristics. We are confident, however, that this estimate is not too far from reality.

The same line of reasoning returns 74, 73, and 3% of recovery in the LMC, SMC, and bulge, respectively, for the Cepheids. While the percentages in the Magellanic Clouds are similar to those of RR Lyrae variables, the percentage in the bulge is very low. This is mainly due to the small number of epochs in this region.

Estimating the recovery percentage throughout the whole MW is rather difficult, since field Cepheids are not easy to detect and there are no systematic and complete surveys over the whole sky to compare with; the *Gaia* survey is indeed the first. To give some rough numbers at least for DCEPs, we compared the SOS Cep&RRL Cepheids published in DR2 with a sample of 417 DCEPs with metallicities measured by Genovali et al. (2013) and found that 68% of the sources are in common. This percentage is in line with what we found for field RR Lyrae variables.

We also note that at the bright extreme of the distribution, we missed some famous RR Lyrae stars and Cepheids. For example, RR Lyr itself is not present in the SOS Cep&RRL RR Lyrae tables because during the validation of the SOS Cep&RRL results, the star was dropped as the value of the ϕ_{21} Fourier parameter placed the star at the limit of a region of heavy contamination ($\phi_{21} \sim 3.14$ rad) where sources were automatically rejected. Although not present in the DR2 variability tables, RR Lyr is present in the DR2 photometry and astrometry catalogues, but with values for magnitudes and astrometry that are to be taken with caution (Arenou et al. 2018 and Gaia Collaboration 2018).

Finally, a comparison of the DR2 SOS Cep&RRL confirmed RR Lyrae stars and Cepheids with the sources released in *Gaia* DR1 show that of the 2595 RR Lyrae stars published in DR1, 2517 are also confirmed in DR2, and 78 are missing. Of the 599 DR1 Cepheids, 533 are also confirmed in DR2 and 66 are missing. This can be due to several reasons. For instance, the increased number of epoch data available for DR2 along with modifications of the general variability pipeline before the SOS processing and updates of the SOS Cep&RRL pipeline (see Sect. 2) may have resulted in a different period estimate, and in turn, in a different position in the diagrams we use for the classification. Specifically, of the 78 missing DR1 RR Lyrae stars, 23 were no longer classified as RR Lyrae by the classifiers in DR2, hence never made it to the SOS Cep&RRL pipeline (11 of them are bona fide RR Lyrae stars known from OGLE), and 55 were rejected during the SOS processing. Of these 55, 49 are bona fide RR Lyrae stars known from OGLE, and 6 are DR1 RR Lyrae that hence are not confirmed. Of the 66 missing DR1 Cepheids, 31 never made it to the SOS pipeline (3 of them are Cepheids confirmed by OGLE), and of the remaining 35, 28 are Cepheids

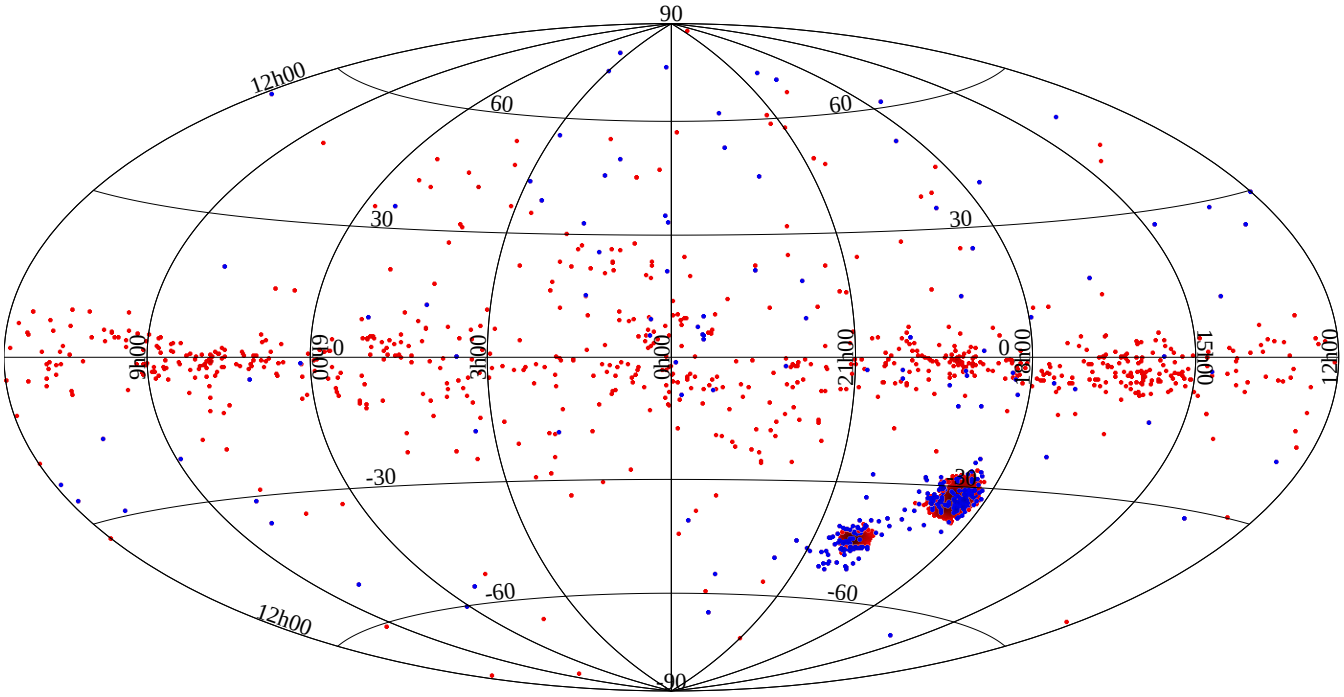

Fig. 39. Spatial distribution of the bona fide Cepheids released in *Gaia* DR2 (about 8900 sources in total). Red filled circles are known Cepheids in the literature (OGLE and other surveys), and blue filled circles are new Cepheids detected by *Gaia* in the LMC, SMC (118 in total), and All-Sky (about 240 in total). The latter sample has been cleaned from other types of variable objects that do not follow the PL and PW relations, see discussion at the end of Sect. 3.2.

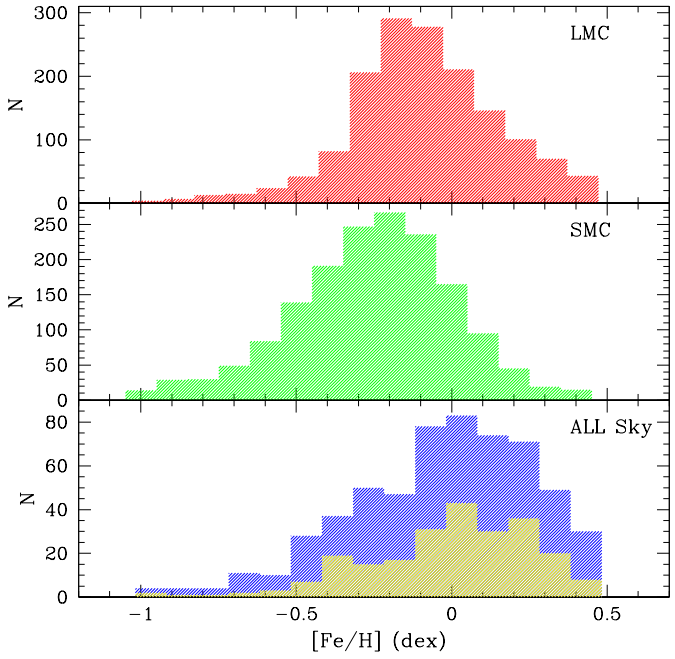

Fig. 40. Metallicity distributions of 3738 fundamental-mode classical Cepheids with periods shorter than 6.3 days published in DR2. *From top to bottom panels*: LMC, SMC, and All-Sky distributions (blue histogram). They have median values of [Fe/H] approximately −0.2, −0.1, and 0.0 dex for the SMC, LMC, and All-Sky samples, respectively. The yellow histogram in the lower panel highlights the metallicity distribution of 235 fundamental-mode classical Cepheids in the All-Sky sample that are known in the literature. They are indistinguishable from the total sample (blue histogram). Ten of these are sources that are located on the faint sequences below the dashed lines in Fig.8.

confirmed by OGLE. Hence, working out the above numbers, the total contamination of the new RR Lyrae stars released in DR1 is lower than 0.7% and that of the DR1 Cepheids is around 5.8%.

After the opening of the *Gaia* DR2 archive on 25 April 2018, we received feedback from the users about RR Lyrae stars and Cepheids that may have been misclassified by the SOS Cep&RRL pipeline. We provide in Appendix C the sourceids of a number of these possibly misclassified sources.

6. Conclusions and future developments

Gaia DR2 represents a significant step forward in our knowledge of the RR Lyrae and Cepheid census in the MW and its close neighbours. The number of RR Lyrae stars confirmed by the SOS Cep&RRL pipeline that have been released in DR2 along with those that are known in the literature provides even now the largest-ever census of RR Lyrae stars in our Galaxy and its close companions. This number will increase with further *Gaia* releases as it is also expected to increase the number and accuracy of the parameters derived for these sources by the SOS Cep&RRL pipeline. In particular, for *Gaia* Data Release 3 (DR3), time-series radial velocities will be processed by the SOS pipeline in addition to the multi-band photometry. This will allow us to improve the source characterisation and will also open the path to deriving additional stellar parameters such as gravities, temperatures, and absolute magnitudes independent of parallaxes.

Several improvements of the SOS Cep&RRL pipeline are being implemented in view of the forthcoming release of *Gaia* (DR3). They include the following:

Gaia SOS

OGLE	ACEP F	ACEP 1O	DCEP F	DCEP 1O	DCEP 2O	DCEP MULTI	T2CEP BL Her	T2CEP W Vir	T2CEP RV Tau	Sum
ACEP F	59 93.6%	4 5.4%								63
ACEP 1O		27 100%								27
DCEP F			3778 93.3%	198 4.9%		75 1.8%				4051
DCEP 1O			32 1.2%	2568 93.0%		161 5.8%				2761
DCEP 2O				17 77.3%	0 0%	5 22.7%				22
DCEP MULTI			10 3.1%	200 62.3%		111 34.6%				321
T2CEP BL Her							24 100%			24
T2CEP W Vir								58 100%		58
T2CEP RV Tau								2 14.3%	12 85.7%	14

Fig. 41. Confusion matrix for Cepheids. As control sample, we used all the variable stars that are classified as ACEP, DCEP, and T2CEP by the OGLE survey in the LMC and SMC that have a cross-match within a radius of 3 arcsec with the Cepheids of the aforementioned types published in *Gaia* DR2, for a total of 7341 objects. Rows refer to literature results and columns to results of the SOS Cep&RRL pipeline. The corresponding success percentage is shown in the diagonal cells.

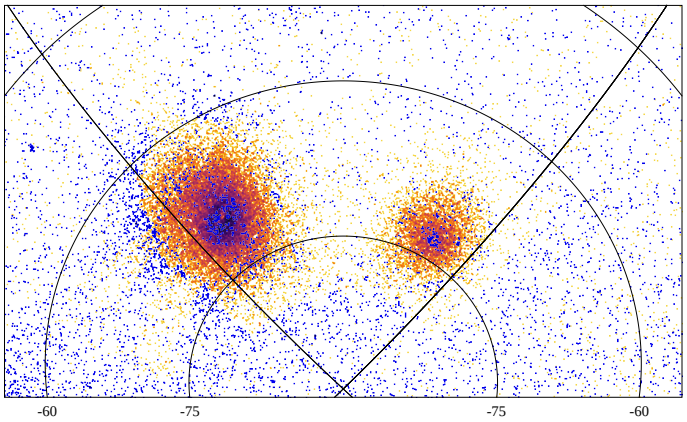

Fig. 42. Spatial distribution of RR Lyrae stars located in the region of the two Magellanic Clouds released in *Gaia* DR2. Orange dots show known RR Lyrae stars from the OGLE surveys, and blue dots represent RR Lyrae stars identified by *Gaia* and confirmed by the SOS Cep&Cep pipeline.

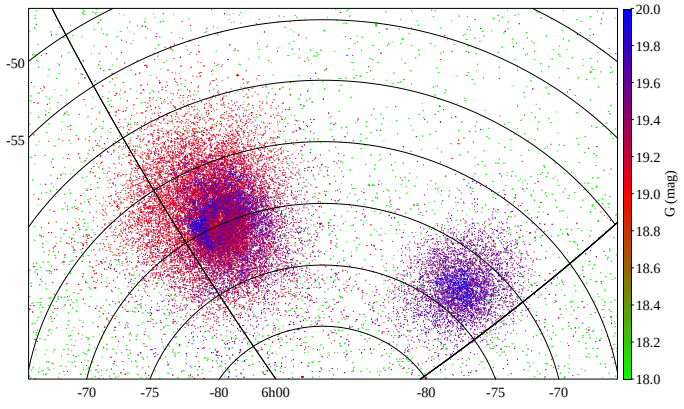

Fig. 43. Same as in Fig. 42, but the RR Lyrae stars are colour-coded according to their apparent magnitude.

1. The relations used to classify and characterise the sources in the SOS Cep&RRL pipeline will be computed directly from the *Gaia* light curves of confirmed Cepheids and RR Lyrae stars.
2. Parts of the SOS Cep&RRL pipeline dedicated to the processing of the radial velocity time series will be activated.
3. Identification and characterisation of double-mode RR Lyrae stars and multi-mode classical Cepheids (F/1O, 1O/2O, etc.) will be activated only for sources with a sufficient number of epochs and the detection algorithm will be

Table 4. Links to *Gaia* archive tables for retrieving the pulsation characteristics: period(s), epochs of maximum light (E), peak-to-peak amplitudes, intensity-averaged mean magnitudes, ϕ_{21}, R_{21}, ϕ_{31}, R_{31} Fourier parameters with related uncertainties, metallicity, and absorption in the G band computed by the SOS Cep&RRL pipeline for the 140 784 RR Lyrae stars confirmed by SOS and released in *Gaia* DR2.

Table URL	http://archives.esac.esa.int/gaia/
RR Lyrae star parameters computed by the SOS Cep&RRL pipeline	
Table name	**gaiadr2.vari_rrlyrae**
Source ID	source_id
Type	best_classification (one of RRC, RRAB or RRD)
$Pf, P1O, P2O, P3O$	p_f, p1_o, p2_o, p3_o: NB p3_o is empty for RRL
$\sigma(Pf, P1O, P2O, P3O)$	pf_error, p1_o_error, p2_o_error, p3_o_error: NB p3_o_error is empty for RRL
E[a](G, G_{BP}, G_{RP})	epoch_g, epoch_bp, epoch_rp
σE(G, G_{BP}, G_{RP})	epoch_g_error, epoch_bp_error, epoch_rp_error
$\langle G\rangle, \langle G_{BP}\rangle, \langle G_{RP}\rangle$	int_average_g, int_average_bp, int_average_rp
$\sigma\langle G\rangle, \sigma\langle G_{BP}\rangle, \sigma\langle G_{RP}\rangle$	int_average_g_error, int_average_bp_error, int_average_rp_error
Amp(G, G_{BP}, G_{RP})	peak_to_peak_g, peak_to_peak_bp, peak_to_peak_rp
σ[Amp(G)], σ[Amp(G_{BP})], σ[Amp(G_{RP})]	peak_to_peak_g_error, peak_to_peak_bp_error, peak_to_peak_rp_error
ϕ_{21}	phi21_g
$\sigma(\phi_{21})$	phi21_g_error
R_{21}	r21_g
$\sigma(R_{21})$	r21_g_error
ϕ_{31}	phi31_g
$\sigma(\phi_{31})$	phi31_g_error
R_{31}	r31_g
$\sigma(R_{31})$	r31_g_error
[Fe/H][b]	metallicity
σ ([Fe/H])	metallicity_error
A(G)[c]	g_absorption
σ A(G)	g_absorption_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

Notes. To facilitate access to the table, we also provide the correspondence between parameter (period(s), E, etc.) and the name of the parameter in the *Gaia* archive table. [(a)]The BJD of the epoch of maximum light is offset by JD 2455197.5 d (= J2010.0). [(b)]Photometric metal abundance derived from the ϕ_{31} Fourier parameter of the light curve for 54 272 fundamental-mode RR Lyrae stars (see Sects. 2.1.1 and 4.1). [(c)]Absorption in the G band computed from a relation that links the intrinsic colour of the star to the period and the amplitude of the light variation (see Sects. 2.1.2 and 4.1).

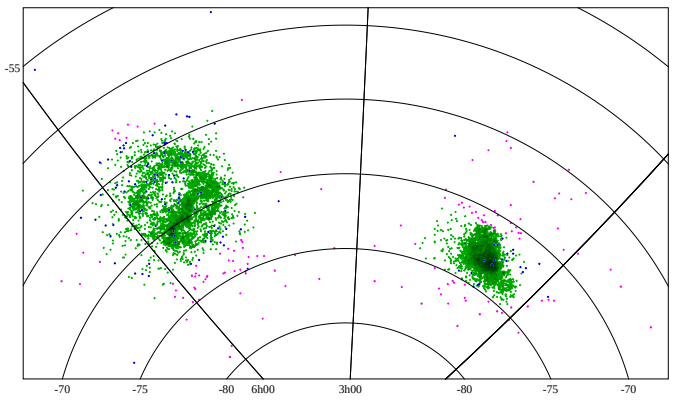

Fig. 44. Spatial distribution of *Gaia* DR2 Cepheids located in the region of the two Magellanic Clouds. Green filled circles show known Cepheids observed in the LMC and SMC regions defined in Sect. 2 by the OGLE survey, magenta filled circles represent All-Sky known (from OGLE or other surveys) Cepheids, and blue filled circles show new Cepheids identified by *Gaia* in the LMC and SMC (118 in total).

improved by properly taking into account the scatter in the folded light curve. This will reduce the number of false positives.

4. A classifier will be developed to optimise the type and subtype classification of Cepheids and RR Lyrae stars that is performed by the SOS Cep&RRL pipeline.
5. Automated validation and cleaning procedures will be put in place.
6. CMDs in absolute magnitude and the comparison with theoretical instability strips for Cepheids and RR Lyrae stars will be used to improve the source classification and the derivation of their intrinsic parameters (e.g. effective temperatures).

To conclude, the results of *Gaia* all-sky Cepheids and RR Lyrae stars obtained with the SOS Cep&RRL processing demonstrate the excellent quality of *Gaia* multi-band photometry released in DR2 and nicely showcase the potential of *Gaia* in the field of variable star studies and for Cepheids and RR Lyrae stars in particular.

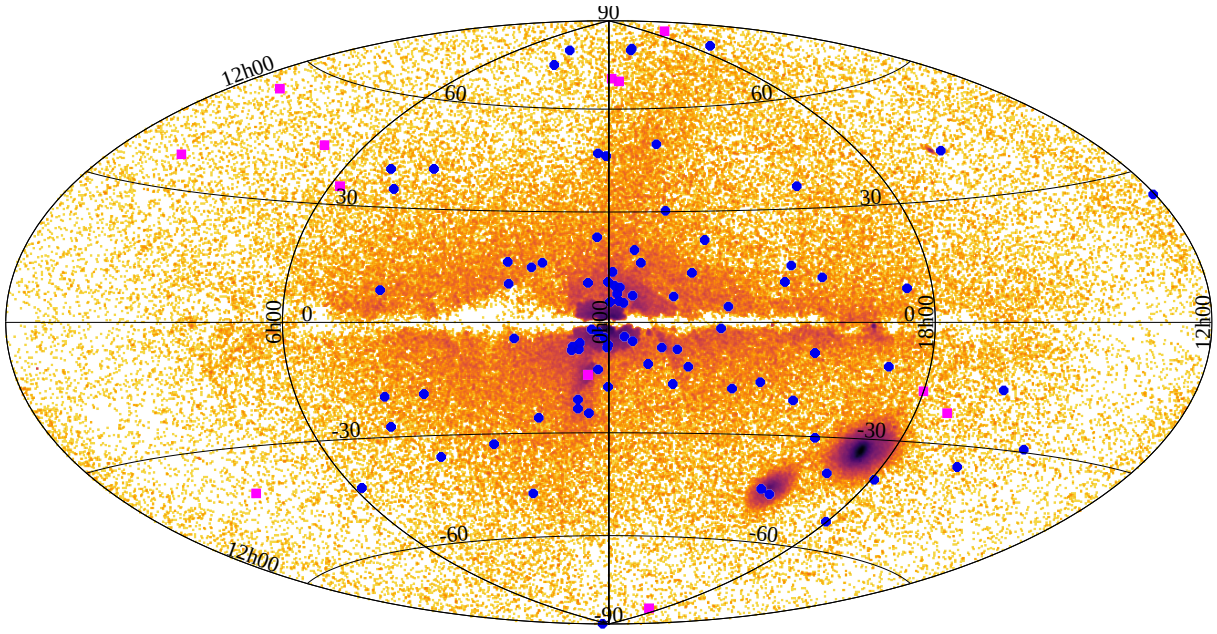

Fig. 45. Distribution on sky in galactic coordinates of RR Lyrae stars within the limiting magnitude of *Gaia* (orange points). The map combines known literature sources with and without a *Gaia* counterpart and new RR Lyrae stars discovered by *Gaia* and confirmed by the SOS Cep&RRL pipeline for more than 223 000 RR Lyrae stars in total. Blue filled dots and magenta filled squares indicate 87 globular clusters and 12 dwarf spheroidal galaxies (classical and ultra-faint) in which *Gaia* has observed RR Lyrae stars that are confirmed by the SOS Cep&RRL pipeline.

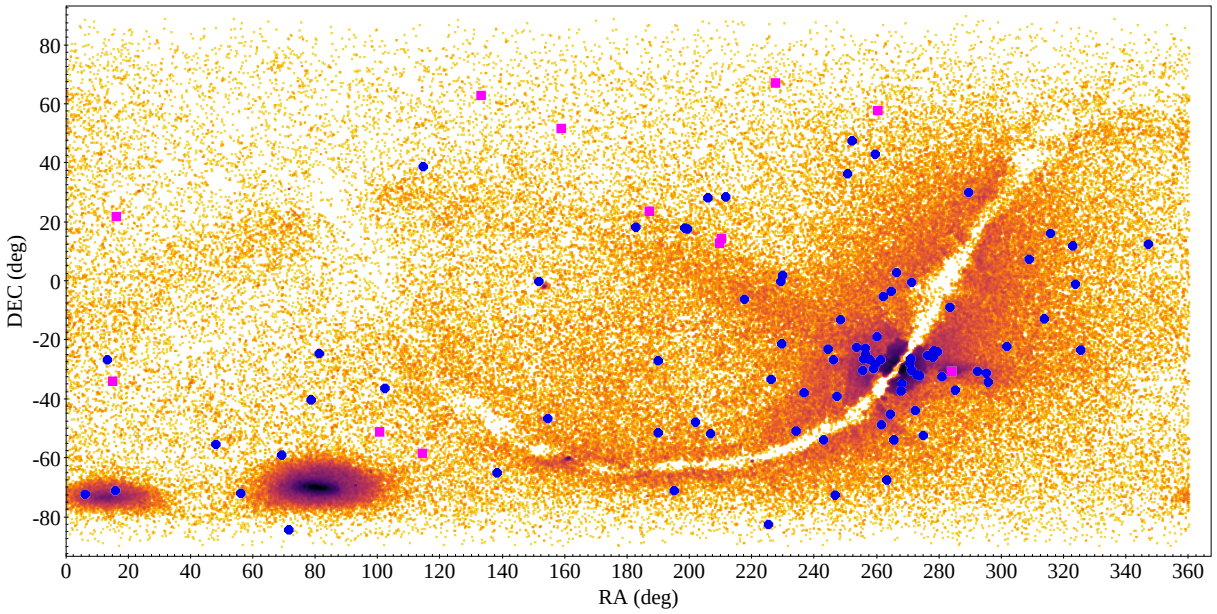

Fig. 46. Same as in Fig. 45, but in equatorial coordinates.

Acknowledgements. This work has made use of data from the ESA space mission *Gaia*, processed by the *Gaia* Data Processing and Analysis Consortium (DPAC). Funding for the DPAC has been provided by national institutions participating in the *Gaia* Multilateral Agreement. In particular, 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, and 2014-025-R.1.2015 to INAF (PI M.G. Lattanzi), the Belgian participation by the BELgian federal Science Policy (BELSPO) through PRODEX grants, the Swiss participation by the Swiss State Secretariat for Education, Research and Innovation through the ESA Prodex program, the "Mesures d'accompagnement", the "Activités Nationales Complémentaires", the Swiss National Science Foundation, and the Early Postdoc Mobility fellowship, the Spanish participation by the Spanish Ministry of Economy MINECO-FEDER through grants AyA2014-55216, AyA2011-24052 and the Hungarian participation through the PECS programme (contracts C98009 and 40001066398/12/NL/KML). 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. The *Gaia* mission website is http://www.cosmos.esa.int/gaia. This research has made use of the International Variable Star Index (VSX) database, operated at AAVSO, Cambridge, Massachusetts, USA (https://www.aavso.org/vsx/index.php) and of the SIMBAD database, operated at CDS, Strasbourg, France. We thank an anonymous referee for carefully reading our paper and for providing comments that contributed to significantly improving the readability and quality of our manuscript. We are indebted to Dante Minniti for sharing in advance of publication a list of RR Lyrae candidates discovered by the VVV survey in the MW bulge and disc. We gratefully acknowledge feedback from S. Cheng and S. Koposov, who identified the misclassified sources listed in Appendix C during a

Table 5. Links to *Gaia* archive tables for retrieving the pulsation characteristics: period(s), epochs of maximum light (E), peak-to-peak amplitudes, intensity-averaged mean magnitudes, ϕ_{21}, R_{21}, ϕ_{31}, R_{31} Fourier parameters with related uncertainties, and metallicity computed by the SOS Cep&RRL pipeline for the 9575 Cepheids confirmed by SOS and released in *Gaia* DR2.

Table URL	http://archives.esac.esa.int/gaia/
Cepheid parameters computed by the SOS Cep&RRL pipeline	
Table name	**gaiadr2.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
$\mathrm{E}^a(G, G_{\mathrm{BP}}, G_{\mathrm{RP}})$	epoch_g, epoch_bp, epoch_rp
$\sigma\mathrm{E}(G, G_{\mathrm{BP}}, G_{\mathrm{RP}})$	epoch_g_error, epoch_bp_error, epoch_rp_error
$\langle G\rangle, \langle G_{\mathrm{BP}}\rangle, \langle G_{\mathrm{RP}}\rangle$	int_average_g, int_average_bp, int_average_rp
$\sigma\langle G\rangle, \sigma\langle G_{\mathrm{BP}}\rangle, \sigma\langle G_{\mathrm{RP}}\rangle$	int_average_g_error, int_average_bp_error, int_average_rp_error
$\mathrm{Amp}(G, G_{\mathrm{BP}}, G_{\mathrm{RP}})$	peak_to_peak_g, peak_to_peak_bp, peak_to_peak_rp
$\sigma[\mathrm{Amp}(G)], \sigma[\mathrm{Amp}(G_{\mathrm{BP}})], \sigma[\mathrm{Amp}(G_{\mathrm{RP}})]$	peak_to_peak_g_error, peak_to_peak_bp_error, peak_to_peak_rp_error
ϕ_{21}	phi21_g
$\sigma(\phi_{21})$	phi21_g_error
R_{21}	r21_g
$\sigma(R_{21})$	r21_g_error
ϕ_{31}	phi31_g
$\sigma(\phi_{31})$	phi31_g_error
R_{31}	r31_g
$\sigma(R_{31})$	r31_g_error
$[\mathrm{Fe/H}]^b$	metallicity
σ ([Fe/H])	metallicity_error
$\mathrm{A}(G)$	g_absorption: empty for Cepheids
σ $\mathrm{A}(G)$	g_absorption_error: empty for Cepheids
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

Notes. To facilitate access to the table, we also provide the correspondence between parameter (period(s), E, etc.) and the name of the parameter in the *Gaia* archive table. [(a)]The BJD of the epoch of maximum light is offset by JD 2455197.5 d (= J2010.0). [(b)]Photometric metal abundance derived from the Fourier parameters of the light curve for 3738 fundamental-mode DCEPs with period shorter than 6.3 days (see Sects. 2.1.1 and 4.2).

project developed in part at the 2018 NYC Gaia Sprint, hosted by the Center for Computational Astrophysics of the Flatiron Institute in New York City. We are also indebted to Hsiang-Chih Hwang from the John Hopkins University, who pointed out another four possibly misclassified sources (one RR Lyrae star and three Cepheids) discovered at the *Gaia* DR2 Experiment Lab (2018) in Madrid and to Tim Bedding and Dan Hey, who informed us of the misclassification as Cepheid of the spotted rotating star KIC 6619830 observed by *Kepler*. In this study we have largely made use of TOPCAT (Taylor 2005).

References

Alfonso-Garzón, J., Domingo, A., Mas-Hesse, J. M., & Giménez, A. 2012, A&A, 548, A79
Andrae, R., Fouesneau, M., Creevey, O., et al. 2018, A&A, 616, A8 (*Gaia* 2 SI)
Arellano Ferro, A., Figuera Jaimes, R., Giridhar, S., et al. 2011, MNRAS, 416, 2265
Arellano Ferro, A., Bramich, D. M., Giridhar, S., et al. 2013, Acta Astron., 63, 429
Arellano Ferro, A., Ahumada, J. A., Calderón, J. H., & Kains, N. 2014, Rev. Mex. Astron. Astrofis., 50, 307
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
Arenou, F., Luri, X., Babusiaux, C., et al. 2018, A&A, 616, A17 (*Gaia* 2 SI)
Bailey, S. I. 1902, Ann. Harvard Coll. Observ., 38, 1
Benkő, J. M., Bakos, G. Á., & Nuspl, J. 2006, MNRAS, 372, 1657
Blazhko, S. 1907, Astron. Nachr., 175, 325
Braga, V. F., Stetson, P. B., Bono, G., et al. 2016, AJ, 152, 170
Cacciari, C., Corwin, T. M., & Carney, B. W. 2005, AJ, 129, 267
Carretta, E., Bragaglia, A., Gratton, R. G., et al. 2009, A&A, 505, 117
Clement, C. M., Muzzin, A., Dufton, Q., et al. 2001, AJ, 122, 2587
Clementini, G., Ripepi, V., Bragaglia, A., et al. 2005, MNRAS, 363, 734
Clementini, G., Ripepi, V., Leccia, S., et al. 2016, A&A, 595, A133 (Paper 1)
Clementini, G., Garofalo, A., Muraveva, T., & Ripepi, V. 2018, ASP Conf. Ser., 514, 89
Contreras, R., Catelan, M., Smith, H. A., et al. 2010, AJ, 140, 1766
Corwin, T. M., Borissova, J., Stetson, P. B., et al. 2008, AJ, 135, 1459
Cuffey, J. 1966, AJ, 71, 514
Dall'Ora, M., Kinemuchi, K., Ripepi, V., et al. 2012, ApJ, 752, 42
Debosscher, J., Blomme, J., Aerts, C., & De Ridder, J. 2011, A&A, 529, A89
Di Fabrizio, L., Clementini, G., Maio, M., et al. 2005, A&A, 430, 603
Drake, A.J., Catelan, M., Djorgovski, S.G., et al. 2013a, ApJ, 763, 32
Drake, A. J., Catelan, M., Djorgovski, S. G., et al. 2013b, ApJ, 765, 154
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
Evans, D., Riello, M., De Angeli, F., et al. 2018, A&A, 616, A4 (*Gaia* 2 SI)

Eyer, L., Clementini, G., Guy, L. P., et al. 2017a, EPJ Web Conf. 152, 02002
Eyer, L., Mowlavi, N., Ewans, D.W., et al. 2017b, A&A, submitted [arXiv:1702.03295]
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 (Clementini, G., et al.) 2017, A&A, 605, A79
Gaia Collaboration (Brown, A.G.A., et al.) 2018, A&A, 616, A1 (*Gaia* 2 SI)
Gaia Collaboration (Eyer, L. et al.) 2019, A&A, in press, DOI: 10.1051/0004-6361/201833304 (*Gaia* 2 SI)
Garofalo, A., Cusano, F., Clementini, G., et al. 2013, ApJ, 767, 62
Genovali, K., Lemasle, B., Bono, G., et al. 2013, A&A, 554, A132
Goranskij, V. P. 1976, Peremennye Zvezdy Prilozhenie, 3, 1
Gran, F., Minniti, D., Saito, R. K., et al. 2016, A&A, 591, A145
Gratton, R. G., Bragaglia, A., Clementini, G., et al. 2004, A&A, 421, 937
Harris, W. E. 1996, AJ, 112, 1487
Holl, B., Audard, M., Nienartowicz, K., et al. 2018, A&A, 618, A30 (*Gaia* 2 SI)
Jayasinghe, T., Kochanek, C. S., Stanek, K. Z., et al. 2018, MNRAS, 477, 3145
Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, 48
Jurcsik, J., & Kovács, G. 1996, A&A, 312, 111
Kaluzny, J., Kubiak, M., Szymanski, M., et al. 1995, A&AS, 112, 407
Kaluzny, J., Krzeminski, W., & Nalezyty, M. 1997, A&AS, 125, 337
Kinemuchi, K., Harris, H. C., Smith, H. A., et al. 2008, AJ, 136, 1921
Kirby, E. N., Lanfranchi, G. A., Simon, J. D., Cohen, J. G., & Guhathakurta, P. 2011, ApJ, 727, 78
Klagyivik, P., Szabados, L., Szing, A., Leccia, S., & Mowlavi, N., 2013, MNRAS, 434, 418
Layden, A. C., & Sarajedini, A. 2003, AJ, 125, 208
Lindegren, L., Hernández, J., Bombrunet A., et al. 2018, A&A, 616, A2 (*Gaia* 2 SI)
Lomb, N. R. 1976, Ap&SS, 39, 447
Madore, B. F. 1982, ApJ, 253, 575
Minniti, D., Dékány, I., Majaess, D., et al. 2017, AJ, 153, 179
Molnár, L., Pál, A., Plachy, E., et al. 2015a, ApJ, 812, 2
Molnár, L., Szabó, R., Moskalik, P. A., et al. 2015b, MNRAS, 452, 4283
Molnár, L., Szabó, R., & Plachy, E. 2016, JAAVSO, 44, 168
Morgan, S. M., Wahl, J. N., & Wieckhorst, R. M. 2007, MNRAS, 374, 1421
Nemec, J. M., Smolec, R., Benkő, J. M., et al. 2011, MNRAS, 417, 1022
Nemec, J. M., Cohen, J. G., Ripepi, V., et al. 2013, ApJ, 773, 181
Osuna, P., Ortiz, I., Lusted, J., et al. 2008, IVOA Astronomical Data Query Language
Palaversa, L., Ivezić, Ž., Eyer, L., et al. 2013, AJ, 146, 101
Piersimoni, A. M., Bono, G., & Ripepi, V. 2002, AJ, 124, 1528
Pietrzyński, G., Graczyk, D., Gieren, W., et al. 2013, Nature, 495, 76
Pojmański, G. 1997, Acta Astron., 47, 467
R Core Team 2018, A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Richards, J. W., Starr, D. L., Miller, A. A., et al. 2012, ApJS, 203, 32
Riello, M., De Angeli, F. , Evanset, D. W., et al. 2018, A&A, 616, A3 (*Gaia* 2 SI)
Rimoldini, L., Holl, B., Audard, M., et al. 2018, A&A, submitted [arXiv:1811.03919] (*Gaia* 2 SI)
Ripepi, V., Molinaro, R. Musella, I., et al. 2018, A&A, submitted [arXiv:1810.10486]
ROTSE Collaboration (Kinemuchi, K., et al.) 2006, AJ, 132, 1202
Salinas, R., Contreras Ramos, R., Strader, J., et al. 2016, AJ, 152, 55
Sartoretti, P., Katz, D., Cropper, M., et al. 2018, A&A, 616, A6 (*Gaia* 2 SI)
Scargle, J. D. 1982, ApJ, 263, 835
Sesar, B., Hernitschek, N., Mitrović, S., et al. 2017, AJ, 153, 204
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2008, Acta Astron., 58, 293
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2010, Acta Astron., 60, 91
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2014, Acta Astron., 64, 177
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2015a, Acta Astron., 65, 297
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2015b, Acta Astron., 65, 233
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2016, Acta Astron., 66, 131
Taylor, M. B. 2005, ASP Conf. Ser., 347, 29
Tonry, J. L., Denneau, L., Heinze, A. N., et al. 2018, PASP, 130, 4505
Torrealba, G., Catelan, M., Drake, A. J., et al. 2015, MNRAS, 446, 2251
Walker, A. R. 1998, AJ, 116, 220
Walker, A. R., & Nemec, J. M. 1996, AJ, 112, 2026
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9

Appendix A: Examples of *Gaia* archive queries

Table A.1. Queries to retrieve DR2 information on the Cepheids and RR Lyrae stars from the *Gaia* archive in the Astronomical Data Query Language (Osuna et al. 2008).

Query to retrieve time series of all Cepheids in the *Gaia* DR2.

```
select gaia.source_id, epoch_photometry_url from gaiadr2.gaia_source as gaia
     inner join gaiadr2.vari_cepheid as cep on gaia.source_id=cep.source_id
```

Query to retrieve time series of all RR Lyrae stars in the *Gaia* DR2

```
select gaia.source_id, epoch_photometry_url from gaiadr2.gaia_source as gaia
     inner join gaiadr2.vari_rrlyrae as rrl on gaia.source_id=rrl.source_id
```

Query to retrieve the number of processed observations and SOS Cep&RRL-computed parameters of all Cepheids in the *Gaia* DR2

```
select cep.*,tsr.num_selected_g_fov,tsr.num_selected_bp,tsr.num_selected_rp from gaiadr2.vari_cepheid cep
     inner join gaiadr2.vari_time_series_statistics tsr on cep.source_id=tsr.source_id
```

Query to retrieve the number of processed observations and SOS Cep&RRL-computed parameters of all RR Lyrae in the *Gaia* DR2

```
select rrl.*,tsr.num_selected_g_fov,tsr.num_selected_bp,tsr.num_selected_rp from gaiadr2.vari_rrlyrae rrl
     inner join gaiadr2.vari_time_series_statistics tsr on rrl.source_id=tsr.source_id
```

Appendix B: Acronyms

Table B.1. List of acronyms used in this paper.

Acronym	Description
ACEP	Anomalous Cepheid
ALL_SKY	The celestial region excluding the LMC and SMC regions
Amp(G)	Amplitude of the light variation in the G band
Amp(G_{BP})	Amplitude of the light variation in the G_{BP} band
Amp(G_{RP})	Amplitude of the light variation in the G_{RP} band
BLHER	BL Herculis class of variables
CMD	Colour Magnitude Diagram
DCEP	Classical Cepheid (Population I)
dSph	Dwarf spheroidal galaxy
DR	Data Release
F	Fundamental mode of pulsation
FO	First overtone mode of pulsation
G	*Gaia* photometric G-band
G_{BP}	*Gaia* photometric G_{BP} band
G_{RP}	*Gaia* photometric G_{RP} band
GC	Globular cluster
LMC	Large Magellanic Cloud
MW	Milky Way
PA	Period–Amplitude
PL	Period–Luminosity
PW	Period–Wesenheit
RRab	RR Lyrae star of ab type
RRc	RR Lyrae star of c type
RRd	Double-mode RR Lyrae star
RVTAU	RV Tauri class of variables
SMC	Small Magellanic Cloud
SOS	Specific Object Study
T2CEP	Type II Cepheid (Population II)
WVIR	W Virginis class of variables

Appendix C: Misclassified sources

Table C.1. List of sources that are likely misclassified as RR Lyrae stars (courtesy of S. Cheng, S. Koposov and H-C. Hwang).

Gaia sourceid	Notes
1754525270341133312	Weird folded light curve
32205593226566016	Galaxy
361431775815909376	Galaxy
361451975047400448	Galaxy
585136968493947264	Galaxy

Notes. This table is available in its entirety at the CDS, only a portion is shown here for guidance regarding its form and content.

Table C.2. List of sources that are likely misclassified as Cepheids (courtesy of H-C. Hwang, T. Bedding and D. Hey).

Gaia sourceid	Literature sourceid	Notes
2077108036182676224	KIC 6619830	Spotted rotating star
2641587994382309632		Fainter and redder than typical Cepheids
5424247204666300032		Fainter and redder than typical Cepheids
5804085561048983552		Fainter and redder than typical Cepheids

Notes. A more complete list of likely misclassified Cepheids is published in Ripepi et al. (2018).