DISSERTATION

submitted to the

COMBINED FACULTY OF NATURAL SCIENCES AND MATHEMATICS

of the

UNIVERSITY OF HEIDELBERG, GERMANY

for the degree of

DOCTOR OF NATURAL SCIENCES

Put forward by

Hitesh Narendra LALA

born in Ambernath, India

Oral examination: 15.11.2023

Cepheids and RR Lyrae Stars in the Milky Way and the Magellanic Clouds

Advisor:

Prof. Dr Eva K. Grebel

Co-advisor:

Dr Bertrand Lemasle

Referees:

Prof. Dr Eva K. Grebel

Priv.-Doz. Dr Sabine Reffert

Co-examiners:

Priv.-Doz. Dr Coryn Bailer-Jones

Prof. Dr Luca Amendola

To Mom and Dad

Abstract

RR Lyrae stars and Cepheids play a pivotal role in the study of the Universe. From Milky Way archaeology to cosmology, these variable stars contribute significantly to a myriad of fields. Many large-scale photometric surveys have classified and characterized these stars into various subtypes. However, no single survey is complete, and many of them suffer from varying degrees of contamination. The main goal of this study is to generate a homogeneous, pure, and complete (to date) catalog of Cepheids and RR Lyrae stars in the Milky Way and in the Magellanic Clouds. To this end, we have homogenized RR Lyrae and Cepheid (classical, Type-II, and anomalous) catalogs from eight wide-field optical surveys. We developed a new Bhattacharyya distance-based algorithm for a probabilistic cross-match of these catalogs, taking into account the full astrometry covariance matrix and variability periods. We compared all surveys against each other and numerous literature catalogs to validate their periods and classification. We complemented our compilation by cross-matching with cleaned catalogs of more than 20 other photometric and spectroscopic surveys. These data were used to compute about 200 newly-defined, highly-precise and accurate period-Wesenheit relations. These relations enable the computation of distances at over 98 percent accuracy and better than two percent precision. We release a catalog of about 330 000 Cepheids and RR Lyrae stars with more than 1000 features, including periods, subtype classification, optical and infrared photometry, line-of-sight velocities, and distances. This catalog, named RRLCep, holds significant potential in the context of Milky Way archaeology and near-field cosmology.

Zusammenfassung

RR-Lyrae-Sterne und Cepheiden spielen eine zentrale Rolle bei der Erforschung des Universums. Von der Archäologie der Milchstraße bis zur Kosmologie tragen diese veränderlichen Sterne erheblich zu einer Vielzahl von Fachgebieten bei. Viele groß angelegte photometrische Umfragen haben diese Sterne in verschiedene Untergruppen eingeteilt und charakterisiert. Dennoch ist keine einzelne Umfrage vollständig, und viele von ihnen leiden unter unterschiedlichen Grad an Kontamination. Das Hauptziel dieser Studie ist die Erstellung eines homogenen, reinen und vollständigen (bis heute) Katalogs von Cepheiden und RR Lyrae Sternen in der Milchstraße und den Magellanschen Wolken. Zu diesem Zweck haben wir RR-Lyrae- und Cepheiden-Kataloge (klassische, Typ-II- und anomale Kataloge) aus acht optischen Weitfelddurchmusterungen homogenisiert. Wir haben einen neuen Bhattacharyya-Distanz-basierten Algorithmus für ein probabilistisches Abgleichen dieser Kataloge entwickelt, unter Berücksichtigung der gesamten Astrometrie-Kovarianzmatrix und der Variabilitätsperioden. Wir haben alle Umfragen miteinander und zahlreiche Literatur Kataloge verglichen, um ihre Perioden und Klassifikation zu validieren. Wir ergänzen unsere Zusammenstellung durch einen Abgleich mit bereinigten Katalogen von mehr als 20 anderen photometrischen und spektroskopischen Umfragen. Diese Daten wurden verwendet, um etwa 200 neu definierte, hochpräzise und genaue Perioden-Wesenheit-Beziehungen zu berechnen. Diese Beziehungen ermöglichen die Berechnung von Entfernungen mit einer Genauigkeit von über 98 Prozent und einer Präzision von besser als zwei Prozent. Wir veröffentlichen einen Katalog von etwa 330.000 Cepheiden und RR Lyrae Sternen mit mehr als 1000 Merkmalen, darunter Perioden, Subtyp-Klassifikation, optische und Infrarot-Photometrie, Geschwindigkeiten entlang der Sichtlinie und Entfernungen. Dieser Katalog mit dem Namen RRLCep birgt erhebliches Potenzial im Kontext der Milchstraßenarchäologie und der Nahfeldkosmologie.

Contents

C. Supplementary information III

Bibliography

Publications of Hitesh Lala

Ackowledgments

List of Figures

List of Tables

I

Introduction

If you don't know, the thing to do is not to get scared, but to learn.

– Ayn Rand

Cepheids and RR Lyrae stars have been at the forefront of many discoveries in astronomy. From establishing the cosmic distance scale and determining H_0 (Hubble, 1929; Madore & Freedman, 1991; Riess et al., 2019a), to studying dark matter (Hattori et al., 2021; Ablimit et al., 2020), from investigating the structure (Skowron et al., 2019; Lemasle et al., 2022; Soszyński et al., 2019c) and evolution of the Milky Way (Ibata et al., 2020; Cassisi & Salaris, 2013; Lemasle et al., 2013) to testing stellar evolution and pulsation models (Bono et al., 1997; Caputo et al., 2005; Sandage & Tammann, 2006; Bono et al., 2020), these stars are credited for multitudes of advancements in these fields.

In this dissertation, I have created the most comprehensive catalog of Cepheids and RR Lyrae stars, which is named RRLCep. This catalog includes Cepheids and RR Lyrae stars that have been identified in the Milky Way and in the Magellanic Clouds over the past few decades. It is the most complete catalog to date, with its every property meticulously curated. As its primary application, I present the results of 192 period-Wesenheit relations computed using the contents of this catalog. In this chapter, I present a brief summary of the concepts pertinent to this dissertation. For an in-depth review of these topics, please refer to the following excellent resources: Catelan & Smith (2015), Cox (1980), Bhardwaj (2020),

1

Riess et al. (2022), Madore & Freedman (1991), Bland-Hawthorn & Gerhard (2016), Naidu et al. (2020), and Carroll & Ostlie (1996, 2017).

I.1 Pulsating Stars

Variable stars are stars whose brightness fluctuates (mostly) cyclically over time. They are categorized into two primary types: intrinsic and extrinsic variable stars. Our target stars, namely, the Cepheids and RR Lyrae stars, are intrinsic variable stars, specifically pulsating stars. These stars undergo rhythmic contractions and expansions, and the period of this variability is determined by the laws of physics. These laws, as described below, also govern the star's effective temperature (T_{eff}) and its radius, R.

In particular, the effective temperature and the radius combine to yield Stefan's law, which provides the bolometric luminosity, L, of a star as:

$$L = 4\pi R^2 \sigma_B T_{eff}^4, \tag{I.1}$$

where σ_B is the Stefan-Boltzmann constant. Here, the first term $4\pi R^2$ is the total emitting surface area of a star, and $\sigma_B T_{eff}^4$ quantifies the areal surface brightness. Expressing this law in terms of the bolometric magnitude, M, where $M = -2.5 \log_{10}(L)$, we get:

$$M = -5 \log_{10}(R) - 10 \log_{10}(T_{eff}) + C. \tag{I.2}$$

Here, C is a star-independent constant. Stellar evolution and the virial theorem impose further constraints on this law, consequently limiting the effective temperature, radius, and thereby, luminosity of a star to a finite range. For instance, a natural outcome of the hydrostatic equilibrium is the period-density relation (or the Ritter relation after Ritter 1879), whereby the variability period P can be computed as:

$$P = Q \cdot \rho^{-1/2}, \tag{I.3}$$

where Q is a pulsation constant (also known as a structural constant in the context of mechanical systems) and ρ parametrizes the mean density of a star. Moreover, the mass ($\mathcal{M}$) of a star is a function of both its radius R and density ρ, i.e., for a sphere, we have:

$$\mathcal{M} = 4/3\pi R^3 \rho. \tag{I.4}$$

Furthermore, the effective temperature of a star can be considered a function of a star's color, specifically the difference between its brightness as measured at two different wavelengths (Planck's law; Planck, 1901):

$$T_{eff} \propto M_{\lambda_1} - M_{\lambda_2}, \tag{I.5}$$

where M_λ is the magnitude measured at a given wavelength λ.

Combining the preceding equations with the stellar evolution-dictated mass-luminosity relation, yields the observational equivalent of the theoretical Stefan's law, also known as the period-luminosity-color relation:

$$M_\lambda = \alpha + \beta \cdot \log_{10}(P) + \gamma \cdot (M_{\lambda_1} - M_{\lambda_2}), \tag{I.6}$$

where α, β, and γ are constants. The period-luminosity (PL) relation is a projection of this period-luminosity-color (PLC) hyperplane onto the luminosity *vs* period plane. Such a PL relation can be considered self-consistent if it is being computed only for stars exhibiting a narrow range of effective temperatures. In such a scenario, we get:

$$M_\lambda = \alpha + \beta \cdot \log_{10}(P), \tag{I.7}$$

where the last term of Eqn. I.6 — assumed as a constant for a sample of stars exhibiting similar T_{eff} — is absorbed by the other constant α.

Another projection of the PLC relation can be seen in the luminosity-effective temperature plane, or the Hertzsprung-Russell diagram (see Fig. I.1). Just like the Stefan's law, the PLC relation holds true for any star and not just variable stars. However, pulsations can be readily detected only for certain types of stars. Such stars are unstable to pulsations owing to the change in opacity (κ) of the partial ionization zones in their atmospheres. Cepheids and RR Lyrae stars' pulsations, in particular, are explained by the κ-mechanism, wherein the opacity of the Helium ionization zone changes with temperature. This mechanism manifests itself in the Hertzsprung-Russell diagram as the instability strip. Specifically, our target stars straddle the classical instability strip labeled 'IS' in Fig. I.1. For a detailed review of the pulsation theory, confer Cox (1980); Sandage (1958); Madore & Freedman (1991).

Empirically, the calibration of a PL or a PLC relation for a given set of stars requires repeated photometric observations. These observations are used to compute the variability periods and the mean apparent magnitude (brightness) of the stars. To convert apparent magnitudes to absolute magnitudes (or luminosity), one needs distance and extinction towards the stars. Both of these quantities are notoriously hard to measure (see e.g., Anders et al., 2022, and references therein).

Due to the star formation cycle, the interstellar medium is interspersed with dust grains. These dust molecules absorb and scatter photons as a function of wavelength. The amount of light obscured by dust lying along the sightline of a star is measured as extinction (A_λ), with $A_\lambda \propto \lambda^{-1}$. The effect of this dust is readily apparent in Fig. I.5, where we see that the stars in the central regions of the Milky Way cannot be detected in optical wavelengths. Extinction A_λ is computed by measuring the reddening or the color excess (say) $E(\lambda_b - \lambda_c)$. Hereafter, λ_x stands for the apparent magnitude measured in band x. If $(\lambda_b - \lambda_c)$ is the observed color
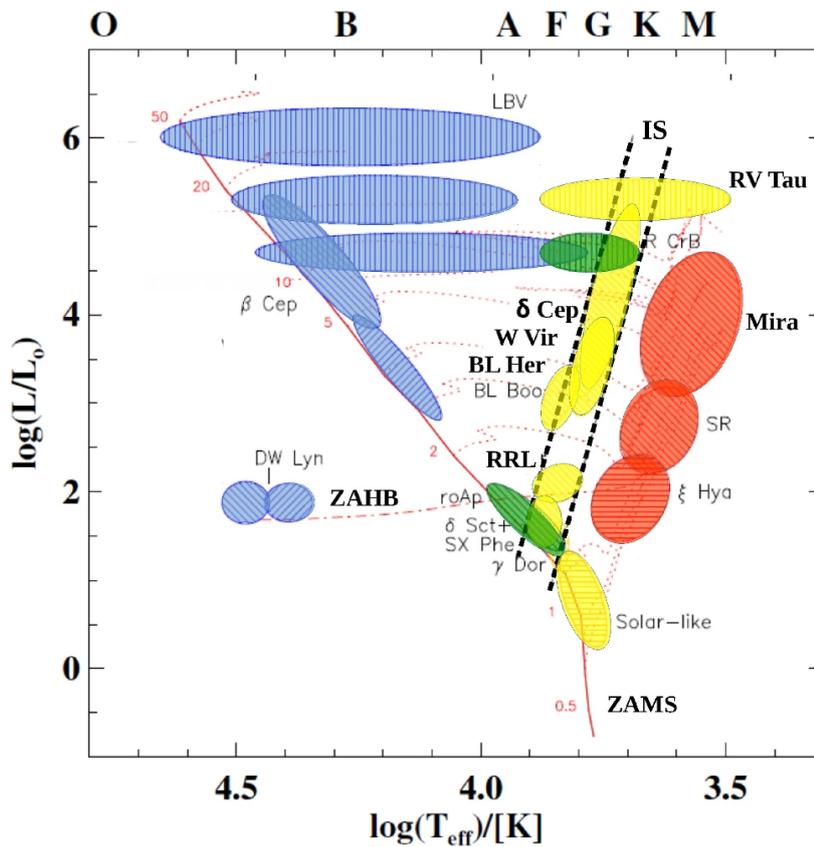
FIGURE I.1: Hertzsprung-Russell diagram with the classical instability strip (IS) and the locations of our target subtypes that are highlighted as follows: BL Boo: anomalous Cepheids, δ Cep: classical Cepheids, RRL: RR Lyrae stars. The Type-II Cepheids subtypes are: BL Her, W Vir, and RV Tau. These subtypes will be introduced below. Here, the *x*-axis represents the decadic logarithm of the effective temperature T_{eff} of the stars and the *y*-axis represents the same for the absolute luminosity of a star normalized by that of the Sun ($L/L_{\odot}$) Credits: Image from Bhardwaj (2020). *Reproduced with permission from Springer Nature.*
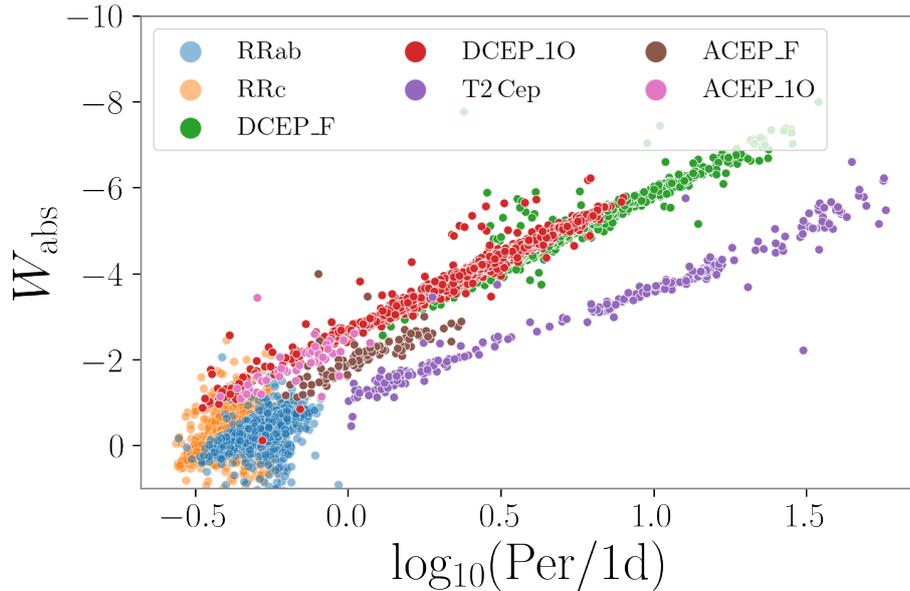
FIGURE I.2: Period-Wesenheit relation for Cepheids and RR Lyrae stars in the Large Magellanic Cloud. The *x*-axis represents the decadic logarithm of the variability period (in days) and the *y*-axis represents the absolute Wesenheit index defined using the *V* and *I*-band magnitudes.

of a star, and if we assume $(\lambda_b - \lambda_c)_0$ as the color corrected for extinction, the color excess can be computed using the following equation:

$$(\lambda_b - \lambda_c)_0 = (\lambda_b - \lambda_c) - \mathrm{E}(\lambda_b - \lambda_c) \tag{I.8}$$

Certain types of stars, like the Red Clump stars have a theoretically-defined constant absolute (or dereddened) color. These stars can be used to determine a reddening map in a region where they are present in copious amounts (c.f. e.g., Skowron et al., 2021; Haschke et al., 2011, and references therein).

To overcome this estimation of reddening on a star-by-star basis, Madore (1976, 1982); Madore & Freedman (1991) defined a Wesenheit index *W* as:

$$W \equiv \lambda_a - R(\lambda_b - \lambda_c), \tag{I.9}$$

where the coefficient *R* is the extinction law, a property of the dust grains inhabiting a given interstellar medium. It is defined as:

$$R = A_{\lambda_a}/\mathrm{E}(\lambda_b - \lambda_c). \tag{I.10}$$

If we now assume λ_a^0 to denote the extinction-corrected magnitude in the band *a*, we get:

$$\lambda_a^0 = \lambda_a - A_{\lambda_a}. \tag{I.11}$$

TABLE I.1: Properties of Cepheids and RR Lyrae stars. Brightness order reflects the absolute luminosity compared at similar periods.

Class	Subtypes	Period (d)	Age	Mass ($M_\odot$)	[Fe/H] (dex)	Brightness Order
Classical Cepheids	DCEP_F, DCEP_1O	0.4 – 80	<300 Myr	1 – 8	-1 – 0.5	1 (brightest)
Anomalous Cepheids	ACEP_F, ACEP_1O	0.3 – 2	1 – 6 Gyr	0.8 – 2.2	-2.5 – -0.5	2
Type-II Cepheids	BL Her, W Vir, RV Tau	1 – 80	>10 Gyr	0.3 – 0.5	-2 – 0.0	3
RR Lyrae	RRab, RRc	0.2 – 1	>10 Gyr	0.6 – 0.7	-2.5 – -0.5	4 (faintest)

Thus, using Equations I.8, I.10, and I.11 in Eqn. I.9, we get that the apparent Wesenheit:

$$
\begin{aligned}
W &\equiv \lambda_a - R(\lambda_b - \lambda_c) \\
W &= \lambda_a^0 + A_{\lambda_a} - R[(\lambda_b - \lambda_c)_0 + E(\lambda_b - \lambda_c)] \\
W &= \lambda_a^0 + R \cdot E(\lambda_a - \lambda_b) - R(\lambda_b - \lambda_c)_0 - R \cdot E(\lambda_b - \lambda_c) \\
W &= \lambda_a^0 - R(\lambda_b - \lambda_c)_0 \equiv W_0.
\end{aligned}
\tag{I.12}
$$

That is, the apparent Wesenheit index is equal to the dereddened Wesenheit index (W_0). Therefore, these Wesenheit indices are reddening-free by construction and only depend on the reddening law i.e., R. Consequently, the absolute Wesenheit index (W^{abs}) can be computed as:

$$
W^{\text{abs}} = W - \mu,
\tag{I.13}
$$

where μ is the distance modulus of the star. This quantity can now be used as proxy for the absolute luminosity in the PL relation defined earlier. This substitution gives rise to the period-Wesenheit (PW) relation, which is formulated as:

$$
W^{\text{abs}} = \alpha + \beta \cdot \log_{10}(P).
\tag{I.14}
$$

The computation of over 200 PW relations of this kind are discussed in the latter chapters of this dissertation. One such relation is presented for all our target variability types in Fig. I.2. The provenance of the data used to generate this figure will be presented in the next chapter.

I.2 RR Lyrae stars and Cepheids

Here, we provide a concise overview of the primary properties and applications of our RR Lyrae stars and Cepheid variables. A selection of their characteristics is listed in Table I.1, with further details elaborated in the text below.

I.2.1 RR Lyrae stars

RR Lyrae stars are old (> 10 Gyr), low-mass, horizontal branch stars with iron abundances ([Fe/H]) ranging from very metal-poor to super solar. They are classified into three main sub-types following Bailey (1902); Bailey et al. (1919); Schwarzschild (1940); Nemec (1984): RRab-type pulsating in the fundamental mode, RRc-type pulsating in the first overtone, and RRd-type stars pulsating simultaneously in both the modes. A sub-group of the latter, called anomalous RRd stars (aRRd), have a different period-ratio as compared to RRd stars (Soszyński et al., 2016b). Moreover, ~ 40% of RRab stars (Jurcsik et al., 2012; Netzel et al., 2018) and ~ 10.7% of RRc stars (Benkő et al., 2023) are predicted to exhibit the Blazhko effect (Blažko, 1907). This effect is the presence of secondary modulation in light curves of these stars with a period ranging from a few days to a few years.

Leveraging their ubiquity and precise distances, RR Lyrae stars are used to trace the Milky Way bulge (Braga et al., 2019; Kunder et al., 2020), disk (Iorio & Belokurov, 2021), halo (Iorio et al., 2018; Iorio & Belokurov, 2019), its substructures (Ramos et al., 2020; Ibata et al., 2020; Vasiliev & Baumgardt, 2021; Koposov et al., 2023), and the Magellanic Clouds (Soszyński et al., 2016b; Jacyszyn-Dobrzeniecka et al., 2016, 2020a).

I.2.2 Classical Cepheids

Classical Cepheids (DCEP) are young, massive, metal-rich variable stars mainly exhibiting radial pulsations in the fundamental (DCEP_F), first-overtone (DCEP_1O), and multiple modes (DCEP_mul). Multi-mode classical Cepheids have been detected pulsating in more than two radial modes simultaneously (Soszyński et al., 2017a). From an evolutionary standpoint, these stars are core-He burning post-red giant branch stars.

Being much younger and much more massive than RR Lyrae stars (< 300 Myr), DCEPs are mainly found in the Milky Way thin disk (Pietrukowicz et al., 2021; Soszyński et al., 2017b), the central regions of the Magellanic Clouds (Soszyński et al., 2017a), and spiral galaxies (Madore & Freedman, 1991; Freedman et al., 2001). Owing to their characteristic light curve shape (Leavitt, 1908, also see Fig. I.3), high luminosities, and well-defined PL relations (Leavitt & Pickering, 1912; Soszyński et al., 2017a), these stars form the fundamental building blocks of the cosmic distance ladder (Baade, 1952; Sandage, 1958; Madore & Freedman, 1991; Bono et al., 2010; Riess et al., 2022).

I.2.3 Anomalous Cepheids

Anomalous Cepheids (ACEP) are the low-mass, metal-poor analogues of classical Cepheids (Caputo et al., 2004). Zinn & Searle (1976) termed these stars "anomalous" as they follow a
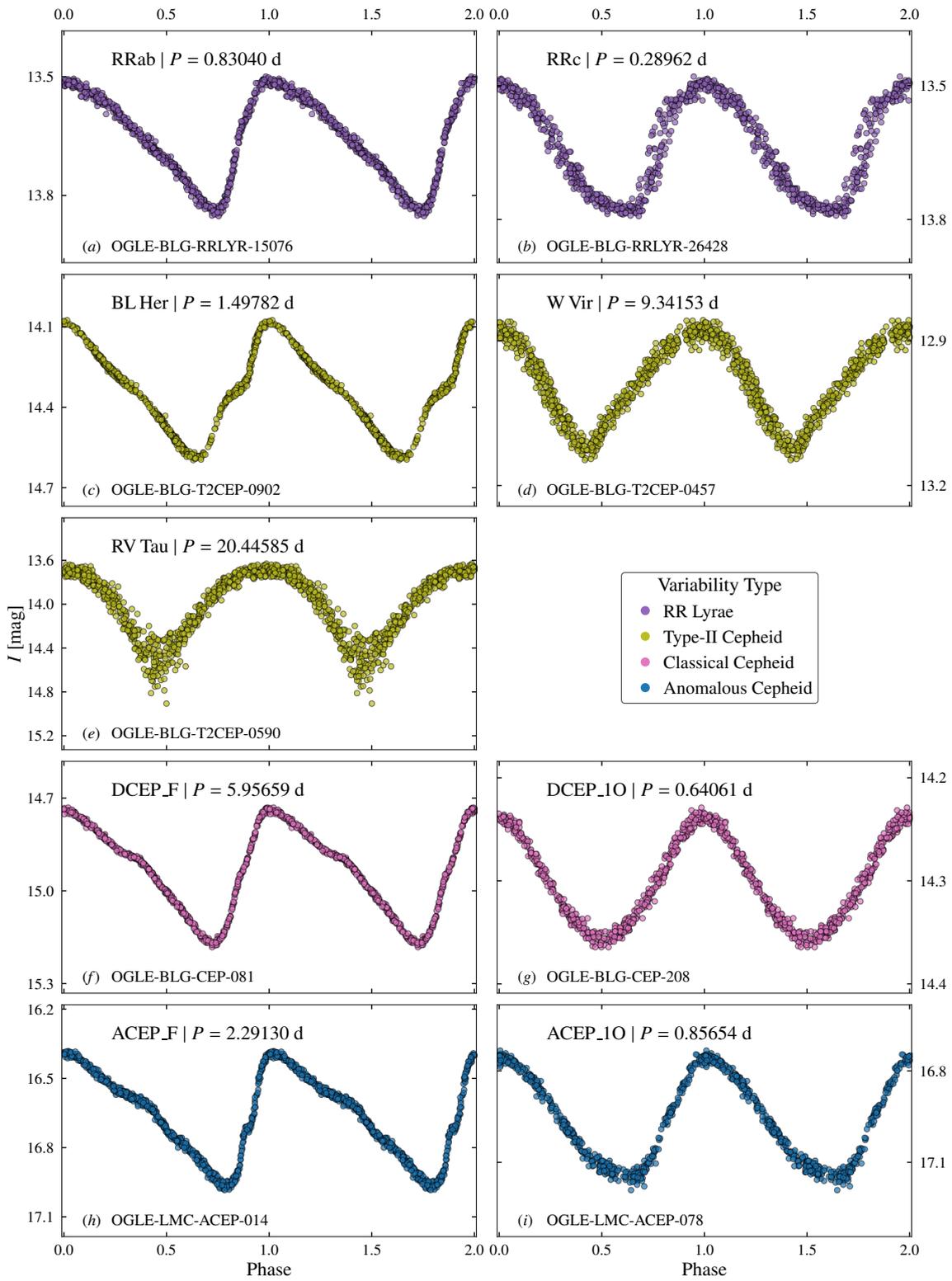
FIGURE I.3: Light curves of a randomly selected member from each of the target subtypes pulsating in a single mode are presented in: Panel (a) for an RRab-type star; (b) RRc; (c) BL Her; (d) W Vir; (e) RV Tau; (f) DCEP_F; (g) DCEP_1O; (h) ACEP_F; (i) ACEP_1O. The star's unique identifier and its variability period (P) are provided on each panel.
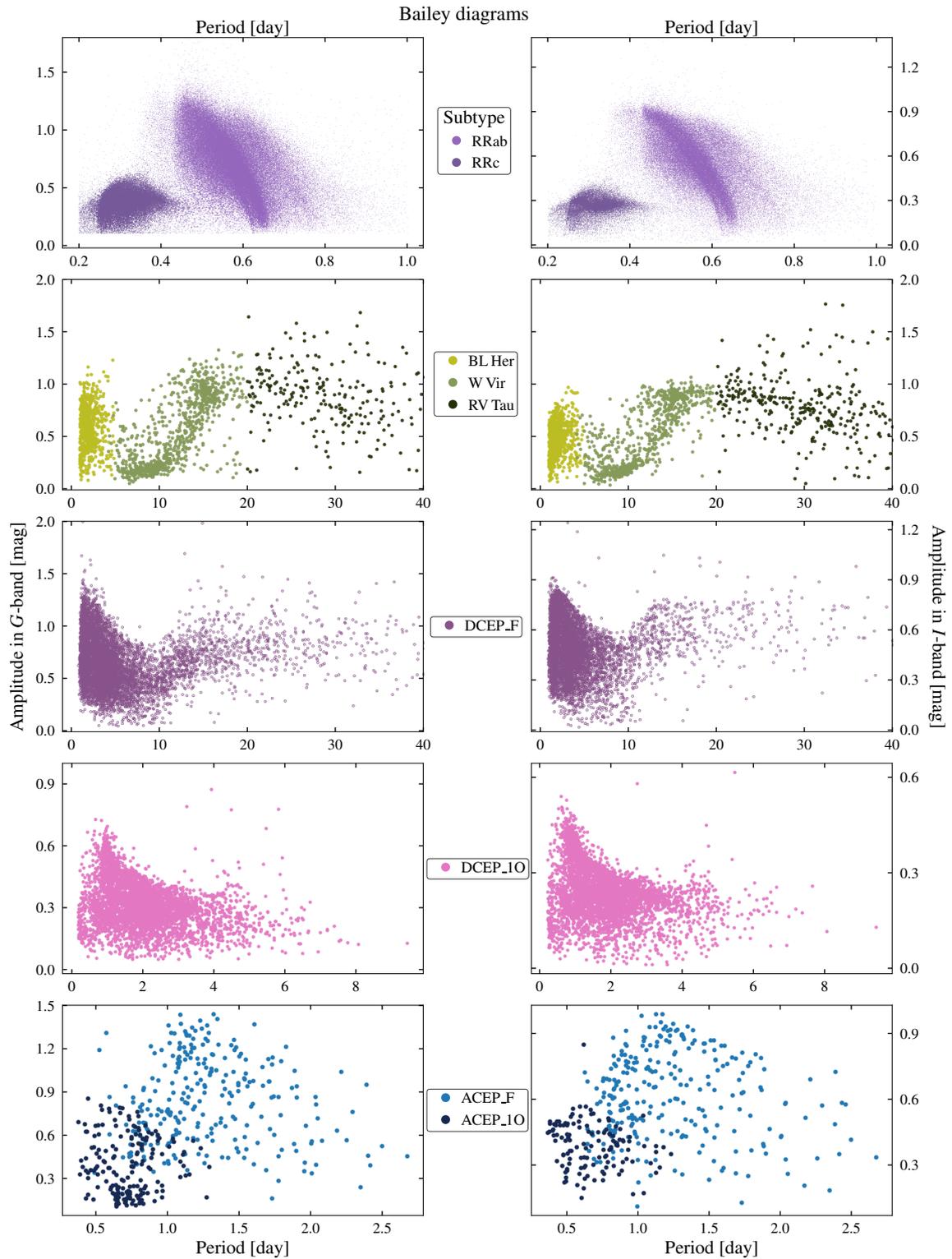
FIGURE I.4: Period vs Amplitude distributions (Bailey diagrams) of all target subtypes. The *y*-axes of left panels represent variability amplitude in the *G*-band and those in the right panels represent amplitude in the *I*-band. Note the different scale of the *y*-axes between the left and right panels as the amplitude decreases with increasing wavelength. The origin of this photometry will be discussed in the next chapter.

PL relation fainter than classical Cepheids but brighter than Type-II Cepheids. These stars are central He-burning with masses in the range of 0.8–2.2 $M_\odot$ (Fiorentino & Monelli, 2012; Caputo et al., 2004). Evolutionary models predict these stars can be formed through single-star evolution of intermediate-age (1–6 Gyr) stars or through mass transfer in older binary systems (Bono et al., 1997; Groenewegen & Jurkovic, 2017). They mainly pulsate in the fundamental (ACEP_F) or first-overtone mode (ACEP_1O), although Soszyński et al. (2020a) observed a possibility of multi-mode pulsations in relatively more metal-rich anomalous Cepheids (ACEP_mul).

I.2.4   Type-II Cepheids

Type-II Cepheids were separated from their classical counterparts by Baade (1952), and their difference was already hinted at by Joy (1937). These stars are old (>10 Gyr), low-mass stars further classified into three subtypes based on their periods: BL Herculis stars (BL Her; $P \lesssim 5\,\mathrm{d}$), W Virginis stars (W Vir; $5\,\mathrm{d} \lesssim P \lesssim 20\,\mathrm{d}$), and RV Tauri stars(RV Tau $P \gtrsim 20\,\mathrm{d}$). In the period-variability amplitude diagrams (called the Bailey diagrams after Bailey 1902) plotted in Fig. I.4, the distinct distributions of the three subtypes are easily discernible.

As Type-II Cepheids are intrinsically brighter and less massive than RR Lyrae stars, their periods are longer. The empirical boundary set between the periods of RR Lyrae stars and shorter-period Type-II Cepheids (BL Her) stars is a consequence of their different evolutionary stages (Braga et al., 2020). While RR Lyrae stars are burning He in their cores, Type-II Cepheids have exhausted their central He content and are double (H, He) shell burning.

Specifically, Bono et al. (2020, by using horizontal branch evolutionary models) predicted that BL Her stars are evolving off of the zero age horizontal branch and are on the first crossing of the instability strip. They termed these stars post-early asymptotic giant branch (PEAGB) stars as they are undergoing double shell (H and He) burning. They also proposed that W Vir stars are a mix of PEAGB and post-AGB stars on their second crossing (red to blue) of the instability strip. Finally, they classified RV Tau stars as post-AGB stars too, with the longer-period RV Tau stars undergoing thermal pulses while evolving towards the white dwarfs graveyard.

All Type-II Cepheids mainly pulsate in the fundamental model, though a few first-overtone (Soszyński et al., 2019a) and double-mode (Smolec et al., 2018) BL Her pulsators have been discovered across the Bulge and the LMC. Furthermore, ∼ 15% of W Vir stars show a slightly different light curve morphology, wherein their rising branch is steeper than the descending one. These "peculiar" W Vir stars (pW Vir) discovered by Soszyński et al. (2008) are brighter and bluer than their "regular" counterparts, with ∼ 50% of them exhibit-

ing signs of either eclipsing or ellipsoidal binarity (Soszyński et al., 2018; Plachy et al., 2021). Their spatial distribution in the LMC, and their absence from Milky Way globular clusters are consistent with them being younger than other Type-II Cepheids (Matsunaga et al., 2009). Lastly, RV Tauri stars exhibit alternating deep and shallow minima in their light curves. They are divided into two sub-groups: the RVa-type which maintain a constant mean brightness over long term, and the RVb stars which exhibit an additional long-period trend in their mean magnitudes (Bódi & Kiss, 2019), possibly due to a circumbinary dusty disk (Kiss & Bódi, 2017).

Except for the thin disk, Type-II Cepheids are found in all regions of the Galaxy (Soszyński et al., 2009; Ripepi et al., 2023; Matsunaga et al., 2006; Dékány et al., 2019), the Magellanic Clouds (Soszyński et al., 2018), and even external galaxies (Majaess et al., 2009; Kodric et al., 2018). Although the PL relation of Type-II Cepheids is systematically fainter than that of classical Cepheids by ∼ 1.5–2 mag, their accurate distances can be used to determine structural properties of old systems, where they closely follow the spatial distribution of RR Lyrae stars.

I.3 Milky Way and the Magellanic Clouds

The Milky Way is a barred spiral galaxy and, along with the Andromeda galaxy (also known as M31), it forms the Local Group of galaxies. Our position in the Milky Way allows us to resolve the stellar populations of the Galaxy finely. However, it also hampers our ability to directly observe its expansive structure in its totality. Nonetheless, the intricate structure of the Milky Way can be appreciated from our vantage point, as shown in Fig. I.5. Based on the 3-d distribution of the stars shown in this figure, an artist's depiction of the complete structure of the Milky Way can be imagined. Figure I.6 presents one such demonstration of a face-on and an edge-on view of the Galaxy.

Highlighted in these figures are the various substructures of our galaxy:

(i) the Galactic bulge, initially characterized by Baade (1946), and whose shape was determined to be boxy/peanut-like by Weiland et al. (1994); (ii) this bulge consists of the Galactic bar (Shen et al., 2010) and also plays host to the central supermassive black hole of the Galaxy, i.e., Sagittarius A (Balick & Brown, 1974)[1]; (iii) the Galactic disk, which is flatter and more spread out as compared to the bulge. This disk comprises two components: the thin disk and the thick disk (e.g., Gilmore & Reid, 1983). The thin disk is concentrated towards the Galactic midplane and consists of metal-rich stars following an ordered rotation pattern

---

[1]In a striking example of modern technological advancements, Event Horizon Telescope Collaboration et al. (2022) presented an image of the shadow of Sagittarius A*. This image stands as a testament to human ingenuity.

i.e., the thin disk is kinematically cold. On the other hand, the thick disk is kinematically hotter and has existed for at least a few Gyr (e.g., Bland-Hawthorn et al., 2019). The many spiral arms of the Galaxy also traverse the Galactic disk (Reid et al., 2019).

(iv) The Milky Way halo consists of older, relatively metal-poor stars. It encompasses both the bulge and the disk. Recent chemo-dynamical investigations (e.g., Naidu et al., 2020, 2021; Bonaca et al., 2021; Malhan et al., 2022) of the Galactic halo give rise to a scenario where over 80% of the halo is formed ex-situ. In other words, a major fraction of the halo is made up of stars that were born not in the Milky Way but in other galaxies that have since been accreted by the Milky Way. A majority of the numerous globular clusters and dwarf galaxies in the Milky Way halo (shown in Fig. I.7) also shares the same ex-situ origin (Malhan et al., 2022, and references therein).

The Large and Small Magellanic Clouds (LMC and SMC) are two of the most easily-identifiable objects on the Southern sky. They are dwarf irregular galaxies that are interacting with each other, while simultaneously continuing to fall towards the Milky Way. Recent analyses (e.g., Vasiliev, 2023) that reconstruct the orbit of the Magellanic system reveal that the Clouds are quite possibly on their second infall towards the Milky Way. Their interplay with the Milky Way, and the Sagittarius dwarf spheroidal galaxy (the remnant of a previous massive merger of the Galaxy, Vasiliev & Belokurov 2020) is well characterized in the literature (see e.g., Vasiliev et al., 2021).

Figure I.8 depicts the number density of stars in the Magellanic Clouds as measured by the *Gaia* satellite (described below). The LMC (on the left in Fig. I.8) and the SMC (on the right) constitute two of the very few galaxies whose stellar populations can be resolved with such great detail. Apart from the readily apparent disk and loose spiral structure of the LMC, other substructures are also discernible. For instance, the Magellanic bridge, which is an almost contiguous stream of stars flowing from the SMC to the LMC, is seen in blue. Due to their interaction with each other and the Milky Way, the Magellanic Clouds are marked by many such substructures. A summary can be found in Gaia Collaboration et al. (2021b) and Jacyszyn-Dobrzeniecka et al. (2020b).

I.3.1 Milky Way Archaeology in the *Gaia* era

Milky Way Archaeology, or the study of the structure and formation of the Milky Way, has been revolutionized by the *Gaia* satellite (Gaia Collaboration et al., 2016). Starting with its second data release (DR2, Gaia Collaboration et al. 2018a), and continuing through its third data release (DR3, Gaia Collaboration et al. 2023c), *Gaia* data has played in instrumental role in reshaping the paradigms of Milky Way Archaeology. *Gaia*'s primary aim is to perform all-sky astrometry, i.e., to measure the positions, proper motions, and parallax (or
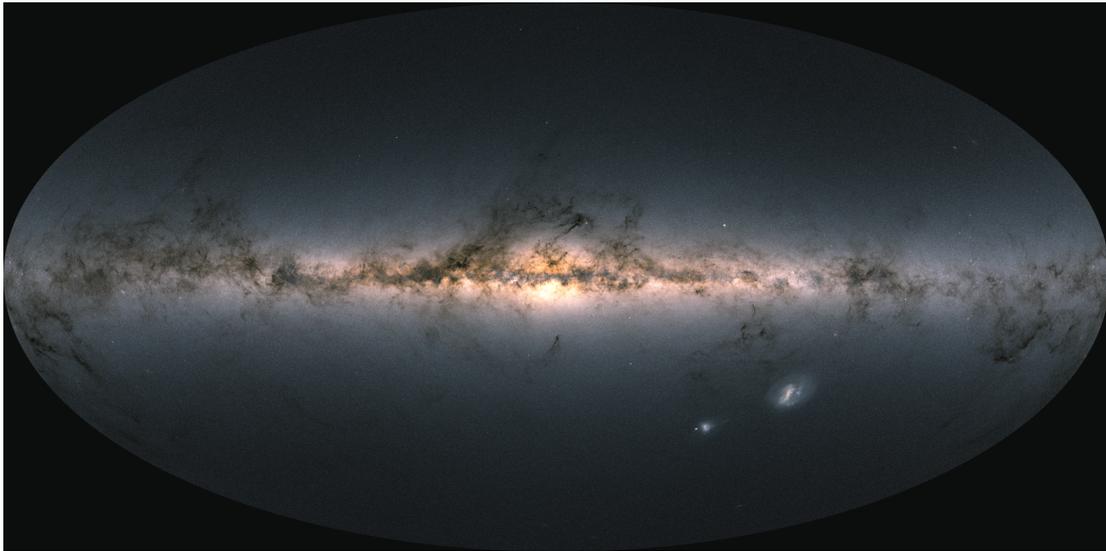
FIGURE I.5: An all-sky map of the Milky Way built using the third data release (DR3) of the *Gaia* satellite. The Large and Small Magellanic Clouds are also visible towards the bottom on the right side. This map is a true composite created by combining the *Gaia* photometry in three different bands and by measuring the number density of stars in each pixel. Credits: ESA/Gaia/DPAC; CC BY-SA 3.0 IGO. Acknowledgement: A. Moitinho.
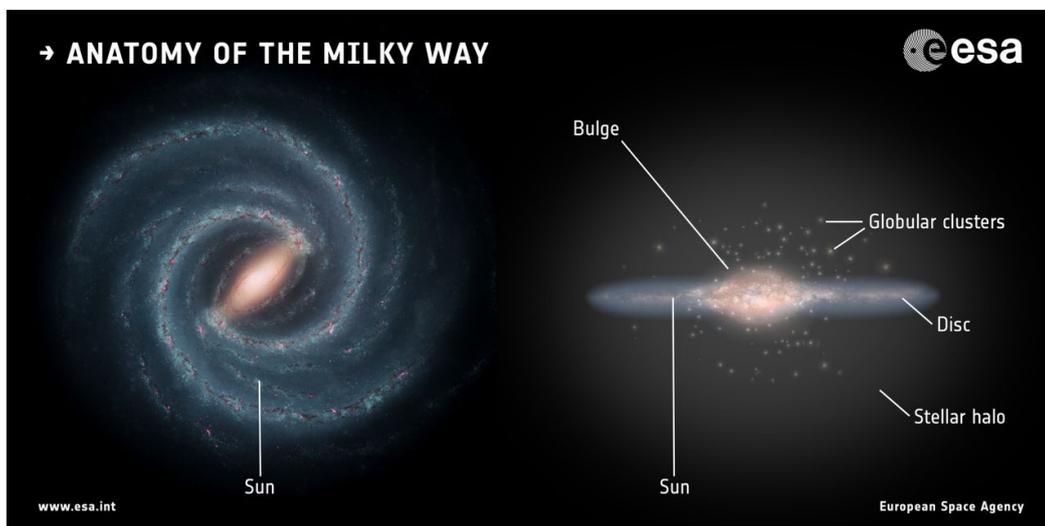
FIGURE I.6: An artist's impression of the structure of the Milky Way. The position of the Sun is highlighted in both the plots. The *left* panel presents an imagined face-on view of the Milky Way, with its spiral structure and the central bar clearly discernible. The *right* panel presents an edge-on view and outlines the various substructures of the Galaxy. Credits: Left: NASA/JPL-Caltech; right: ESA; layout: ESA/ATG medialab.

FIGURE I.7: The Milky Way subsystems. The positions of Galactic globular clusters (in blue) and dwarf galaxies (in red) are highlighted on a gray-scale number density map of stars. Credits: ESA/Gaia/DPAC, CC BY-SA 3.0 IGO. Acknowledgement: Gaia Data Processing and Analysis Consortium (DPAC); A. Helmi / M. Breddels, Kapteyn Astronomical Institute, University of Groningen, The Netherlands; F. van Leeuwen, University of Cambridge, UK; P. McMillan, Lund Observatory, Sweden.

FIGURE I.8: A depiction of the number-density of stars in and around the Magellanic Clouds. The Magellanic bridge can also be seen emanating from the Small Magellanic Cloud. Credits: ESA/Gaia/DPAC – CC BY-SA 3.0 IGO. Acknowledgments: This image was created by Laurent Chemin using the data from Gaia Collaboration et al. (2021b).

distance) of the stars. It undertakes this task with an unparalleled level of precision, pushing the boundaries to achieve micro-arcsecond astrometry (see, e.g., Lindegren et al., 2021b).

In addition to its groundbreaking astrometric data, *Gaia* also measures or provides data to estimate various other stellar parameters. A handful of these properties are illustrated in Fig. I.9. For an overview of the complete *Gaia* inventory, please refer to Gaia Collaboration et al. (2023c). These data have been used to reconstruct the merger history of the Milky Way (e.g., Bonaca et al., 2021; Naidu et al., 2021, 2020; Malhan et al., 2022, also see Fig. I.10), study the internal kinematics of the Magellanic Clouds (e.g., Gaia Collaboration et al., 2021b, see Fig. I.11), and find members of the Milky Way subsystems like the globular clusters, dwarf galaxies, and streams (c.f. e.g., Vasiliev & Baumgardt, 2021; Gaia Collaboration et al., 2018b; Malhan & Ibata, 2018, and references therein).

The distance of our stars, used in synergy with their *Gaia* data, can lead to a whole host of applications that will be explored throughout this dissertation. As a cursory example, Fig. I.12 demonstrates the power of *Gaia* data leveraged in concert with the distance of our stars. The left panel of this figure presents the Galactocentric distribution of classical Cepheids. These Cepheids reside in the disk of the Milky Way and their distance was computed based on the PW relations and the methods outlined in this thesis. Using a manifold learning algorithm (presented in Lemasle et al., 2022), these Cepheids were categorized into various groups. By modeling the fiducial ridge lines (shown in green) of the contiguous groups, an analytical approximation of the Milky Way spiral arms can be generated. This spiral arm model is compared with the widely-accepted model of the Milky Way spiral structure from Reid et al. (2019) in the right panel. The Lemasle et al. (2022) model is not only in good agreement with the Reid et al. (2019) model, but it is also able to trace the spiral arms at greater distances.

Along with their distance, intrinsic parameters of these stars can also be measured using photometry alone. The model parameters of their Fourier-fit light curves can be used to measure their effective temperature, luminosity, mass, radius, and metallicity (Dékány & Grebel, 2022; Plachy et al., 2021; Bellinger et al., 2020). Naturally, with new findings (of the *Gaia* era), come unanswered questions. The properties of the progenitors of many newly-discovered Milky Way mergers remain shrouded in mystery (e.g., Bonaca et al., 2021; Prudil et al., 2021). Through a spectroscopic analysis of stars attributed to a certain merger, its progenitor can be found using chemical tagging (Li et al., 2019; Ji et al., 2020, and references therein). However, the aforementioned properties of our stars facilitate the undertaking of a preliminary chemo-dynamical investigation of these mergers using photometry alone.
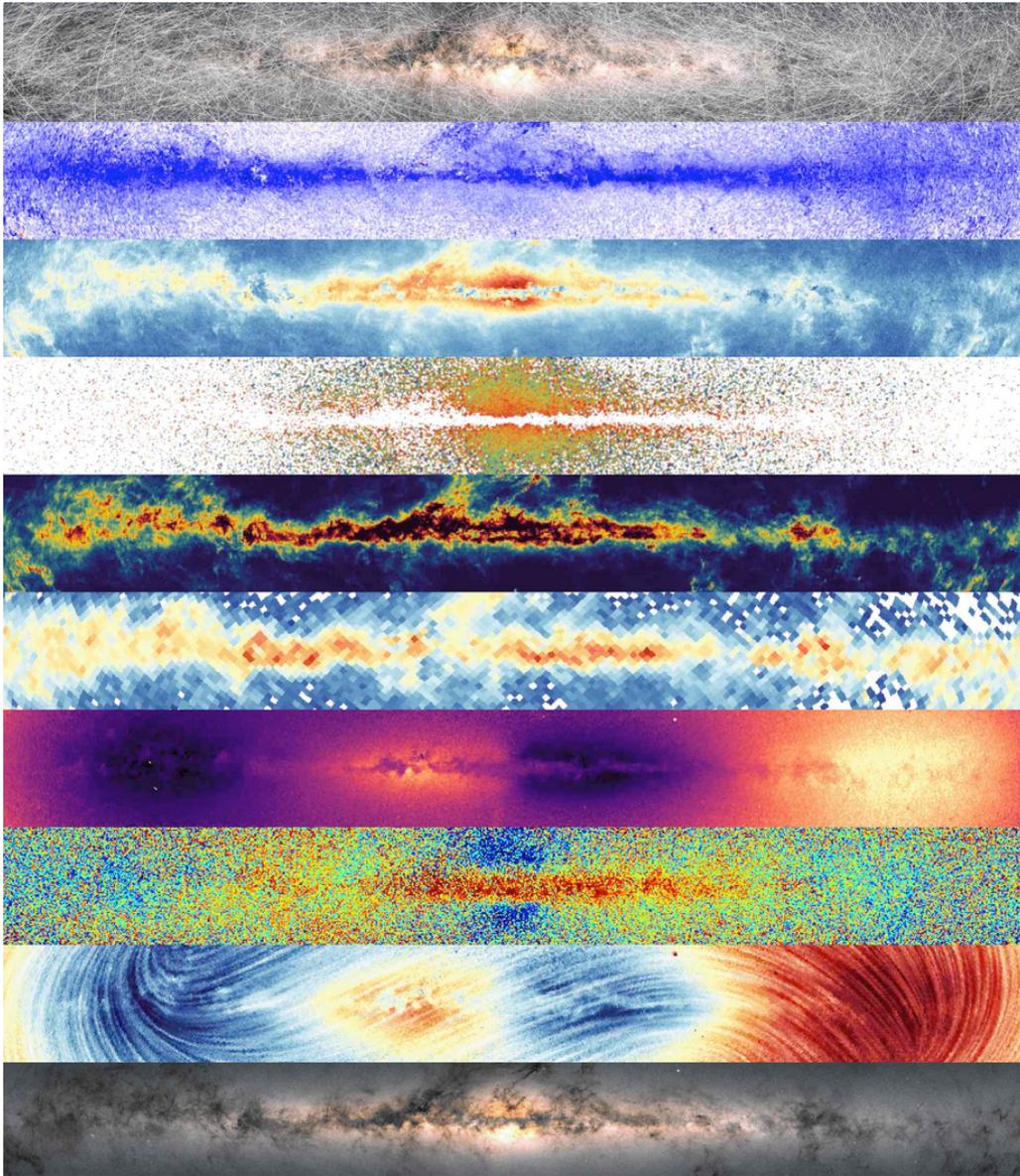
FIGURE I.9: Composition of *Gaia* sky maps where projections of the central regions of the Milky Way are overlaid with the following *Gaia*-based properties (from top to bottom): (1). Integrated orbits of a subset of stars with 6d phase space information measured in *Gaia* DR3. (2). Stellar ages computed by *Gaia* DPAC. (3). Color of the stars in the *Gaia* filters. (4). The *Gaia* sample of variable stars. (5). Extinction as measured by *Gaia*. (6). Distribution of the diffuse interstellar bands (used to study the interstellar medium). (7). Heliocentric radial velocities measured by *Gaia*. (8). Metal content of the stars. (9). 3-d map of the stellar motions built using *Gaia* proper motions and radial velocities. (10). Number density of stars detected by *Gaia* (also presented in full in Fig. I.5). Many of these properties will be used throughout this thesis. Credits: ESA/Gaia/DPAC - CC BY-SA 3.0 IGO. Created by Tineke Roegiers, based on different Gaia sky maps created by ESA/Gaia/DPAC - CC BY-SA 3.0 IGO and inspired by NASA's Multiwavelength Milky Way Images.
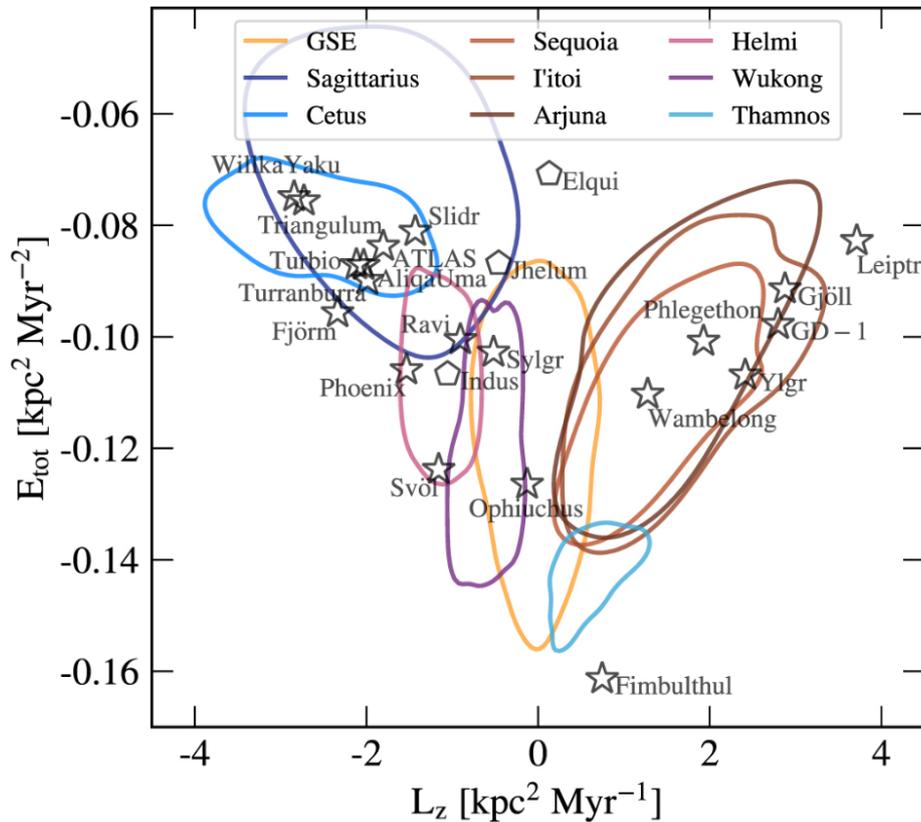
FIGURE I.10: Integrals-of-motion projection of stars in the the Milky Way halo. The *x*-axis represents the angular momentum of stars along the Galactocentric *z*-direction, while the *y*-axis represents their total energy. The "star" and "pentagon" markers denote the Milky Way stellar streams, while the solid contours represent the merger events of the Milky Way. The stars, now part of the Milky Way halo, due to the numerous mergers of the Milky Way, retain information of their progenitor's integrals of motion. The various distributions outlined here point to an active past of the Milky Way and many substructures in its present-day halo. Credits: Image taken from Bonaca et al. (2021). *Reproduced by permission of the AAS and the author.*
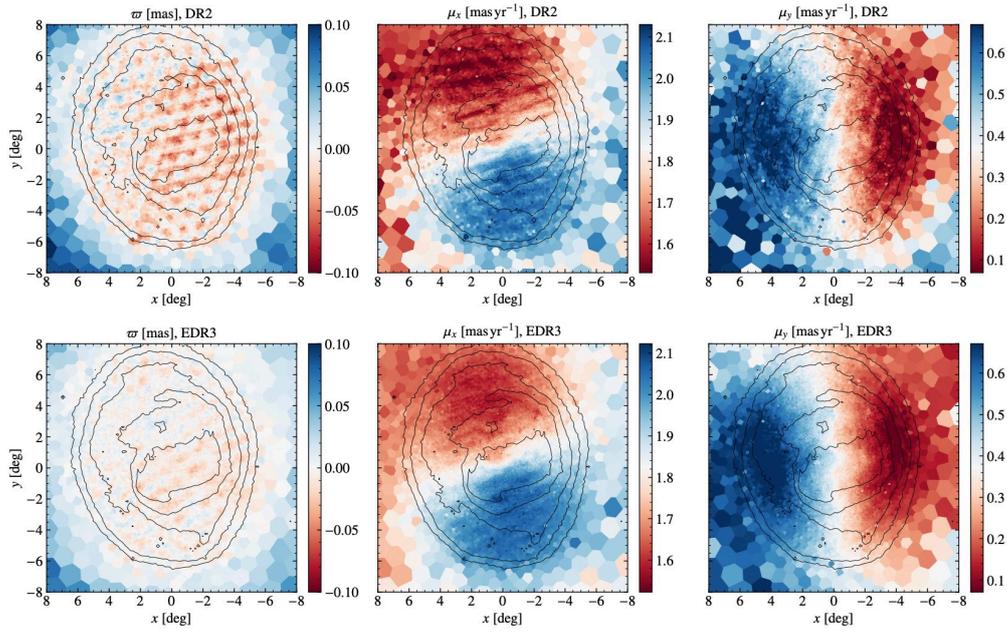
FIGURE I.11: The distributions of *Gaia* astrometry for the stars in the Large Magellanic Cloud. The LMC-centric proper motion distributions are presented in the *middle* and *right* panels, while the parallax distribution is plotted in the *left* panel. The *top* row contains data from *Gaia* DR2, while the *bottom* row plots these quantities from *Gaia* DR3. A marked improvement in the handling of systematic errors in *Gaia* DR3 data is readily apparent. Image taken from Gaia Collaboration et al. (2021b).
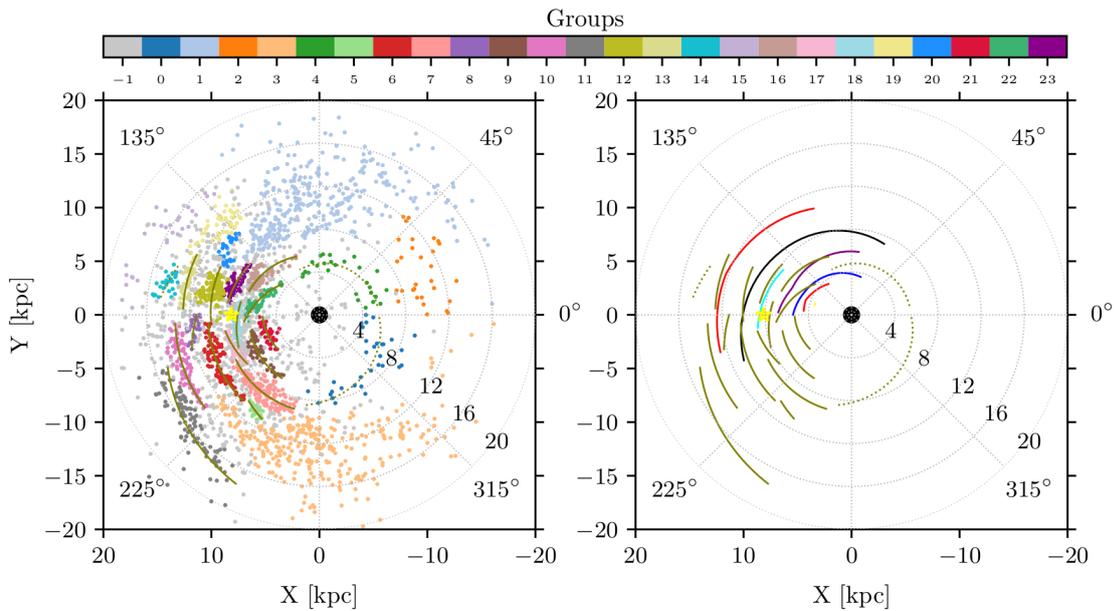
FIGURE I.12: The Milky Way spiral arms modeled using classical Cepheids from our compilation. Credits: Image adapted from Lemasle et al. (2022).

I.4   The Hubble Tension

The Hubble constant (H_0, Hubble 1929) constitutes one of the six cosmological parameters that describe the properties and evolution of the universe (Planck Collaboration et al., 2020). The value of H_0 sets one of the estimates of the age (i.e., the Hubble time, defined as $1/H_0$) and the size (i..e., the Hubble radius $= c/H_0$, where c is the speed of light) of the universe. Over the past two decades, modern observations of the cosmic microwave background (CMB) have led to a precise determination of $H_0 = 67.4 \pm 0.5\,\mathrm{km\,s^{-1}Mpc^{-1}}$ (see Fig. I.13). H_0 estimates of the local universe are based on the distances of supernovae-host galaxies. These distances are in turn, computed using PL relations of classical Cepheids. This mode of computation leads to a value of $H_0 = 73.04 \pm 1.04\,\mathrm{km\,s^{-1}Mpc^{-1}}$, which deviate by more than 5σ from the CMB-based measurement (c.f. e.g., Riess et al., 2022, 2020).

These Cepheids-based determinations rely mostly on the Hubble Space Telescope (*HST*) data and many independent analyses of it (e.g., the Hubble Key project, Freedman et al. 2001; the SH0ES collaboration, Riess et al. 2022). Many possible sources of systematic errors in both the *HST* data and the properties of classical Cepheids have been analyzed (see Riess & Breuval, 2023, for a recent review). However, the tension lingers on. The latest determinations of the Cepheids-based local H_0 computed using data from the James Webb Space Telescope (*JWST*) do not resolve the tension either (Riess et al., 2023).

Other distance indicators like the tip of the red giant branch (TRGB, Lee et al. 1993) do lead to an H_0 value (shown in red in Fig. I.13) much closer to that computed by Planck Collaboration et al. (2020). This result, coupled with the preliminary results from the *JWST*, practically singles out classical Cepheids as the main culprit of the Hubble tension. Thus, a better understanding of the systematic errors plaguing the PL relations of classical Cepheids is needed in earnest urgency.

Population-II stars like the RR Lyrae stars and Type-II Cepheids are much older than the classical Cepheids. Thus, they can be detected in elliptical as well as spiral galaxies. Their application as extragalactic distance indicators has been theorized (e.g., Beaton et al., 2018, and references therein) but seldom undertaken. This reluctance is based on the fact that these stars are fainter than the classical Cepheids and hence, their detection in supernovae-host galaxies is challenging. Moreover, the PL/PW relations of Type-II Cepheids are not as well calibrated as those of the classical Cepheids (c.f. e.g., Ripepi et al., 2023; Ngeow et al., 2022c; Bhardwaj, 2020). However, with better data and contemporary fitting methods, RR Lyrae stars (Savino et al., 2022) and Type-II Cepheids (Ngeow et al., 2022c, Pipwala et al., in prep.) have been used to estimate the distance to M31. Thus, these analyses have taken the first steps in probing the potency of RR Lyrae stars and Type-II Cepheids as anchors of H_0.
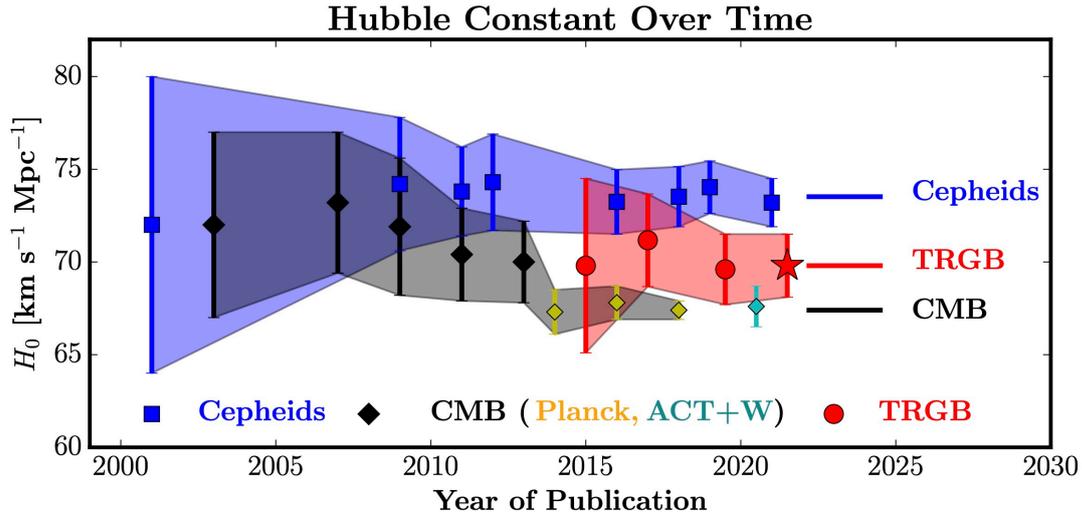
FIGURE I.13: The Hubble tension. The *y*-axis represents the measured H_0 values and the *x*-axis denotes the years in which these values were released. Credits: Image taken from Freedman & Madore (2023).

Furthermore, in the infrared wavelengths, RR Lyrae stars and Type-II Cepheids likely follow the same PL relations (e.g., Braga et al., 2020). Thus, the combination of these stars can be used as an even more potent distance indicator. The work presented here will lay the groundwork of testing these hypotheses empirically.

I.5   This thesis

I.5.1   Motivation

With advancements in instrumentation and data processing, hundreds of thousands of RR Lyrae stars and Cepheids (called "target stars" henceforth) have been classified within ~ 200 kpc. Our objective is to compile a comprehensive list of all objects classified as RR Lyrae or Cepheids in the Milky Way and in the Magellanic system. The Optical Gravitational Lensing Experiment (OGLE) survey (Udalski et al., 2008, 2015) is a monumental effort in variable star classification and forms the cornerstone of this collection. The survey monitors over 3000 deg^2 of the Southern sky, and achieves exceptional accuracy in classifying our target stars through visual inspection of well-populated light curves. As discussed earlier, the *Gaia* mission (Gaia Collaboration et al., 2016) has been revolutionary for Galactic astronomy. In addition to the properties already discussed, *Gaia* also performs all-sky photometry, which has been used to discover millions of variable stars (Eyer et al., 2023), including hundreds of thousands of our target stars (Clementini et al., 2023; Ripepi et al., 2023). However, limited by the relatively fewer number of observations in its third data

release (DR3, Gaia Collaboration et al. 2023c), it does not reach the completeness or purity levels of OGLE for several subtypes. In our present undertaking, we have combined RR Lyrae stars and Cepheids from both OGLE and *Gaia* with those from six other surveys.

We focused exclusively on optical surveys due to inherent challenges in classifying these subtypes in the infrared (IR, Braga et al. 2019). However, these optical surveys are mutually complementary and supplementary in various aspects. For instance, OGLE provides extensive coverage of the Magellanic Clouds, the Galactic bulge, and disk but does not cover the Galactic halo. On the other hand, the Zwicky Transient Facility (ZTF, Bellm et al. 2019) and PanSTARRS1 (PS1, Chambers et al. 2016) respectively observe the brighter and fainter stars in the Northern hemisphere, while the Dark Energy Survey (DES, Abbott et al. 2021) monitors the Southern hemisphere but not the Magellanic Clouds, which are anyway already covered by OGLE. *Gaia*, the Catalina Sky Survey (CSS, Drake et al., 2012), the All-Sky Automated Survey for Supernovae (ASAS-SN, Shappee et al. 2014; Kochanek et al. 2017), and the All Sky Automated Survey (ASAS, Pojmanski 1997) do provide full-sky coverage but have different detection efficiencies at different magnitudes.

Even when surveys have overlapping regions of observations, they may have different number of epochs and hence, varying degrees of confidence in the classification of the same star. Or due to their differing detection limits, they may not be able to photometer the star altogether. Besides, their classification pipelines and period-finding algorithms are largely different and may lead to an incorrect classification or period determination of poorly-observed stars. The motley of surveys collected here and the systematic steps taken to homogenize them help minimize the impact of such differences among surveys.

Furthermore, the distance to all these subtypes has never been computed under the umbrella of a single analysis or using the same method. The computation of their period-luminosity and period-Wesenheit relations requires a careful consideration of numerous aspects like the photometry, distance calibrators, periods, classification, and their uncertainties. By leveraging the properties of our catalog, an unprecedented number of such PL/PW relations can be computed homogeneously. These relations lead to the determination of the distance to our target stars at a high level of accuracy and precision.

These distances (and many other properties developed over the course of this dissertation) of our stars can be used as a stepping stone to inch closer towards answering the following fundamental questions: (i) How did the Milky Way form? What is its current structure and composition? (ii) How fast is the local Universe expanding? How do we resolve the Hubble tension?

I.5.2 Structure

The structure of the thesis is as follows.

- In Chapter II, I describe the data I used and the new method that I developed to construct the most homogeneous and the largest ever catalog of Cepheids and RR Lyrae stars.

- In Chapter III, I outline the results of both an internal and an external validation of the classification and the periods of these stars.

- With this validation in place, I cross-matched this catalog with numerous literature surveys to accrue photometric and spectroscopic data. These ingredients, imperative for a 6-d analysis of these stars, are presented in Chapter IV.

- In Chapter V, I demonstrate one application of the immense potential of the catalog by computing a $\sim 99\%$ accurate distance to the LMC, derived using ~ 200 new period-Wesenheit relations.

- I summarize the results of this dissertation and discuss future applications of the catalog in Chapter VI.

II

The Catalog

> *You can find the entire cosmos lurking in its least*
> *remarkable objects.*
>
> – WISLAWA SZYMBORSKA

In this chapter, we present the creation of the RRLCep catalog. This catalog is the result of the implementation of numerous rigorous measures. These efforts ensured consistent cross-matches, classification, periods, and distance determination of all the stars. This chapter sheds light on the first two steps of our endeavor. We describe all eight surveys in Sect. II.1. In Sect. II.2, we illustrate the performance of a new cross-match algorithm we developed to merge all eight catalogs.

II.1   Survey description and data requisition

In each subsection, we present an overview of the survey and their specific catalogs that we utilized. All the catalogs discussed here were standardized as per the following procedure — we ensured their equatorial coordinates (α, δ) and magnitudes were within physical ranges. We transformed the classification name given by each survey to follow this naming scheme: DCEP_F, DCEP_1O, DCEP_mul, ACEP_F, ACEP_1O, BL HER, W VIR, RV TAU, RRab, RRc, RRd (these subtypes have been introduced earlier). We ensured subtypes from each catalog adhered to their respective period ranges listed in Table II.1. Furthermore, variable

23

TABLE II.1: Empirical boundaries on variability periods for various subtypes.

Class	Subtype	Period (d)
RR Lyrae	RRab	≤ 0.998
	RRc	$0.2 < P \leq 0.6$
Classical Cepheids (DCEP)	DCEP_F	> 0.8
	DCEP_1O	$0.2 < P < 10$
Type-II Cepheids	BL Her	$1.002 \leq P < 5$
	W Vir	$5 \leq P \leq 20$
	RV Tau	> 20
Anomalous Cepheids (ACEP)	ACEP_F	≤ 2.5
	ACEP_1O	$0.4 \leq P \leq 1.2$

stars with periods (in days) in the interval [0.998, 1.002] detected using ground-based observations alone might exhibit false periodicity induced by the Earth's rotation. Hence, we removed such variables.

We placed emphasis on retrieving or estimating uncertainties in coordinates, periods, and photometry of each survey, as these factors are used to cross-match all survey catalogs (elaborated in Sect. II.2). The correct estimation of uncertainties in magnitudes and periods is especially important in the context of distance determination through period-luminosity (PL) or period-Wesenheit (PW) relations. Additionally, we computed the number of observations per star (N_{obs}) for all surveys, as it plays a crucial role in the internal validation of periods (discussed in Sect. III.1.1) and in the following steps.

For a few surveys, the unique identifier associated with each star had been constructed using the star's equatorial coordinates. In some cases, the same identifier was assigned to more than one star, plausibly due to a rounding off limit of the coordinates. We dealt with these duplicated stars by retaining the one with the higher number of observations.

Since many surveys included here combined data from several small telescopes, we performed an internal cross-match to check for duplicate stars within each catalog. Using the average FWHM (full-width at half maximum of the image PSF) of a survey's images as the cross-match radius, we created an on-sky neighborhood for every star in its catalog. Pairs with similar periods, magnitudes, and classification were considered duplicated and only the star with the higher number of observations was retained. This step was performed on the catalogs from Catalina (Sect. II.1.2), ASAS-SN (Sect. II.1.4), and ASAS (Sect. II.1.6) surveys. A summary of the relevant features of all surveys is presented in Table. II.2.

TABLE II.2: Properties of surveys whose RR Lyrae and Cepheid catalogs are combined in this work.

Notes. (a) Number of telescopes used by the survey and their hemispheric position on Earth: N: North; S: South. For *Gaia*, L2: second Lagrangian point for the Earth–Sun orbit. (b) Diameter of the mirror (or lens in case of ASAS-SN). For Catalina, the mirror sizes are listed in the order: SSS, CSS, MLS. (c) Field of view given in x degrees. The effective field of view can be computed from this as $x \times x \deg^2$. (d) Filters used for observation. Measurements for filters listed in parentheses are either not accessible or not included here. (e) Observation period of the data included here. (f) Period-finding algorithm employed: LS: Lomb-Scargle periodogram; GLS: Generalized Lomb-Scargle; BLS: Box Least Squares; AFD: Adaptive Fourier Decomposition. (g) RF: Random Forests; Cuts: Constraints on color, periods, and Fourier parameters; VI: visual inspection; MKDE: Multi-variate kernel density estimation.

References. (1) Pojmanski (1997); (2) Pojmanski (2002); (3) Richards et al. (2012); (4) Shappee et al. (2014); (5) Kochanek et al. (2017); (6) Jayasinghe et al. (2018); (7) Jayasinghe et al. (2020); (8) Christy et al. (2023); (9) Drake et al. (2012); (10) Drake et al. (2013b); (11) Drake et al. (2013a); (12) Drake et al. (2014); (13) Torrealba et al. (2015); (14) Drake et al. (2017); (15) Dark Energy Survey Collaboration et al. (2016); (16) Abbott et al. (2021); (17) Stringer et al. (2019); (18) Stringer et al. (2021); (19) Gaia Collaboration et al. (2016); (20) Rimoldini et al. (2023); (21) Clementini et al. (2023); (22) Ripepi et al. (2023); (23) Udalski et al. (2015); (24) Udalski et al. (2018); (25) Soszyński et al. (2017a); (26) Chambers et al. (2016); (27) Flewelling et al. (2020); (28) Magnier et al. (2020a); (29) Magnier et al. (2020b); (30) Sesar et al. (2017); (31) Bellm et al. (2019); (32) Graham et al. (2019); (33) Chen et al. (2020);

Survey	$N_{\text{telescope}}$[a]	Size[b] (m)	FOV[c] (deg)	Filters[d]	Obs. Per.[e]	Astrometry Epoch	FWHM ('')	Photometric Range (mag)	PF Alg.[f]	Classification Pipeline[g]	Ref.
ASAS	2 (S)	0.7	8.5	$V, (I)$	2000-03	J2000.0	30	$8 \leq V \leq 14$	Fast LS	RF / Cuts	1–3
ASAS-SN	20 (N, S)	0.14	4.5	V, g	2013-20	J2000.0	16	$10\text{-}11 \leq V \leq 17$ / $12 \leq g \leq 18.5$	GLS / BLS	RF / VI	4–8
Catalina	3 (N, S)	0.5 / 0.7 / 1.5	4.2 / 8.2 / 1.1	V_{CSS}	2004-15	J2000.0	3–6	$11 < V_{\text{CSS}} < 20$	LS / AFD	MKDE / VI	9–14
DES	1 (S)	4	3	$griz y_{\text{DES}}$	2013-19	J2015.5	1	$15 < g_{\text{DES}} < 24$	Templates	Cuts / RF	15–18
Gaia	2 (L2)	0.7	0.7	G $G_{\text{BP}}, G_{\text{RP}}$	2014-17	J2016.0	0.2	$6 < G < 21$	GLS / Period04	Cuts / XGBoost	19–22
OGLE	1 (S)	1.3	1.4	V, I	2010-20	J2000.0	1.3–1.5	$10 < I < 21.7$	FnPeak / Tatry	VI / Cuts	23–25
PS1	1 (N)	1.8	3	$griz_{\text{P1}}$	2010-14	J2012.5	0.7–1.3	$14 < gri_{\text{P1}} < 21.5$	Gatspy Template	XGBoost	26–30
ZTF	1 (N)	1.2	6.9	$g_{\text{P1}}, r_{\text{P1}}, (i_{\text{P1}})$	2018-19	J2015.0	2	$12.5 \leq r_{\text{P1}} \leq 20.5$	GLS	DBSCAN / Cuts	31–33

II.1.1   OGLE-IV

The Optical Gravitational Lensing Experiment (OGLE) has been surveying the sky for more than three decades. Started in 1990, the survey was aimed to build on ideas developed by Paczynski (1986) for the detection and characterization of microlensing events in the Milky Way. Since then, it has also been a widely used resource for the study of variable stars. We only focused on the catalogs generated in the fourth phase of the survey (OGLE-IV, Udalski et al. 2015). In principle, one can concatenate the light curves collected in OGLE-II and OGLE-III (Udalski et al., 2008) with those in OGLE-IV. However, one needs to take into account the systematic differences between the mean brightness of stars across these phases. Hence, we only focused on OGLE-IV catalogs, which, anyway, are the most comprehensive of the three.

The survey monitors over 3000 deg^2 across the Galactic bulge, disk, and the Magellanic system using a 1.3 m telescope situated at the Las Campanas Observatory in Chile. It undertakes observations using *V*- and *I*-band filters with the latter getting considerably more time. For our target stars in the Magellanic system, the average number of observations in the *I*-band is almost five times that of the *V*-band (~ 480 vs ~ 90). The *I*-filter light curves were mainly used for the classification of variable stars, and *V* magnitudes were used in conjunction to estimate the color.

The OGLE team performed a preliminary selection of variable stars (and their subtypes) using Fourier decomposition and template-fitting of *I*-band light curves. However, purity of their catalog stems not only from the high quality of their light curves but also from the fact that each light curve is classified manually after visual inspection. In ambiguous cases, the final decision is taken after taking into account other parameters of the stars like their position in the CMD and other diagnostic plots.

We combined RR Lyrae stars and Cepheid catalogs from the following OGLE studies: Soszyński et al. (2015a,b, 2016b, 2017a,b); Udalski et al. (2018); Soszyński et al. (2019b,c, 2020b). They cover the Magellanic Clouds, the Galactic disk and bulge. The Magellanic Clouds' catalogs contain line-of-sight halo RR Lyrae stars as well, given that there is no natural luminosity boundary between the two systems (Soszyński et al., 2016b). The OGLE *I*-band light curves of randomly selected stars belonging to each target subtype are plotted in Fig. I.3.

After combining the catalogs, we determined the number of observations (N_{obs}) and the photometric uncertainty from the light curves of individual stars. The final uncertainties in the *I* and *V* magnitudes are computed as the median of the errors on epoch magnitudes provided in the light curves. For ~ 1500 stars that are missing an OGLE-IV light curve, we determined the uncertainty in their photometry by fitting the *I* vs I_{error} distribution of
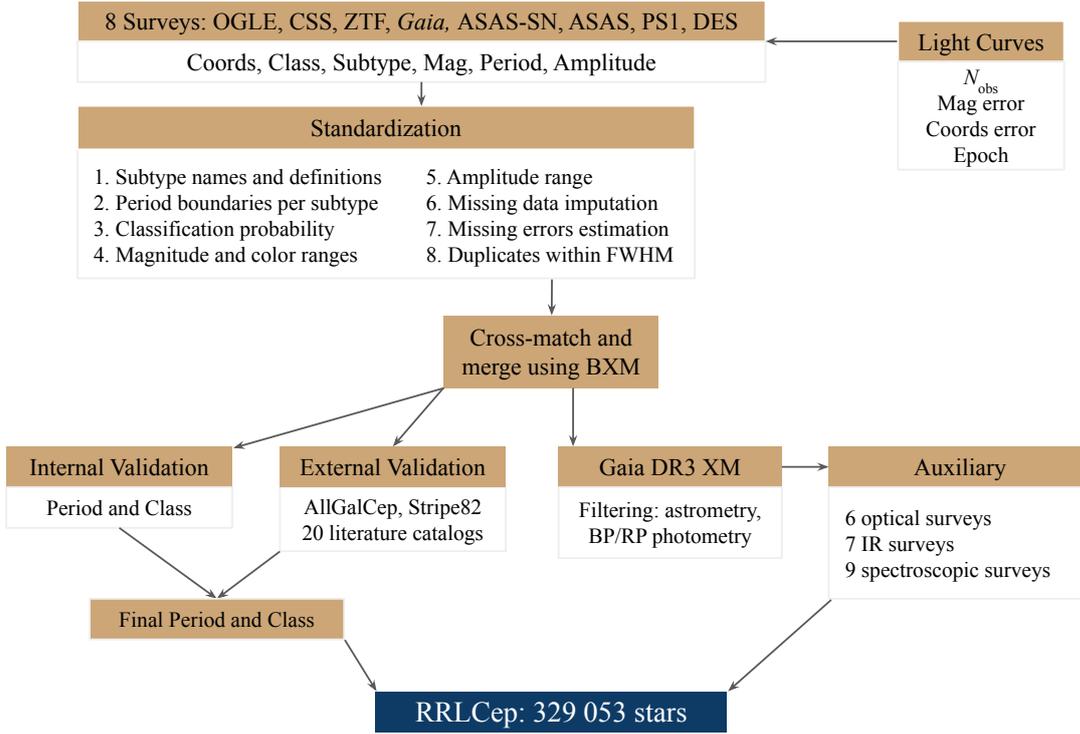
FIGURE II.1: Schematic representation of the workflow. The surveys, their contents, and the standardization of their properties are described in Sect. II.1. Post-standardization, all eight catalogs were merged using Bhattacharyya distance-based cross-match (BXM, described in Sect. II.2). The periods and classification of all stars were first validated internally (Sect. III.1.1 and III.1.2), and then, using literature catalogs (Sect. III.2). The merged catalog, called RRLCep, was cross-matched with the *Gaia* DR3 source catalog. Its coordinates were used to cross-match against auxiliary surveys to get optical, IR photometry and radial velocities (Sect. IV).
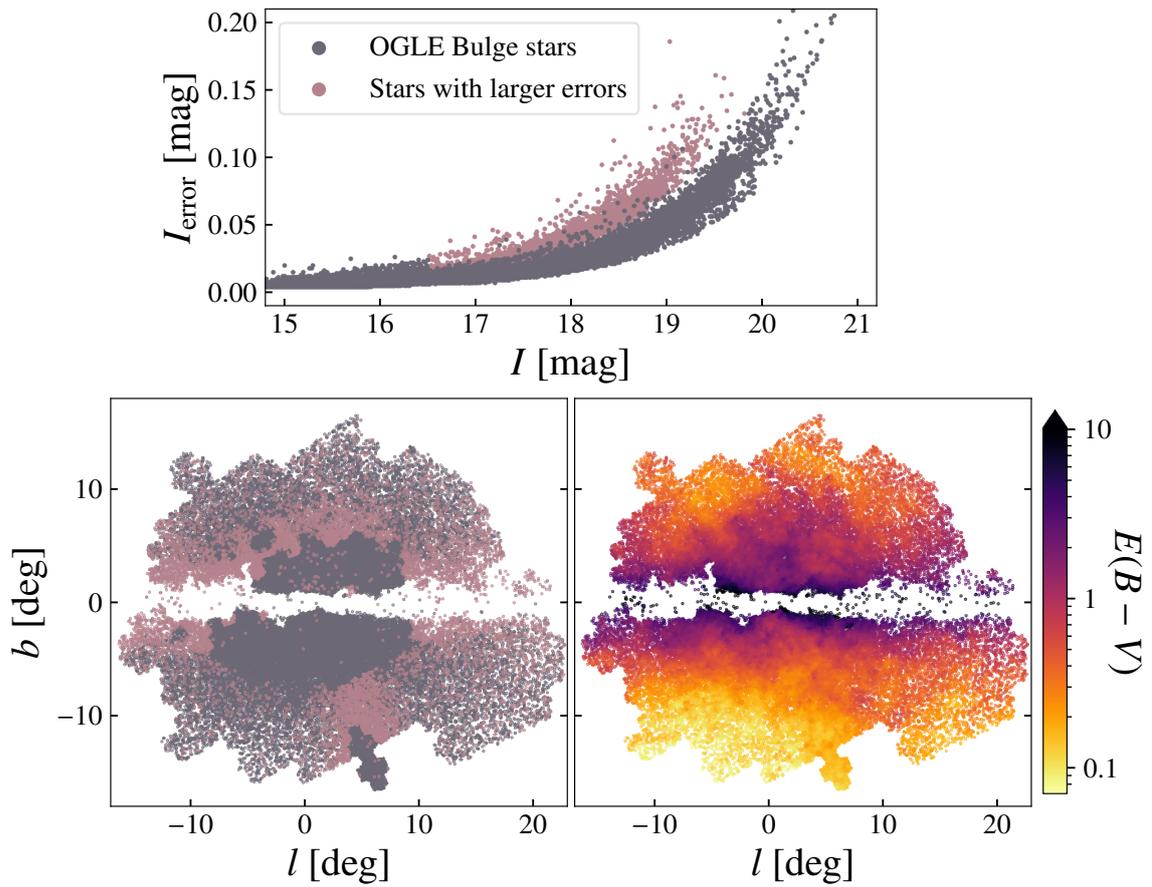
FIGURE II.2: Errors in *I*-band photometry of OGLE bulge stars (*top panel*). A subsample of stars had larger errors for a given *I* magnitude. In the *bottom row*, the on-sky distribution of the same stars is plotted in Galactic coordinates. In the bottom right panel, the points are colored according to reddening, E(B-V), taken from the maps of Schlegel et al. (1998) and recalibrated using Schlafly & Finkbeiner (2011).

the remaining stars with a fourth-degree polynomial. This polynomial was fit on a region-by-region basis. The I vs I_{error} distribution for stars in the OGLE Galactic bulge fields is plotted in Fig. II.2. The polynomial was fit in the range $15 \leq I \leq 21$, as the uncertainty floor is ~ 0.007 mag for brighter stars.

We investigated the reason behind the two trends seen in this phase space in the top panel of Fig. II.2. A subset of the stars have larger photometric uncertainties (colored in red). We did not find any strong association of this trend with the number of observations or with the type of star. A weak correlation with extinction can be seen in bottom panels of Fig. II.2. However, we realized that a majority (>95%) of these sources are from Soszyński et al. (2019c, S19) and these stars were observed using shallower exposures of 25 s as opposed to $\gtrsim 100$ s exposure time of other OGLE stars. Our main conclusion from this investigation is that it is essential to avoid assuming a single, underestimated value of photometric uncertainty across magnitude ranges and substructures (as usually done while computing period-luminosity relations).

Soszyński et al. (2015b) estimated OGLE astrometric uncertainty, in comparison to 2MASS (Two Micron All Sky Survey, Skrutskie et al. 2006), to be $0.2''$. This value seems adequate based on our comparison of OGLE coordinates with those of *Gaia* DR3, as discussed in Sect. II.2.

While combining the OGLE catalogs, we recorded the remarks associated with each of them. The flags and the number of stars they affect are given in Table II.3. These stars show one or several of the following phenomena: Blazhko effect, period doubling, irregular changes in period and/or amplitude, secondary periodicity, optical doubling. These stars, along with those with uncertain classification, comprise 356 RR Lyrae stars, 359 classical, 130 Type-II, and 4 anomalous Cepheids. Most of the flagged stars are RR Lyrae stars with uncertain classification towards the Galactic bulge or the LMC. Overall, the final number of OGLE stars in various classes and subtypes is shown in Table II.4.

II.1.2   Catalina Sky Survey

The Catalina Sky Survey (Drake et al., 2012) is a combination of three different sub-surveys, namely, the Catalina Schmidt Survey (CSS), the Mount Lemmon Survey (MLS), and the Siding Spring Survey (SSS). Started in 2003, the survey's main aim is to detect near-earth objects and potentially hazardous objects. Each sub-survey uses eponymous telescopes that have slightly different specifications, as listed in Table II.2). Nevertheless, the photometry between these sub-surveys is consistent as the cameras on these telescopes are nearly identical and the data were processed using the same Catalina pipeline. The cameras capture the images unfiltered to improve throughput and detection efficiency.

TABLE II.3: OGLE flags.

Flag	Information	N_{star}
1	Class uncertain	403
2	Shows secondary period	282
3	In an eclipsing/ellipsoidal binary	42
4	Blended	36
5	Blazkho modulations	22
6	RVb-type	21
7	Double star	18
8	Variable amplitude and/or period	14
9	First-overtone BL Her candidate	5
10	No longer pulsating	3
11	δ Scuti candidate	3

II.1.2.1   Catalogs

Catalina variability catalogs have been released in three stages. First, Drake et al. (2013a,b) collectively used data from all three sub-surveys (predominantly from the Catalina Schmidt survey) to classify RRab stars in the Galactic halo. Later on, Torrealba et al. (2015) used data only from the Siding Spring survey and discovered a large number of new RRab stars in Southern hemisphere. Finally, Drake et al. (2014, 2017) put forth general variable star catalogs classifying $\sim 85\,000$ stars into 17 subtypes which included $\sim 18\,000$ of our target subtypes.

All these studies provided mean magnitudes (and light curves) on the Catalina photometric scale V_{CSS}, i.e., not transformed to Johnson V-band scale. This photometry has a precision and accuracy of ~ 0.1 mag (Drake et al., 2012, 2013a). The best variability period and average magnitude of the stars were determined using the Adaptive Fourier Decomposition algorithm, described in Torrealba et al. (2015).

While combining the RR Lyrae catalogs, we realized that the object identifiers given to the stars were non-unique. $\sim 15\%$ stars had duplicated identifiers. After correcting for this by generating new identifiers, we checked for duplicate observations of stars within catalogs of all three sub-surveys. We ran an internal cross-match algorithm that checked for pairs of stars in $4''$ neighborhoods and flagged those pairs with an angular separation of less than $4''$. All stars within $0.5''$ of each other were considered duplicates, and the one with the lower N_{obs} in a pair was removed. This limit was decided based on the separation distribution of all pairs — these had the highest probability of being duplicates. For the remaining pairs, with separation between $0.5''$ and $4''$, we calculated the absolute difference between their

V magnitudes (ΔV), variability amplitudes in V (Δ_{Amp}^V), and periods (ΔPer). Those with $\Delta V < 0.3$ mag and $\Delta_{Amp}^V < 0.4$ mag and $\Delta Per < 5 \times 10^{-5}$ d were flagged and the rest were assumed to be truly different close-by stars (chance overlaps) and retained. These upper limits were set close to three times the average uncertainty in the quantities. For the flagged pairs, only stars with greater N_{obs} were retained.

Due to the higher spatial density of the Catalina general variable star catalogs (presented in Drake et al. 2014, 2017), duplicate stars, if any, needed to be treated even more carefully. Applying the same treatment as described above, we found ~ 350 pairs with separation between 0.5″ and 4″. After inspecting the distribution of their (ΔV, Δ_{Amp}^V, ΔPer), the limits to rule them out as possible duplicates were set at more restrictive criteria (0.3 mag, 0.5 mag, 10^{-3} d). ~ 320 pairs satisfied the above limits and were treated as true duplicates. The remaining 30 pairs had $\Delta V < 0.05$ mag and $\Delta_{Amp}^V < 0.1$ mag but ΔPer greater than 10^{-3} d. To investigate further, we compared their variability types. ~ 20 pairs had the same type and we treated them as duplicates, too. The remaining ~ 10 pairs had differing variability types, but similar light curve morphologies. For instance, 9 pairs had one star classified as RRc, and the other as a W Uma-type/Algol-type binary. We decided to drop both the stars in these 10 pairs owing to the confusion in their classification. In total, these steps removed ~ 360 stars from the Drake et al. (2014) catalog.

Repeating these steps for the catalogs presented in Torrealba et al. (2015), Drake et al. (2017) did not lead to any removals. Combining all catalogs resulted in over 110 000 stars in 22 subtypes with > 43 000 of our target subtypes.

Except for this astrometry- and photometry-related filtering, we also removed those stars that were either blended or exhibited unusual light curve morphology (as per the flags computed by the respective studies). This removed a further 53 stars. We also discarded 24 stars that were listed as non-RR Lyrae stars in the Catalina documentation.

II.1.2.2 Light curves

For all Catalina stars whose light curves were accessible, we computed the astrometry covariance matrix $Cov(\alpha, \delta)$ using the coordinates provided for each observation of the stars. We also computed the average uncertainty in V as the median of the magnitude errors over all observations in the light curves. The error distributions for every Catalina catalog are presented in Fig. II.3 (bottom panel). For stars whose light curves were not accessible, we imputed these values as follows.

For the RRab stars in Torrealba et al. (2015), we computed V_{error} by fitting a fourth-degree polynomial to the distribution V vs V_{error} of Drake et al. (2017) RRab stars. Since, these observations were taken using the same 0.5 m Siding Spring telescope and observing setup, we used this fourth-degree polynomial to estimate V error for Torrealba et al. (2015)
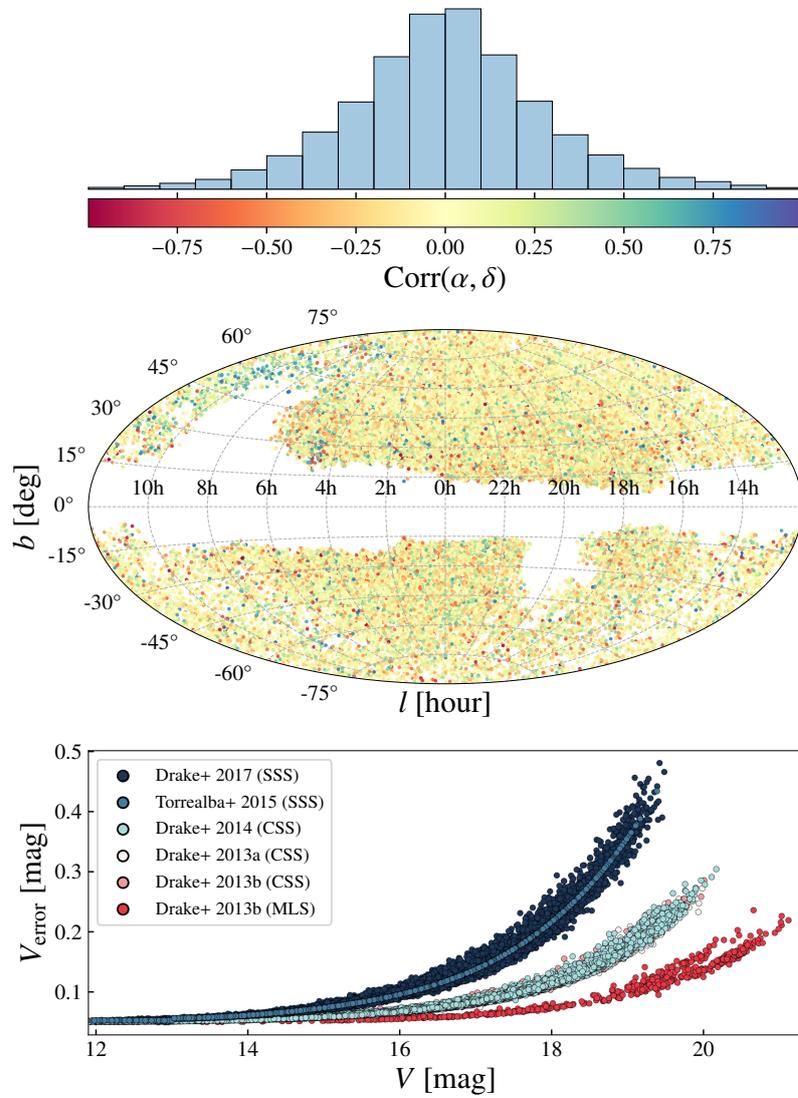
FIGURE II.3: Errors in *V*-band photometry of Catalina stars (*bottom*). The photometric uncertainty is inversely proportional to the size of the telescope mirror (given in Table II.2). In the *middle panel*, the on-sky distribution of all Catalina stars is plotted in Galactic coordinates. They are colored according to the correlation between the coordinates $\mathrm{Corr}(\alpha, \delta)$. The normalized histogram of $\mathrm{Corr}(\alpha, \delta)$ is plotted in the *top* panel. This smooth distribution is in contrast to what is seen for ZTF in Fig. II.5.

stars.

It was possible to compute the astrometry covariance matrix only for the light curves published by Drake et al. (2013a, 2017). As shown in the middle panel of Fig. II.3, there is no strong correlation (Corr(α, δ)) between right ascension (α) and declination (δ) for these stars in any particular direction. Keeping this in mind, coupled with the telescope size (given in Table II.2), we assumed median error in α and δ of SSS (MLS) astrometry to be 1σ greater (smaller) than the CSS median of these distributions. To estimate these missing covariances, we drew Monte Carlo samples from the CSS Cov(α, δ).

Drake et al. (2013a) estimated that the AFD algorithm calculates variability periods with an uncertainty of $\sim 0.005\%$. We used this value to compute the period uncertainty for our target stars. Hereafter, we refer to the combined Catalina catalog, containing our target subtypes, as the CSS catalog.

II.1.3   ZTF DR2

The Zwicky Transient Facility has been designed to be "a stepping stone for LSST" (Bellm et al., 2019, Legacy Survey of Space and Time). It maps the entire northern sky once every night. Its second data release (DR2) was based on 470 observing nights, which resulted in a median of ~ 150 observations each in the PanSTARRS g- and r-bands. We used the classification catalog presented by Chen et al. (2020). In their analysis, Chen et al. used the clustering algorithm DBSCAN (Density-Based Spatial Clustering of Applications with Noise, Schubert et al. 2017) to classify stars. The input feature space included the period, Fourier parameters of phased light curves, absolute Wesenheit in the gr bands, etc. Lastly, they incorporated empirical cuts on the aforementioned parameters for subtype classification (see their Table 1).

Since Chen et al. did not provide subtype classification for classical Cepheids in their catalog, we cross-matched their classical Cepheids sample with the Galactic classical Cepheids catalog of Pietrukowicz et al. (2021, AGC; discussed in detail in Sect. III.2.1). Out of ~ 1200 ZTF classical Cepheids, we found a match for $\sim 92\%$ of them and adopted the validated subtype from AGC. We removed the remaining $\sim 8\%$. Additionally, we discarded 60 Cepheids and RR Lyrae stars that were obvious outliers in the period vs amplitude diagram (Bailey diagram) (see Table II.1 for the limits on periods).

Given the one-day cadence of the survey, the Chen et al. catalog (referred to as the ZTF catalog hereafter) suffered heavily from period aliasing. They took steps to rectify this problem by visually inspecting phased light curves of those stars whose periods in g- and r-band were multiples of each other. They coupled this inspection with the examination of other properties of these stars (for instance, comparing their Fourier parameters and variability
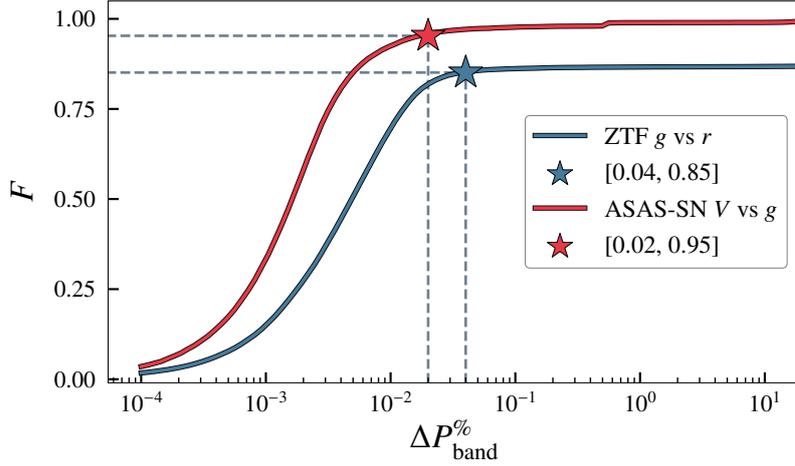
FIGURE II.4: Level of agreement between g- and r-band periods in the ZTF catalog (in blue) and V- and g-band periods from ASAS-SN (in red). F (y-axis) is the fraction of stars with periods in the aforementioned bands within $\Delta P^{\%}_{\text{band}}$ (x-axis). $\Delta P^{\%}_{\text{band}}$ is defined in Sect. II.1.3. The marked points highlight the uncertainty assumed in the ZTF (0.04%) and ASAS-SN (0.02%) periods.

amplitude with those of their subtypes' distributions) to estimate the best period (P_{best}) for every star. All in all, they provide three period estimates for every star: periods in the g- and r-bands, and P_{best}. The difference between P_{best} and one of the band periods was $\lesssim 1\%$ for $\sim 99.5\%$ of stars. However, for ~ 275 RRab and RRc stars, P_{best} was $\gtrsim 30\%$ shorter compared to both g- and r-band periods. Additionally, for ~ 60 stars P_{best} was equal to either ~ 0.5 or 1 d but their periods in the g- and r-bands were $\gtrsim 10$ days. Collectively, these ~ 330 stars were the ones that suffered from aliasing the most and, hence, were removed from our analysis.

To estimate an uncertainty in the ZTF periods, we computed the relative percentage difference between the periods of the g- and r-band light curves (P_g, P_r) defined as: $\Delta P_{\text{band}} = P_{\text{abs}}/P_{\text{band}} \times 100$, where band $\in (r, g)$ and $P_{\text{abs}} = |P_g - P_r|$. When we computed the fraction of stars (F) with $\Delta P_r < x$ or $\Delta P_g < x$, where $x \in [0.001, 20]$, we noticed that this cumulative distribution flattened out at $\sim 0.04\%$ (see Fig. II.4). This meant that for $\sim 85\%$ of the stars, P_g and P_r agree within $\sim 0.04\%$. Therefore, we assumed this to be the uncertainty in the final period of all ZTF stars.

We computed the astrometric covariance matrix and imputed the photometric uncertainty for stars with missing light curves as explained in Sect. II.1.2.2 for the CSS catalog. The importance of incorporating the former is underlined by Fig. II.5. The bottom panel shows the impact of the observational jitter pattern and how the stars at the edge of observing fields have correlated coordinates, possibly due to duplicated observations between two adjacent fields. Even though a majority of stars have Corr(α, δ)=0, the peaks seen in the
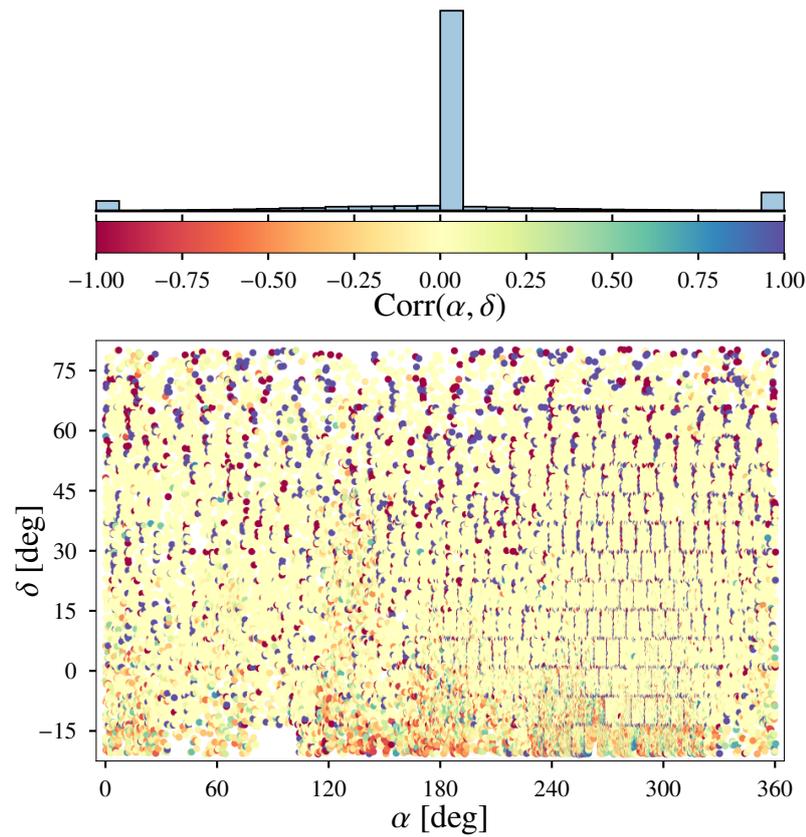
FIGURE II.5: On-sky distribution of stars from the Chen et al. (2020) catalog shown in celestial coordinates and colored according to the correlation between the coordinates $\mathrm{Corr}(\alpha, \delta)$. The coordinates are plotted in rectilinear projection to highlight the observational jitter pattern. Stars on the edges of overlapping fields have strongly correlated or anti-correlated coordinates (also seen as peaks at $\mathrm{Corr}(\alpha, \delta)$=-1 and 1 in the histogram in the *top* panel).

histogram (top panel of Fig. II.5) of $\text{Corr}(\alpha, \delta)$ at 1 and -1 contribute non-negligibly to the cross-match of this catalog with others (discussed in Sect. II.2).

II.1.4 ASAS-SN

The All-Sky Automated Survey for Supernovae (ASAS-SN, Shappee et al. 2014; Kochanek et al. 2017) uses twenty telescopes located across both the hemispheres to conduct a wide-field, all-sky photometric survey with the aim of detecting supernovae and other transient events. Originally, it collected images in the V- band with a cadence of ~ 2–3 days up to a limiting magnitude $V < 17$ mag (Jayasinghe et al., 2018). Since 2018, it has started collecting g-band data with a higher cadence of $\lesssim 24$ h and a fainter detection limit ($g \lesssim 18.5$ mag, Christy et al. 2023).

Jayasinghe et al. (2018, 2019a,b, 2020) used a three-step procedure to classify ASAS-SN variable stars into various subtypes. In the first step, they employed upsilon (Kim & Bailer-Jones, 2016) to classify stars using V-band light curves into broad variability types and reject non-variable ones. In the second step, the light curves of the variable stars were visually inspected to confirm or rearrange their broad upsilon-based classes. Finally, they trained a random forest (RF) classifier on 16 features (such as the period, variability amplitude, etc.) to classify stars into 10 different subtypes. Moreover, the subtype classification was also performed manually using well-motivated filters. Additionally, the RF classifiers of the g-band variable star catalog (Christy et al., 2023, 2022) were trained on data cleaned using visual inspection by volunteer citizen scientists.

The internal cross-match step, when applied to the V-band catalog, removed 400 stars that had a duplicate star in their $10''$ neighborhoods. These duplicates also had periods within 1% and magnitudes within 2% of each other. The duplicates arise as the FWHM for ASAS-SN images is $\sim 16''$, and by removing them, we diminished the effects of blending. For the g-band catalog, we did not find any stars within $10''$ of each other.

For the V-band catalog, we selected only those target stars with classification probability ≥ 0.5. For the g-band catalog, Christy et al. only selected those stars with a classification probability ≥ 0.89. Furthermore, ~ 475 (65) stars were removed from the V-band (g-band) catalog based on inconsistent periods (see Table II.1) or abnormally large amplitudes.

We merged the V- and g-band catalogs using a matching radius of $5''$ resulting in a total of $\sim 52\,000$ stars. For $\sim 21\,000$ stars that are common to both catalogs, we preferred the classification of the catalog with the higher N_{obs}. However, the classification in the Christy et al. g-band catalog for stars of a few subtypes differed from that of the V-band catalog (see Fig. II.6). Most of the disagreement arises from stars classified as Type-II Cepheids in the g-band catalog, but classified as classical Cepheids by Jayasinghe et al. (2018). Moreover,

FIGURE II.6: Confusion matrix for stars common between the ASAS-SN *g*- and *V*-band catalogs. Precision×100 is shown in every cell below the number of stars. These values have been rounded to the first decimal and sum to 100 along every column. The disagreeing subtypes between the two catalogs have been dealt with in Sect. III.1.2.

most of the RRd stars in the *V*-band catalog are classified as RRab in the *g*-band catalog. These cases are dealt with in Sect. III.1.2.

Based on the difference in periods of common stars with identical classification in both catalogs, we estimated the uncertainty in their periods to be $\sim 0.02\%$ (see Fig. II.4). This merged catalog forms the ASAS-SN sample of our study and consists of 52 427 stars, of which 46 534 (27 227) were from the *V* (*g*)-band catalog.

II.1.5 *Gaia* DR3

Rimoldini et al. (2023) used a subsample of variable stars compiled by Gavras et al. (2023) for supervised classification of *Gaia* DR3 (Gaia Collaboration et al., 2023c) light curves into over 20 types of variable stars. This formed the General Supervised Classification module of the *Gaia* variability analysis pipeline. This module uses *Gaia* G-band light curves and
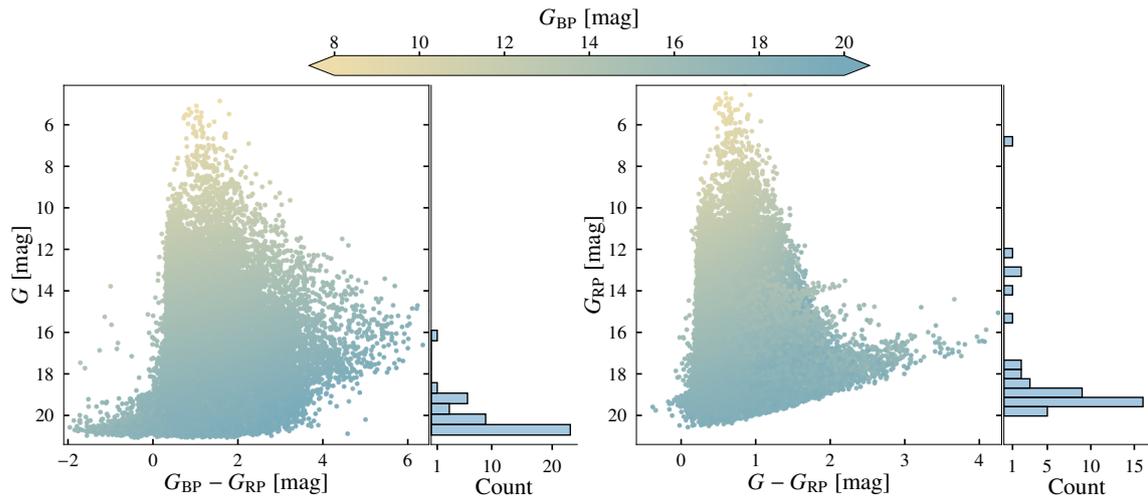
Figure II.7: Color-magnitude diagrams of the cleaned *Gaia* SOS Cep&RRL catalog. The points are colored according to their G_{BP} magnitude. The G magnitude distribution of stars whose BP/RP photometry is discarded on the basis of the color cuts (described in Sect. II.1.5) is plotted in the *second from left* and *right most* panels.

integrated G_{BP}, G_{RP} magnitudes from low-resolution prism spectra. Ripepi et al. (2023) and Clementini et al. (2023) validated and characterized the Cepheid and RR Lyrae catalogs as part of Specific Object Study (SOS Cep&RRL) pipeline. These catalogs contain the largest number of Cepheids (~ 15 000) and RR Lyrae stars (~ 270 000) classified by a single survey.

Ripepi et al. (2023) reclassified 1160 SOS Cepheids (listed in Tables 5, 6 of their paper) after visual inspection and comparison with literature light curves. 507 of these were initially identified as classical Cepheids and later classified as Type-II Cepheids (142 the other way around). 209 stars of the reclassified sample were new discoveries made by *Gaia*. We noticed that a few of these had incorrect periods for their new classes. For instance, a star reclassified from WVIR to BLHER had a period > 5 d. Since the recomputed periods of these reclassified stars were not provided, we flagged these stars. If their renewed classification agreed with other surveys, we assumed the periods measured in other surveys to be the final one. Three stars initially classified as ACEP, but without any mode identification, were reclassified to ACEP_F. However, these stars did not have a period estimate and, hence, were removed.

Additionally, we removed 888 stars from the SOS RR Lyrae sample that are listed in Tables 5–8 of Clementini et al. (2023). These were flagged to be either galaxies, QSOs, or binaries during SOS validation. Moreover, there are a number of RR Lyrae stars in the SOS catalog that have null or extremely large errors in their BP and RP photometry. A majority of these reside in high-extinction regions towards the Galactic bulge or the LMC. We decided to remove the BP/RP photometry of these stars. Moreover, a few stars have BP/RP $\lesssim$ -21, which was also noted by Babusiaux et al. (2023). Consequently, we discarded

BP/RP photometry of stars that have $G_X < 4$ mag or error in $G_X < 10^{-5}$ or > 2 mag, for X in [BP, RP]. This rendered the BP (RP) photometry useless for ~ 150 (80) stars.

There were still ~ 40 stars left that occupied an unnatural space in the CMD. We handled them by discarding the BP/RP photometry of those stars which satisfied these criteria: $(G_{BP} - G) < -5$ or $(G_{BP} - G_{RP}) < -2$ or $(G - G_{RP}) > 5$ mag. Figure II.7 shows color-magnitude diagrams of the cleaned sample in various colors. The removed stars are the faintest (their G magnitude distribution is plotted in the second and fourth panel of Fig. II.7) and severely affected by extinction.

II.1.6 ASAS-MACC

The All Sky Automated Survey (ASAS) was one of the first all-sky surveys designed to find variable stars (Pojmanski, 1997). Initially, it observed the sky using two 0.7 m telescopes situated at the Las Campanas Observatory, Chile. Subsequently, two more telescopes were added for ultra-wide field observations. Out of ~ 17 million objects that were observed in the V-band magnitude range [8, 14], Pojmanski (2002) found $\sim 50\,000$ to be variable. However, their broad classification was found to be marginally inconsistent for Cepheids. For instance, Schmidt et al. (2009) were able to confirm only $\sim 10\%$ of ASAS Type-II Cepheid-candidates in their study. Therefore, Richards et al. (2012) reclassified these stars using random forest classifiers trained on a set of 67 features and five optical-IR colors (see Richards et al. 2011 for the list of variability features used for classification). Their variable stars catalog, referred to as MACC (Machine-learned ASAS Classification Catalog), supplied the brighter variable stars in our collection, which saturated the CCDs of other surveys.

We realized that MACC astrometry was imprecise for many stars and, hence, we retrieved corrected coordinates for these stars from the original ASAS catalog stored on Vizier [1]. However, when we cross-matched these stars with Simbad and *Gaia* DR2/DR3, $\sim 10\%$ of stars did not have any counterpart within $1'$. We omitted these stars from any further consideration, as without accurate astrometry, this sample would lead to the presence of spurious duplicates in the merged catalog. We found a neighbor within $15''$ in Simbad for the remaining $\sim 90\%$ of the stars. We assumed $3''$ as the uncertainty in these coordinates, which corresponds to 0.2 pixels or one-fifth of the FWHM (c.f. Sect. 4 of Pojmanski 1997 and Sect. 2 of Pojmanski et al. 2005). Furthermore, based on the photometric errors of epoch magnitudes that were included in the light curves, we assumed 0.25 mag as the photometric uncertainty floor for MACC.

We retained only those target stars from MACC that had a classification probability ≥ 0.5 and an anomaly score ≤ 10. Figure II.8 shows the distribution of these features for

---

[1] https://vizier.cds.unistra.fr/viz-bin/VizieR-3?-source=II/264/asas3
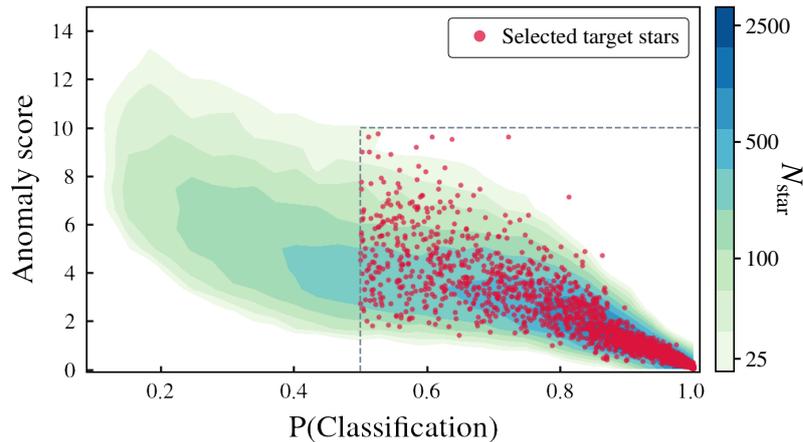
FIGURE II.8: Anomaly score (y-axis) plotted against classification probability (x-axis) for MACC variable stars. The underlying contours represent the entire sample (colored according to number of stars per level), while the red dots denote the selected RR Lyrae stars and Cepheids.

all $\sim 50\,000$ variable stars from MACC in the underlying contours. Our selected target stars (shown in red) are mainly clustered towards bottom right. Richards et al. (2012) set up the anomaly score to measure the dissimilarity between the features of the target star against those of the training set. A higher anomaly score means the star is too dissimilar from the training set. The reason for this dissimilarity might be inherent to the star (i.e., it might belong to a subtype not included in the training set), or it could be external (i.e., due to bad or incomplete data). This step reduced the total number of stars to $\sim 33\,000$.

Since Schmidt (2013) found irregularities with MACC classification, we cross-matched this catalog with AGC catalog (described in Sect. III.2.1) to measure Galactic classical Cepheids accuracy for MACC. We recovered 380 AGC stars within $30''$ of MACC stars. Out of these 380 OGLE-verified Galactic classical Cepheids, 363 were classical Cepheid candidates in the MACC catalog and $\geq 50\%$ of them had a probability of classification ≥ 0.9. The remaining 17 contaminants did not belong to any single dominant class. Another sample of ~ 30 stars considered Galactic Cepheid candidates in MACC were absent from AGC and more than ~ 20 of these had classification probability ≤ 0.6. Therefore, we established that the MACC classification accuracy for Galactic classical Cepheids with probability ≥ 0.5 and anomaly score ≤ 10 is $\sim 90\%$. The number of stars of each subtype from MACC is presented in Table II.4.

II.1.7   PanSTARRS-1

PanSTARRS-1 3π survey (PS1, Chambers et al. 2016; Flewelling et al. 2020) was designed for a precise photometric survey of the northern hemisphere. Using a $1.8\,\mathrm{m}$ telescope in

Hawaii, PS1 observes the sky in $grizy$ bands with a 5σ detection limit below 23 mag for gri bands. Sesar et al. (2017, hereafter S17) presented a catalog of RRab and RRc stars using data from PS1 DR1. This catalog was the widest and deepest when published. Over the 4.5-year observation period (see Table II.2), PS1 acquired ~ 72 observations per star, which after quality cuts, decreased to ~ 67 observations per star over all bands.

S17 implemented the XGBoost classifier (a gradient tree boosting algorithm, Chen & Guestrin, 2016) in every step of their three-stage classification model. In the first stage, they trained an XGBoost classifier on a set of ten variability features computed using PS1 and WISE photometry (the latter used to help differentiate RR Lyrae stars from QSOs). In the second stage, they computed multi-band periodograms of PS1 light curves and appended the first 20 possible periods (and their powers) to the initial feature set and re-trained on this augmented feature set. In the final step, they used multi-band SDSS $griz$ templates of Sesar et al. (2010) to fit PS1 light curves. They extracted 20 features from the phased PS1 light curves and trained the final classifier on 70 features. Consequently, this model computed two probabilities for every star: $score_{ab}$ ($score_c$), the probability that a star was of the type RRab (RRc). We selected stars with $score_{ab} > 0.8$ and $score_c > 0.55$ as the final PS1 sample. This sample — comprised of 44 208 RRab and 16 936 RRc stars — is $\geq 90\%$ pure and $\sim 77\%$ complete at ~ 80 kpc (see Tables 2 and 4 in S17).

We noticed a few RRab stars in this sample that had a relatively large variability amplitude compared to the overall distribution. In total, we removed 29 stars with $Amp_r > 1.4$ mag. These stars had relatively large amplitudes across all bands and much larger than the High Amplitude Short Period RR Lyrae from Fiorentino et al. (2015).

S17 showed that the accuracy of template-fit periods is within 2 s for $\gtrsim 80\%$ of the stars. Therefore, we conservatively assumed period_error to be 0.001% of the measured periods. Since S17 did not provide individual object identifiers of the stars, we cross-matched their catalog with the main source table of PS1 survey using a $1''$ matching radius. We obtained errors in magnitudes and coordinates from the source table for all but ~ 10 stars. We calculated the uncertainties in coordinates using eqns. 8–10 of Pineau et al. (2017). The astrometric uncertainty floor of 10 mas (c.f. Magnier et al. 2020b) was added to the coordinates' uncertainties (which have a median of ~ 5 mas). For stars without a photometric uncertainty from the PS1 source table, we assumed an uncertainty of 15 mmag, following Magnier et al. (2020b).

The epoch of the survey was computed to be J2012.5 based on the average epochs of the individual stars. We assumed the total number of observations per star as $\sim 95\%$ of the total number of single epoch detections across all filters. This was done to account for the quality control steps taken by S17. The resultant N_{obs} distribution could be approximated with a normal distribution with mean $N_{obs} = 67$, which mimics the median value stated in

S17. In total, this catalog (referred to as the PS1 catalog hereafter) consisted of $\sim 44\,000$ RRab and $\sim 17\,000$ RRc stars.

II.1.8   DES DR2

The Dark Energy survey (DES, Dark Energy Survey Collaboration et al. 2016) was designed to determine the dark energy equation of state and, to this end, it has mapped ~ 300 million galaxies across the Southern sky. However, the deep sky coverage of DES can also be leveraged to probe the Milky Way halo. Stringer et al. (2019, 2021) used the high-quality photometric data (5σ detection limit between 23.5-24 mag) from DES Data Release 2 (DR2) to identify RR Lyrae stars (of type ab). This dataset includes six years of photometry in $grizy$ bands obtained using the Dark Energy Camera situated at the Cerro Tololo Inter-American Observatory.

Stringer et al. (2021) applied cuts in color-color space to limit their set to RRab-like stars. They trained a two-stage random forest-based model on a set of ten variability features. The first-stage RF classifiers were used to reject non-variable stars and QSOs. In the second-stage, they fitted multi-band RRab templates of Sesar et al. (2010) to DES $griz$ light curves. A set of photometric features computed using phased light curves were fed to train another RF classifier, which computed the final probability of a star being of the type RRab ($score_{ab}$). Consequently, they selected only those stars with score ≥ 0.605.

We noticed a global minimum in the distribution of these probabilities at ~ 0.8 (see top panel of Fig. II.9). The period vs amplitude distributions (Fig. II.9, bottom row) of stars on either side of this minimum show that the higher-probability stars follow the Oosterhoff dichotomy[2] quite closely, while majority of the others do not. In the distribution of lower-probability members ($score_{ab} < 0.8$), one can see a peak at ~ 1 day, which possibly arose due to variability induced by the Earth's rotation in non-RR Lyrae stars. We will further discus this sample of stars with DES $score_{ab} < 0.8$ in Sect. III.1.1 and III.1.2.

We used the median of the rescaled uncertainties in single-epoch magnitudes — as included in the light curves — as proxy for the photometric uncertainty. The median of these uncertainties is ~ 5 mmag, reflecting the high quality of both the DES data and the template-fitting algorithm of Stringer et al. (2021). Abbott et al. (2021) estimated the astrometric precision of DES coadded images to be ~ 27 mas with respect to *Gaia* DR2. We assumed this value as the uncertainty floor for the equatorial coordinates of the 5664 high-probability RRab stars. For 700 stars common between their catalog and OGLE LMC sample, Stringer et al. compared the periods estimated by the two surveys and found that ~ 70 % of the

---

[2]The Ootserhoff dichotomy is the presence of two different trends in the period distribution of RR Lyrae stars in Galactic globular clusters (Oosterhoff, 1939). For a recent review, see Fabrizio et al. (2021).
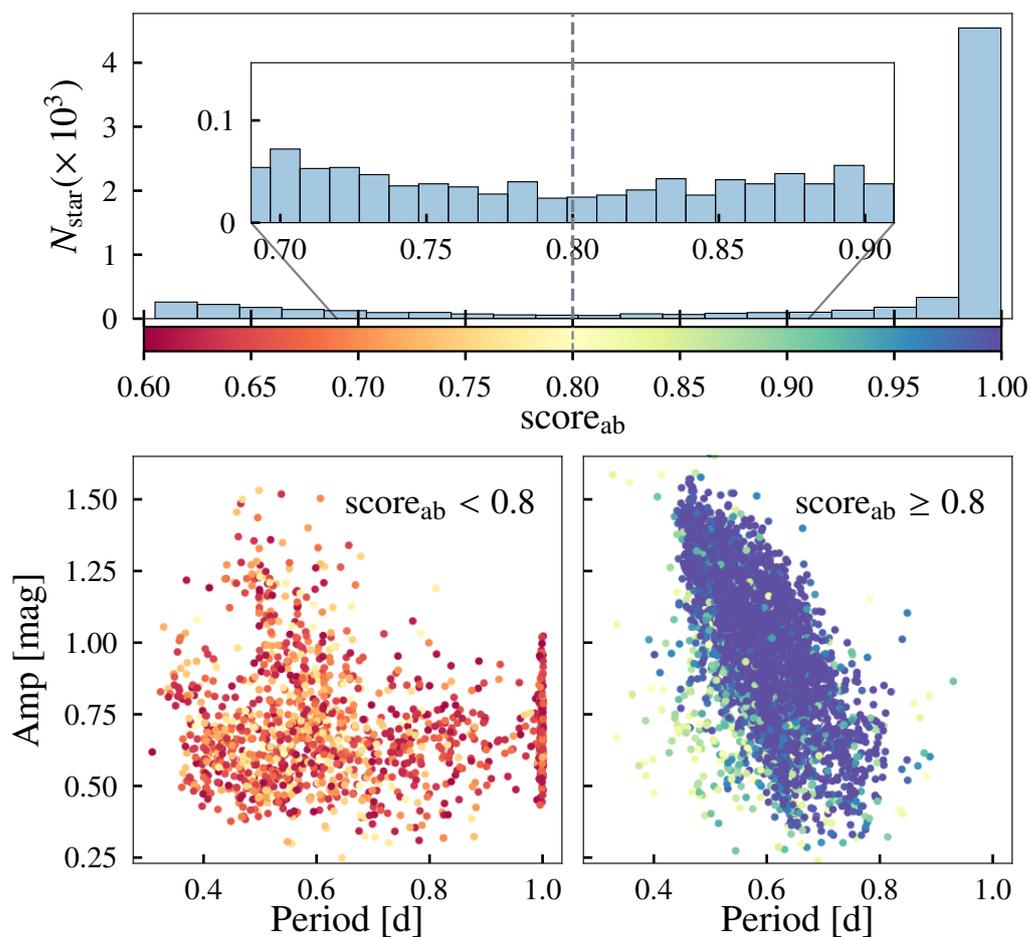
Figure II.9: DES RRab stars classification probability (score_ab; *top* panel; from Stringer et al. 2021). The global minimum of the distribution at 0.8 is marked across the *top* panel and its zoomed inset. The period vs amplitude distribution of stars with score_ab < 0.8 is plotted in the *bottom left* panel and those with score_ab ≥ 0.8 are plotted on the *bottom right* panel. The points are colored according to their classification probability.

sample had periods within $\lesssim 5$ s. Nonetheless, we assumed 0.05% (~ 25s for a 0.5 d RRab variable) as the uncertainty in DES periods. This slightly conservative assumption is justified in Sect. III.1.1.

II.2 Cross-matching catalogs

As all eight catalogs have different angular resolutions, photometric systems, and astrometric epochs (given in Table II.2), it is imperative to ensure that the cross-matches between these catalogs are handled carefully. Most of the commonly used algorithms for the cross-match of stellar catalogs employ a k-d tree (with k=3) for an efficient nearest neighbor search (Bentley, 1975). For lower dimensions, this algorithm performs $O(\log n)$ nearest neighbor searches. For instance, for cross-match of two stellar catalogs, equatorial coordinates (α, δ) are first converted into Cartesian coordinates (x, y, z) (to avoid the problems of spherical geometry at the poles). Then, for a given star, its closest neighbor on the 3-d binary tree is searched for recursively and the best match is the one with the shortest Euclidean distance (Maneewongvatana & Mount, 1999).

Although this method is extremely efficient, it has a few limitations. It requires either unitless quantities or quantities expressed in the same units. As a result, the use of quantities from multiple phase spaces is prohibited. Additionally, it cannot encompass uncertainties on coordinates unless the distance metric is changed. Therefore, for our cross-match, we developed a new algorithm we call Bhattacharyya distance-based cross-match (BXM).

II.2.1 Bhattacharyya distance-based cross-match (BXM)

Instead of using the Euclidean distance as a metric to find the closest neighbor, we employ the Bhattacharyya distance (Bhattacharyya, 1943, 1946). In essence, the Bhattacharyya distance measures the true distance between two n-dimensional distributions, while taking their shapes into account (Kailath, 1967; Aherne et al., 1998). For two multivariate normal distributions $p_i = \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i); i \in [1, 2]$, the Bhattacharyya distance between them can be calculated as:

$$D_B(p_1, p_2) = \frac{1}{8}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$
$$+ \frac{1}{2} \ln \left(\frac{\det(\boldsymbol{\Sigma})}{\sqrt{\det(\boldsymbol{\Sigma}_1)\ \det(\boldsymbol{\Sigma}_2)}} \right), \quad \text{(II.1)}$$

where det(**M**) denotes the determinant of a matrix **M**, and the joint covariance $\boldsymbol{\Sigma} = \frac{1}{2} \times (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)$. In case the two distributions are identical, $D_B = 0$, however, in general,

D_B follows a $\Gamma(n/2, 1/4)$ distribution, where n is the dimension of the distributions. Properties of the Bhattacharyya distance and its comparison with the Mahalanobis distance are discussed in Appendix 1.2.

As D_B involves normalization using the covariance matrices, it is independent of units. This allows us to incorporate photometry, periods, and the full astrometric covariance matrix in the process of finding the best neighbor. As shown in Fig. II.5, a non-negligible number of stars in the ZTF catalog (and catalogs from other surveys in general) suffer from a large correlation in coordinates. Omitting this correlation leads to information loss, underestimated probabilities of cross-match, and ultimately, to incorrect cross-matches. Taking $Cov(\alpha, \delta)$ into account when cross-matching is critical when: (i) surveys with large photometric FWHM are involved or (ii) catalogs cover crowded regions or regions of varying number density or (iii) a systematic difference exists between astrometric epochs of the surveys. These cases and broader applications of this algorithm are discussed in Lala et al. (in preparation). We briefly outline the core steps of the algorithm here.

For an astrometry-only cross-match (i.e., using the coordinates, their covariance matrix, and proper motions), the mean vector in Cartesian coordinates is composed as $\boldsymbol{\mu}_1 = (x_{j1}, y_{j1}, z_{j1})\big|_{j=1}^{N}$. It is an $N \times 3$ matrix, where N is the number of stars in the catalog.

Given the size of our catalogs (and stellar catalogs in general), it is impractical to compute D_B for every star against all others. Hence, we first build a k-d tree to get all stars in the on-sky neighborhood of a given star. In order, to determine the optimal cross-match radius, d_{match} (the largest angular separation between two stars to classify them as a match), we performed a purity vs completeness test. We define purity as the ratio of number of unique one-to-one matches to the total number of matches found between two catalogs. Completeness is defined as the ratio of number of unique one-to-one matches to the total number of stars in the smaller catalog. We run multiple cross-matches for a series of increasing d_{match} in the range $0.1''$ to $30''$. Here, $0.1''$ was chosen as half of the *Gaia* FWHM, and $30''$ was chosen as the largest FWHM out of all surveys being matched (the MACC FWHM). d_{match} is selected as one of the values along the purity-completeness curves as described in Sect. II.2.2.

Once the on-sky neighborhood for a given star (say star X) is determined, we compute probabilities for every neighbor being its match and the one with the highest probability is selected as its 'best' match. The probability that a certain star Y (with coordinates μ_y and covariance Σ_y) is a match of star X (with coordinates μ_x and covariance Σ_x) is defined as:

$$p(H/D) = \frac{p(D/H) \times p(H)}{p(D)}, \tag{II.2}$$

where H is the hypothesis that two stars are a match, D is the observed data $(\mu_x, \mu_y, \Sigma_x, \Sigma_y)$. $p(D)$ is the probability of observing the data (also termed evidence) and considered a nor-

malizing factor. The prior probability, $p(H)$, is defined as

$$p(H) = \begin{cases} 1, & \text{if } d \leq d_{\text{match}}; \\ 0, & \text{otherwise}, \end{cases} \tag{II.3}$$

where, d is the great-circle distance between the stars coordinates (calculated using the formula given in Vincenty 1975). The likelihood function in Eqn. II.2, $p(D/H)$, uses the Bhattacharyya distance as defined in Eqn. II.1. This likelihood is computed as:

$$p(D/H) = 1 - \mathcal{F}_{\Gamma}(D_B, n), \tag{II.4}$$

where $\mathcal{F}$ is the cumulative distribution function of a Γ-variate with shape $= n/2$ and scale $= 1/4$. Thus, combining Eqns. II.2, II.3, and II.4, the posterior probability of star Y being a match of star X (p_{match}) can be computed as:

$$p_{\text{match}} \equiv p(H/D) \propto (1 - \mathcal{F}_{\Gamma}(D_B, n)) \times p(H). \tag{II.5}$$

To summarize, the algorithm works as follows for a run in astrometry-only mode:

1. Propagate coordinates to the same epoch if proper motions are available.

2. Perform purity vs completeness test to determine d_{match}.

3. Build a kd-tree and form an on-sky neighborhood of all stars within d_{match}.

4. Select the neighbor with the highest p_{match} (or the shortest D_B) as the best neighbor.

II.2.2 Employing BXM

For our cross-match, we used the *Gaia* SOS catalog (described in Sect. II.1.5) as the base survey to leverage its high-accuracy astrometry and cross-matched every catalog against it. For all eight catalogs considered here, their spherical astrometric covariance matrices were converted to the Cartesian frame following the equations given in Sect. 1.5.5 of ESA (1997) and Sect. 3.1.7 of Hobbs et al. (2018). We used the same equations to propagate *Gaia* DR3 coordinates to the epoch of every other catalog (see Table II.2).

The results of the cross-match between *Gaia* and three catalogs (namely DES, OGLE, and ASAS-SN) are shown in Fig. II.10. The purity-completeness curve for each pair is plotted in the second column (panels g, h, i) of Fig. II.10. For each survey, d_{match} was selected from the top right corner, as a trade-off between purity and completeness.

The determination of d_{match} was performed on a region-by-region basis to limit the effects of extinction and varying stellar density. The samples were divided into the following
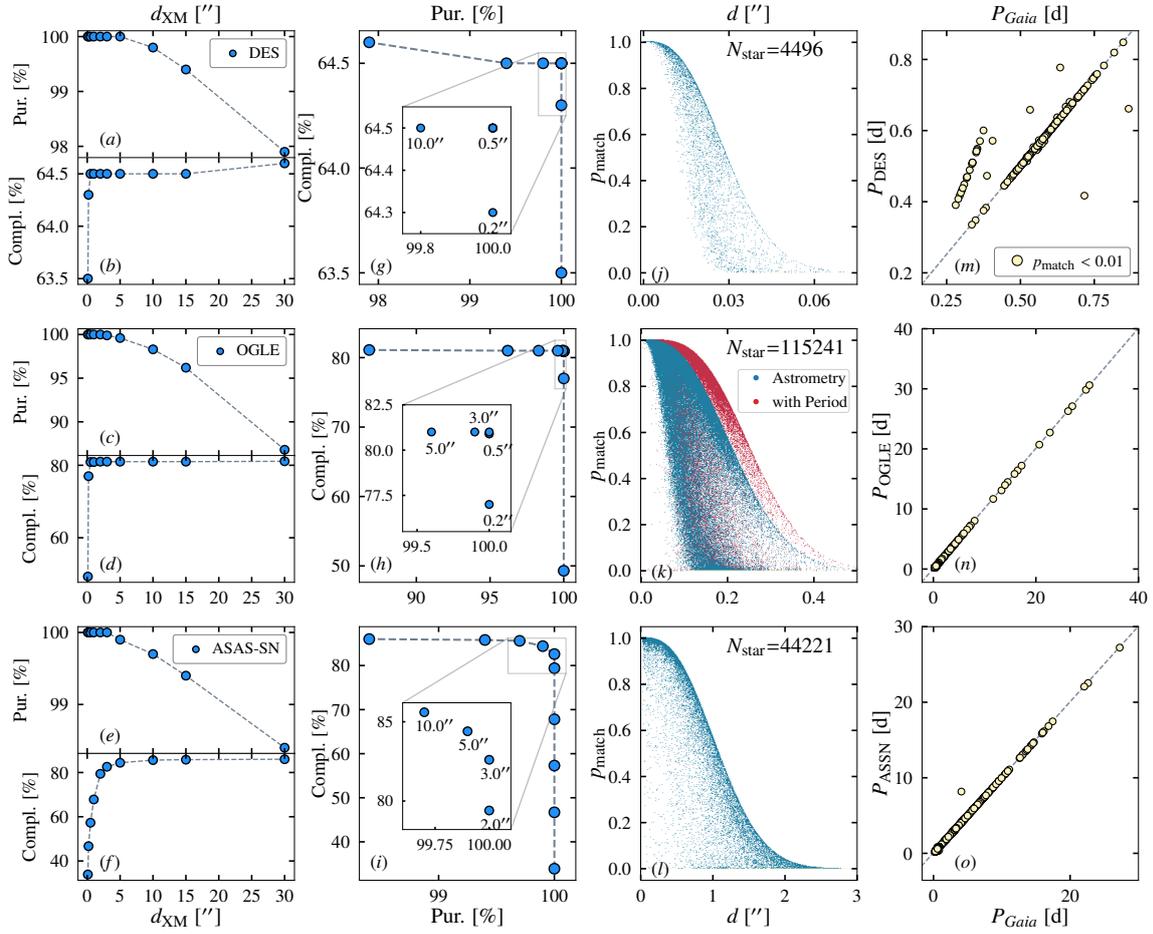
FIGURE II.10: Results of Bhattacharyya-distance based cross-match between *Gaia* and DES (*top*), OGLE (*middle*), and ASAS-SN (*bottom*). The first two columns (panels *a–i*) depict the results from the purity-vs-completeness test. Purity and completeness (as defined in Sect. II.2) are plotted against various cross-match radii (d_{XM} in ″) in the panels (a)−(f) of the first column. With increasing cross-match radius, completeness (Compl.) increases and purity (Pur.) decreases. This pattern is plotted in the second column (panels g, h, i) for each survey. The insets zoom into the region from which the optimal cross-match radius (d_{match}) is chosen. In the third column (panels j, k, l), the probability of two stars being a match (p_{match}, as defined in Eqn. II.5) is plotted against their angular separation (d in ″). This probability was computed using the astrometry-only mode of BXM (in blue). For OGLE (panel k), we also present the probability computed when taking variability periods into account (in red). For a few stellar pairs whose $p_{match} \leq 0.01$, we compare their periods in the last column (panels m, n, o).

regions: Galactic bulge, disk, halo, and the Magellanic clouds. Their on-sky demarcations are provided in Appendix 1.3. These divisions distribute the stars according to their on-sky positions only and do not ascribe full membership to the regions. Naturally, d_{match} is inversely proportional to the on-sky number density of stars in a region — the more crowded the region the smaller the d_{match}. Nevertheless, for the sake of concision, we present the results of all-sky matches in Fig. II.10.

The distributions of p_{match} against the on-sky angular separation (d) are plotted in panels (j), (k), (l) Fig. II.10. As expected, pairs at larger distances have a relatively low p_{match}. These probabilities were computed using just the astrometric part of BXM (i.e., without periods and photometry but using proper motions). $\sim 1\%$ of matched stars have $p_{\text{match}} \leq 0.01$. To confirm whether they are likely to be the same star or not, we compared their periods and magnitudes. In the last column (panels m, n, o) of Fig. II.10, we compare their periods. For OGLE and ASAS-SN, periods of matched stars share excellent agreement. The low p_{match} value for these pairs possibly arises from underestimated astrometric uncertainty. For OGLE, we also show the astrometry+period p_{match} values in red. By incorporating periods into BXM, the overall p_{match} values increase and very few stars retain a $p_{\text{match}} \leq 0.01$ within d_{match}.

However, for DES, we see the periods are off the identity line for a substantial number of stars. Stars in these pairs are within $0.08''$ of each other and also share the same classification. However, their periods in DES are inaccurate. Validation of these stars is discussed in Sect. III.1.1. Regardless, stars with incorrect period estimates in one survey (due to fewer N_{obs}) can lead to genuine matches having very low values of p_{match}. Hence, we do not use periods to compute p_{match} for pairs with $d \leq 0.5''$. For assessing matches with $d > 0.5''$, we used both astrometry and periods.

We collected 330 747 unique stars after cross-matching all seven catalogs against the *Gaia* SOS catalog. The number of stars contributed by each survey are shown in Fig. II.11. Additionally, the fraction of stars that are unique to each survey is shown in the inset panel of Fig. II.11. *Gaia* and OGLE contribute the most stars, while MACC and DES contribute the fewest. In Fig. II.12, we plot Gaussian kernel density estimates of magnitudes of unique stars from each survey. This figure underlines the importance of this compilation. The unique stars contributed by each survey lie along different magnitude ranges, especially so those from MACC and DES. The former provides the brightest stars that were probably too saturated in other surveys, while the latter contributes the faintest RRab stars in the Southern hemisphere. Overall, $\sim 40\%$ of stars are detected in just one survey (i.e., they are uniquely contributed stars from each survey). This detection frequency of stars is shown in Fig. II.13. It displays the number of surveys (N_{survey}) in which stars are detected and classified. No star is detected in more than six surveys due to varied observation footprints, detection limits,
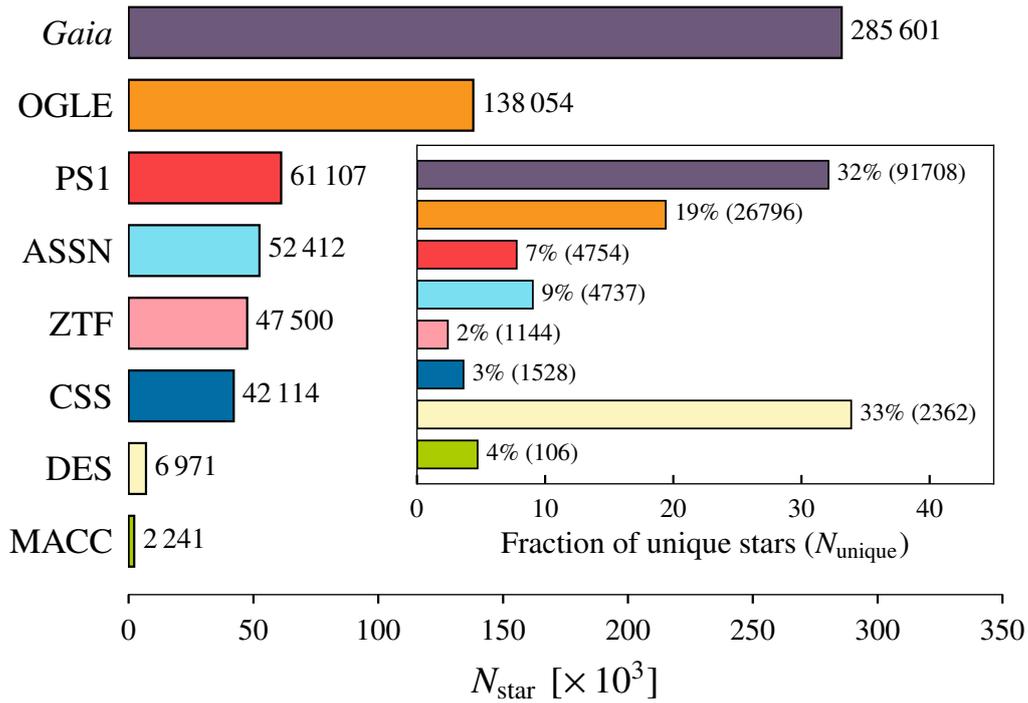
FIGURE II.11: Number of variable stars classified by each survey. The *Inset panel* shows the fraction of stars from each survey that are unique to it.

and locations of all survey telescopes (given in Table II.2).

II.2.3 Cross-match with *Gaia* DR3 source table

$\sim 45\,000$ stars from the seven other surveys were not classified as RR Lyrae stars or Cepheids in *Gaia* DR3. These stars most likely have very few observations in DR3 and, therefore, were not processed by the SOS Cep&RRL pipeline. We cross-matched this non-SOS sample with *Gaia* DR3 source table (`gaia_source`) using CDS XMatch (Boch et al., 2012; Pineau et al., 2020). We used a cross-match radius of $10''$ to get a wide on-sky neighborhood of every star. This ensured that we did not miss out on true counterparts of high proper motion stars. Firstly, *Gaia* coordinates were transformed to epochs of all other surveys as described in the previous section. Then, we used the astrometry-only mode of BXM to find the best *Gaia* neighbor.

This step found a *Gaia* match for all but ~ 2800 stars ($\sim 0.8\%$ of all stars). The missing stars either belong to crowded regions where *Gaia* DR3 has few observations (~ 1000 stars in the Galactic bulge) or they are in the Galactic halo and outside the detection limits of *Gaia* (~ 1100 DES-only RRab stars with $g_{\mathrm{DES}} \gtrsim 21.5$ mag). Out of $\sim 42\,200$ stars with a match in *Gaia* source table but not in *Gaia* SOS catalogs, $\sim 14\,000$ have *Gaia* `phot_variable_flag` set to 'VARIABLE'. However, they did not have epoch data in DR3 for detailed classification.
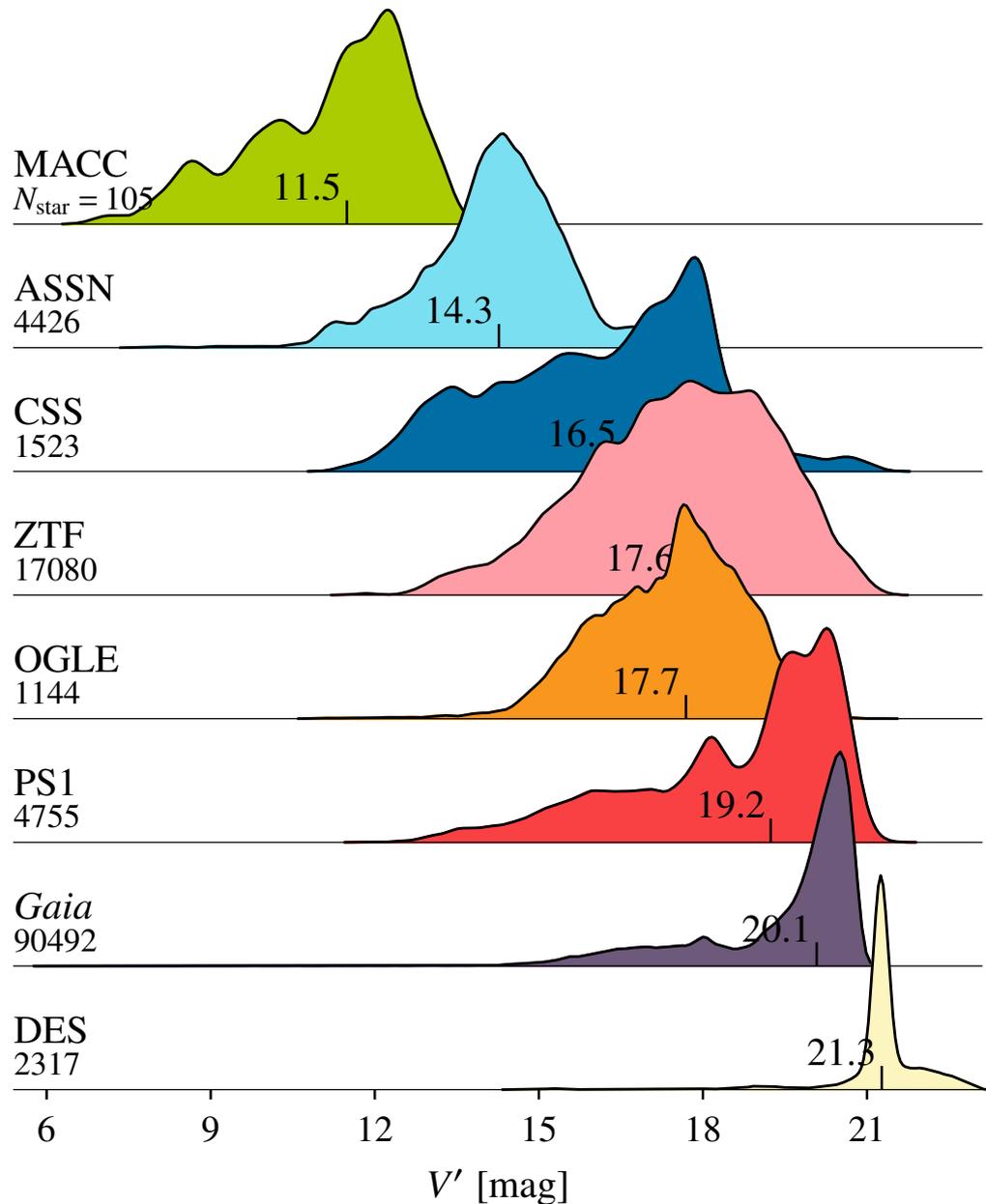
FIGURE II.12: Magnitude distribution of unique stars from each survey. Only the Milky Way stars are considered to avoid biases from the Magellanic Clouds sample. V' is a proxy magnitude. V mag from MACC, ASAS-SN, CSS, OGLE were included. r-band magnitudes from ZTF, PS1, and DES, and G from *Gaia*. The median of every distribution is highlighted and the total number of unique stars with measured respective magnitudes is displayed below the name of each survey. The different range of magnitudes in which every survey detects and classifies variable stars highlights the utility of this compilation.
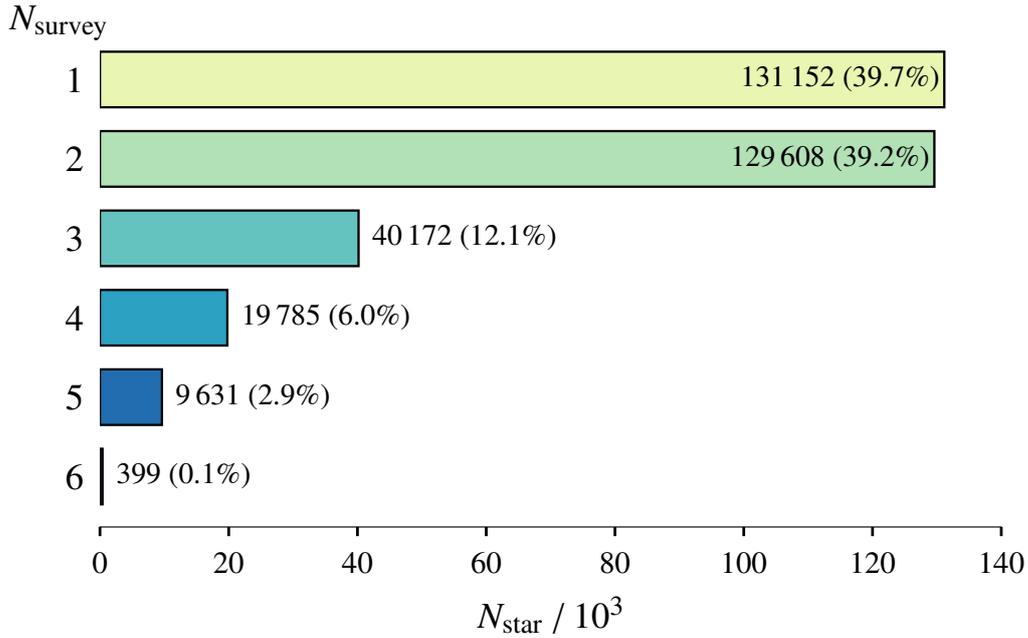
FIGURE II.13: Detection frequency of stars. Number of stars (N_{star}; x-axis) that are classified in N_{survey} number of surveys (y-axis).

We use this homogeneous *Gaia* astrometry to cross-match our catalog against literature surveys in Sect. IV.

II.2.4 Filtering *Gaia* DR3

Except for the tables and properties discussed already, *Gaia* DR3 also provides an abundance of other data along with their caveats. Here we describe the steps we took to select the best sample of *Gaia* properties.

II.2.4.1 Non-SOS G-band photometry

Given that $\geq$ 99.1% of stars have a *Gaia* DR3 match, *Gaia* can be used as a source of homogeneous photometry for our catalog. There are two such photometric sources possible (Hambly et al., 2022): `phot_g_mean_mag` from `gaia_source` and `mean_mag_g_fov` from `vari_summary` (similarly for BP/RP photometry as well). The former is computed as weighted average of the star's flux over all observations and then converted into magnitude. The latter is the unweighted mean of all magnitudes recorded in a star's time-series. Eyer et al. (2023, c.f. their Fig. 8.) noted a 0.2 mag bias in `phot_g_mean_mag` because it is wieghted with flux uncertainty. This comes about because: when a variable star is fainter (at a minimum of its light curve), its Poisson flux uncertainty is lower. Therefore, the weight (inverse of flux uncertainty) will be larger for lower fluxes. Thus, a star's flux will
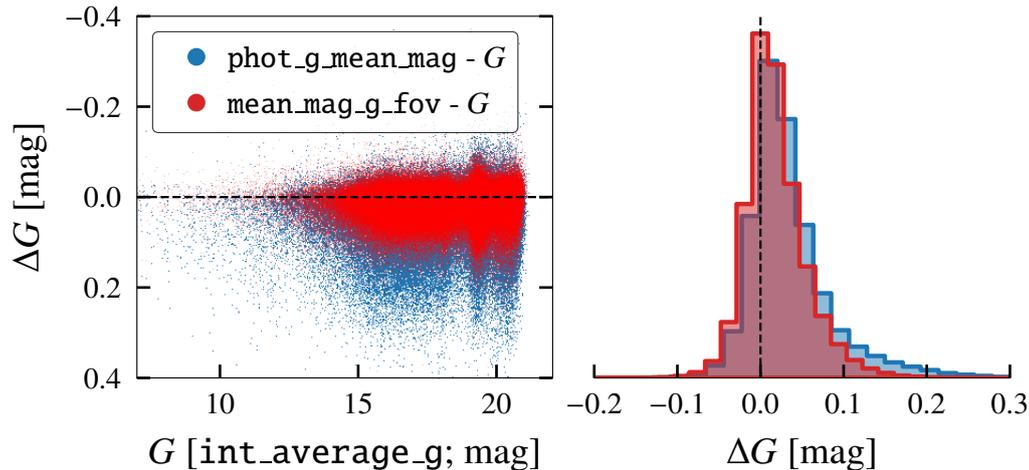
FIGURE II.14: Comparison of flux uncertainty-weighted magnitudes (`phot_g_mean_mag`, in blue) and unweighted mean magnitudes (`mean_mag_g_fov`, in red) against Fourier-fit intensity-averaged magnitudes (`int_average_g` or G). The residuals ΔG are plotted against G magnitude on the *left* panel and their normalized histograms are plotted on the *right* panel.

be biased towards lower values and, as a result, these stars will have overestimated (fainter) `phot_g_mean_mag` (or underestimated flux).

To confirm this bias, we computed the residuals of both these magnitudes against the intensity-averaged magnitudes computed by SOS Cep&RRL pipeline (plotted in Fig. II.14). The latter are computed as described in Clementini et al. (2019). First, a Fourier model is fit to a star's light curve (taken in magnitudes). Then, this model is converted into intensity (or flux) for computation of its mean. This mean is then converted back into magnitude and reported as `int_average_g` in the DR3 tables `vari_rrlyrae` and `vari_cepheid`. We note that the residuals for `phot_g_mean_mag` are indeed greater than those for `mean_mag_g_fov`, i.e., overall, `mean_mag_g_fov` is a better approximation to `int_average_g` than `phot_g_mean_mag`. We noticed a similar pattern for BP/RP photometry, too.

II.2.4.2 BP/RP photometry

While validating photometric content of *Gaia* EDR3, Riello et al. (2021) discovered that about half of the stars fainter than $G_{BP} \sim 20.3$–20.9 mag are potentially 0.2 mag fainter than their G_{BP} value given in DR3 `gaia_source`. This occurs due to an on-board minimum flux threshold set on epoch photometry of stars (in order to avoid extreme outliers). Given that a bias of 0.2 mag in G_{BP} has considerable effects on the distance of a star computed using BP photometry, we discarded BP photometry for all stars fainter than $G_{BP}=20.3$ mag.

The parameter `phot_bp_rp_excess_factor` can identify stars with incoherent G-band

vs BP/RP photometry. It measure the excess flux in BP and RP as compared to that in G-band. Riello et al. introduced a corrected excess factor, C^*, for *Gaia* DR3. Under this new formalism, objects with positive C^* values (i.e. abnormally higher flux in BP/RP) could belong to one of these subsets: extended objects, those with anomalous SEDs (quasars or planetary nebulae), stars whose background flux has been overestimated, high-probability blended stars. However, they also note that a stringent filter on C^* can remove variable stars, too. Based on its distribution for our sample, we flagged BP/RP photometry of stars with $C^* > 3$.

Riello et al. defined a blending fraction β to tackle the effects of crowding on BP/RP photometry. We did not see any strong correlation between β and BP/RP photometry of our stars. Hence, we flagged BP/RP photometry only for those stars whose $\beta = 1.0$, i.e., all of their DR3 BP and RP transits were registered as blended.

Gaia Collaboration et al. (2021a) and Torra et al. (2021) introduced `transit_ratio`, which measures the fraction of *Gaia* DR2 transits for an object that were retained in DR3. For ~ 700 stars with `transit_ratio` < 1, we flagged their BP/RP photometry. These are mostly very bright stars ($G \lesssim 5$ mag) or faint stars in crowded regions.

All in all, we flagged BP/RP photometry of stars with:

$$
\begin{aligned}
&G_{\mathrm{BP}} > 20.3 \\
&\mathrm{OR}\ C^* > 3 \\
&\mathrm{OR}\ \texttt{transit\_ratio} < 1 \\
&\mathrm{OR}\ \beta < 1.0.
\end{aligned}
\tag{II.6}
$$

II.2.4.3 Astrometry

Following Fabricius et al. (2021); Lindegren et al. (2012), we filter on `astrometric_excess_noise` to remove stars with larger than expected residuals from astrometric fitting. We flagged astrometry for ~ 9000 stars that satisfied the following conditions (tailored to our sample):

$$
\begin{aligned}
&(\texttt{astrometric\_excess\_noise} > 5 \\
&\mathrm{OR}\ \texttt{astrometric\_excess\_noise\_sig} > 2) \\
&\mathrm{AND}\ (1 \le \texttt{astrometric\_excess\_noise} \le 5 \\
&\mathrm{OR}\ \texttt{astrometric\_excess\_noise\_sig} > 100).
\end{aligned}
\tag{II.7}
$$

Similarly, we followed their prescription to identify possible binaries/duplicity signa-

tures in the processing of the transits. We flagged ~ 400 stars with:

$$
\begin{aligned}
&\texttt{ipd\_frac\_multi\_peak} > 2 \\
&\text{AND } \texttt{ipd\_gof\_harmonic\_amplitude} > 0.4 \\
&\text{AND } \texttt{ruwe} > 1.4.
\end{aligned}
\tag{II.8}
$$

$\texttt{ipd\_frac\_multi\_peak}$, $\texttt{ipd\_gof\_harmonic\_amplitude}$ are computed by the Image Parameter Determination (IPD) module of *Gaia*. They indicate the detection of resolved stellar pairs and asymmetry in the spatial structure of an object's image, respectively. $\texttt{ruwe}$ is the renormalized unit weight error and together with the IPD parameters, it helps in filtering out non-single objects whose *Gaia* astrometry is unideal.

We combined Eqns. II.6, II.7, II.8 into a single flag we term GAIA_FLAG. This flag is set to 0 for stars with neither astrometry nor photometry flagged and set to 3 for both flagged. It is set to 1 (2) if only astrometry (photometry) is flagged. The on-sky distribution of stars with GAIA_FLAG = 3 is shown in the middle panel of Fig. II.15. Indeed, a majority of them lie along the Galactic disk and bulge. Impressions of the *Gaia* scanning law can also be perceived from the map. In the top and bottom panels of Fig. II.15, we demonstrate that the flagged stars have lower N_{obs} in all three bands as compared to those with GAIA_FLAG <3. With *Gaia* DR4, most of these stars will have enough observations and, thereby, astrometry and photometry of better quality.
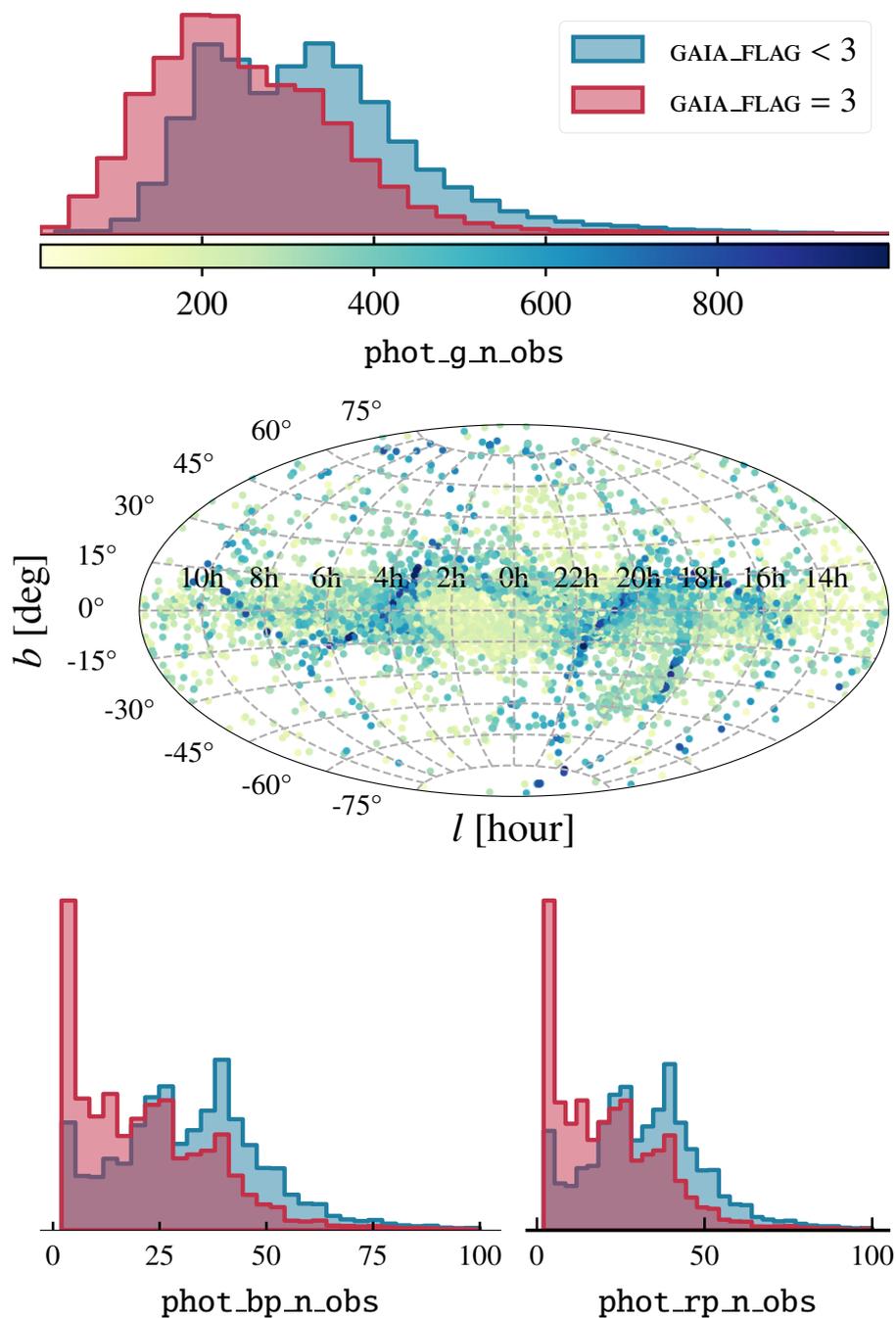
FIGURE II.15: Distribution of number of observations (in *G*-band, *top*; BP/RP *bottom*) of stars with GAIA_FLAG = 3 (i.e., astrometry and photometry flagged, in red) against those with GAIA_FLAG < 3 (in blue). The *middle* panel shows an Aitoff projection of the on-sky distribution of stars with GAIA_FLAG = 3 (in galactic coordinates). The points are colored according to the number of *G*-band observations.

TABLE II.4: Number of stars of different subtypes in all survey catalogs and in the merged and validated versions of our catalog (RRLCep).
Notes. (*a*) This row presents the preliminary number of stars from the BXM-merged catalog. This information is included here for the sake of completeness. In this version, the classification has been provisionally selected as the one from the survey with the highest number of observations. (*b*) This row lists the final number of stars of every subtype in our catalog for which the classification has been validated in Sections III.1.2 and III.2. (*c*) This column lists the number of stars of following subtypes: DCEP_2O, aRRd, pW Vir, ACEP_und (anomalous Cepheid without mode identification).

Number of → Catalog↓	RRab	RRc	RRd	DCEP_F	DCEP_1O	DCEP_mul	BLHER	WVIR	RVTAU	ACEP_F	ACEP_1O	Other*c*	Total
OGLE	92 103	32 831	3251	6435	4169	970	798	732	379	262	126	307	142 363
Gaia	174 931	93 933	2006	7319	4858	363	684	884	335	286	235	31	285 865
CSS	31 264	9151	999	10	0	0	80	97	61	214	0	240	42 116
ZTF	32 217	13 773	0	777	330	57	187	112	48	0	0	0	47 501
ASAS-SN	35 401	12 920	889	1552	553	0	349	424	339	0	0	0	52 427
MACC	1326	302	50	352	68	15	22	20	86	0	0	0	2241
PS1	44 179	16 936	0	0	0	0	0	0	0	0	0	0	61 115
DES	6971	0	0	0	0	0	0	0	0	0	0	0	6971
RRLCep pre-validation*a*	202 454	105 164	4875	8030	4829	1001	1264	1316	840	413	189	372	330 747
RRLCep post-validation*b*	202 148	104 637	4462	7985	4638	1132	1033	1243	779	453	234	309	329 053

III

Validation

When it is not in our power to determine what is true, we ought to follow what is most probable.
– RENE DESCARTES

Given the varying properties of all the surveys and their distinct period-finding and classification algorithms, it was essential to cross-check these labels for every survey. We compare and validate variability periods and classification of all catalogs with each other in Sect. III.1.1 and III.1.2, respectively. In Sect. III.2, we validate our catalog's properties using other literature catalogs.

III.1   Internal validation

III.1.1   Period validation

Here, we discuss how we computed the final period for stars observed in multiple surveys. We also measure the accuracy of periods by comparing all surveys against each other. As depicted in Fig. II.13, $\sim 40\%$ of stars were classified in just one survey (i.e., $N_{\text{survey}} = 1$). We compare periods of the other $\sim 60\%$ (i.e., $\sim 196\,000$ stars with $N_{\text{survey}} \geq 2$) to establish an internal accuracy estimate of the former sample. Moreover, this comparison only includes

single-mode pulsating stars and leaves out $\sim 5\,000$ stars classified as RRd or multi-mode Cepheids.

The significance of measuring accurate periods (and validating them) can be appreciated by comparing the light curves presented in Fig. I.3 with those in Fig. A.1. In Fig. I.3, we phased the light curves with the "true" period, while in Fig. A.1, the light curves were phased with an altered period. We changed the "true" periods by 0.05% and the resultant light curves are drastically different. The effect is more prominent on the shorter-period variables and less so on the longer-period variables. Therefore, in the following validation, we have defined different accuracy limits for stars in three distinct period ranges.

We calculated the relative percentage difference (defined for two surveys 'a' and 'b' as $\Delta P_{ab}^{\%} = |P_a - P_b|/P_a \times 100$) between periods of all possible combinations of the eight surveys $\binom{8}{2}$ $(= 28)$. For all 28 pairs, we flagged the stars if PERIOD_FLAG $=$True, which is defined as:

$$\text{PERIOD\_FLAG} = \begin{cases} \text{True,} & \begin{aligned} &\text{if } \Delta P_{ab}^{\%} > 0.1 \quad \& \quad P_a \leq 1\,\mathrm{d}; \\ &\text{or } \Delta P_{ab}^{\%} > 0.5 \quad \& \quad 1\,\mathrm{d} < P_a \leq 20\,\mathrm{d}; \\ &\text{or } \Delta P_{ab}^{\%} > 1.0 \quad \& \quad P_a > 20\,\mathrm{d}; \end{aligned} \\ \text{False,} & \text{otherwise.} \end{cases}$$ (III.1)

The limits on $\Delta P_{ab}^{\%}$ are set such that, for a star with $P_a \sim 0.5\,\mathrm{d}$, P_b is considered close enough if $|P_a - P_b|$ is ≤ 1 min. Similarly, for a star with $P_a \sim 10\,\mathrm{d}$ (or $45\,\mathrm{d}$), the absolute difference has to be $\leq 1\,\mathrm{h}$ (or $10\,\mathrm{h}$). This step flagged $\sim 4.5\%$ of $\sim 196\,000$ stars with $N_{\text{survey}} \geq 2$ (2.7% of all stars), indicating that at least one survey has a period estimate over the aforementioned limits for these stars. 3312 (2.5%) out of $\sim 129\,000$ stars with $N_{\text{survey}} = 2$ have PERIOD_FLAG $=$ True. We selected the final periods for these stars from the survey with greater N_{obs}. We call this sample of stars 'sample A' and discuss it in Sect. III.1.1.2. Moreover, out of $\sim 70\,000$ stars with $N_{\text{survey}} \geq 3$, 5227 ($\sim 7.5\%$) stars have PERIOD_FLAG $=$True. We describe here how we assessed periods of these stars.

III.1.1.1 Computing final periods

We encountered the following specific combinations in the PERIOD_FLAG $=$True sample.

There are ~ 50 cases of a star observed by exactly three surveys and its three estimates are very close to each other, say, $P_a < P_b < P_c$. This means, it has $\Delta P_{ab}^{\%}$ and $\Delta P_{bc}^{\%}$ within the threshold set in Eqn. III.1 but $\Delta P_{ac}^{\%}$ is not and, hence, the star was flagged. For these stars, we computed the N_{obs}-weighted mean of their periods and checked whether $\Delta P_{\text{mean,x}}^{\%}$ for $x \in (a, b, c)$, was within the threshold. All fifty cases were satisfactory — therefore, we assumed this weighted average as the final period (P_{RRLCep}) for these stars.

A total of X stars with $N_{\text{survey}} = n$, for $n \in [3, 4, 5, 6]$, have $(n - 1)$ periods consistent within the limits in Eqn. III.1, while the nth survey has an incorrect or aliased period. The

TABLE III.1: Stars with $N_{survey} \geq 3$ that had 1 incorrect period estimate.

Notes. n is the number of surveys that X stars are detected in and have PERIOD_FLAG =True. Only the top three most frequent discrepant surveys for every (n, X) pair are listed here. A row in this table reads: out of 2982 stars with $N_{survey}=3$ and PERIOD_FLAG =True, the discrepant period is contributed by PS1 for 1661 stars.

n	X	Most frequent discrepant survey	# of stars per discrepant survey
3	2982	PS1	1661
		Gaia	660
		ASAS-SN	213
4	1651	PS1	817
		Gaia	277
		CSS	220
5	538	PS1	279
		CSS	81
		Gaia	77
6	38	DES	27
		PS1	5
		Gaia	4

number of stars X for every n is given in Table III.1. The surveys that are most frequently the nth survey (the one with the discrepant period) are PS1, *Gaia*, CSS, DES (in descending order of frequency).

Another sample of Y stars observed in exactly n surveys, for $n \in [4, 5, 6]$, have $(n-2)$ periods consistent among each other and 2 discrepant estimates. Y and n are given in Table III.2. The two most frequent surveys causing the discrepancy are PS1 and CSS. Furthermore, Z stars have $n=(2, 3, 4)$ period estimates consistent with each other and $m = 2$ period estimates consistent among each other. However, the median periods of both groups are considerably off. Z for every n is given in Table III.3. The pair of surveys giving rise to the discrepant group of m periods most frequently are PS1 and DES.

For all X, Y, and Z stars, we computed their final period as the N_{obs}-weighted mean of their consistent periods. The source surveys of these 5227 flagged stars are shown in Fig. III.1. Since *Gaia* contains the most stars, it has substantial representation in this flagged sample. It has periods estimated for ~ 5100 of these 5227 stars. However, the discrepant period is the one from *Gaia* only for $\sim 20\%$ (995) of these. On the other hand, DES has a period estimate for 542 out of these 5227 stars, and it is the source of discrepancy for $\sim 84\%$

TABLE III.2: Stars with $N_{survey} \geq 4$ that had 2 inconsistent period estimates.

Notes. n is the number of surveys that Y stars are detected in and have PERIOD_FLAG =True. Only the top three most frequent discrepant surveys for every (n, Y) pair are listed here. A row in this table reads: out of 42 stars with $N_{survey}=4$ and PERIOD_FLAG =True, the discrepant periods are contributed by PS1 and CSS for 12 stars.

n	Y	Discrepant pair	# of stars per pair
4	42	PS1, CSS	12
		Gaia, CSS	8
		Gaia, DES	4
5	15	PS1, CSS	5
		Gaia, CSS	4
		PS1, *Gaia*	2
6	2	DES, PS1	1
		DES, CSS	1

TABLE III.3: Stars with $N_{survey} \geq 4$ whose periods were consistent among two groups of surveys but inconsistent with each other.

Notes. n is the number of surveys that Z stars are detected in and have PERIOD_FLAG =True. Only the top two most frequent survey combinations for every (n, Z) pair are listed here. N_{obs}^{sum} is the sum of N_{obs} in each survey of the combination. A row in this table reads: out of 27 stars with $N_{survey}=4$ and PERIOD_FLAG =True, for 7 such stars, ASAS-SN - CSS is the pair with the greater number of observations, while *Gaia*-ZTF has fewer observations.

n	Z	Combination with greater N_{obs}^{sum}	Combination with fewer N_{obs}^{sum}	# of stars per combination
4	27	ASAS-SN - CSS	*Gaia* - ZTF	7
		CSS - PS1	*Gaia* - ZTF	5
5	22	ASAS-SN - *Gaia* - ZTF	CSS - PS1	6
		ASAS-SN - CSS - PS1	*Gaia* - ZTF	2

(456) of them. The same value for PS1 is $\sim 70\%$.

To investigate this high discrepancy ratio, we plotted the period from DES against the final period for these stars (see Fig. III.2). We noticed the same trend that was illustrated in Fig. II.9. Stars with $0.6 < \mathrm{DES\ score_{ab}} < 0.8$ form the bulk of the discrepant sample (219 out of 456 stars). Moreover, 221 stars from this sample are classified as RRc stars in other surveys. Their classification is discussed in detail in Sect. III.1.2. Consequently, we removed ~ 900 unique DES RRab stars from our catalog that have $0.6 < \mathrm{DES\ score_{ab}} < 0.8$. We performed a similar investigation for the PS1 sample but did not notice any trend with their N_{obs} in PS1, classes, or classification probabilities.

Finally, 92 stars have three period estimates all inconsistent with each other (~ 85 included a *Gaia* estimate and have low N_{obs}). We flagged these stars and provisionally accepted the estimate with the highest N_{obs} as the best period. Overall, $\sim 188\,000$ stars detected in more than one survey have PERIOD_FLAG =False. Their final periods were computed as the N_{obs}-weighted average of the survey periods.

III.1.1.2 Results of validation

We present a summary of the validation results in Table III.4. As indicated in column (c) of this table, at least $\gtrsim 65\%$ stars from each survey were compared against at least one other survey during our validation. We combined the stars with discrepant period estimates (i.e., PERIOD_FLAG =True) listed in Tables III.1–III.3 and computed the error rate for all surveys. For a given survey S, this rate is defined as the number of stars for which it has a discrepant period divided by the total number of stars in this survey that are also detected in at least one other survey (i.e., $N_{\mathrm{survey}} > 1$). In other words, it quantifies the fraction of stars for which survey S has an incorrect period (based on the limits set in Eqn. III.1) compared to other surveys. As displayed in Table III.4, the error rates are notably higher for DES and PS1, highlighting the challenges of computing periods using sparse, multi-band light curves.

Multiplying these error rates by the number of unique stars from each survey, $N_{\mathrm{star}}^{\mathrm{unique}}$, yields the expected number of stars with discrepant periods ($N_{\mathrm{bad}}^{\mathrm{unique}}$) for each specific survey. We calculated this expected $N_{\mathrm{bad}}^{\mathrm{unique}}$ for every survey's unique population ($N_{\mathrm{survey}} = 1$). Similarly, we computed the expected $N_{\mathrm{bad}}^{\mathrm{A}}$ based on stars belonging to 'sample A' (defined in Sect. III.1.1). These values are presented in Table III.4. Consequently, we anticipate that the periods of 927 stars from the $N_{\mathrm{survey}} = 1$ sample of $\sim 129\,000$ stars may be inaccurate according to the thresholds set in Eqn. III.1. Similarly, we expect 58 stars in 'sample A' to have inaccurate periods based on the error rates of the surveys from which their period was selected. As a result, out of a total of $\sim 325\,000$ single-mode pulsating stars in our catalog, we estimate ~ 1000 stars (0.3%) may have inaccurate final periods at accuracy levels set in Eqn. III.1.
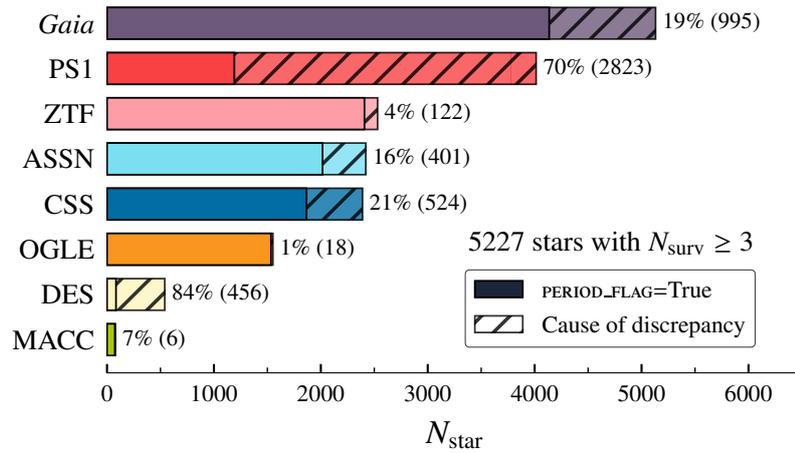
FIGURE III.1: Provenance of 5227 stars with $N_{survey} \geq 3$ that were flagged by Eqn III.1. The crossed portion of every bar denotes the percentage of stars for which a certain survey contributes the discrepant period estimate. For instance, out of these 5227 flagged stars, *Gaia* has a period estimate for ~5 000 of them. However, the discrepant estimate is from *Gaia* only for 995 of them. This percentage and the number of stars it reflects is shown next to every bar.
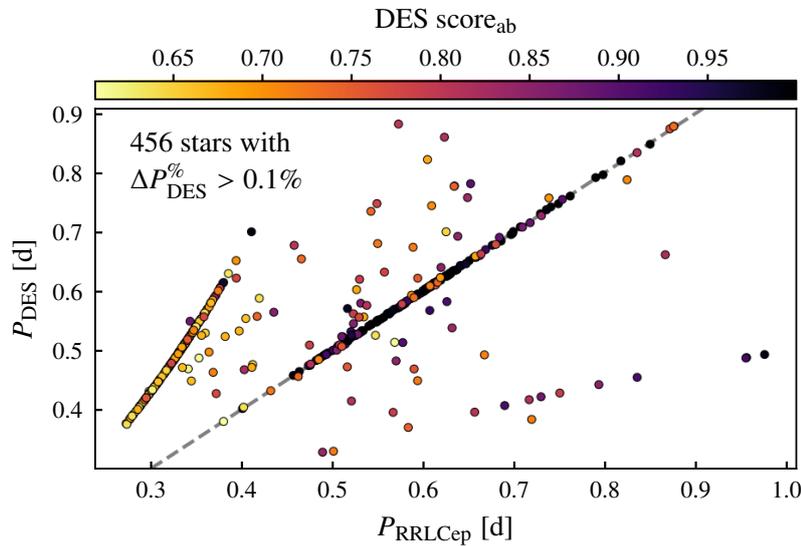
FIGURE III.2: DES period (P_{DES}) versus final periods in our catalog (P_{RRLCep}) for 456 stars, where DES period is inaccurate with respect to at least two other surveys. The points are color-coded according to their DES RRab classification probability, score$_{ab}$. For details behind the structures of the off-identity line distributions (beat or aliased frequencies), refer VanderPlas & Ivezić (2015).
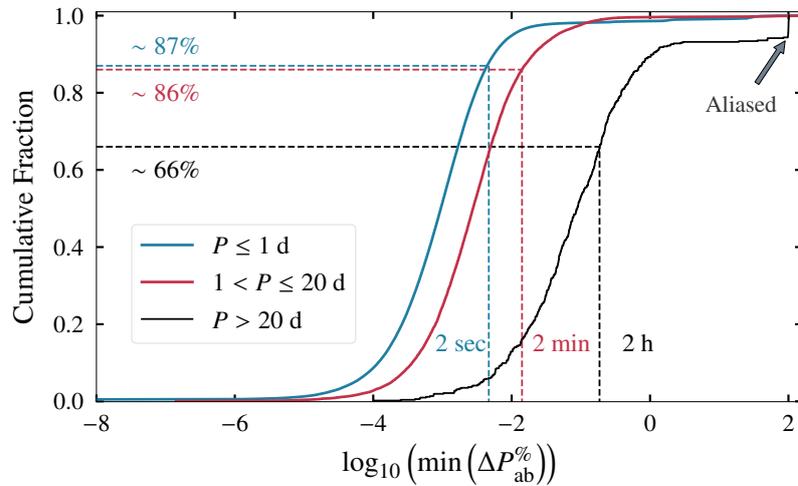
FIGURE III.3: Cumulative distribution of the logarithm of minimum $\Delta P_{ab}^{\%}$ for $\sim 196\,000$ stars with $N_{survey} \geq 2$. Stars are divided into three period ranges shown in the figure: short (in blue), intermediate (in red), and long (in black). The dashed lines indicate the corresponding limits for the following representative variables of period ranges: short: 0.5 d; intermediate: 10 d; long: 45 d variable.
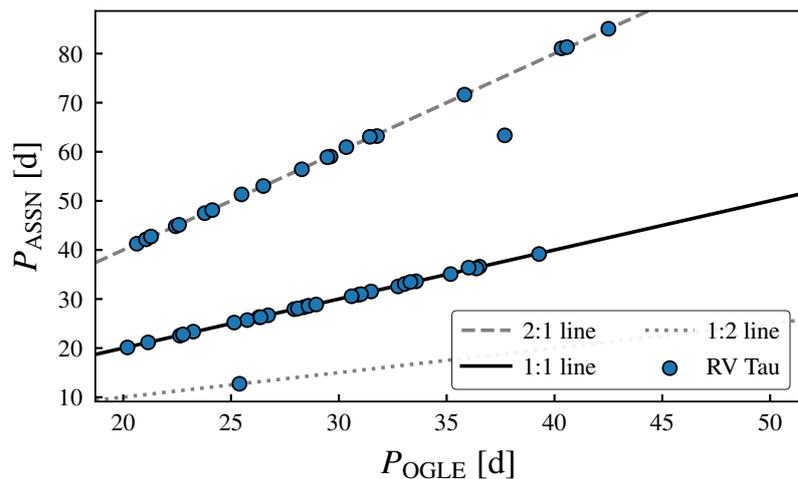
FIGURE III.4: Periods in ASAS-SN catalog (P_{ASSN}) versus those in OGLE (P_{OGLE}) for RV Tau stars.

Table III.4: Summary of the internal period validation.

Notes. Numbers shown here only include single-mode pulsators. The table is sorted according to the error rate. (*a*) Number of stars in the survey's sample included here. (*b*) Number of stars with $N_{survey} > 1$, therefore, their periods could be compared with other surveys. (*c*) Fraction computed as $(b)/(a) \times 100$ (shown as percentages). (*d*) Number of stars with discrepant periods. (*e*) Error rate defined as $(d)/(b) \times 100$. (*f*) Number of unique stars contributed by the survey. (*g*) Expected number of stars with discrepant periods from this survey in its unique sample, computed as: Error rate/100 $\times N_{star}^{unique}$. (*h*) Number of stars in 'sample A' whose periods are from this survey. (*i*) Expected number of discrepant periods in 'sample A', computed as Error rate/100 $\times N_{star}^A$.

Survey	N_{star}	N_{star} with $N_{survey} > 1$	% with $N_{survey} > 1$	Discrepant Periods	Error Rate (%)	N_{star}^{unique}	Expected N_{bad}^{unique}	N_{star}^A	Expected N_{bad}^A
	(*a*)	(*b*)	(*c*)	(*d*)	(*e*)	(*f*)	(*g*)	(*h*)	(*i*)
DES	6933	4556	~65.7	456	~10.01	1458	145	15	1
PS1	61 079	55 340	~90.6	2823	~5.1	4754	242	985	50
CSS	41 934	40 222	~95.9	524	~1.3	1495	19	217	2
ASAS-SN	52 099	47 274	~90.7	401	~0.85	4629	39	196	1
Gaia	281 695	190 419	~67.6	995	~0.52	91 122	476	154	1
MACC	2166	2076	~95.8	6	~0.29	80	1	10	1
ZTF	47 281	45 995	~97.3	122	~0.27	1099	2	187	1
OGLE	137 974	112 585	~81.6	18	~0.02	23 841	3	1548	1
Total						128 478	927	3312	58

Moreover, to measure the period accuracy with greater stringency, we computed the minimum value of $\Delta P_{\text{ab}}^{\%}$ for each star across all 28 survey combinations, denoted as $\min\left(\Delta P_{\text{ab}}^{\%}\right)$. This minimum value represents the absolute difference between the two closest period estimates for a star divided by one of the estimates. It measures the highest possible internal accuracy achievable for these periods, assuming that one of the estimates is the 'true' period. We computed this quantity for $\sim 196\,000$ stars with at least two period estimates ($N_{\text{survey}} \geq 2$). These stars were categorized into three period ranges: short, intermediate, and long. The boundaries are identical to those set in Eqn. III.1 and are illustrated in Fig. III.3. The cumulative distribution of $\log_{10}\left(\min\left(\Delta P_{\text{ab}}^{\%}\right)\right)$ for each period range is plotted in Fig. III.3.

This analysis reveals that for $\sim 87\%$ of short-period stars, the minimum value of $\Delta P_{\text{ab}}^{\%}$ is within $\sim 0.004\%$ (or $\leq 2\,\text{s}$ for a representative $0.5\,\text{d}$ variable). For stars with periods in the intermediate range, $\sim 86\%$ have $\min\left(\Delta P_{\text{ab}}^{\%}\right) \leq \sim 0.01\%$ (or $\leq 2\,\text{min}$ for a $10\,\text{d}$ variable). However, for ~ 500 long-period stars, only $\sim 66\%$ have $\min\left(\Delta P_{\text{ab}}^{\%}\right) \leq \sim 0.2\%$ (or $\leq 2\,\text{h}$ for a $45\,\text{d}$ variable). This lower accuracy is attributable mainly to two factors.

Firstly, in general, long-period stars require greater N_{obs} and a longer baseline of observations to have their variability phase sampled adequately. As a result, surveys with incomplete light curves may measure an alias of the 'true' period. Secondly, the formal periods of RV Tau stars are computed as the time interval between two successive deeper minima. Surveys usually report half of these formal periods, which represent the time interval between two successive minima, one shallower and one deeper (also called "single" periods). However, the ASAS-SN catalog reports the formal periods for a subset of RV Tau stars, as demonstrated in Fig. III.4 (those lying along the 2:1 line). These factors contribute to the 'aliased' peak highlighted in Fig. III.3 at $\log_{10}\left(\min\left(\Delta P_{\text{ab}}^{\%}\right)\right) \sim 2$ (i.e., $P_{\text{a}} = 2 \times P_{\text{b}}$).

To improve the overall accuracy of periods even further, we intend to undertake a homogeneous analysis, wherein periods will be computed by aggregating all light curves per star across all bands and surveys.

III.1.2 Classification

This validation has two primary objectives: (i) to assign a final classification to stars with conflicting subtypes across two or more surveys (ii) to minimize the number of misclassified stars in each survey's unique sample ($N_{\text{survey}}=1$).

With more observations for a variable star, the probability of estimating its correct variability period increases. However, this is not the case with classification. As multiple subtypes share the same period range (listed in Table II.1), a survey can compute the period accurately yet classify the star incorrectly. This occurs because stars exhibit similar light curve morphologies and overlapping distributions of Fourier parameters, which hinders subtype

identification based solely on light curves. Moreover, the different classification pipelines employed by these surveys (given in Table II.2) are better at classifying a few subtypes than at others. Therefore, for a star with different classifications in two or more surveys, we did not select its final class from the survey with higher N_{obs} for it. Instead, we implemented a three-step strategy presented here.

Step 1: Initially, we used OGLE as the base survey and compared every other survey against it. For 304 stars with uncertain classification or blending issues in OGLE (OGLE_FLAG = 2 or 4), we considered their OGLE class as flexible. Conversely, for the remaining $\sim 140\,000$ OGLE stars, their classification was regarded as absolute. This decision was made since OGLE has well-populated light curves for these stars and all of them are classified following visual inspection (Udalski et al., 2018; Skowron et al., 2019). Furthermore, OGLE has a substantial number of stars in common with most of the subtypes from all surveys, thereby enabling an initial assessment of a significant fraction of stars. Out of the former 304 stars, *Gaia* (ASAS-SN) confirmed OGLE classification for 245 (20) stars. For 16 of these stars, their *Gaia* classification differed from that of OGLE, while the remaining 43 were only detected in OGLE. We did not take into account the latter two samples (59 stars in total) when assessing the classification accuracy of other surveys. The implications of this comparison with OGLE are discussed below for every survey.

Step 2: Once the OGLE-based changes were incorporated, we repeated the steps we took to validate the periods, as discussed in Sect. III.1.1. Briefly, for each star detected in at least two surveys (out of 200K stars with $N_{\text{survey}} \geq 2$), we conducted a total of 28 comparisons (equal to $\binom{8}{2}$) and calculated a corresponding CLASS_FLAG for all stars. For a star, CLASS_FLAG is set to true if any survey's subtype classification differs from that of the others. This step flagged 5928 stars in total, i.e., $\sim 2.9\%$ of stars with $N_{\text{survey}} \geq 2$, or $\sim 1.8\%$ of all stars in our catalog. ~ 3000 stars from this sample were classified by OGLE and we retained their OGLE classification (as justified above).

Step 3: To determine the final classification of the remaining ~ 3000 stars with CLASS_FLAG = true, we inspected their light curves and compared their positions in the Bailey diagram (period-vs-amplitude distribution) or against period-Wesenheit relations[1]. This reclassification is outlined for every survey below.

To provide an overview of the implications of Step 1, we present confusion matrices comparing the original classification of each survey with that of OGLE in Appendix 1.4. Similarly, we summarize the results of this validation through confusion matrices plotted in Appendix 1.5. The latter set of matrices illustrate the comparison between the classification of the validated and cleaned sample of each survey against the final class reported in our

---

[1]We derived new period-Wesenheit relationships for all subtypes and their computation and validation will be discussed in the upcoming paper of this series.

$N_{star} = 115243 \mid F_{0.5} = 99.8\% \mid Acc_{bal} = 88.2\%$

FIGURE III.5: Confusion matrix comparing broad classification between *Gaia* and OGLE. Precision $\times$ 100 (purity) is shown in every cell below the number of stars in percentage. The lower percentages have been rounded to the second decimal. The purity percentages sum to 100% along every column. N_{star} is the number of stars common to both catalogs. $F_{0.5}$ and Acc_{bal} are classification accuracy metrics defined in Sect. III.1.2.6.

catalog (Class$_{RRLCep}$) for $\sim 200\,000$ stars detected in at least two surveys.

For every survey, we discuss the broad and subtype classification purity. The former pertains to RR Lyrae stars, classical, Type-II, and anomalous Cepheids and the latter to their various subtypes. We also outline the changes made to their classification, if any.

III.1.2.1 CSS

The CSS catalog overlaps with OGLE mainly towards the outskirts of the Magellanic Clouds and retains a $\geq 99\%$ purity in the RR Lyrae sample ($N_{star} \sim 300$). It has very few Cepheids in common with OGLE to constitute a fair comparison, though 8 DCEP_F stars that it shares with OGLE are classified as DCEP_F by OGLE as well.

CSS subtypes that have the highest fraction of flagged stars are RRd stars and anomalous Cepheids. $\sim 75\%$ of stars identified as RRd in Catalina are classified as RRc by at least two other surveys. After inspecting their light curves in CSS as well as other surveys (mainly *Gaia*, ZTF, ASAS-SN), we realized that $\sim 15\%$ of Catalina observations for most of these

FIGURE III.6: Confusion matrix comparing the initial subtype classification between *Gaia* and OGLE. Blocks of broad variability types are highlighted. The properties of the matrix are similar to those in Fig. III.5.

stars had a lower/higher photometric zero-point. This probably caused the CSS algorithm to misclassify these stars as RRd, confusing the biased observations for secondary periodicity. Out of ~ 1000 stars classified as RRd by CSS, we retained 151 stars and reclassified the others as RRc stars. Similarly, ~ 95% of stars classified as Blazhko stars by CSS were identified as RRab in other surveys. Along with the aforementioned zero-point issue, a subset of these stars had incorrect periods in CSS and were therefore misclassified as Blazhko stars (as predicted by Drake et al. 2017). Nonetheless, we changed the classification for all CSS Blazhko stars to RRab and flagged them as Blazhko candidates.

94 out of 180 CSS ACEP_F stars are classified as non-anomalous Cepheids by other surveys. 85 (5) stars out of these 94 stars are classified as RRab stars (BL Her) by more than two surveys. Given the similar light curve shapes of these subtypes, we flagged these stars as possible ACEP_F candidates, but classified them as RRab. Additionally, we flagged the classification of ~ 30 CSS-only ACEP_F stars as uncertain. The classification of a few of these stars is confirmed during external validation (Sect. III.2).

III.1.2.2   ZTF

The broad classification purity for ZTF RR Lyrae sample is ≥99%. Specifically, ZTF RRab classification shared a 99% agreement with that in OGLE, however the RRc sample suffered from ~ 2% contamination of RRd stars. The purity for Type-II Cepheids is 97%, with the 3% contaminants being RR Lyrae stars. As discussed already in Sect. II.1.3, Chen et al. (2020) did not provide subtype classification for classical Cepheids. Hence, we had used AGC (Pietrukowicz et al., 2021) to determine DCEP subtypes. Therefore, we do not report DCEP classification metrics for ZTF.

Furthermore, ZTF did not classify any anomalous Cepheids, but we recovered two OGLE ACEP stars classified as RR Lyrae in ZTF. Additionally, out of 148 stars that are classified as BL Her by ZTF, 41 are classified as ACEP_F by at least one other survey (*Gaia* or CSS). Our visual inspection could not disentangle their classification. No artifacts were found in the light curves either. Regardless, we flagged the classification of all 32 ZTF-only BL Her stars in our collection as uncertain, suspecting a contamination level of ~30%.

III.1.2.3   ASAS-SN

The broad classification purity for ASAS-SN RR Lyrae sample is ≥99%. However, focusing on RR Lyrae subtypes, over 99% of ASAS-SN RRd stars are classified as RRab by more than one of *Gaia*, PS1, and OGLE. We visually inspected the light curves of these stars and a significant number of them had noisy ASAS-SN light curves. A fraction of their observations, phased with the same variability period, followed the same shape as the majority of the

light curve but was shifted in magnitude (similar to RRd stars in Catalina). Since ASAS-SN combined data from 20 telescopes, a subset of them might have had a different photometric zero-point for a period of time, and this bias was possibly detected by their classification algorithm as secondary periodicity. This affected the fainter RRab stars in the ASAS-SN V-band catalog the most. However, their periods and photometry across surveys are largely consistent and additionally, Christy et al. (2022) using deeper g-band data classified these stars as RRab. Hence, we reclassified the ASAS-SN RRd stars as RRab.

The subtype classification purity for ASAS-SN DCEP_F with respect to OGLE is ≥95% and the same for DCEP_1O is 91%. The main contamination sample ($\sim$ 5%) in the latter is multi-mode Cepheids.

$\sim$ 240 ASAS-SN Type-II Cepheids in the Magellanic Clouds are identified as classical Cepheids in OGLE (see Fig. A.2). The ASAS-SN random forest classifier was trained on photogeometric distances computed using a Milky Way model (Bailer-Jones et al., 2018, 2021). As a result, the distance to the Magellanic Clouds sample was probably underestimated and the farther DCEP_F were classified as intrinsically fainter Type-II Cepheids. Additionally, due to the large ASAS-SN FWHM, the distant sources' light curves are noisier and fuzzier (T. Jayasinghe, in priv. comm.). Hence, we changed the classification of these $\sim$ 240 stars to that of OGLE. Furthermore, we dropped all ASAS-SN-only LMC/SMC Type-II Cepheids.

The photometric uncertainty in ASAS-SN light curves also affected the faintest Type-II Cepheids (BL Her) in the Milky Way. $\sim$ 35% of ASAS-SN BL Her are classified as DCEP_F by OGLE (and other surveys). Subsequently, we removed 79 ASAS-SN-only Milky Way BL Her stars as we could not confirm their classification (due to large uncertainties either in their *Gaia* parallax or ASAS-SN photometry). The subtype confusion matrix of the cleaned ASAS-SN sample against our final classification is presented in Fig. A.8.

III.1.2.4 *Gaia*

There are $\sim$ 115 000 common stars between OGLE-IV and *Gaia* DR3 SOS catalog, out of which only 275 ($\leq$ 0.23%) stars have dissimilar variability types (see Fig. III.5). When taking into account the subtypes, the number of disagreements increases to 3919 ($\sim$ 3.4%, see Fig. III.6).

The *Gaia* RR Lyrae sample has excellent purity with respect to OGLE (≥99.9%; Fig. III.5). Only 4 *Gaia* RR Lyrae stars are classified as Cepheids in OGLE-IV. Taking subtypes into consideration, the purity for RRab is $\geq$ 99%, however, the RRc sample suffers a 5% contamination from RRd stars and a further 1% from RRab stars (see Fig. III.6). The latter sample has stars with periods around 0.35 d, which is the range where the light curve morphology of these two subtypes is quite similar.

$$N_{star} = 193974 \mid F_{0.5}^{RRLCep} = 97.9\% \mid Acc_{bal} = 93.0\%$$

Final class in our catalog	RRab	RRc	RRd	DCEP_F	DCEP_1O	DCEP_mul	BL Her	W Vir	RV Tau	ACEP_F	ACEP_1O
RRab	128680 99%	693 1%	83 4%	1 0.01%	4 0.09%	4 1%				4 1%	4 3%
RRc	442 0.34%	47333 95%	16 0.86%		28 0.62%	1 0.28%					
RRd	131 0.1%	1391 2%	1763 94%		1 0.02%						
aRRd	5	9 0.02%				6 1%					
DCEP_F				6511 97%	157 3%	21 5%	12 2%	7 0.97%	2 0.8%	3 1%	
DCEP_1O	1			67 1.0%	3841 85%	13 3%	10 1%	3 0.41%			5 4%
DCEP_2O				69 1%							
DCEP_mul	1			47 0.7%	381 8%	288 82%				1 0.48%	1 0.95%
BL Her					6 0.13%	1 0.28%	564 94%		1 0.4%	2 0.95%	3 2%
W Vir				22 0.33%			5 0.84%	668 92%	2 0.8%		
pW Vir				8 0.12%	2 0.04%		3 0.5%	44 6%			
RV Tau				8 0.12%			1 0.17%	2 0.28%	246 98%		
ACEP_F				21 0.31%	1 0.02%	13 3%	1 0.17%			200 95%	
ACEP_1O	2				12 0.27%	4 1%					92 87%

Class in *Gaia* DR3

FIGURE III.7: Confusion matrix comparing the subtype classification of the cleaned *Gaia* sample versus the final class in our catalog. This comparison includes 194 042 stars detected in *Gaia* as well as at least one other survey (i.e., $N_{survey} \geq 2$). Blocks of broad variability types are highlighted for convenience. The purity percentages sum to 100% along every column and $F_{0.5}$ and Acc_{bal} are accuracy metrics defined in Sect. III.1.2.6.

TABLE III.5: Results of the classification validation of RR Lyrae stars detected in at least two surveys.

Number of → Catalog↓	Same Type	Same Subtype	Different RR Lyrae subtype	Cepheids	Unique RR Lyrae stars	Expected misclassified RR Lyrae	Expected Cepheid contaminants
Gaia	180 637	177 856	2781	10	90 222	1388	4
ASAS-SN	44 864	44 465	399	93	4246	37	8
DES	4223	4134	89	3	1438	30	1
PS1	56 274	55 956	318	79	4754	26	6
CSS	40 267	39 747	520	26	1359	17	1
ZTF	44 924	44 317	607	43	1022	13	1
MACC	1617	1615	2	5	56	1	1

TABLE III.6: Results of the classification validation of Cepheids detected in at least two surveys.

Number of → Catalog↓	Same Type	Same Subtype	Different Cepheid subtype	RR Lyrae stars	Unique Cepheids	Expected Misclassified Cepheids	Expected RR Lyrae contaminants
Gaia	13 472	12 421	1051	53	1377	107	5
ASAS-SN	2700	2444	256	40	384	35	5
CSS	206	185	21	87	136	9	40
ZTF	1397	1334	63	7	75	3	1
MACC	515	500	15	0	48	1	0

Additionally, 280 stars in the SMC classified as RRc by *Gaia*, show signs of secondary periodicity in OGLE light curves and are classified as RRd by the latter. We adopted the OGLE classification for this sample.

Over ∼ 25% of the *Gaia* RRd sample is classified as RRc stars in ZTF, CSS, PS1, and ASAS-SN. After incorporating the updated periods of *Gaia* RRd stars from Table 2 of Clementini et al. (2019), we compared the ratio of these periods against the periods of RRc stars from the above listed surveys. It became evident that these ground-based telescopes were unable to detect the weaker fundamental modes in light curves of these stars (with fundamental mode period ∼ 0.48d). However, the first overtone periods of *Gaia* RRd stars and the periods of RRc stars from the other surveys were found to be in agreement. As a result, we retained the *Gaia* classification for these stars.

The *Gaia* DCEP sample is ≥99% pure with only about 30 contaminant stars each from RR Lyrae, Type-II Cepheids, and ACEP samples from OGLE. While the DCEP_F subtype classification purity is 98%, for DCEP_1O it stands at ∼ 85%. The main source of contami-

nation in the latter is multi-mode Cepheids ($\sim 8\%$). Similarly, 10% of *Gaia* DCEP_mul are classified as either DCEP_F or DCEP_1O by OGLE. As discussed in Ripepi et al. (2023), with more observations in *Gaia* DR4, the detection efficiency of multi-mode pulsations in Cepheids will improve significantly.

The broad classification purity for *Gaia* Type-II Cepheid sample is > 97%. Most of the contaminants are identified as DCEP in OGLE. Regarding the subtype classification, OGLE is the only survey (of the ones included here) which identifies pW Vir stars and based on their samples, it seems ~ 10-20% of Milky Way-LMC W Vir stars show strong signatures of pW Vir variability. Therefore, unsurprisingly, we found that $\sim 7\%$ of *Gaia* W Vir stars matched with OGLE pW Vir stars. Furthermore, the *Gaia* BL Her and RV Tau samples share a $\geq 98\%$ agreement with their respective OGLE samples.

In the *Gaia* SMC sample, we recovered 115 anomalous Cepheids that are identified as DCEP in OGLE. The periods of these stars from both surveys agree to within 0.03%, confirming they are highly likely to be true matches. Thus, the ACEP classification in the SMC is sub-optimal (as can be confirmed in Fig. F.3 of Ripepi et al. 2023). As discussed by Ripepi et al. (2023), the line-of-sight elongation of the SMC led to inaccurate absolute magnitude estimates of these stars and hence, the SOS pipeline confused them as ACEP. We accepted the OGLE classification for these stars. Besides, OGLE, on average, has 10 times more observations than *Gaia* for these stars (470 vs 43).

III.1.2.5   MACC, PS1, DES

Barring one OGLE ACEP_F and two RRc stars, MACC and OGLE shared a 100% agreement for broader and subtype classification (see Fig. A.7). However, we found six out of 18 BL Her stars in MACC identified as ACEP_F in *Gaia* and CSS. By checking their position in period-Wesenheit diagrams, we adopted ACEP_F as their class for three of these stars. Two other BL Her stars are classified as ACEP_F by literature studies using space-based data (discussed in the next section). Therefore, we flagged ten MACC-only BL Her stars as possible ACEP_F candidates.

Out of 6784 common stars between PS1 and OGLE, only 16 have disagreeing classes (see Fig. A.4). While Sesar et al. (2017) identified these 16 as RR Lyrae stars, OGLE-IV classified them as Cepheids (8 Type-II Cepheids, 6 ACEP, 2 DCEP). For the 6768 stars identified as RR Lyrae stars by both surveys, only 32 stars have dissimilar subtypes. 12 (9) OGLE RRc (RRab) stars are identified as RRab (RRc) by PS1, while 11 of PS1 RRab/c stars were classified as RRd by OGLE. With respect to the final classification, the false discovery/subtype confusion rate of PS1 is $\sim 1\%$ for RRab and $\sim 2\%$ for RRc stars.

The broad classification purity for the DES catalog is $\geq 99\%$ with just 2 out of 801 common stars with OGLE identified as ACEP by the latter. Although Stringer et al. (2021)

only identified RRab stars, we found $\sim 7\%$ (1%) contamination from RRc (RRd) stars (see Fig. A.3). The DES RRab stars classified as RRc stars in OGLE occupy the expected location of the latter in the Bailey diagram. $\sim 55\%$ of these stars have a DES classification probability between 0.6 and 0.8. These numbers largely remain the same when we extend the comparison to the final classification sample. Therefore, regarding the DES RRab stars with DES score$_{ab}$ < 0.8, we either removed them if they were detected only in DES or adopted the classification of other surveys in which they are detected. As depicted in Fig. II.9, this cut did not unfairly omit many true RRab stars. The confusion matrix of the cleaned DES sample against our final classification is shown in Fig. A.9.

Consequently, the classification of ~ 100 stars remains uncertain, they could belong to one of these classes: ACEP_F, BL Her, RRab. These stars make up $\sim 5\%$ of all Cepheids with CLASS_FLAG =true in our catalog ($\sim 0.6\%$ of all Cepheids) and have not been included in the numbers listed in Table II.4. A few of these stars are assigned a definitive class in the next section. For the rest of them, their uncertainty persists. During our analysis, we did not re-fit any light curves using a Fourier model. However, with a detailed Fourier analysis of the rich light curves of *Gaia* DR4 or other space-based observatories like *TESS* and *Kepler*, their classification can be disentangled.

III.1.2.6 Summarizing results of classification validation

With every confusion matrix, we report F_β as the global accuracy measure of a survey's classification with respect to either OGLE or the final classification in our catalog (shown in Appendices 1.4, 1.5, respectively). The multiclass F_β score is selected as the accuracy metric because our comparisons are a case of multiclass classification, with heavily imbalanced classes. This imbalance arises as RR Lyrae stars vastly outnumber the Cepheids. The F_β score (Rijsbergen, 1979) is defined as:

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{\left(\beta^2 \cdot \text{precision}\right) + \text{recall}}, \tag{III.2}$$

where β is the weighting parameter. We use $\beta = 0.5$, which means we assign double the weight of recall to precision. For our catalog, per-survey purity (precision) is more important than completeness (recall). While a survey could be less complete for a particular subtype, our compilation ensures that the final catalog achieves the maximum possible completeness with respect to the literature. We also computed the measure Balanced Accuracy (Acc$_{bal}$) defined as the average of recall and true negative rate for every survey. However, this score gives equal weight to both precision and recall (its shortcomings are illustrated in Fig. III.10). Nonetheless, we report both values for all surveys.

These scores estimate the classification accuracies of each survey for the

$N_{survey} \geq 2$ sample of $\sim 200\,000$ stars. We retrieve $F_{0.5}^{RRLCep} \geq 97.5\%$ for the cleaned samples of all catalogs except DES, for which $F_{0.5}^{RRLCep} \sim 96.3\%$.

As done while validating periods, we use the contamination rates found in this $N_{survey} \geq 2$ sample to estimate the minimum number of contaminants or misclassified stars in the $N_{survey}=1$ sample, i.e., the unique population of each survey. These findings are presented in Table III.5 for the RR Lyrae stars and in Table III.6 for the Cepheids. Based on these estimates, we anticipate a minimum of ~ 1512 stars out of $\sim 125\,000$ RR Lyrae stars with $N_{survey}=1$ may have the wrong subtypes, i.e., confusion between RRab, RRc, and RRd classification. Since *Gaia* contributes the highest number of unique RR Lyrae stars in the halo $\sim 90\,000$, it has the largest presence in this estimation. We expect a contamination of ~ 1400 stars in its unique RR Lyrae sample, with most of the confusion caused by undetected RRd stars (whose classification is impeded by low N_{obs} in *Gaia* DR3).

Similarly, out of ~ 4100 unique Cepheids contributed by all surveys, we estimate ~ 155 may be misclassified as the wrong Cepheid subtype and a further ~ 51 may be RR Lyrae stars. Thus, we estimate a maximum classification accuracy for the sample of $N_{survey}=1$ Cepheids to be $\sim 95\%$ and the same for the $N_{survey}=1$ RR Lyrae sample to be $\sim 98.5\%$. These estimates are corroborated through comparisons with external catalogs, as discussed in the following section.

Additionally, we are currently working on a stacked ensemble of recurrent and deep neural networks trained on light curves and newly derived light curve-related features of these stars. Based on preliminary results, we achieve $\sim 97\%$ accuracy in automatic classification of Cepheids of various subtypes. This architecture will be presented in a future work of this series.

Figures III.8, III.9 depict the on-sky distributions of $\sim 311\,000$ RR Lyrae stars and $\sim 18\,000$ Cepheids included in our catalog. The light curves of a randomly selected member of each subtype are presented in Fig. I.3. Figure I.4 shows the Bailey diagrams plotted with amplitudes in the *Gaia G*-band (left panels) and the OGLE *I*-band (right panels) for our target subtypes.

III.2 External validation

Here, we compare the internally-validated classification and periods in RRLCep against literature studies focused on RR Lyrae stars and Cepheids. We also cross-match against catalogs of variable stars of subtypes not considered here in Sect. III.2.5.
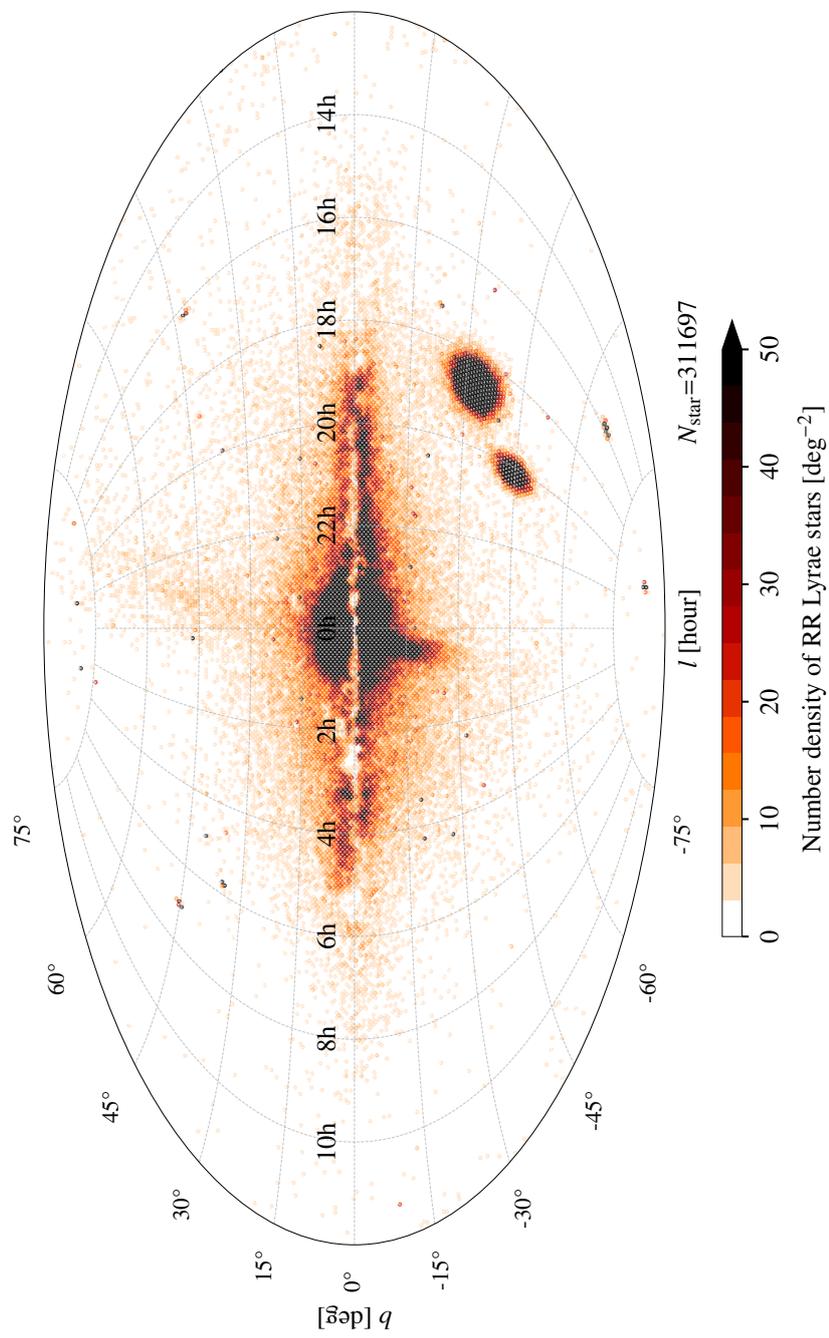
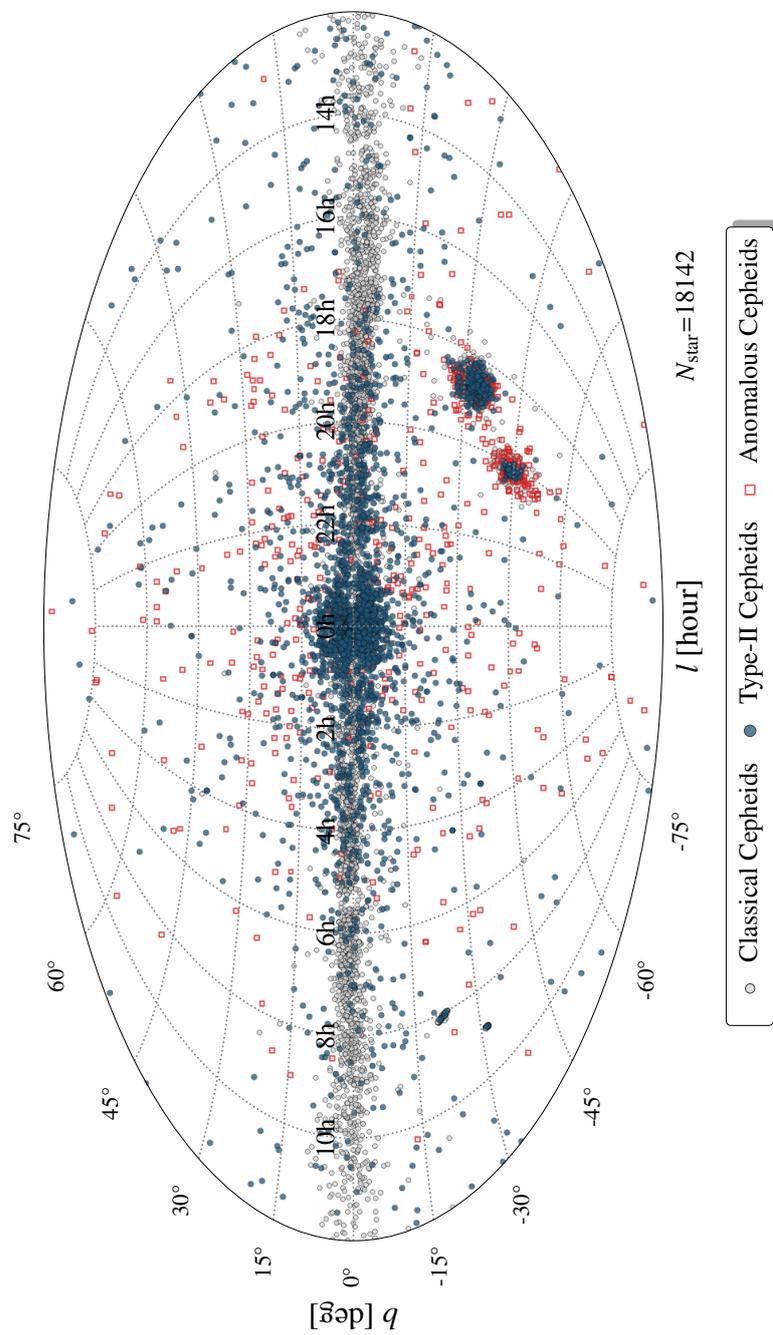**FIGURE III.8:** RR Lyrae stars on-sky map shown in Aitoff projection at HEALpixel level 6.

FIGURE III.9: Cepheids on-sky map shown in Aitoff projection.

III.2.1   Galactic Classical Cepheids List

Along with their own detections, the OGLE collaboration has verified and collated Galactic classical Cepheids from multiple other surveys (Pietrukowicz et al., 2021). They reclassified these stars after inspecting the light curves. This catalog is an excellent sample to assess the Cepheid classification accuracy of the surveys included here. This sample mainly helps in estimating survey-specific classification accuracy for Cepheids in the regions where OGLE itself has no data. Their collection (called AGC henceforth[2]) includes 3666 Cepheids, out of which 1651 are primary OGLE-IV detections and a further 379 that have been first detected by other surveys but contain OGLE-IV data. In total, this list contains 1636 stars that have not been observed by OGLE. While Ripepi et al. (2023) did find possible contamination by binaries (c.f. their Fig. 18) in the DCEP_1O sample of AGC, the classification accuracy for AGC is expected to be quite high ($\geq$97%).

III.2.1.1   Cross-match

We noticed that the AGC star with DR3 `source_id=2163847439043169408` was included twice, once each as DCEP_F and DCEP_1O. In its source catalog (i.e., Ripepi et al. 2023), it was identified as DCEP_1O, hence we considered that as its final classification and removed its other entry from AGC. We cross-matched 3665 AGC stars against our catalog and recovered all but 15 Cepheids using a 10″ radius[3]. 3642 out of the 3650 matched stars have the same *Gaia* DR3 counterparts as found by Pietrukowicz et al. and our algorithm. Six more have a *Gaia* match in both AGC and our catalog but the matched `source_ids` are different. For instance, for the star V890 Ara, AGC has `source_id=5948221021378490496` as its counterpart, while we have `source_id=5948221017072011776`. Though the star matched by us is 4.3″ away, it has the same period (in MACC and ASAS-SN) and magnitude (in *Gaia*) as V890 Ara. On the other hand, the AGC-matched star is 7 magnitudes fainter. Based on similar comparisons for the remaining five stars with conflicting `source_ids`, we retained the star matched by us as the *Gaia* DR3 counterpart (although the differences were less stark). Finally, three faint stars in the bulge (with I > 19.5) do not have a *Gaia* counterpart in either AGC or our catalog.

Out of the 3650 matched stars, there are seven stars that are classified as Type-II Cepheids by OGLE but identified as classical Cepheids by AGC. Six out of these seven stars do not have any properties flagged by OGLE, while one (a W Vir-type star) has uncertain clas-

---

[2]as short for AllGalCep — which is the name of this list as stored on the OGLE database.

[3]Pietrukowicz et al. (2021) used 0.5″ as the common matching radius to construct the AGC catalog, and out of 3650 matches with our catalog, only eight had an angular separation > 0.5″. We confirmed these matches were indeed true by comparing their variability periods and mean magnitudes. All eight of these originated either from MACC or ASAS-SN, whose FWHMs are $\geq$ 15″.

sification. However, ASAS-SN and *Gaia* both have classified this star as W Vir, too. We retained the original OGLE classification for all seven stars and removed them from AGC (P. Pietrukowicz, in priv. comm. confirmed that these stars are non-DCEP and will be removed from AGC).

III.2.1.2 Stars present only in AGC

We investigated the plausible reasons for missing 15 stars that were unique to AGC. One of these 15, the star with OGLE-ID=OGLE-GD-CEP-1834 ($I \sim 13.3$ mag) has `source_id=` 5546611475768394112 as its *Gaia* DR3 counterpart in AGC. This *Gaia* star is 74″ away and 5 magnitudes fainter (`phot_g_mean_mag` ~ 18.85 mag). On the other hand, we have `source_id`=5546611514435107328 as its *Gaia* DR3 match, which is less than 0.1″ away and of comparable brightness (`phot_g_mean_mag` ~ 13.7 mag, `phot_rp_mean_mag` ~ 13.4 mag). Therefore, we assumed that this star's coordinates in AGC are probably inaccurate and, hence, it was missed in our cross-match. Subsequently, we were left with 14 unmatched AGC stars.

Four of the remaining 14 stars are from MACC (Richards et al., 2011) and we had rejected them while collating the survey catalogs as their classification probability in MACC was < 0.5. This also reaffirmed that the cuts we had applied in Sect. II.1.6 did not lead to the omission of many true positives. Three of the remaining 11 stars were identified only in ASAS-SN as variables of subtypes other than RR Lyrae and Cepheids. Out of the seven stars left, six are brighter than G=6.5mag, which meant they were probably too saturated in the surveys we have included. The last star present in AGC but missing from our catalog is a multi-mode disk Cepheid discovered by Turner et al. (2009) as a Polaris analogue. Consequently, we included all 14 of these into our catalog.

III.2.1.3 Period Validation

We computed PERIOD_FLAG (as defined in Eqn. III.1) for all AGC matches with our catalog and found excellent agreement. Only four stars (out of ~ 3600 single mode pulsators) have PERIOD_FLAG = True and the Median($\Delta P_{\mathrm{ab}}^{\%}$) for all stars is $\leq 0.001\%$ (or ≤ 5 s for a 4-day variable).

III.2.1.4 Classification

Table III.7 shows the results of the comparison of the classification in our catalog against that in AGC and we describe its details below.

Gaia: 3047 *Gaia* SOS stars are included in AGC, out of which 167 stars' classification has been altered (i.e., $\sim 5.4\%$ of the sample). 44 of these were classified as non-DCEP by

TABLE III.7: Comparison of the classification of each catalog with AGC.

Catalog	Common N_{star}	N_{star} with same subtype	N_{star} with different DCEP subtype	# of non-DCEP
RRLCep	3643	3600 ($\sim 98.9\%$)	24 (0.6%)	19 (0.5%)
ASAS-SN	960	838 ($\sim 87.2\%$)	74	48
CSS	5	0	0	5
Gaia	3047	2880 ($\sim 94.5\%$)	123	44
OGLE	2030	2027 ($\sim 99.9\%$)	2	1
ZTF	1195	1164 ($\sim 97.4\%$)	0	31

FIGURE III.10: Confusion matrix of our classification against the Galactic classical Cepheids catalog of Pietrukowicz et al. (2021, AGC). Here the percentages shown in every cell are recall values per subtype. 3640 out of 3643 stars common between the catalogs share the same subtype classification. This comparison also illustrates how F_β is a better metric than Acc_{bal} for assessing imbalanced multiclass classification in case of uneven number of classes.

SOS. The rest were classified as DCEP but with differing subtypes. A further 191 stars classified as classical Cepheids only by Ripepi et al. (2023) and located close to the Galactic disk are not included in AGC. Pietrukowicz et al. (in priv. comm.) confirmed that they visually inspected the light curves of these stars and their unambiguous classification requires greater N_{obs} in *Gaia*. These stars could be anomalous Cepheids or spotted stars which are easily confused with DCEP_1O due to their sinusoidal light curves. Ultimately, this means that the possible contamination/subtype misclassification in the *Gaia* Galactic DCEP sample is $\sim 11\%$, i.e., (167 + 191) / 3238.

ASAS-SN: There are 838 common stars between AGC and the stars we selected from ASAS-SN. $\sim 87.2\%$ of the stars are identified as the same subtype in both, while 122 stars have a different classification. Moreover, 157 of these 838 stars are classified only in ASAS-SN and not in any other survey. AGC has reclassified 29 (18%) of them.

Besides these 838 common stars, there are 14 ASAS-SN-only stars included in AGC, for which we did not include any ASAS-SN data. We did include these stars as Cepheids from other surveys but did not include the ASAS-SN properties of these stars. The reasons behind this are: (a) 5 out of 14 were classified as Cepheids (3 stars) or RR Lyrae stars (2 stars) in ASAS-SN but their classification probability is <0.45. As discussed in Sect. II.1.4, we only selected variables with classification probability > 0.5 and, given that this omitted just 3 out of 838 true positives, this only reaffirmed our decision. (b) The remaining 11 are not identified as Cepheids in ASAS-SN.

Furthermore, 79 ASAS-SN-only stars are not present in AGC. Pietrukowicz et al. (in priv. comm.) confirmed these are non-DCEP stars, as explained above for the *Gaia* sample. Therefore, the Galactic Cepheid classification accuracy in ASAS-SN amounts to $\sim 76\%$.

Others: ZTF classical Cepheid subtypes were adopted from AGC as mentioned in Sect. II.1.3. Hence, we only investigated whether any stars classified as RR Lyrae or Type-II Cepheid by ZTF have been reclassified in AGC as DCEP and there are none. CSS did not observe the Galactic disk but has 5 non-DCEP in common with AGC at high Galactic latitudes. We adopted AGC classes for all of these. Pietrukowicz et al. also reclassified three OGLE stars (2 DCEP_1O, 1 RRab) as multi-mode DCEP by analyzing their ZTF photometry. We also removed ten OGLE-only Cepheids that Pietrukowicz et al. found to be artifacts as their OGLE photometry was contaminated by nearby brighter stars. Finally, there are no PS1- or DES-discovered RR Lyrae stars identified as DCEP in AGC.

RRLCep: While the aforementioned component catalogs have their respective classification accuracies, overall, our validated catalog (RRLCep) shares $\sim 98.9\%$ agreement with AGC. Out of 43 stars with disagreeing classification between our catalog and AGC, 24 are classified as classical Cepheid of different subtypes and 19 are non-DCEP stars. This results in overall classification accuracy for our catalog $F_{0.5} = 99.2\%$ (see Fig. III.10). We adopted

AGC classification for 43 stars, consequently resulting in a highly pure collection of Galactic classical Cepheids. As discussed in Sect. III.2.1.2, RRLCep initially contained all but 14 stars from AGC, resulting in a completeness of 99.7%. These missing stars were added to our catalog to make it as complete as possible.

We also cross-matched our catalog against the literature misclassified Classical Cepheids list maintained by OGLE[4]. This list consists of 46 stars and we found 31 stars in common with our catalog. Only four stars have a different class in our catalog — two classical Cepheids from *Gaia* are identified as rotational variables and two ASAS-SN DCEP_F are classified as ACEP_F. We changed the class of the latter two and removed the former two stars from our catalog. The remaining 15 stars from this list, for which we did not find a match in our catalog, are neither Cepheids nor RR Lyrae stars.

III.2.2   Stripe 82 RR Lyrae stars

To validate the classification of the RR Lyrae sample, we cross-matched our catalog against the highly pure RR Lyrae catalog in the SDSS Stripe 82 region (Sesar et al., 2010). This catalog, called S82 henceforth, contains 483 RR Lyrae stars with at least 15 observations in the SDSS g-band and selected using color cuts (see Eqns. 1–4 in Sesar et al. 2010). They visually inspected light curves of all 483 stars and confirmed their classification. They also estimated the catalog to be 100% complete within ~ 120 kpc.

We found 477 RR Lyrae stars in common with S82 (within $2''$) and no contamination from any Cepheids. The confusion matrix between S82 and our classification is shown in Fig. III.11. 473 stars share the same subtype classification. One star identified as RRc by S82 is identified as RRab by five surveys in our catalog. This star has the lowest number of observations in S82 and, therefore, was likely misclassified by them. Similarly, three stars identified as RRab by S82 are classified as RRc by at least three surveys in our collection. We retained the original classification of these stars.

We computed PERIOD_FLAG (as defined in Sect. III.1.1) for all 477 stars and found excellent agreement (see Fig. III.12). $\Delta P_{\mathrm{ab}}^{\%}$ for 95% of stars is $\leq 0.001\%$ (or $\leq 0.5\,\mathrm{s}$ for a 0.5-day variable). However, there are 21 stars that have PERIOD_FLAG = True. The number of observations in S82 for these stars lies in the first quartile of the N_{obs}distribution. After inspecting their light curves, we decided that the final period in our catalog (estimated by multiple surveys) was accurate for 19 stars. The remaining two stars (circled in Fig. III.12) were detected only in PS1 in our catalog, and S82 had the greater number of observations for these stars as well as a better-fitting period. Except for the latter two, we retained the

---

[4]http://www.astrouw.edu.pl/ogle/ogle4/OCVS/Cepheid_Misclassifications/Other/allGalCep.txt; their coordinates were retrieved from Simbad and Vizier.
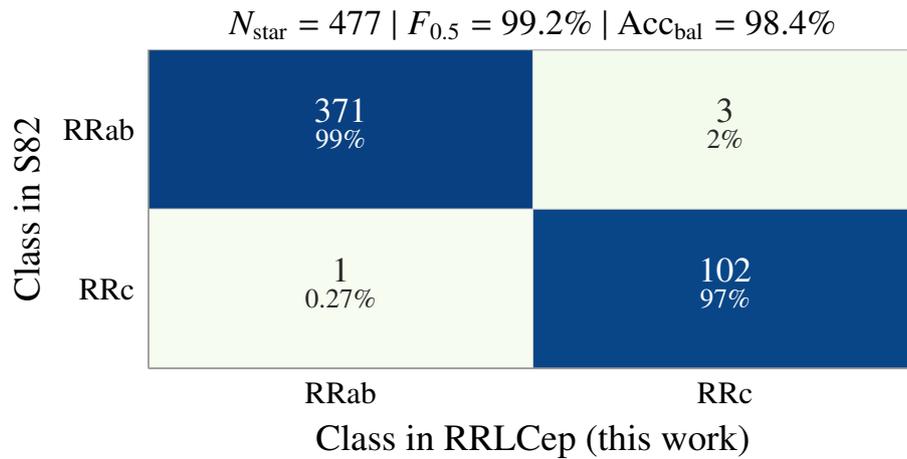
Figure III.11: Confusion matrix of our classification against the RR Lyrae catalog of Sesar et al. (2010, S82).
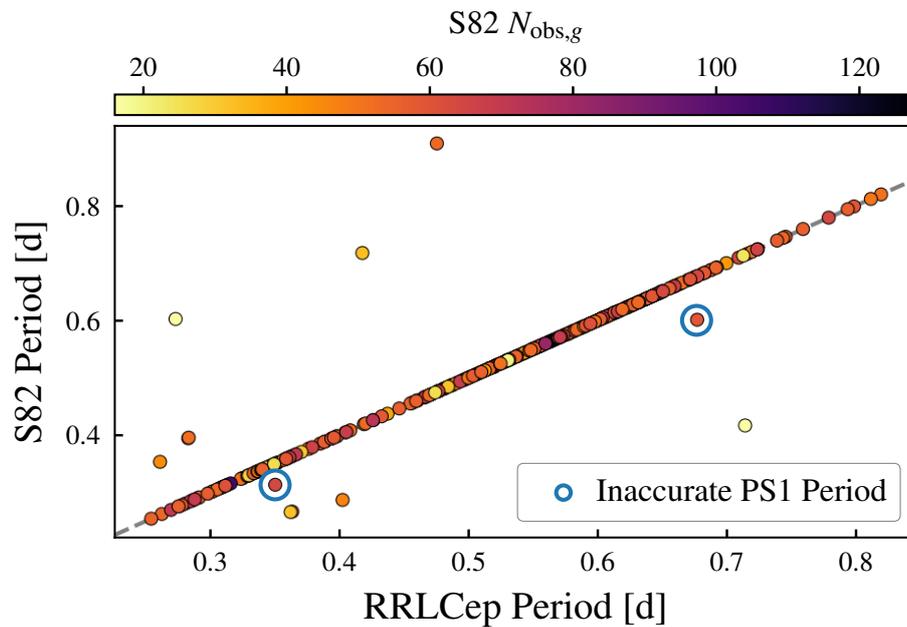

Figure III.12: Comparison of periods from our catalog (RRLCep, on the x-axis) against those from the RR Lyrae catalog of Sesar et al. (2010, S82; on the y-axis). Two stars highlighted in blue had inaccurate period in our catalog and the source survey for both of these stars was PS1. The rest of the stars off the identity line (in gray) have inaccurate periods in S82. A possible reason for this is that these stars have fewer g-band observations in S82 ($N_{\mathrm{obs},g}$).

original periods for all 475 stars.

Six stars were present in S82 but are missing from our catalog. Four of these are located beyond 105 kpc, one of them at ~ 91 kpc, and one at ~ 16 kpc. Since all surveys have different observation footprints and detection limits, the completeness of our catalog is highly inhomogeneous (generally worse in high-extinction regions). Nevertheless, when compared with S82, our RR Lyrae catalog is essentially complete within ~ 100 kpc and 99.9% complete at ~ 130 kpc.

However, our catalog consists of 137 RRab and RRc stars in Stripe 82 that are missing from the Sesar et al. (2010) catalog. 50 of these have been classified as RR Lyrae stars by more than one survey and are unlikely to be false detections. We examined their light curves and confirmed their original classification. At the time of the release of S82 catalog, these stars probably did not have enough $N_{\rm obs}$ in SDSS data to qualify for the assessment of Sesar et al. (2010).

III.2.3 Space-based observations

While there are no large literature catalogs of Type-II and anomalous Cepheids (beyond the ones we have collected), we combined the below listed catalogs to validate their classification. They employed rich light curves from either *TESS* or *Kepler*/K2 and undertook a high-order Fourier analysis.

Plachy et al. (2021) analyzed 26 literature variable stars using well-populated *TESS* light curves and provided new classifications and period estimates for them. They concluded that the star RV Men, classified as RRab in MACC, is actually a rotational variable. We removed this star from our catalog. Except for this star, there were four stars whose classification in our catalog was uncertain as all their parent surveys had classified them differently (either RRab, BL Her, or DCEP_F). All four stars were classified as anomalous Cepheids by Plachy et al.. Overall, the highly precise periods calculated by Plachy et al. agreed well with the final periods in our catalog, with $|P_{TESSS} - P_{\rm RRLCep}| \leq 10^{-4}$ d (~ 8 s) for all stars.

Jurkovic (2018) used literature data to analyze 59 stars previously classified as BL Her stars. By undertaking Fourier analysis of their light curves, they published new classifications and periods for them. We recovered 58 stars in their catalog and have the same classification for 55 of them. The three stars with new classification are one anomalous Cepheid and two pW Vir stars. Periods for all 58 stars agreed within 10^{-3} d. Finally, we added the missing BL Her star V617 Ara from Jurkovic (2018) to our catalog.

Similarly, Jurkovic et al. (2023) used precise light curves from *Kepler* and K2 to perform a higher-order Fourier analysis of 12 Type-II and anomalous Cepheids. We found all 12 stars in our catalog and the classification for 11 of these agreed with that of Jurkovic et al. (2023).

They reclassified one ASAS-SN DCEP_F as ACEP_F and we adopted their classification for this star. The periods for all stars agreed within 10^{-3} d.

III.2.4 Blazhko modulations and multiple periodicity

Several literature studies using OGLE-IV/*TESS* observations have investigated the presence of long-period trends and period/amplitude modulations in Cepheids and RR Lyrae stars. We flagged stars that exhibited Blazhko modulations or multiple periodicity in these studies, according to the scheme laid down in Table III.9.

Netzel et al. (2018) discovered the Blazhko effect and/or additional frequencies in ∼ 700 OGLE RRc stars in the Galactic bulge. 53 of these stars even exhibited multiperiodic Blazhko modulations. We set EXT_CLASS_FLAG = 1 for these stars. Similarly, Netzel & Smolec (2019) analyzed an updated sample of OGLE bulge RR Lyrae stars and detected additional frequencies in a further 1034 RRc and 11 RRd stars. Furthermore, Rathour et al. (2021) performed detailed frequency analysis of OGLE-IV Milky Way classical Cepheid light curves. They discovered 35 candidates of multi-mode Cepheids in the single-mode sample (i.e., DCEP_F, DCEP_1O). Given the low amplitudes of the additional modes, they suspected some of them might have been triggered by the 1-day rotation period of the Earth. Hence, we did not change the classification of these stars but proceeded to set EXT_CLASS_FLAG = 2 for them.

Smolec et al. (2023) performed a high-order frequency analysis of OGLE Magellanic Clouds DCEP_1O. They discovered a dominant secondary radial mode in 71 stars and weak secondary non-radial modes in ∼ 1000 stars. We changed the classification of the former sample to DCEP_mul. In the latter sample, Soszyński et al. (2019b) had flagged ∼ 200 of them for secondary periodicity during their analysis as well. Moreover, Smolec et al. also detected periodic modulation, akin to the Blazhko effect, in 27 DCEP_1O stars. All these stars can be retrieved from our catalog by selecting stars with EXT_CLASS_FLAG = 1 or 2.

Benkő et al. (2023) used two years of *TESS* data to analyze light curves of 670 Galactic RRc stars. They detected the Blazhko effect in 68 stars, thereby setting a 10.4% incidence rate of this phenomenon in Galactic field RRc stars. They detected additional frequencies in ≳ 82% of their sample, preferentially in stars with longer periods. We removed 23 stars that were blended in their data and cross-matched the remaining 647 against our catalog. These consisted of 633 RRc, 3 RRab, and 3 RRd stars, and 8 stars they classified as non-RR Lyrae. We recovered 643 stars within 3″ in our catalog, out of which four are classified as eclipsing binaries by Benkő et al.. We removed these stars from our catalog. Two stars identified by them as RRd stars were classified as RRc in our catalog. We accepted their classification for these stars. The remaining 637 stars' subtype classification is in complete agreement.

Moreover, $\sim 96\%$ of the stars' periods agreed within 10^{-4} d. Stars exhibiting the Blazhko effect or non-radial pulsation modes in their data were flagged accordingly in our catalog (see Table III.9).

Skarka (2013) collected 407 Blazhko RR Lyrae stars in the Galactic field from the literature. We found 400 matches with periods agreeing within 10^{-4} d. All 400 are classified as the same class in our catalog. We flagged these stars appropriately.

III.2.5   Variable stars of similar light curve shapes

We collected literature catalogs of other variable subtypes whose light curve morphology is similar to that of RR Lyrae stars and Cepheids. These variable stars can be distributed into the following broad types: short-period pulsating stars, long-period variables (LPV), eclipsing binaries, and rotational variables. Their subtypes are listed in Table III.9. Here we describe the provenance of the external catalogs and their stars found in common with our catalog. While establishing the classification purity of these external catalogs is beyond the scope of this study, this step helps us flag possible contaminants. The flagging scheme is outlined in Table III.9.

III.2.5.1   *Gaia* variables from other SOS pipelines

We cross-matched against the *Gaia* long-period variables catalog validated by Lebzelter et al. (2023) and recovered 131 matches. Interestingly, 33 out of these 131 are RRab stars discovered by OGLE. The separation between their coordinates is $<1''$, so we flagged them for future follow up. Given the difference between the periods of these subtypes, these stars might be exhibiting Blazhko modulations. We also found 12 stars in our catalog which have a close counterpart in the *Gaia* rotation modulation stars catalog (Distefano et al., 2023). As expected, there are ~ 4000 RRc (and ~ 500 RRab) stars which are identified as eclipsing binaries by *Gaia* (Mowlavi et al., 2023). Furthermore, there are 22 DCEP_1O stars identified only by the *Gaia* SOS pipeline that are also classified as eclipsing binaries by *Gaia* (as opposed to 80 that were found by Eyer et al. 2023, c.f. their Fig. 6). Finally, there are 19 matches between our catalog and the *Gaia* close companion (Gomel et al., 2023) and main-sequence oscillator catalogs (Gaia Collaboration et al., 2023b). All flags from these comparisons are encoded into EXT_CLASS_FLAG as listed in Table III.9.

III.2.5.2   OGLE Collection of Variable Stars

Except for the RR Lyrae stars and Cepheids included here, OGLE also monitors and classifies other variable stars. Specifically, we focused on the following subtypes: δ Scuti variables (DSCT), eclipsing binaries (ECL), and long period variables (LPV). We recovered

TABLE III.8: Provenance of variable stars collected from the OGLE Catalog of Variable stars. The classes are DSCT: δ Scuti, ECL: eclipsing binaries, and LPV: Long Period Variables.

References.
(1) Poleski et al. (2010); (2) Pietrukowicz et al. (2020); (3) Soszyński et al. (2021); (4) Pietrukowicz et al. (2022); (5) Soszyński et al. (2022); (6) Pietrukowicz et al. (2013); (7) Pawlak et al. (2016); (8) Soszyński et al. (2016a); (9) Soszyński et al. (2009); (10) Soszyński et al. (2011); (11) Soszyński et al. (2013); (12) Iwanek et al. (2022).

Class	Subtypes/Modes	N_{star}	Source
DSCT	Single mode	30 204	1–5
	Multi mode		
ECL	Contact	510 782	6–8
	Non-contact		
	Ellipsoidal		
LPV	Mira	177 360	9–12
	Semi-regular		
	OSARG		

their catalogs from the OGLE Collection of Variable Stars (OCVS) and assimilated $\sim 30\,000$ DSCT, $\sim 510\,000$ ECL, and $\sim 177\,000$ LPV stars. These samples were gathered from both OGLE-III and OGLE-IV (Udalski et al., 2008, 2015) and their provenance is given in Table III.8. The number of common stars found with these catalogs and how to retrieve them are illustrated in Table III.9.

III.2.5.3   Other variable star catalogs

The ASAS-SN catalogs (Jayasinghe et al., 2018; Christy et al., 2023), Catalina catalogs (Drake et al., 2014, 2017), and the ZTF catalog (Chen et al., 2020) consist of variable stars of other subtypes, too. We cross-matched against each of them and flagged the common stars following Table III.9. These stars were not removed altogether as the false discovery rates of variable stars of these subtypes have not been validated by us. Nonetheless, the flagged stars should be treated cautiously.

Green et al. (2023) used *TESS* data to classify and characterize $\sim 15\,000$ ellipsoidal binaries. Based on spectroscopic observations of a subset, they estimated the purity of their catalog to be $83 \pm 13\%$. Given their high completeness for systems with orbital periods less than 1 d, we cross-matched our catalog to find any RR Lyrae stars classified as ellipsoidal binaries by them. We recovered ~ 200 such RR Lyrae stars (190 of type RRab), out of which
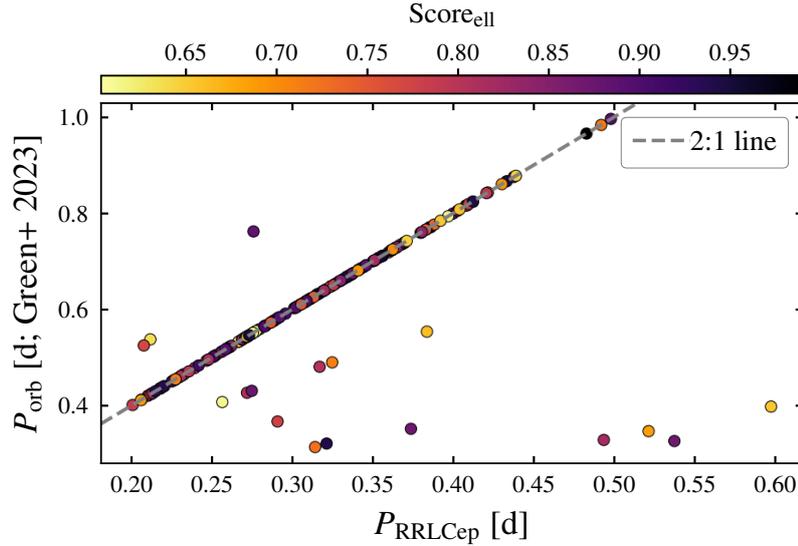
FIGURE III.13: Comparison of variability periods from our catalog (RRLCep, on the x-axis) against ellipsoidal binaries orbital periods from Green et al. (2023, on the y-axis). The points are colored according to the probability of being ellipsoidal binaries (Score$_{ell}$). 95% of stars lie along the 2:1 line (shown in gray). These stars were removed from our catalog and the rest were flagged as candidate ellipsoidal binaries.

~ 160 (18) are ASAS-SN-only (CSS-only) detections. $\sim 95\%$ of ASAS-SN-only sample's variability period is equal to half of the orbital period of the ellipsoidal binary, as shown in Fig. III.13. We removed all ASAS-SN and CSS-only detections that lie along the 2:1 line (168 stars) in Fig. III.13 and flagged the rest (31 stars). The latter sample, anyway, has relatively lower probability of being ellipsoidal binaries (Score$_{ell}$) in Green et al. (2023).

III.2.6 Non-stellar contaminants

Due to the shape of their light curves, RR Lyrae stars can sometimes be confused with Active Galactic Nuclei (AGN). To rid our catalog of possible non-stellar objects masquerading as variable stars, we retrieved and cross-matched against the following catalogs.

Gaia Collaboration et al. (2023a) described the extragalactic content released in *Gaia* DR3 as part of the tables `galaxy_candidates` and `qso_candidates`. We used the queries listed in Tables 11 and 12 of their article to generate pure samples of high-probability galaxy and QSO (Quasi-stellar objects) candidates. However, Storey-Fisher et al. (2023) found numerous non-QSO objects in the latter sample. By combining *Gaia* data with unWISE photometry (Schlafly et al., 2019, described in detail in Sect. IV.1.1.1), Storey-Fisher et al. published a cleaner sample of $\sim 750\,000$ quasars brighter than $G \sim 20$ mag. They named this catalog Quaia.

Gaia DR3 also contained the first *Gaia* catalog of variable AGNs which was validated by

Figure III.14: Confusion matrix of our catalog (RRLCep) against the non-stellar objects catalogs described in Sec. III.2.6. Var. AGN here stands for the variable AGNs from the GLEAN catalog (Carnerero et al., 2023).

the AGN Specific Object Study module (Carnerero et al., 2023). They cleaned this sample by using the AGNs that form the *Gaia* Celestial Reference Frame 3 (Gaia Collaboration et al., 2022, Gaia-CRF3) and the cleaned sample (called GLEAN) is expected to be $\lesssim$ 95% pure. Except for these four all-sky samples, we also collected the region-limited but highly pure and complete SDSS DR16 quasar catalog (Lyke et al., 2020). This collection was formed using data from the extended Baryon Oscillation Spectroscopic Survey of SDSS-IV and consists of ∼ 700 000 QSOs. It suffers only ∼ 1% contamination and is 99.8% complete.

We found a total of 304 non-stellar objects in our catalog (∼ 0.09%) and their confusion matrix is shown in Fig. III.14. As expected, 83% of the matched stars are RR Lyrae stars. We removed 246 RR Lyrae stars from our catalog (∼ 0.07%) that matched against at least one of Quaia, GLEAN, *Gaia*-CRF3, `galaxy_candidates`, or SDSS quasar catalog. Finally, we flagged 58 stars (comprised of 50 Cepheids, 7 RRab, and 1 RRc stars) that only had a counterpart in the table `qso_candidates`. Due to the relatively low purity of this sample, we did not remove these stars, but instead chose to flag them. Interestingly, this flagged sample includes 46 DCEP_F stars identified by *Gaia*. These stars are split between Andromeda (M31) and Triangulum (M33) galaxies (see panels *f* and *g* in Fig. III.15). After inspecting their *G*-band light curves, we decided to retain them as Cepheids in our catalog.
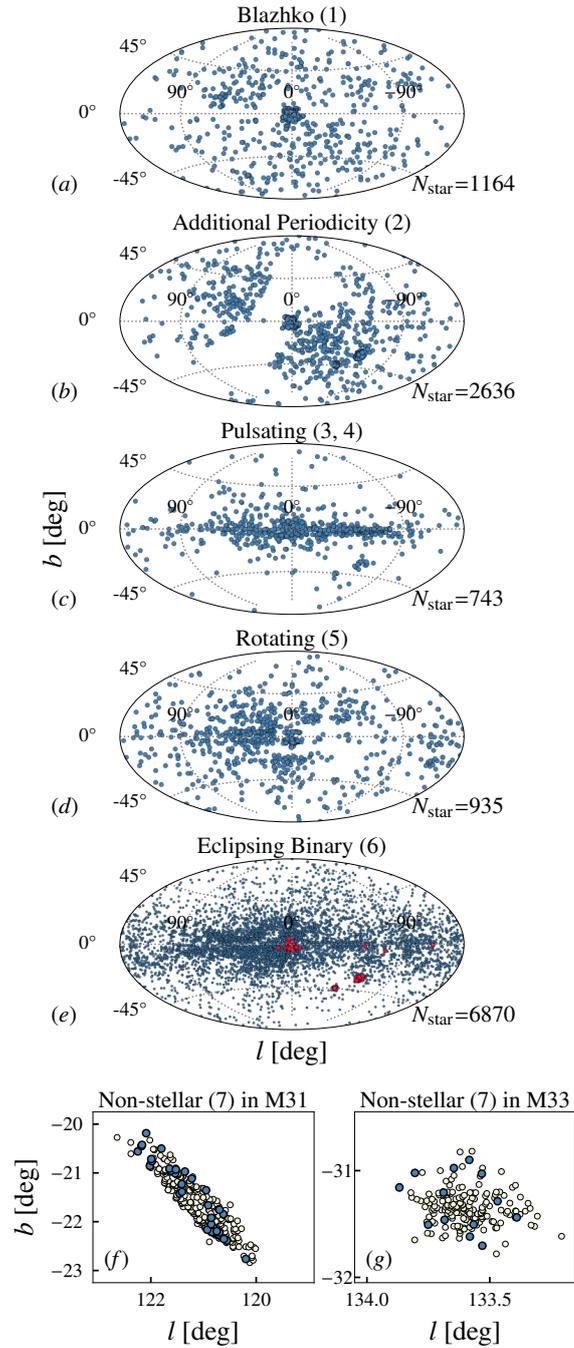
FIGURE III.15: Sky maps (in galactic coordinates) of stars flagged into various classes by external validation. Blue dots in every panel denote the flagged stars. The value of EXT_CLASS_FLAG is given in parenthesis above every plot. In panel *e*, the red dots are those Cepheids and RR Lyrae stars that are highly likely to be in eclipsing binaries and have EXT_CLASS_FLAG = 61 (i.e., from Hajdu et al. 2021) or OGLE_FLAG=2. In panels *f* and *g*, the white dots are non-flagged Cepheids in M31 and M33 from *Gaia* and blue dots denote those DCEP_F that are flagged as QSOs only by qso_candidates.

TABLE III.9: Provenance of objects used for external validation of our catalog.

Notes.

(*a*) The subtypes not already discussed are: HADS: High-amplitude δ Scuti; OSARG: OGLE small-amplitude red giants; Rot.: Rotational variable without subtype specification; Ell.: Ellipsoidal binary; EA: Algol-type binary; EB: β Lyrae-type binary; EW: W UMa-type binary; Var. AGN: variable AGN. (*b*) Number of objects collected from external catalogs. (*c*) Number of objects that matched with stars in our catalog.

Source.

(1) Netzel et al. (2018); (2) Benkő et al. (2023); (3) Skarka (2013); (4) Smolec et al. (2023); (5) Rathour et al. (2021); (6) Netzel & Smolec (2022); (7) Chen et al. (2020); (8) Jayasinghe et al. (2018); (9) Jayasinghe et al. (2020); (10) Christy et al. (2023); (11) Drake et al. (2014); (12) Drake et al. (2017); (13) Poleski et al. (2010); (14) Pietrukowicz et al. (2020); (15) Soszyński et al. (2021); (16) Pietrukowicz et al. (2022); (17) Soszyński et al. (2022); (18) Gaia Collaboration et al. (2023b); (19) Soszyński et al. (2009); (20) Soszyński et al. (2011); (21) Soszyński et al. (2013); (22) Iwanek et al. (2022); (23) Lebzelter et al. (2023); (24) Distefano et al. (2023); (25) Gomel et al. (2023); (26) Green et al. (2023); (27) Pietrukowicz et al. (2013); (28) Pawlak et al. (2016); (29) Soszyński et al. (2016a); (30) Mowlavi et al. (2023); (31) Hajdu et al. (2021); (32) Gaia Collaboration et al. (2023a); (33) Storey-Fisher et al. (2023); (34) Lyke et al. (2020); (35) Carnerero et al. (2023).

Object type	Subtypes[a]	Source	N_{object}[b]	N_{matched}[c]	EXT_CLASS_FLAG
Blazhko stars	RRab, RRc, RRd	1–3	1164	1164	1
Additional periodicity	DCEP_1O, RRab, RRc, RRd	1,2, 4–6	2636	2636	2
Short-period pulsators	δ Scuti, HADS, γ Dor	7–18	115 252	291	3
Long-period pulsators	LPV, Mira, SR, OSARG	7–12, 19–23	2 423 216	452	4
Rotating variables	Rot., RS CVn, BY Dra, Ell.	7–12, 24–26	177 820	935	5
Eclipsing binaries	EA, EB, EW	7–12, 27–31	1 443 786	6850	6
Non-stellar objects	AGN, Var. AGN, Quasar, Galaxy	32–35	13 869 996	304	7
		Total	18 030 070	12 632	

III.2.7 Summarizing external validation

In summary, we collected ∼ 18 million objects from various literature surveys and flagged 12 632 stars in our catalog. The complete provenance of this external validation is listed in Table III.9. On-sky distributions of flagged stars are shown in Fig. III.15. A total of 1164 RR Lyrae stars (EXT_CLASS_FLAG = 1) are flagged as Blazhko candidates and ∼ 2600 stars (EXT_CLASS_FLAG = 2) are flagged as stars showing additional periodicity. We intend to study their impact on period-luminosity(-metallicity) relations and distances of these subtypes (see for e.g., Prudil & Skarka 2017).

291 RR Lyrae stars are flagged as short-period variables (mainly δ Scuti-type, EXT_CLASS_FLAG = 3). We reiterate that we did not validate the classification of short-period variables' external parent surveys. Nevertheless, assuming a perfectly clean sample, and given the RR Lyrae population in our catalog (∼ 311 000), we estimate that the maximum contamination from other short-period variables in our RR Lyrae sample may be ∼ 0.08%.

The long-period variables (EXT_CLASS_FLAG = 4) mainly overlapped with DCEP_F and RV Tau stars in our catalog. Similarly to the short-period variables, we anticipate that the maximum contamination from other long-period variables in our sample of DCEP_F and RV Tau stars may be ∼ 2%.

∼ 30 non-OGLE W Vir stars are flagged as rotating variables (ellipsoidal-type, EXT_CLASS_FLAG = 5). These stars could belong to pW Vir class, as we expect ∼ 10% unidentified pW Vir stars in the non-OGLE W Vir sample. ∼ 6500 RR Lyrae stars and ∼ 300 Cepheids are flagged as eclipsing binary candidates (EXT_CLASS_FLAG = 6). As RRc stars are easily confused as eclipsing binaries (due to their sinusoidal light curves), they form the bulk of this sample (∼ 5700). Using light-travel time effects and observed–calculated diagrams, Hajdu et al. (2021) found 87 RRab stars in binary systems in the Galactic bulge region (these stars have EXT_CLASS_FLAG = 61 in our catalog). A similar investigation of the remaining ∼ 1100 non-RRc stars can confirm whether these stars are misclassified as pulsating stars or components of binary systems. However, disentangling pulsations from eclipses of these stars is beyond the scope of this work. Nevertheless, they make for an excellent follow-up sample, as pulsating stars in eclipsing binaries are a hotbed for employing asteroseismology and orbital dynamics simultaneously (Murphy, 2018; Lampens, 2021). Given their presence in binary systems, dynamical masses and radii of these pulsating stars can be calculated (independent of stellar models) to a high level of precision and accuracy. These parameters, in turn, can be used to compute precise asteroseismology models for these stars, enabling determination of their ages and atmospheric parameters (such as effective temperature, surface gravity, metallicity) with great accuracy.

∼ 70% of the 246 RR Lyrae stars that we removed from our sample as likely non-stellar

objects are classified only in PS1 or DES. Therefore, non-stellar objects are preferentially a source of contamination in stars classified using sparse multi-band light curves. This finding is especially relevant for LSST, since in its initial data releases, the fainter and hitherto undetected objects will only have a few observations. Hence, it becomes imperative that before running a RR Lyrae classification pipeline on sparse multi-band light curves, the sample should be cleaned of AGN candidates.

IV

Auxiliary Data

*Laziness is nothing more than the habit of resting
before you get tired.*

– JULES RENARD

Computation of stellar orbits in the Milky Way requires six stellar parameters: coordinates
(α, δ), proper motions $(\mu_\alpha \cos(\delta), \mu_\delta)$, distance, and radial velocity (v_{los}). With *Gaia* DR3,
we get coordinates for $\gtrsim 99.1\%$ of stars in our catalog, all at the same epoch. It also pro-
vides proper motions for $\gtrsim 96\%$ of them. After applying the quality control cuts described
in Sect. II.2, we get high quality proper motions for $\sim 90\%$ stars. In this chapter, we describe
how ancillary data from other literature surveys complement our catalog and help us com-
plete the aforementioned 6-d phase space. Additionally, for every auxiliary cross-match, we
outline the quality control steps we took to ensure retention of high-quality data.

In Sect. IV.1, we outline how we accrued complementary optical and NIR photometry
from multi-band, wide-field surveys. In Sect. IV.2, we discuss our cross-matches with low-
and high-resolution spectroscopic surveys. We describe the reddening maps that we cross-
matched with to measure extinction to our stars in Sect. IV.4. Finally, we present a detailed
algorithm with novel modifications for determining LMC/SMC membership in Sect. IV.3.

IV.1   Photometry

IV.1.1   IR photometry

As discussed in Chapter I, RR Lyrae stars and Cepheids are excellent distance indicators, having been used as standard candles for over a century. However, no single survey in our complement has observed all the stars. Hence, the distance estimates for these stars computed using respective period-luminosity relations will be heterogeneous. Furthermore, for the fainter stars from every survey and the stars affected by reddening, photometric uncertainty can be $\geq$ 0.05 mag, thereby reducing the precision of distances. Therefore, getting photometry from other optical surveys with homogeneous coverage of large parts of the sky and substantial number of observations can be highly beneficial.

Similarly, utilizing IR photometry, the dependence on metallicity (Bono et al., 2010; Bhardwaj et al., 2023) and reddening uncertainties can be minimized (Freedman & Madore, 2010; Matsunaga et al., 2006, 2011). For instance, the classical Cepheid sample from an early version of this catalog was used to trace the Milky Way spiral arms (Lemasle et al., 2022). To minimize the effect of extinction in the Galactic disk, we had used WISE photometry to determine distances of that sample. This enabled us to model spiral arms in an hitherto unexplored region of the galaxy.

An important outcome of collecting IR photometry for these stars is the construction of their light curve templates. NIR templates of a few subtypes of these stars have been computed (Braga et al., 2019; Inno et al., 2015; Bhardwaj et al., 2017). However, given the overlap between WISE and JWST (James Webb Space Telescope) filters, WISE MIR templates can be used to phase JWST observations for RR Lyrae stars and Cepheids (Gavrilchenko et al., 2014). For stars with multiple observations in JWST, such templates can be used to identify these variable subtypes. Therefore, we augment our catalog with photometry from the following surveys.

IV.1.1.1   WISE

Schlafly et al. (2019) reduced five years of data from WISE/NEOWISE missions (Wright et al., 2010) using `crowdsource` and produced the unWISE W1, W2 band-merged co-added catalog. This catalog is significantly deeper and better-resolved than the AllWISE all-sky atlas (Cutri et al., 2021). We cross-matched our catalog with unWISE using a 6.5″ cross-match radius (corresponding to unWISE FWHM), and recovered 90% matches within 3″. We retained unWISE photometry only for those stars that satisfied the following criteria: $-1 \leq W1 - W2 \leq 1.3$ (loosely based on the width of the classical instability strip), the PSF quality factor $\geq$ 0.5 (which ensures at least half of the pixels covered by the PSF contributed

to this detection). We removed those stars whose `W1(2)_info_flag` indicated a problem with the PSF when being processed by `crowdsource` (corresponding to bits 0, 3-7, for details, c.f. Schlafly et al.). Additionally, we also discarded stars that were either saturated or contaminated by a bright neighbor. These stars were identified as the ones that had at least one of the unWISE flag bits set to true. Consequently, we amassed unWISE W1, W2 photometry for $\sim 175\,000$ stars. They all have at least ~ 65 observations in both W1 and W2 bands, with a median $N_{obs} \sim 130$.

We also cross-matched our catalog with the CatWISE2020 catalog of Marocco et al. (2021). A couple of differences between unWISE and CatWISE catalogs are explained here. While unWISE performs astrometry and photometry of WISE/NEOWISE images independently in W1 and W2 bands, CatWISE performs it jointly for both the bands. On one hand, CatWISE2020 is based on 6-year full-depth unWISE coadds and its photometry was performed using the AllWISE pipeline. On the other hand, the PSF photometry algorithm `crowdsource` was used by Schlafly et al. (2019) for unWISE photometry. Hence, we decided to cross-match with both and assess their accuracy for our target stars. We cross-matched with CatWISE2020 stored at Vizier, since source coordinates in this version have been corrected for the positional offsets discovered by Marocco et al. (2021). We recovered a CatWISE candidate for $\sim 255\,000$ stars within $5.5''$ (corresponding to two WISE pixels). Out of these, we selected $\sim 191\,000$ stars that met the following criteria: (a) their photometry was not contaminated either by artifacts or nearby sources and, (b) they had ≥ 4 photometric measurements of S/N>3 in both W1 and W2 coadds.

To estimate the difference between these two catalogs, we compared their photometry with the Fourier-fitted mean magnitudes computed by the WISE variable stars catalog (Chen et al. 2018)[1]. We recovered ~ 4300 stars in common with the Chen et al. (2018) catalog.

We compare unWISE and CatWISE photometry against that of Chen et al. (C18) in Fig. IV.1. We notice that unWISE magnitudes compare positively with C18 Fourier-fit photometry as compared to CatWISE photometry, especially for W1 (top panel) and the color W1 - W2 (bottom panel). For W2, the scatter in CatWISE W2 - C18 W2 (ΔW2, the residuals) is fairly low as well. Photometry for brighter stars is missing from unWISE but it is available and well-behaved in CatWISE. However, for both W1 and (W1-W2), the residuals for fainter stars in unWISE are considerably smaller as compared to those in CatWISE. For this reason, we had utilized unWISE photometry to compute distances to a sample of classical Cepheids in the Milky Way, and used their Galactocentric positions to trace the Galactic

---

[1]This catalog was not used as one of the input catalogs in our study as they did not provide subtype classification for our target stars. Moreover, the OGLE team suggested that there might be $\sim 39\%$ contamination in the classification of the Chen et al. Galactic Classical Cepheid sample. Refer here: http://www.astrouw.edu.pl/ogle/ogle4/OCVS/Cepheid_Misclassifications/Chen_et_al_Nature_Astronomy/NATURE_ASTR.dossier

spiral arms (Lemasle et al., 2022).

IV.1.1.2   2MASS

Skrutskie et al. (2006) released the Two Micron All Sky Survey (2MASS) Point Source Catalog which included all-sky NIR photometry for over 470 million stars. We procured 2MASS counterparts for our stars using the *Gaia* DR3x2MASS neighborhood table (`tmass_psc_xsc_best_neighbour`). We recovered $\sim 136\,000$ matches out of which $\sim 29\,000$ were flagged out based on quality control flag recommendations prescribed in the 2MASS user guide.[2] We briefly discuss them here. $JHKs$ photometry was retained only for those stars whose band-specific flags followed the following criteria: `Read flag = 2` (to keep only those stars for which PSF photometry was performed), `Photometric Quality flag = A, B, or C` (to ensure photometric S/N ≥ 5), `Blend flag = 1` (to select isolated sources which were fitted using a single profile), `Contamination flag = 0` (to avoid sources affected by nearby artifacts or stars), `Galaxy contamination flag = 0`, and `Ext Key = null` (to avoid contamination from sources from the 2MASS Extended Source Catalog). More than 75% of these rejected stars were located across the Galactic bulge and disk, indicating that crowded regions were affected the most (probably due to the 2.5–3″ FWHM of 2MASS images).

Consequently, we only selected $\sim 84\,000$ stars that were identified in ≥ 5 frames in every band, and in ≥ 2 of which they had a $\geq 3\sigma$ measurement. These steps ensured that the $JHKs$ magnitudes were of the highest quality and as close to the mean magnitudes of stars as possible, given the arbitrarily-timed observations across the variability phase.

IV.1.1.3   VISTA

The VISTA Hemisphere Survey (VHS, McMahon et al. 2013), the VISTA survey of the Magellanic Clouds (VMC, Cioni et al. 2011), the VISTA Variable in the Via Lactea survey (VVV, Minniti et al. 2010) collectively observe the entire Southern celestial hemisphere using the ESO VISTA telescope. These surveys provide much deeper coverage in the $JHKs$ bands as compared to 2MASS. We gathered both Fourier/template-fit mean magnitudes from the corresponding studies listed below as well as the unphased mean magnitudes that are included in the source table of these surveys.

Specifically, we assimilated template-fit mean magnitudes for LMC Cepheids and RR Lyrae stars from Ripepi et al. (2022) and Cusano et al. (2021), respectively. These studies used VMC observations of literature variable stars to measure $YJKs$-band magnitudes for $\sim 33\,000$ stars in our catalog. While Ripepi et al. (2022) provide Fourier-modeled

---

[2]See https://www.ipac.caltech.edu/2mass/releases/allsky/doc/sec1_6b.html#origphot.

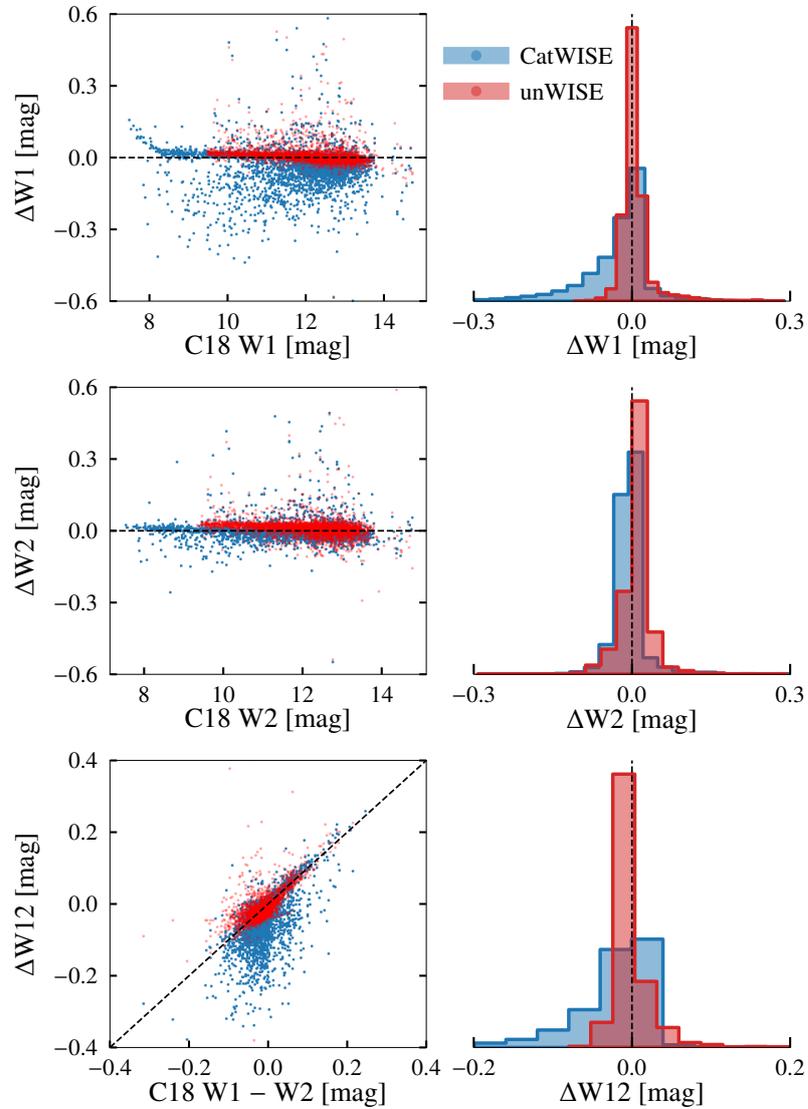**FIGURE IV.1:** Comparison of CatWISE (in blue) and unWISE (in red) photometry against Chen et al. (2018, C18) magnitudes. The *top left* panel displays the distribution of the residual ΔW1 = unWISE/catWISE W1 - C18 W1 as a function of C18 W1. Similarly, the *middle left* and *bottom left* panels show the same distributions for W2 and color W12 (= W1 - W2), respectively. Normalized histograms of these residuals are shown for both catWISE (in blue) and unWISE (in blue) in the *right* panels.

variability amplitudes, Cusano et al. (2021) do not. We measured these quantities from the light curves used by them. On average, both these samples contain light curves with $N_{\rm obs} \sim (5, 5, 15)$ in (Y, J, Ks) bands, which are enough to determine the mean magnitudes with an uncertainty < 10 mmag (discussed in detail in Sect. V.1.1).

Similarly, for the stars in the VVV footprint (the Galactic bulge and disk), we accrued their Fourier-fit $ZYJHKs$-band magnitudes and amplitudes from the VIVACE catalog (VIrac VAriable Classification Ensemble, Molnar et al. 2022). We found $\sim 38\,000$ stars in common with VIVACE, with $\geq 98\%$ classified as RR Lyrae or Cepheids in VIVACE as well. However, this sample includes only $\sim 50\%$ of the RR Lyrae stars in VIVACE and only $\sim 10\%$ of the Cepheids. Due to inherent challenges in classifying these stars using NIR light curves, we intend to include the VIVACE-only stars into our catalog after visual inspection.

Except for these $\sim 70,000$ stars from the aforementioned catalogs, we collected VISTA photometry from merged catalogs of VVV DR5 and VMC DR6[3] for a further $\sim 30\,000$ stars, totaling to $\sim 110\,000$ stars with VISTA photometry.

Furthermore, we gathered $YJHKs$ photometry from VHS DR5 for $\sim 88\,000$ stars in the Southern (celestial) halo. While VMC and VVV tie their photometry to the Vega system, the VHS photometry is released as AB magnitudes. We transformed the latter to the Vega system using the conversions listed in Appendix D of González-Fernández et al. (2018).

We retained NIR photometry only for those stars that did not contain any bad pixels in their default aperture or exhibit any other grave problems (like saturation) during their photometry. This was ensured by selecting only those stars which had post-processing error flag `XppErrBits` ≤ 16, where X is the filter being used. Furthermore, we selected only those sources that were primary photometric objects and had a high probability of being a star. With this, we assembled VISTA photometry for $\sim 134\,000$ stars.

As depicted in Fig. IV.2, 2MASS and the VISTA surveys complement each other by contributing NIR photometry for brighter and fainter stars, respectively. Stars brighter than ~ 12 mag in $JHKs$ are saturated in the VISTA images but measured accurately by 2MASS. However, the detection limit of 2MASS is ~ 16.5 mag, while the VISTA surveys go ~ 3 mag deeper.

IV.1.2   Optical Photometry

The Rubin Observatory Legacy Survey of Space and Time (LSST) will observe the sky in SDSS *ugrizy* filters (Ivezić et al., 2019). Therefore, Cepheid and RR Lyrae period-luminosity(-Wesenheit) relations in these filters can be applied to LSST data. Our input surveys provide photometry in the filters listed in Table II.2. Only three surveys compute

---

[3]https://www.eso.org/rm/api/v1/public/releaseDescriptions/194
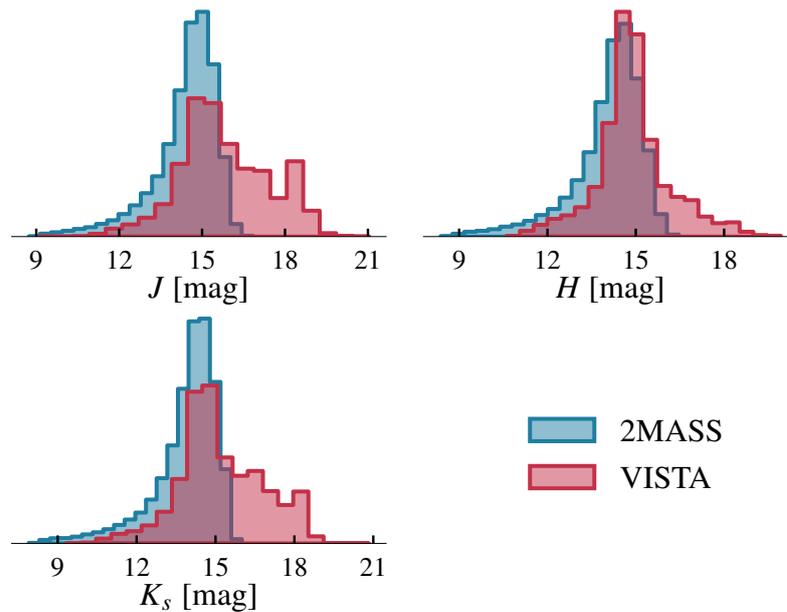
FIGURE IV.2: Normalized histograms of 2MASS (blue) and VISTA photometry (red).

Fourier-fit light curve magnitude in *griz* bands, namely PS1, DES, and ZTF. However, the PS1 catalog (Sesar et al., 2017) included here only contains RRab and RRc stars, and the same applies to the DES sample (Stringer et al., 2021). While the ZTF catalog (Chen et al., 2020) does contain Cepheids, it is limited to the *g*- and *r*-bands only. Therefore, to accumulate photometry in *ugrizy* bands for as many stars as possible, we cross-matched against the following catalogs. These enable the computation of PL/PW relations and template creation in a filter complement similar, and therefore, easily applicable to the LSST.

IV.1.2.1   DECam Surveys

Beyond the Dark Energy Survey, the Dark Energy Camera (DECam, Flaugher et al. 2015) is used to conduct multiple other surveys. They all observe complementary parts of the sky in all or a subset of *ugrizy* bands. We include data from DELVE, DECaPS, SMASH, and the NOIRLab source catalog for our stars. The descriptions of these DECam surveys and the steps taken to clean their data are outlined below.

The DECam Local Volume Exploration Survey (DELVE, Drlica-Wagner et al., 2021) reprocesses archival DECam images and reduces new observations in regions with $|b| > 10°$. The DECam Plane Survey (DECaPS, Saydjari et al. 2023) is complementary to DELVE DR2 (Drlica-Wagner et al., 2022) as it contains observations of the Galactic plane in *grizy* bands. However, since DELVE DR2 is a catalog level coadd, its detection limit towards the Magellanic Clouds is shallow as compared to SMASH (Survey of the MAgellanic Stellar

History, Nidever et al. 2017)[4]. Hence, we also cross-matched against SMASH and preferred their observations for stars common between SMASH and DELVE DR2.

The NOIRLab Source Catalog (NSC, Nidever et al. 2021), similar to DELVE, re-processed most of the public images available in the NOIRLab astronomy data archive[5]. DELVE, DECaPS, and SMASH combined, cover almost the same region in the sky as NSC DR2. However, there are a few differences between them. While DELVE also contains new observations, the NSC does not. The NSC DR2 contains photometry in $ugrizy$ and DECam VR bands, as compared to $griz$ bands for DELVE. Nidever et al. (2021) processed the NSC DR2 data using a custom pipeline built by Nidever et al. (2018), while DELVE DR2 (Drlica-Wagner et al., 2022) used the pipeline built by the DES Data Management team (Morganson et al., 2018). In light of these differences, we cross-matched against all the aforementioned surveys to assess their performance against Fourier-fit magnitudes.

For all matched stars from DELVE, DECaPS, and SMASH, we applied the following filters. We retained a star's photometry in a band X (from $ugrizy$) if:

- PSF FWHM $\leq 10''$;

- ≥ 3 detections with S/N ≥ 5;

- $\mathtt{fracflux_X} > 0.75$ (to ensure $\geq 75\%$ of the PSF-weighted flux originates from the source and not its neighbors);

- $\mathtt{quality\_factor} \geq 0.99$ (to ensure $\geq 99\%$ of the pixels with a non-zero weight over a $5'' \times 5''$ region are not affected by any artifacts;

- $\mathtt{parent\_flag = false}$ (to remove unresolved stars);

- $\mathtt{extended\_class = 0}$ (to remove contamination from extended objects).

In addition to the cuts mentioned above, for NSC DR2, we avoided highly-reddened regions by retaining only those stars with Schlegel et al. (1998) E(B-V) < 1 mag (as recommended by Nidever et al. 2021). Consequently, we recovered $\sim 191\,000$ stars from DELVE+DECaPS+SMASH and $\sim 131\,000$ stars from NSC DR2. Approximately, $100\,000$ stars are common to both catalogs. The unique stars in the former are mostly located in regions of newer observations, which were unavailable to Nidever et al. (2021).

---

[4]However, DELVE-MC, a sub-survey of DELVE not included in DR2, will increase this depth beyond that of SMASH (A. Drlica-Wagner, in priv. comm.).

[5]https://datalab.noirlab.edu/index.php

IV.1.2.2 PS1 and ATLAS

To complement the DECam surveys in the Northern sky, we cross-matched our catalog against Pan-STARRS1 3π and Medium-Deep surveys (PS1 DR2). This cross-match also allows us to garner PS1 photometry for stars absent from Sesar et al. (2017) RR Lyrae catalog. We retained PS1 magnitudes for stars with `quality_flag` > 0.99 and ≥ 5 observations in one of $grizy$ bands. Ultimately, PS1 DR2, in conjunction with DECam surveys and NSC DR2, provide $griz$ magnitudes for the entire sky. However, to account for their limited $N_{\rm obs}$, we cross-matched against the ATLAS variable stars catalog.

The Asteroid Terrestrial-impact Last Alert System (Tonry et al., 2018, ATLAS) uses four telescopes of 0.5 m diameter each to detect asteroids up to a limiting magnitude of ~ 19.5 mag. Heinze et al. (2018) published the first catalog of variable stars based on high-$N_{\rm obs}$ATLAS data. They classified stars into broad variability types after Fourier-fitting their ATLAS light curves. We cross-matched against this catalog to gather their Fourier-fit o and c magnitudes, which are based on PS1 filters. We selected stars with at least 40 observations in both o- and c-bands and at least 30 of which were identified as 'good' by their Fourier-fitting routine. Furthermore, we confirmed that stars' photometry was minimally affected by their neighbors, by selecting only those stars with `proxSTAT = 1`. Following Eqn. 2 of Tonry et al. (2018), we converted the ATLAS o, c magnitudes to PS1 gri magnitudes. We will assess these magnitudes and those from DECam surveys against gri from Sesar et al. (2017) in a future work.

IV.2 Spectroscopy

IV.2.1 Radial Velocity

The last dimension required to complete 6d phase space data for our stars is the line-of-sight velocity ($v_{\rm los}$, also commonly referred to as radial velocity). For pulsating stars, the measured $v_{\rm los}$ contains two components: one arising due to the systemic motion of the star with respect to the solar system barycenter (v_γ), and the other due to expansion and contraction of the stellar atmosphere, which is driven by the pulsations. By measuring a star's $v_{\rm los}$ multiple times over its phase, a radial velocity curve (RVC) can be computed and the mean of this RVC estimates the star's barycentric velocity, v_γ. To compute v_γ using just a few observations, RVC templates are employed to phase these measurements. Such templates are readily available for RR Lyrae stars (Sesar, 2012; Braga et al., 2021) and classical Cepheids (Kovacs et al., 1990; Marconi et al., 2017; Gallenne et al., 2019), but very rarely for Type-II or Anomalous Cepheids. We have not applied any template correction to the velocities we amassed from the surveys discussed below. We foresee incorporating these templates into
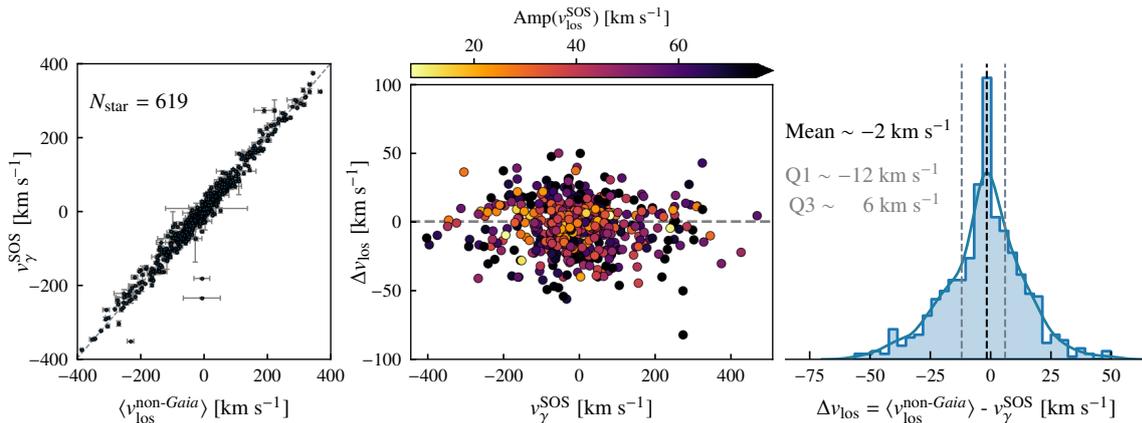
FIGURE IV.3: *Left*: Comparison of mean v_{los} from surveys other than *Gaia*, $\langle v_{\mathrm{los}}^{\mathrm{non\text{-}Gaia}} \rangle$, and the average v_{los} from *Gaia* SOS Cep&RRL pipeline ($v_{\gamma}^{\mathrm{SOS}}$). *Middle*: The difference Δv_{los} between the velocities is plotted against the latter. The points are color-coded according to the amplitude of radial velocity curve, Amp(v_{los}). *Right*: Normalized distribution and Gaussian KDE fit of Δv_{los}. The mean, first quartile (Q1), and third quartile (Q3) of the distribution are marked with dashed lines.

an advanced version of the catalog later. The epoch velocities we accumulate here can also aid in the construction of the first RVC templates for several subtypes.

IV.2.1.1 Data Collection

We gathered v_{los} estimates from eight spectroscopic surveys. Here, we describe the quality control steps we undertook to ascertain that the gathered v_{los} values are of high quality.

The Radial Velocity Spectrometer (RVS) onboard *Gaia* measures v_{los} for stars brighter than $G_{\mathrm{RVS}} \sim 14$ mag, (Katz et al., 2023). We discarded RVS v_{los} estimates for stars with $(G_{\mathrm{RVS}} - G) < -3$ as their spectra are likely to suffer from contamination by a nearby source or their images have underestimated background fluxes (Babusiaux et al., 2023). Similarly, following Eqn. 1 of Babusiaux et al. (2023), we rescaled v_{los} uncertainties for both the bright ($8 < G_{\mathrm{RVS}} < 12$) and faint stars ($G_{\mathrm{RVS}} > 12$). In total, RVS v_{los} estimates are available for $\sim 10\,600$ stars in our catalog.

Gaia DR3 also contains epoch v_{los} measurements for ~ 1800 stars from this sample which have ≥ 7 epoch v_{los} measurements. Ripepi et al. (2023); Clementini et al. (2023) phase-folded and Fourier-fitted RVC of these stars and computed the average v_{los}. This average is indicative of the barycentric velocity, any systematics notwithstanding. We used these SOS v_{los} values over those from DR3 `source_table` for the common stars.

LAMOST low- (Cui et al., 2012; Deng et al., 2012; Zhao et al., 2012) and medium-resolution (Liu et al., 2020) surveys obtain spectra for millions of stars in the Northern sky.

We gathered v_{los} determinations from the eighth data release (DR8) of LAMOST for ~ 6200 stars. For this sample, we selected low-resolution v_{los} measurements for those stars with S/N ≥ 3 in R-band spectra and medium-resolution v_{los} estimates for stars with zero flags in R-band spectra (`rv_r_flag = 0`).

As part of SDSS-I and SDSS-II, the Sloan Extension for Galactic Understanding and Exploration (SEGUE) obtained low-resolution spectra covering almost the entire optical range. Lee et al. (2008) developed the SEGUE Stellar Parameter Pipeline (SSPP) to measure v_{los} and atmospheric parameters. We procured SSPP v_{los} estimates for ~ 4100 stars in our catalog after limiting our sample to stars with S/N > 3 and `zwarning = [0, 1, 4]`, to limit low-quality spectra and bad template fits.

The Apache Point Observatory Galactic Evolution Experiment (APOGEE), as part of SDSS DR16, released NIR spectra for $\sim 400\,000$ stars located across both the hemispheres (Jönsson et al., 2020; Ahumada et al., 2020). Following Price-Whelan et al. (2018), we discarded v_{los} for those stars for which either `starflags` contained one of: `PERSIST_HIGH`, `PERSIST_JUMP_POS`, `PERSIST_JUMP_NEG`, `VERY_BRIGHT_NEIGHBOR`, `LOW_SNR`, `BAD_PIXELS` or `aspcapflags` was set to `STAR_BAD`. These flags make sure that the spectra were calibrated without any major problems and that the measured stellar parameters are reliable. We accrued APOGEE v_{los} values for ~ 200 stars in our catalog.

To remove stars whose epoch v_{los} values were derived with low confidence from the RAVE DR6 sample (Radial Velocity Experiment; Steinmetz et al. 2020), we followed Kordopatis et al. (2013) to limit the sample to only those stars that have S/N > 5, uncertainty(v_{los}) $\leq 8\,\mathrm{km\,s^{-1}}$, absolute value of zero-point correction to $v_{los} < 10\,\mathrm{km\,s^{-1}}$, and Tonry-Davis R correlation coefficient < 10. These cuts eliminated RAVE stars with erroneous mean v_{los} estimates, and ultimately returned high-confidence v_{los} for ~ 150 bright stars in our catalog.

The Galactic Archaeology with HERMES (GALAH) survey obtains high-resolution spectra of the Southern sky. With its third data release (Buder et al., 2021), it released $\sim 700\,000$ spectra of bright stars. From this sample, we only selected stars with v_{los} computed either by Zwitter et al. (2021) or by the GALAH SME pipeline. We further limited our sample to stars with un-flagged stellar parameters by setting the constraints recommended by Buder et al. (2021): `flag_sp = 0`, `flag_feh = 0`, and `snr_c3 > 30`.

Li et al. (2019) carried out targeted spectroscopic follow-up of Galactic streams using the Anglo-Australian Telescope. We accrued v_{los} measurements from this catalog for ~ 300 stars. A detailed analysis ascribing these stars to their parent streams will be pursued in a future work.

The *Gaia*-ESO Public Spectroscopic survey (GES; Gilmore et al. 2012; Randich et al. 2013) used the ESO spectrographs UVES and GIRAFFE to obtain high-resolution spectra
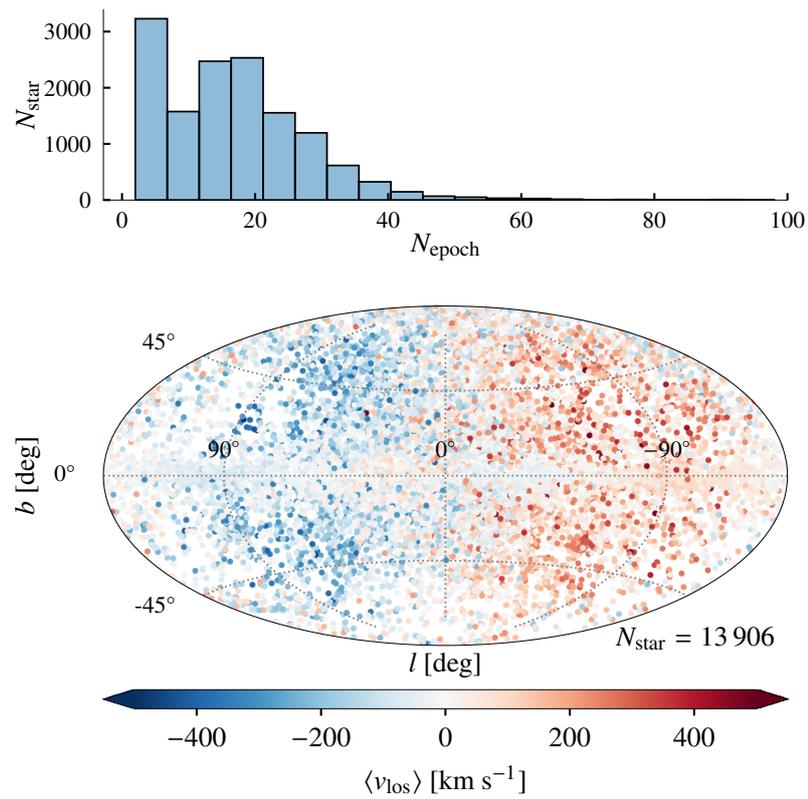
FIGURE IV.4: *Top*: Distribution of number of epoch v_{los} combined per star (N_{epoch}) for stars with $N_{\mathrm{epoch}} \geq 2$. *Bottom*: On-sky distribution of the same stars colored according to the mean of their epoch velocities $\langle v_{\mathrm{los}} \rangle$.
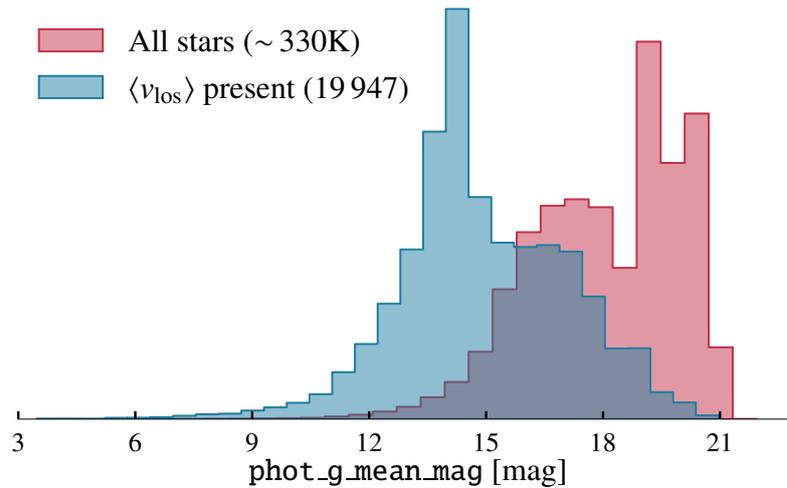
FIGURE IV.5: Normalized histograms of `phot_g_mean_mag` for all stars in our catalog (in red) and for those with a $\langle v_{los} \rangle$ estimate (in blue).

suited for chemical abundance analysis. In its final data release[6] (Gilmore et al., 2022, DR 5.1), GES released v_{los} measurements for $\sim 115\,000$ Milky Way stars. We retrieved GES v_{los} estimates with S/N ≥ 3 for 125 stars in our catalog.

We also accumulated v_{los} from the general star catalog hosted by Simbad in its `BASIC` table. Eliminating v_{los} estimates from the aforementioned surveys already collated, and those with the poorest quality measurements (`rvz_qual = E`), we amassed v_{los} for ~ 1900 stars from Simbad. The survey that contributed the highest number of v_{los} estimates is the Bulge Radial Velocity Assay of RR Lyrae stars (BRAVA-RR; Kunder et al. 2012, 2020).

IV.2.1.2   Combining and validating v_{los}

For all surveys, we removed v_{los} estimates for stars with $|v_{los}| \geq 600\,\mathrm{km\,s^{-1}}$, to limit the number of extragalactic contaminants (see, e.g., Prudil et al., 2022). For stars with multiple v_{los} measurements across surveys, we computed the uncertainty- and N_{obs}-weighted mean, $\langle v_{los} \rangle$, and median of all epoch v_{los} measurements. We also report the standard deviation in their epoch velocities, $\sigma(v_{los})$, and the uncertainty on the median. As done for *Gaia* DR2 (Sartoretti et al., 2018) and DR3 (Hambly et al., 2022), the uncertainty on the median, $\sigma_{v_{los}}$, is computed as:

$$\sigma_{v_{los}} = \sqrt{\frac{\pi}{2}} \frac{\sigma(v_{los})}{\sqrt{N_{epoch}}}, \tag{IV.1}$$

where N_{epoch} is the number of epochs across surveys.

---

[6]https://www.eso.org/rm/api/v1/public/releaseDescriptions/210

Consequently, we accrued $\langle v_{\mathrm{los}} \rangle$ for 19 947 stars. To validate these, we use the *Gaia* SOS Fourier-fit sample whose average v_{los} measurements are good estimates of v_γ (as discussed earlier). We computed $\langle v_{\mathrm{los}}^{\mathrm{non}\text{-}Gaia} \rangle$ by taking into account all surveys except *Gaia* (neither the SOS v_{los} nor the `source_table` estimates were considered). We found 619 stars in common between this non-*Gaia* sample and the *Gaia* SOS sample. Their 1:1 comparison is shown in Fig. IV.3 (left panel). As expected, given the content of both velocities, there are a few outliers whose uncertainties do not account for the difference between the two quantities. However, the overall agreement is good. This agreement is measured in the middle panel of Fig. IV.3. Here, we plot the residuals Δv_{los}, computed as $\langle v_{\mathrm{los}}^{\mathrm{non}\text{-}Gaia} \rangle$ - v_γ^{SOS}, against v_γ^{SOS}. The points are colored according to the amplitude of their RVC, Amp(v_{los}), as computed by the SOS Cep&RRL pipeline and provided as `peak_to_peak_rv`. We see the expected pattern in this figure, wherein $\langle v_{\mathrm{los}}^{\mathrm{non}\text{-}Gaia} \rangle$ for stars with smaller amplitudes is relatively closer to v_γ^{SOS} than those with larger Amp(v_{los}), i.e., their Δv_{los} are relatively tightly distributed around $0\,\mathrm{km\,s^{-1}}$ (though not symmetrically).

The normalized distribution of the residuals, Δv_{los}, is shown in the right panel of Fig. IV.3. It has a mean of $\sim -2\,\mathrm{km\,s^{-1}}$ and is biased more towards negative values. This bias results from the interplay between the sampling bias in our collection and the out-of-phase nature of Cepheids and RR Lyrae stars' radial velocity curves with respect to their light curves. Specifically, when a Cepheid or RR Lyrae star is brightest along its variability phase, it exhibits the lowest v_{los}. As the spectroscopic surveys are S/N-limited, a star's randomly timed observations will be biased towards the times when it is brighter. Consequently, the average of arbitrarily collected epoch v_{los} of a star will be biased towards lower values. Therefore, in our sample, we find that the distribution of Δv_{los} is not symmetrically distributed around $0\,\mathrm{km\,s^{-1}}$, but biased towards negative values.

We also computed the variability index proposed by Katz et al. (2023) to measure the RVC coverage of stars from the DR3 `gaia_source` sample. This variability criterion is defined as `rv_chisq_pvalue` ≤ 0.01 and `rv_renormalised_gof` > 4. This index marked $\sim 90\%$ of the stars in our sample (with at least 10 transits in *Gaia* DR3) as variable, thereby suggesting good coverage across the variability phase.

We emphasize that the final velocities we report include *Gaia* velocities as well — both SOS velocities for ~ 1800 stars and RVS velocities from Katz et al. (2023) for a further $\sim 8\,000$ stars. We also reiterate that the final velocities have not been phased using templates. Nevertheless, the on-sky distribution of these velocities readily reflects the expected bipolar distribution due to the Sun's peculiar motion in the Milky Way. This distribution is shown in the bottom panel of Fig. IV.4 for $\sim 14\,000$ stars with at least 2 epoch measurements.

The magnitude distribution of stars with a $\langle v_{\mathrm{los}} \rangle$ estimate in our catalog is shown in Fig. IV.5. Naturally, brighter stars are preferentially observed in spectroscopic sur-

veys and this is reflected in our collection. Moreover, as Cepheids are intrinsically brighter than RR Lyrae stars, a greater proportion of the former's sample is present in our $\langle v_{\mathrm{los}} \rangle$ complement. Specifically, we collected $\langle v_{\mathrm{los}} \rangle$ for 5% of RR Lyrae stars, while the fractions for classical Cepheids, Type-II Cepheids, and anomalous Cepheids are $\sim 19\%$, $\sim 35\%$, $\sim 15\%$, respectively. Following a similar trend, more than half of the brightest Type-II Cepheids (RV Tau) have a $\langle v_{\mathrm{los}} \rangle$ determination in our catalog ($\sim 56\%$), while only $\sim 15\%$ of the faintest (BL Her) are covered.

IV.2.2   Iron-Abundances

We also collected iron-abundances ([Fe/H]) measured by these surveys using high-resolution spectra. However, the added challenges in measuring [Fe/H] as compared to v_{los} , warrant a greater degree of quality control than the one we undertook. Nevertheless, we include the unvalidated [Fe/H] estimates for ~ 8300 stars in our catalog.

IV.3   Substructure Membership

The satellite galaxies of the Milky Way, that is, the Large and Small Magellanic Clouds, have been used as primary calibrators of PL/PW laws for over a century (Leavitt, 1908; Leavitt & Pickering, 1912). Independently, the distance to these Irregular galaxies has been computed using geometric methods (detached eclipsing binaries, in particular) to $1 - -2\%$ accuracy (see Pietrzyński et al. 2019 for the LMC and Graczyk et al. 2020 for the SMC). As a result, the Magellanic Clouds' classical Cepheids are a significant rung of the extragalactic distance ladder (Riess et al., 2019a). In addition to the classical Cepheids, the Magellanic Clouds are hosts to all our target subtypes, as shown in Chapter II. The extensive observations of the OGLE survey have resulted in an almost complete census of our target subtypes in the Magellanic Clouds.

Except for the Magellanic Clouds, the Milky Way subsystems, namely, open clusters (OCs), globular clusters (GCs), and dwarf spheroidal galaxies (DGs) are also hosts of our target stars (Bailey, 1902; Ceraski, 1905; Irwin, 1955; Zinn & Dahn, 1976; Clement et al., 2001; Dias et al., 2002). Given their small line-of-sight depth (Baumgardt & Vasiliev, 2021; Martínez-Vázquez et al., 2017; Hunt & Reffert, 2023), coupled with the fact that the distance to these subsystems is usually known to a precision and accuracy of 1% or better (Vasiliev & Baumgardt, 2021; Cantat-Gaudin et al., 2020), they can be used as independent calibrators of PL/PW relations of our target subtypes (and references therein Bono et al., 2007; Bhardwaj, 2022; Ngeow et al., 2022b,c,a). Furthermore, they can also be employed to validate PL/PW relations trained on other calibrators like the LMC or *Gaia* parallax sample.

The membership lists of the OCs, GCs, and DGs are readily available (discussed below). Furthermore, the use of different methods and membership criteria lead to few differences in the attribution of certain stars to a given cluster. Hence, provisionally, we rely on literature membership lists to ascribe subsystem membership to our stars. Since most of these lists are primarily based on *Gaia* (E)DR3, they are generally incomplete towards the centers of highly-dense systems (particularly GCs and DGs). To increase this completeness, in Lemasle et al. (in prep.), we present a compilation of variable stars detected in dedicated photometric follow-up surveys that are present in the literature. In this quoted study, we perform a precursory cleaning of literature catalogs of GCs and DGs and use BXM (developed in Sect. II.2) to cross-match these dense catalogs with each other. However, in the current work, we used the membership lists only from the studies that are discussed here.

IV.3.1   Open Clusters

We accrued open cluster members from (Castro-Ginard et al., 2022; Cantat-Gaudin & Anders, 2020; Cantat-Gaudin et al., 2020), (Reyes & Anderson, 2023), (Hunt & Reffert, 2023), and Medina et al. (2021). They provide membership lists of OCs in the Milky Way, along with the mean distance, age, extinction, and radial velocity of the cluster. We summarize these analyses here.

Cantat-Gaudin & Anders (2020) fed *Gaia* DR2 data to the unsupervised membership assignment algorithm called UPMASK (Krone-Martins & Moitinho, 2014). Cantat-Gaudin et al. (2020) built on the Cantat-Gaudin & Anders (2020) membership lists of 1481 open clusters by combining them with literature catalogs and characterizing them with deep learning. They released a list of 1867 open clusters with reliable ages and distance modulus.

Castro-Ginard et al. (2022) employed a custom algorithm they call `OCfinder` (Castro-Ginard et al., 2018) on *Gaia* EDR3 data to detect and characterize 1274 new open clusters. This list includes the new discoveries made by them using *Gaia* DR2 data (Castro-Ginard et al., 2020). In essence, they utilized a clustering algorithm termed DBSCAN (Density-based spatial clustering of applications with noise) to detect overdensities in the *Gaia* astrometric photometric space. Consequently, they leverage *Gaia* photometry (with deep learning) to characterize the detected open clusters using synthetic isochrones.

Hunt & Reffert (2021) compared the performance of DBSCAN, Hierarchical DBSCAN (HDBSCAN) and Gaussian mixture models (GMM) for unsupervised cluster detection in *Gaia* DR2 data. They concluded that HDBSCAN is the most sensitive algorithm albeit it requires post-processing to eliminate false positives. Consequently, in Hunt & Reffert (2023), they employed HDBSCAN on *Gaia* EDR3 data and validated their results using a Bayesian neural network. They released a list of 4105 high-confidence clusters (including

open clusters, moving groups, and globular clusters), out of which 739 were new discoveries. A distinct advantage of their algorithm over that of Cantat-Gaudin et al. (2020) is the ability to recover true members even away from the cluster core. This is borne out of the fact that the UPMASK algorithm used in the latter analysis has the highest detection efficiency towards the central regions.

We cross-matched RRLCep with these catalogs and recovered 58 classical Cepheids in 54 open clusters. We were able to recover 13 pairs out of 19 high-probability pairs of Reyes & Anderson (2023). The on-sky distribution of these clusters is shown in Fig. As shown in Reyes & Anderson (2023) these cluster Cepheids make for an excellent calibrator of PL/PW relations.

IV.3.2 Globular Clusters and Dwarf Galaxies

Vasiliev & Baumgardt (2021) published membership lists of Galactic globular clusters (GGCs) using data from *Gaia* DR3 providing a homogeneous set of distances to ~ 170 GGcs. Baumgardt & Vasiliev (2021) combined these distance estimates with those available in the literature to compute highly precise and accurate GGC distances and their respective members.

Vasiliev (2023) predicted that 7 classical MW DGs (c.f. e.g, McConnachie 2012 and Martínez-Vázquez et al. 2017 for properties of these galaxies and Gallagher & Wyse 1994; Grebel et al. 2003 for a broader review) were formed in the Magellanic Clouds and were caught in Milky Way's potential during the first infall of the Clouds $\gtrsim 5$ Gyr ago. Nevertheless, for our analysis, we considered these seven galaxies as part of the Milky Way halo. We relied on the membership lists of Gaia Collaboration et al. (2018b) and gathered distances to individual galaxies from the compilation of Huang & Koposov (2021).

In total, we found 2168 RR Lyrae stars and 72 Type-II and anomalous Cepheids in 63 Milky Way GCs and DGs. We reiterate that this list is by no means complete. For instance, our catalog includes ~ 100 RR Lyrae stars in w Centauri (NGC 5139, presented in Fig. IV.6). However, in the same system, Braga et al. (2016) classified ~ 180 RR Lyrae stars by leveraging over two decades of photometry. Similarly, using a 4 m telescope, Navarrete et al. (2017) classified secen Type-II Cepheids in w Cen, five of which are also recovered in our catalog (see Fig. IV.7).

This incompleteness in our catalog is due to the crowded nature of w Cen's central regions. As a result, the membership list of Baumgardt & Vasiliev (2021), which is based on *Gaia* (E)DR3 itself, is incomplete. Furthermore, the high stellar density also affects the ability of other ground-based, all-sky surveys to probe the interior regions of w Cen, thereby hampering their variable star identification. A comprehensive membership list per cluster
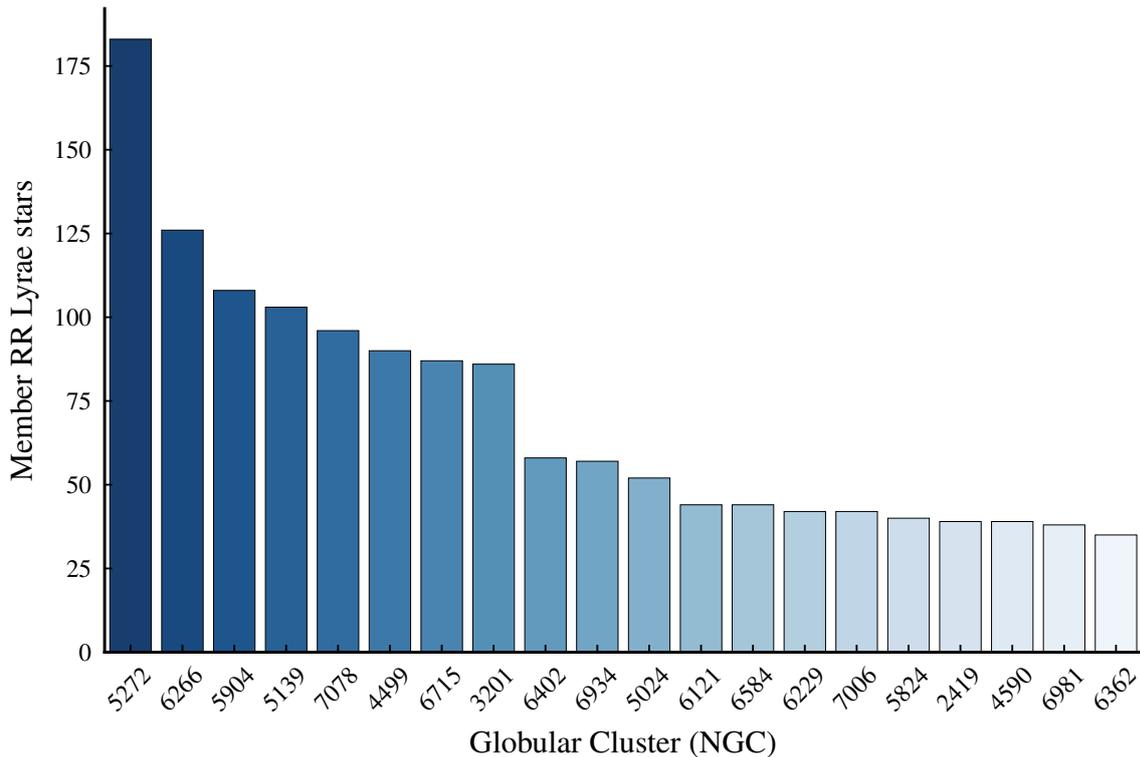
FIGURE IV.6: Number distribution of RR Lyrae stars in the globular clusters of the Milky Way. The *x*-axis represents the name of the cluster in the format NGC XXXX. Only the top 20 clusters with the largest RR Lyrae population are shown here.

and dwarf galaxy will be presented in Lemasle et al. (in prep.). Nevertheless, these member stars present an invaluable preliminary sample for an independent calibration and validation of PL/PW relations (see, e.g., Ngeow et al. 2022a,b,c).

IV.3.3 The Magellanic Clouds

In contrast with these stellar systems, the Magellanic Clouds membership lists are few and far between (see discussion in Gaia Collaboration et al., 2021b).

To ascertain which stars in our catalog are part of the Magellanic clouds, we compared their positions and proper motions against those of the Clouds. We employed the strategy outlined here to assign XMC (LMC/SMC) membership.

Step 1: We initially selected XMC stars based only on their coordinates as discussed in Chapter II and detailed in Appendix 1.3.

Step 2: From these preliminary samples, we removed all the stars with $\varpi/\sigma_\varpi \geq 8$ deeming them foreground contaminants, where ϖ, σ_ϖ are *Gaia* DR3 parallax and its uncertainty. This limit in parallax S/N did not remove any stars from the central regions of the clouds, while it eliminated ~ 220 (70) stars towards the LMC (SMC) outskirts. More than $\sim 95\%$
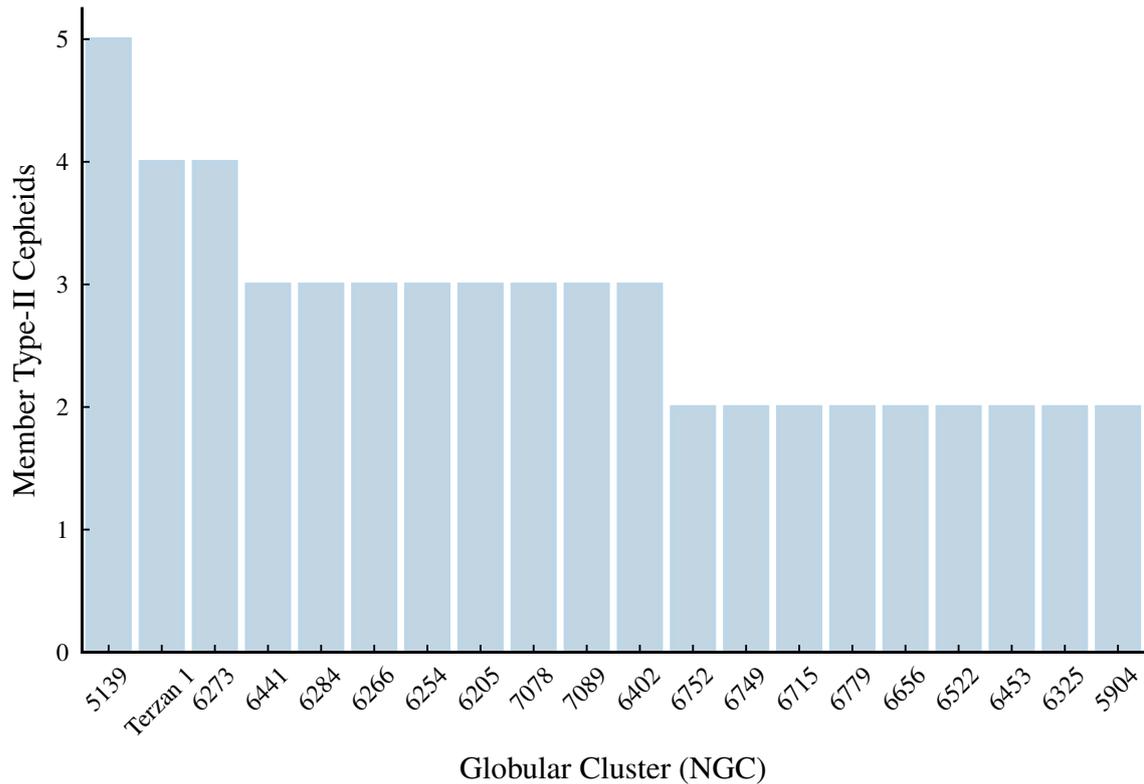
FIGURE IV.7: Same as Figure IV.6 but for Type-II Cepheids.

of these stars are RR Lyrae and they have $G \lesssim 16$ mag, while the core LMC RR Lyrae population is fainter than $G \sim 18$ mag.

Step 3: We employed the Minimum Covariance Determinant estimator (MCD, Rousseeuw 1984; Rousseeuw & Driessen 1999; explained below) to estimate the on-sky center (α_0, δ_0) of XMC.

Step 4: To avoid measuring distances in spherical geometry, we transformed equatorial coordinates (α, δ) to orthographic cloud-centric coordinates (x, y). Given the limited angular size of the Clouds, coupled with the absence of stellar distances, it is adequate to use a 2-d orthographic projection over a 3-d (x, y, z) gnomonic projection (c.f. Gaia Collaboration et al. 2018b). Following Jacyszyn-Dobrzeniecka et al. (2016); Iwanek et al. (2018); Gaia Collaboration et al. (2021b), we computed x, y as:

$$
\begin{aligned}
x &= \cos \delta \sin(\alpha - \alpha_0) \\
y &= \sin \delta \cos \delta_0 - \cos \delta \sin \delta_0 \cos(\alpha - \alpha_0).
\end{aligned}
\tag{IV.2}
$$

Here, α_0, δ_0 is the center of the XMC as determined in the previous step.

Step 5: Similarly, to leverage *Gaia* astrometry, we converted the stellar proper motions along RA and Dec $(\mu_{\alpha^*}, \mu_\delta)$ to their orthographic components (μ_x, μ_y) as follows:

$$\mu_x = \mu_{\alpha^*} \cos(\alpha - \alpha_0) - \mu_\delta \sin\delta \sin(\alpha - \alpha_0)$$

$$\mu_y = \mu_{\alpha^*} \sin\delta_0 \sin(\alpha - \alpha_0) \tag{IV.3}$$

$$+ \mu_\delta \left(\cos\delta \cos\delta_0 + \sin\delta \sin\delta_0 \cos(\alpha - \alpha_0) \right).$$

Step 6: For the final selection of members, we calculated the Mahalanobis distance (D_M) between every star and the XMC center in both phase spaces: the coordinates (x, y) and proper motions (μ_x, μ_y). We determined the centers and shapes (covariance matrices) of these distributions using the MCD estimator. Essentially, the MCD aims to identify the largest subset of stars that contribute to the lowest covariance within the sample. This estimator is immune to outliers and considered highly robust when $N_{\text{star}} > 5 \times d$, where d is the dimensionality.

Step 7: To estimate the center $(\mu_{x,0}, \mu_{y,0})$ and the shape $(\Sigma_{\mu_{xy},0})$ of the XMC (μ_x, μ_y) distribution, we updated the `scikit-learn` (Pedregosa et al., 2011) implementation of the MCD algorithm (Rousseeuw & Driessen, 1999). This updated version incorporates the uncertainty covariance matrix ($\Sigma_{\mu_{xy}}$) for each star and is, consequently, more robust as compared to the original implementation.

Step 8: The measured proper motion uncertainty covariance matrix $\Sigma_{\mu_{\alpha^*\delta}}$ was transformed to $\Sigma_{\mu_{xy}}$ using the following Jacobian matrix:

$$J = \begin{pmatrix} \cos(\alpha - \alpha_0) & -\sin\delta \sin(\alpha - \alpha_0) \\ \sin\delta_0 \sin(\alpha - \alpha_0) & \cos\delta \cos\delta_0 + \sin\delta \sin\delta_0 \cos(\alpha - \alpha_0) \end{pmatrix}, \tag{IV.4}$$

in this equation:

$$\Sigma_{\mu_{xy}} = J \, \Sigma_{\mu_{\alpha^*\delta}} \, J^T \tag{IV.5}$$

Step 9: We tried using the updated MCD estimator to compute the center (x_0, y_0) and the shape $(\Sigma_{xy,0})$ of the XMC orthographic coordinates. This approach took into account both the uncertainties on coordinates and the correlation between coordinates and proper motions (i.e, the full 4×4 covariance matrix). However, this did not lead to a significant difference as the *Gaia* coordinates are measured to high precision, and the correlation between coordinates and proper motions is only marginal. Hence, the final estimation of (x_0, y_0) and $\Sigma_{xy,0}$ did not include the per-star uncertainties.

Step 10: Finally, we computed $D_{M,j}$ (for j from 1 to N_{star}) as:

$$\begin{aligned} D_{M,xy}^j &= \Delta_{xy,j} \, \Sigma_{xy,0}^{-1} \, \Delta_{xy,j}^T \\ D_{M,\mu_{xy}}^j &= \Delta_{\mu_{xy},j} \, \Sigma_{\mu_{xy},j,\text{joint}}^{-1} \, \Delta_{\mu_{xy},j}^T \end{aligned}, \tag{IV.6}$$
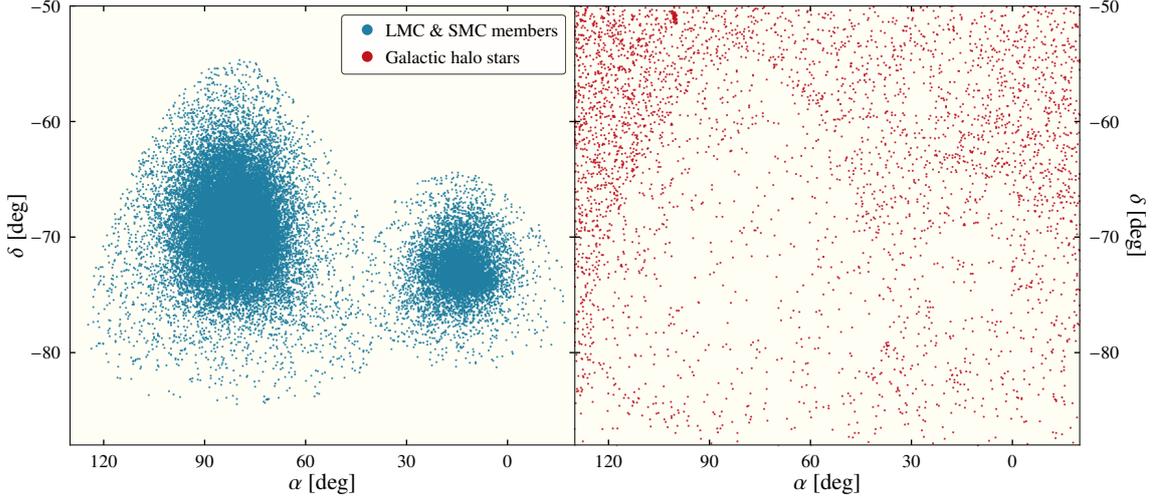
FIGURE IV.8: On-sky map of stars towards the Magellanic Clouds (in equatorial coordinates). *Left*: Stars selected as members of the Magellanic Clouds (blue). *Right*: Subtracted map of the same regions depicting the Galactic halo stars (in red).

where,

$$
\begin{aligned}
\Delta_{xy,j} &= [x_j,\, y_j]^T - [x_0,\, y_0]^T \\
\Delta_{\mu_{xy},j} &= [\mu_{x,j},\, \mu_{y,j}]^T - [\mu_{x,0},\, \mu_{y,0}]^T \\
\Sigma_{\mu_{xy,j},\text{joint}} &= \frac{(\Sigma_{\mu_{xy,j}} + \Sigma_{\mu_{xy,0}})}{2}.
\end{aligned}
\tag{IV.7}
$$

Here, D_M^j follows a χ^2 distribution with 2 degrees of freedom, therefore, the probability of membership p^j can be computed as: $p^j = 1 - \mathcal{F}_{\chi^2}(D_M^j, 2)$, where F is the cumulative distribution function. We removed any star with either $p_{xy}^j \leq 0.001$ or $p_{\mu_{xy}}^j \leq 0.001$. After removing the stars on the basis of proper motions, we noticed that $\sim 5\%$ of stars less than $1°$ from the XMC center were classified as non-XMC members. A possible reason for this is that in these crowded central regions, the proper motion solutions in *Gaia* DR3 are inaccurate. Therefore, we excluded stars less than $\sim 1°$ from the XMC center from proper motion filtering.

Consequently, we selected $\sim 48\,000$ stars as LMC members and $\sim 11\,000$ as SMC members. Their on-sky and orthographic maps are presented in Figures IV.8 and IV.9. In Fig. IV.8 this, we display the XMC subtracted maps of the Magellanic regions, which prove that our selection is quite pure and complete. Furthermore, this selection will be made even more robust by taking individual stellar distances into account in Chapter V.
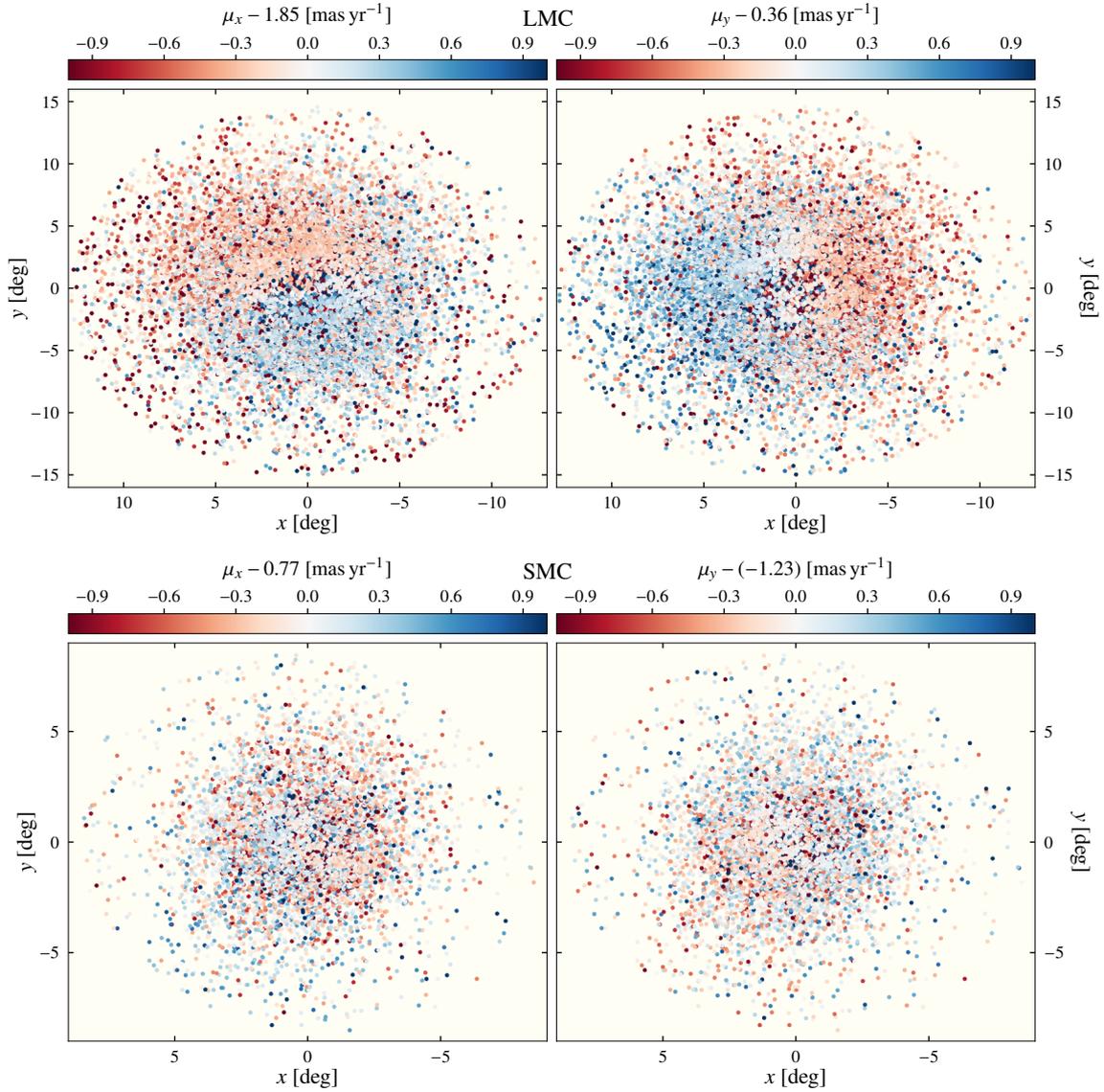
FIGURE IV.9: Orthographic projections of the LMC (*top row*) and the SMC (*bottom row*) member stars. The stars are colored according to their proper motions along the orthographic components: μ_x (*left column*) and μ_y (*right column*). These proper motions have been centered and their mean values are listed above the color bar of each panel. The maps in the *top* row are consistent with the maps produced by Gaia Collaboration et al. (2021b) for the entire LMC population and were discussed in Fig. I.11.

IV.4   Extinction

An essential component in the computation of period-luminosity relations is the extinction to each star (A). Since their release, the 2-d reddening map of Schlegel et al. (1998) — commonly referred to as the SFD map — is the *de facto* reddening map used by most studies. However, with better data and modern methodologies, 3-d maps focused on a specific Galactic region (like the disk) or 3-d all-sky maps have been produced. Here, we describe the latest reddening maps that we used to garner an extinction estimate for our stars.

IV.4.1   The Milky Way midplane

By combining *Gaia* DR3 parallax and 2MASS photometry, Lallement et al. (2019, 2022) estimated extinction within 0.8 kpc of the Galactic midplane. As discussed by Babusiaux et al. (2023) and presented in Ulla et al. (2022), this monochromatic extinction A_0 at 5500 approximates A_V fairly well for regions with $A_0 \leq 2.5$ mag. Hence, for stars near the disk (mainly DCEP), we assimilated extinction from Lallement et al. (2022) 3-d map[7]. The distance we used to retrieve this 3-d extinction was derived using PW relations computed in the next chapter.

IV.4.2   The Magellanic Clouds

Red Clump stars exhibit a narrow range of intrinsic luminosities and effective temperatures borne out of their evolutionary stages. By leveraging the theoretically calculable intrinsic colors of such Red Clump stars, Skowron et al. (2021) produced a detailed reddening map of the Magellanic Clouds. This map, which quantifies reddening as E($V - I$), yields greater accuracy and resolution towards the central regions of the Clouds compared to the SFD map. We extracted E($V - I$) estimates from this map for $\sim 50\,000$ stars in the Magellanic Clouds.

IV.4.3   All-sky map

For the rest of the sky, we used the 3-d extinction map from Bovy et al. (2016). This map is a combination of the following maps: Green et al. (2019) map based on Pan-STARRS and 2MASS photometry which covers three-quarters of the sky, Marshall et al. (2006) map based on 2MASS data towards the Galactic disk, and the COBE-DIRBE-based Drimmel et al. (2003) map for stars at higher latitudes in the Southern hemisphere.

We converted reddening measurements from these maps to extinction in various bands using Table 6 of Schlafly et al. (2016) and Table 3 of Wang & Chen (2019). We used the latter

---

[7]https://explore-platform.eu/sda/g-tomo

table to compute the total-to-selective extinction ratios that constitute the various Wesenheit indices of our PW relations. These relations and Wesenheit indices are presented in the next chapter.

V

Distance

One of the major applications of our catalog is the computation of highly precise and accurate period-luminosity (PL) and period-Wesenheit (PW) relations (also called Leavitt laws in honor of the pioneering works of Henrietta Leavitt, Leavitt 1908; Leavitt & Pickering 1912), thereby determining high-quality distances to RR Lyrae stars and Cepheids. For a recent *Gaia* DR3-based review of the Leavitt laws, refer Groenewegen (2023).

To compute such relations for our stars, the required inventory includes: mean magnitudes (preferably Fourier-fit or phase-corrected using templates), variability periods, reddening law, reddening map (only for PL relations), calibrators (stars with known distance or parallax), and correctly estimated uncertainties on each of these quantities. Given this inventory, the PL/PW laws computed from them also depend on the model used to fit the relations. Many traditional methods either do not take the uncertainties into account or they can only afford an incomplete treatment of these errors.

In this chapter, we present both the ingredients (in Sections V.1–V.3) and the model that were used to fit over 500 different PL/PW relations. Each one of these relations is based

on exclusive ingredients (like the photometry) and hence, warrants a detailed validation. Within the time-frame of this thesis, 200 period-Wesenheit relations have been validated. We present the results of these 200 precise PW relations in Sect. V.5 and also discuss their internal validation in Sect. V.6.

V.1   Mean magnitudes

Through the data aggregation performed in Chapters II and IV, our catalog contained mean magnitudes of stars in over 50 bands. A subset of these were used to construct Wesenheit indices described in the next section. Here, we outline the steps that we took to ensure appropriate uncertainty estimation of these quantities.

V.1.1   Uncertainty on mean magnitude

We estimated the total formal uncertainty, σ_{m_x}, on the apparent mean magnitude in band x (m_x) following Madore & Freedman (2005), Scowcroft et al. (2011), and Neeley et al. (2015) as:

$$\sigma_{m_x} = \sqrt{\sigma_{\text{phot}}^2 + \sigma_{\text{fit}}^2},$$

where,

$$\sigma_{\text{phot}} = \frac{\sqrt{\sum_{i=1}^{N_{\text{obs},x}} \sigma_{i,\text{phot},x}^2}}{N_{\text{obs},x}} \quad \text{and} \quad \sigma_{\text{fit}} = \frac{\text{Amp}_x}{\sqrt{12 N_{\text{obs},x}}}.$$

(V.1)

Here, $\sigma_{i,\text{phot},x}$ is the photometric uncertainty associated with the ith observation in band x. In cases where the epoch observations were inaccessible, and only a single average estimate ($\sigma_{\text{avg, phot}}$) of the photometric uncertainty was available, we calculated σ_{phot} as $\frac{\sigma_{\text{avg, phot}}}{\sqrt{N_{\text{obs},x}}}$. Furthermore, we assumed that the apparent magnitude of a star follows a uniform (or rectangular) distribution $\mathcal{U}\left(m_x - \frac{\text{Amp}_x}{2}, m_x + \frac{\text{Amp}_x}{2}\right)$, where Amp_x is the peak-to-peak amplitude in band x. Under this assumption, the uncertainty on the mean magnitude, $\sigma_{\text{fit},x}$, is computed as given in Eqn. V.1. We justify this assumption below.

The eight constituent surveys of our catalog provide either Fourier-fit (OGLE, *Gaia*, ZTF, ASAS-SN, CSS, MACC) or template-fit (PS1, DES) mean magnitudes. Thus, σ_{m_x} for these surveys is equal to $\sigma_{\text{phot},x}$ (as defined in Eqn. V.1). This treatment was adopted for all surveys except *Gaia*, as the bootstrapped uncertainty estimation (Andrae, 2010) as implemented by the SOS Cep&RRL pipeline (Clementini et al., 2023; Ripepi et al., 2023) already accounts for the photometric uncertainty.

However, in Chapter IV, we cross-matched our catalog with ancillary surveys such as 2MASS, VISTA surveys, WISE, and DECAM surveys. These surveys provide NIR
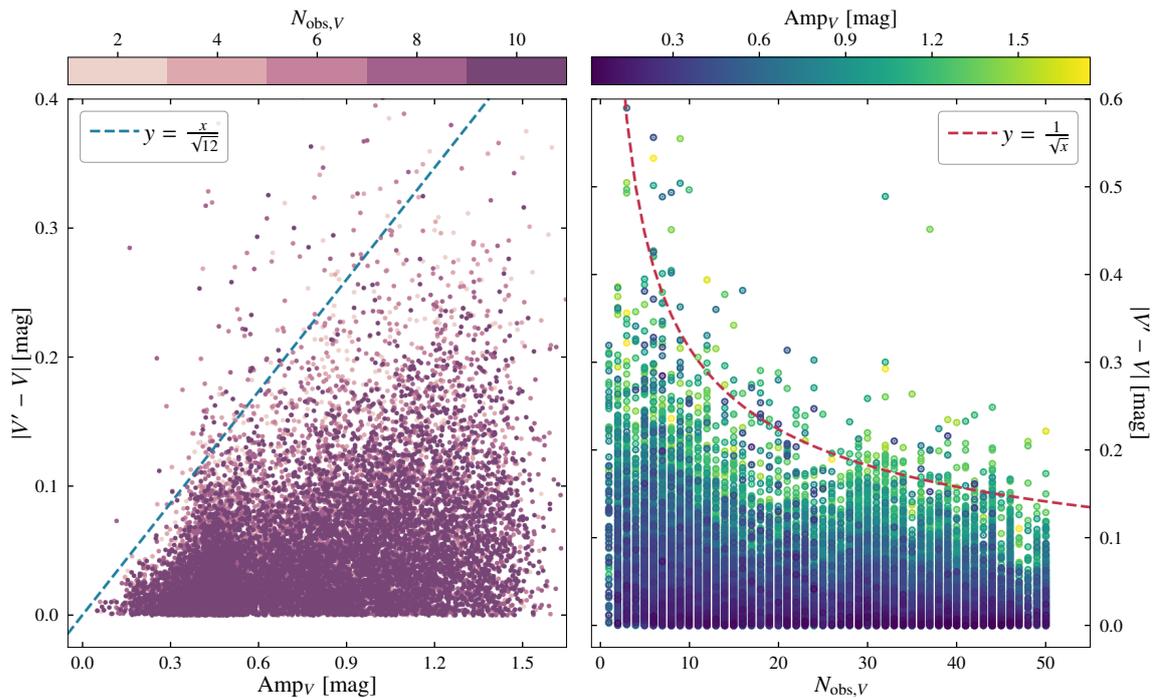
Figure V.1: Uncertainty on mean magnitude as a function of variability amplitude (Amp$_V$, *left*) and number of observations ($N_{\mathrm{obs},V}$, *right*) in OGLE V-band. Here, V is the Fourier-fit intensity-averaged magnitude and V' is the arithmetic mean of the observed magnitudes. The absolute difference of these two quantities (plotted on the y-axis) quantifies $\sigma_{\mathrm{fit},V'}$, i.e., the error on mean magnitudes that are not corrected for phase (using templates). $\sigma_{\mathrm{fit},V'}$ is defined in Eqn. V.1 and its dependence can be decomposed into Amp$_V$ and $N_{\mathrm{obs},V}$, with upper limits computed as: (i) $\frac{\mathrm{Amp}_V}{\sqrt{12}}$ (shown in blue) for stars with $N_{\mathrm{obs},V} = 1$ and (ii) $\frac{1}{\sqrt{N_{\mathrm{obs},V}}}$ (red) for stars with Amp$_V$ = 1 mag.

($ZYJHKs$), MIR (W1, W2), and optical ($ugrizy$) magnitudes that are complementary to the Fourier/template-fit magnitudes included in our catalog. However, these magnitudes are not corrected for phase (using templates or Fourier models) but are computed either using co-added images or those that are an arithmetic mean of all observations. This crude averaging induces another source of uncertainty (on top of the photometric uncertainty $\sigma_{\mathrm{phot},x}$) in the photometry of variable stars. This uncertainty is modeled as $\sigma_{\mathrm{fit},x}$ as defined in Eqn. V.1, and its dependence on the amplitude and number of observations is presented in Fig. V.1, using OGLE V-band photometry as an example.

For stars with fewer than 10 V-band observations in OGLE-IV, their mean V-band magnitudes (hereafter denoted as V for concision) are computed using templates generated from their I-band Fourier fits (I. Soszynski, in priv. comm.). For a star with ≥ 10 photometric epochs, its V-band magnitude is calculated by fitting a Fourier model to the V-band light curve itself. These mean magnitudes (V) are intensity-averaged. Specifically, their Fourier model (in magnitudes) is converted to a flux scale, and the average of this flux is converted back to magnitudes and reported as V. In Fig. V.1, we compare this intensity-averaged V magnitude to V', where we computed V' as a simple mean of all measured magnitudes for a star (uncorrected for phase).

We demonstrate here that the observed absolute difference $|V' - V|$ can be used to approximate the theoretically defined $\sigma_{\mathrm{fit},V'}$. Here, $\sigma_{\mathrm{fit},V'}$ represents the error that should be incorporated into V' under the assumption that the Fourier-fit V is the "true" magnitude of the star and follows the uniform distribution defined above. Under this assumption, $\sigma_{\mathrm{fit},V'}$, specifically, can be computed as (following Eqn. V.1) $\frac{\mathrm{Amp}_V}{\sqrt{12 N_{\mathrm{obs},V}}}$. Therefore, $|V' - V|$ should exhibit a similar dependence on the amplitude and number of observations.

This dependence is readily apparent from Fig. V.1. Here, we plot the distribution of $|V' - V|$ against Amp_V (in the left panel) and $N_{\mathrm{obs},V}$ (right panel). Naturally, with increasing amplitude, the error on uncorrected-for-phase V' magnitude increases. The blue line in this panel has a slope of $\frac{1}{\sqrt{12}}$, and it sets the upper limit on $|V' - V|$ (for a given amplitude), assuming $N_{\mathrm{obs},V} = 1$. With increasing number of observations per star, one expects the difference between V' and V to diminish. This effect is discernible in the left panel itself, where, for stars with ~ 10 observations, we observe $|V' - V| \leq 0.1$.

However, this effect becomes more evident in the right panel, where we see that $|V' - V|$ scales down with $\sqrt{N_{\mathrm{obs},V}}$ (depicted as the red line). The points in this panel are color-coded based on their Amp_V values. This pattern reaffirms the finding from the left panel that stars with larger amplitudes exhibit larger $|V' - V|$ values, for a given number of observations. Combining these two dependencies, yields $|V' - V| \approx \frac{\mathrm{Amp}_V}{\sqrt{12 N_{\mathrm{obs},V}}} (= \sigma_{\mathrm{fit},V'})$.

Thus, in the absence of Fourier-fit intensity-averaged mean magnitudes, $\sigma_{\mathrm{fit},x}$ (equal to $\frac{\mathrm{Amp}_x}{\sqrt{12 N_{\mathrm{obs},x}}}$) should be included as a significant source of uncertainty for magnitudes computed

as a mean of randomly-timed observations.

V.1.2   Amplitude ratios

To compute σ_{m_x}, both Amp_x and $N_{\mathrm{obs},x}$ are required. In cases where neither the amplitude nor the epoch photometry were accessible, we used newly-derived amplitude ratios to estimate Amp_x and compute σ_{fit}. Inaccurate amplitude ratios can potentially be a source of systematic bias arising due to blending/crowding (see, for e.g., Riess et al., 2020; Sharon et al., 2023) and therefore, their computation warrants a detailed report which is beyond the scope of the current analysis. Nevertheless, one such amplitude ratio is discussed below and the rest (along with their dependence of metallicity and period) will be presented in a future work.

As discussed in Chapter IV and Lemasle et al. (2022), the unWISE co-added W1, W2 photometry (Schlafly et al., 2019) offers a better approximation of Fourier-fit mean magnitudes of Chen et al. (2018), compared to CatWISE photometry Marocco et al. 2021. However, the Chen et al. (2018) variable star catalog only contains ~ 4300 stars in common with our catalog, while the unWISE 5-year coadds provide high-quality W1, W2 measurements for $\sim 180\,000$ stars (see Paper I for a description of the quality control steps). Therefore, to estimate σ_{fit} for the unWISE sample, we computed the amplitude ratio $\frac{\mathrm{Amp_{W1}}}{\mathrm{Amp}_x}$, where $\mathrm{Amp_{W1}}$ represents the amplitude of the Fourier model employed by Chen et al. (2018) to fit WISE light curves. We computed this ratio independently for all variability types in our catalog, and for all available Amp_x measurements. We found these ratios to be consistent across variability types, with a standard deviation of about 0.1 magnitudes.

For instance, we obtained $\frac{\mathrm{Amp_{W1}}}{\mathrm{Amp}_I} \sim 0.49 \pm 0.11$ (standard deviation) for ~ 800 stars common between the Chen et al. (2018) and OGLE catalogs. Additionally, when calculated for ~ 3300 stars in common with the *Gaia* catalog, we get $\frac{\mathrm{Amp_{W1}}}{\mathrm{Amp}_G} \sim 0.37 \pm 0.11$. These estimates are consistent with $\frac{\mathrm{Amp}_G}{\mathrm{Amp}_I} \sim 1.32 \pm 0.18$, which was calculated using $\sim 112\,000$ stars common between the OGLE-IV and *Gaia* DR3 catalogs. When not possible to compute them using our data, we used literature amplitude ratios like those from Braga et al. (2018) and Bhardwaj et al. (2020).

V.1.3   Count-Rate Nonlinearity

Count-Rate Nonlinearity (CRNL, c.f. Riess et al. 2019b) affects NIR observations performed using HgCdTe detectors, like those on HST WFC3. Riess et al. (2019a) take this effect into account (estimated to be 0.75% per unit dex or flux ratio) when applying HST NIR PL relations trained on the LMC Cepheids to SNe Ia host galaxies. However, for the VISTA NIR observations included here, we do not correct for this effect on the basis of two

factors (J. M. Irwin, priv. comm.). Firstly, first-order CRNL corrections are performed by the photometric reduction pipeline of VIRCAM (Irwin et al., 2004; Neeser et al., 2016). Secondly, post-correction residual nonlinearity is below the measured photometric uncertainty.

V.2   Wesenheit Indices and Reddening Law

We used the apparent magnitudes to construct the Wesenheit indices, W, that are listed in Table V.1. Madore & Freedman (1991, in their Appendix B) illustrated the theoretical advantages of using the Wesenheit index (Madore, 1976, 1982). In essence, W, by definition, is independent of reddening towards a star and only depends on the reddening law of the parent system.

Under a three-band formalism, with magnitudes (say) $\lambda_1, \lambda_2,$ and λ_3, (where λ, the effective wavelength of a filter, represents the magnitude in that band), the Wesenheit index is defined as (Madore, 1982):

$$W_{\lambda_1,\lambda_2,\lambda_3} = \lambda_1 - R \cdot (\lambda_2 - \lambda_3)$$
$$R = A_{\lambda_1}/E(\lambda_2 - \lambda_3).$$

(V.2)

Here, A_{λ_1} is the extinction in band 1 and $E(\lambda_2 - \lambda_3)$ is the reddening or the color excess in band 2 with respect to band 3. R, also termed as the total-to-selective extinction ratio, is called the Wesenheit coefficient hereafter. Its value is defined by the reddening law and the effective wavelengths of the filters. While such an index is usually denoted as $W_{\lambda_2,\lambda_3}^{\lambda_1}$, we chose an alternative simpler representation owing to the number and variety of Wesenheit indices included here. These indices, their component magnitudes, and the respective Wesenheit coefficients, R, are listed in Table V.1.

We follow the Wesenheit formalism wherein for all indices used here, we have $\lambda_2 > \lambda_1 > \lambda_3$. In the event that three different magnitudes are unavailable, we defined a 2-band Wesenheit index with $\lambda_1 = \lambda_3$. These conditions are satisfied by all the indices listed in Table V.1, except for $W_{G,G,\mathrm{RP}}$. For this index, by design we set $\lambda_1 = \lambda_2$, to assess its performance against the other *Gaia* index, i.e., $W_{G,\mathrm{BP},\mathrm{RP}}$.

We constructed W for our stars following the extinction law from Wang & Chen (2019, hereafter WC19) with the total-to-selective extinction ratio $R_V = 3.16 \pm 0.15$. The effect of using different reddening laws on the computation of distances was investigated by Riess et al. (2022). They concluded that: when applying PW relations trained on one galaxy to another, it is better to use a Wesenheit system based on the same reddening law across different galaxies (due to large uncertainties in data). In order to use a host-specific reddening

Table V.1: Wesenheit indices used in this thesis.

Notes. $W_{\lambda_1,\lambda_2,\lambda_3}$ is the Wesenheit index constructed using the magnitudes $\lambda_1, \lambda_2, \lambda_3$ listed in the column 'Magnitudes'. The coefficient R and its uncertainty σ_R are used to computed W as defined in Eqn V.2

W	Magnitudes $(\lambda_1, \lambda_2, \lambda_3)$	R	$\sigma_R \times 10^3$
$W_{G,\mathrm{BP,RP}}$	G, BP, RP	1.91	15
$W_{G,G,\mathrm{RP}}$	G, G, RP	3.945	21
$W_{I,V,I}$ (OGLE)	I, V, I (OGLE)	1.506	15
$W_{r,g,r}$ (DES)	r, g, r (DES)	2.048	18
$W_{r,g,i}$ (DES)	r, g, i (DES)	1.325	9
$W_{i,r,z}$ (DES)	i, r, z (DES)	1.664	16
$W_{i,r,y}$ (DES)	i, r, y (DES)	1.437	14
$W_{r,g,r}$ (ZTF)	r, g, r (ZTF)	2.048	25
$W_{r,g,r}$ (PS1)	r, g, r (PS1)	2.702	32
$W_{r,g,i}$ (PS1)	r, g, i (PS1)	1.6	14
$W_{i,r,z}$ (PS1)	i, r, z (PS1)	1.764	14
$W_{i,r,y}$ (PS1)	i, r, y (PS1)	1.402	11
$W_{Ks,J,Ks}$ (2MASS)	Ks, J, Ks (2MASS)	0.473	6
$W_{H,J,Ks}$ (2MASS)	H, J, Ks (2MASS)	0.794	38
$W_{Ks,J,Ks}$ (VISTA)	Ks, J, Ks (VISTA)	0.503	9
$W_{J,Y,Ks}$ (VISTA)	J, Y, Ks (VISTA)	0.951	31
$W_{H,J,Ks}$ (VISTA)	H, J, Ks (VISTA)	0.845	45
$W_{\mathrm{W2,W1,W2}}$ (unWISE)	W2, W1, W2 (unWISE)	2.0	14

law, one first needs to take into account the intrinsic color of the stars. However, the small width of the instability strip, coupled with the uncertainties on the observed quantities, leads to only marginal improvements in distance determination. Hence, we used the WC19 law and the subsequent Wesenheit coefficients listed in V.1 for stars in the Milky Way and the Magellanic Clouds.

V.2.1   Selection of Wesenheit indices

The number of two-band or three-band W that can be constructed using all the different magnitudes included in our catalog is quite large. In our current undertaking, we set up 25 different W that are a mix of commonly used indices as well as those indices that have not been explored before. Furthermore, a PW relation for many of our subtypes in most of these indices has never been computed. Thus, this analysis presents the first ever PW relations for

a large number of (variability subtype-Wesenheit index) combinations.

V.2.1.1 *Gaia* **Indices**

We recollect the different nature of the three sets of *Gaia* magnitudes (initially described in Sect. II.2). Firstly, the Fourier-fit mean magnitudes from SOS Cep&RRL pipeline that are intensity-averaged. Secondly, the `mean_x_mag` from `vari_summary` table computed as an unweighted arithmetic mean of the epoch fluxes. Lastly, the `phot_x_mean_mag` from `gaia_source` calculated as an uncertainty-weighted average of the epoch magnitudes.

The first set of magnitudes is available only for the stars classified as RR Lyrae stars or Cepheids by Clementini et al. (2023) or Ripepi et al. (2023). However, the second set of magnitudes is available for all the stars in which any variability is detected by the *Gaia* general variability pipeline (Eyer et al., 2023). Furthermore, the `gaia_source` magnitudes are available for all the stars detected by *Gaia* with a large enough S/N (Gaia Collaboration et al., 2023c). As a result, the availability of these sets of magnitudes is successively greater, that is many more stars have a `gaia_source` magnitude than an SOS magnitude. However, the accuracy of these magnitudes is also successively worse.

With future *Gaia* data releases, almost all the stars in our catalog will have a Fourier-fit magnitude. However, until then, for stars classified as variable in the literature but not by *Gaia*, can the `gaia_source` magnitudes (or the `vari_summary` magnitudes) be used as a proxy for the 'true' apparent magnitudes of these stars? To answer this question, we fit two PW relations in all three systems. Specifically, we constructed $W_{G,\mathrm{BP,RP}}$ and $W_{G,G,\mathrm{RP}}$ using each of the three sets of magnitudes resulting in six different *Gaia* indices.

V.2.1.2 **VISTA Indices**

Much like the *Gaia* photometry, we have both the Fourier/template-fit and the unphased VISTA NIR magnitudes that were discussed in Sect. IV.1.1.3. Thus, just like the *Gaia* indices, we computed two sets of PW relations for the three VISTA indices listed in Table V.1. The ones that are based on the phase-corrected magnitudes are suffixed with 'fit'.

All in all, we constructed 25 different Wesenheit indices for all the stars in our catalog.

V.3   Calibrators

We employed two independent calibrators for computing our PW relations, namely, the Large Magellanic Cloud and the *Gaia* DR3 parallax sample.

V.3.1   LMC members

The LMC members of our catalog were selected based on the methodology presented in Sect. IV.3. The distance to these objects was adopted as the one computed by Pietrzyński et al. (2019) using detached eclipsing binaries, i.e., $d_{\mathrm{LMC}} = 49.591 \pm 0.616$ kpc. Alternatively, the distance modulus of the LMC (μ_{LMC}) = 18.477 ± 0.026 mag. We used this distance modulus, along with the apparent Wesenheit index (referred to as W), to compute the absolute Wesenheit W^{abs} as:

$$W^{\mathrm{abs}} = W - \mu_{\mathrm{LMC}}. \tag{V.3}$$

Furthermore, for the LMC DCEP sample, we used the geometric model from van der Marel & Cioni (2001), Jacyszyn-Dobrzeniecka et al. (2017) to correct for the inclination of the LMC disk. The inclination and the angle of the line of nodes of the LMC disk were adopted from Ripepi et al. (2022). Due to the dispersed (and symmetric) 3-d distribution of the other subtypes, this model was not applied to their photometry.

V.3.2   Parallax

For stars in the Milky Way, we leveraged their *Gaia* DR3 astrometry. Specifically, following Arenou & Luri (1999) and Luri et al. (2018), we computed the astrometry-based luminosity (ABL) of the stars using their *Gaia* parallax. ABL is defined as:

$$\mathrm{ABL} = \varpi \cdot 10^{0.2\,W - 2}, \tag{V.4}$$

where ϖ is the parallax (in mas). While *Gaia* (specifically its DR3) presents significant advancements in the computation of stellar parallaxes, the accuracy of these parallaxes dwindles quickly beyond 5 kpc (see, for e.g., Gaia Collaboration et al., 2023c; Bailer-Jones et al., 2021). Moreover, some stars with a good astrometric solution in *Gaia* can contain negative parallaxes (Lindegren et al., 2021b). This necessitates the use of the ABL formalism when dealing with *Gaia* parallaxes. An additional advantage of this formalism is that the ABL is a linear function of the parallax and thus, the Gaussian uncertainty on the parallax can be propagated conveniently (Luri et al., 2018).

We used this ABL as a proxy for the absolute Wesenheit for the Milky Way stars. Moreover, we limited our sample to only those stars with GAIA_FLAG< 3 (see II.2), which ensured parallaxes of only astrometrically well-behaved stars were used.

V.4   Summarizing the combinations

Collectively, the ingredients described in the preceding sections allowed us to compute $25 \times 2 = 50$ different PW relations for each subtype in our catalog. We derived these PW relations
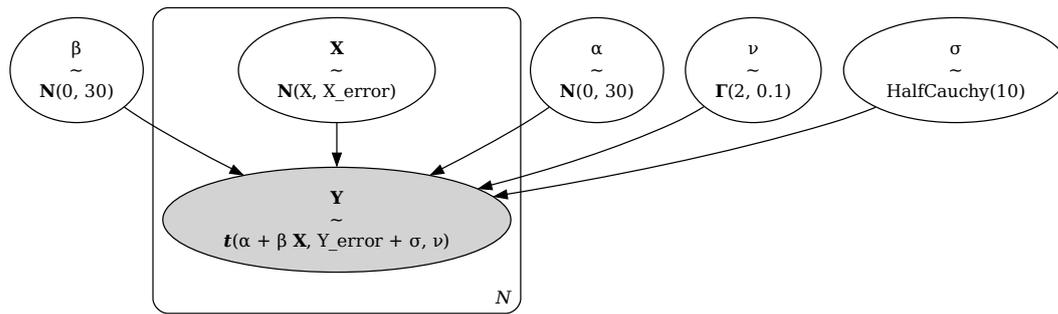
FIGURE V.2: The Bayesian regression model used to compute the PW relations. Here, X is the predictor variable, i.e, the decadic logarithm of the period and X_error is its measurement uncertainty. Similarly, Y and Y_error are the absolute Wesenheit and its uncertainty. N is the number of stars and **N** denotes a normal distribution. The other model parameters are defined in Sect. V.5.

only for the following single-mode pulsators: DCEP_F, DCEP_1O, RRab, RRc, (BL Her + W Vir), all Type-II Cepheids (i.e, BL Her + W Vir + RV Tau), ACEP_F, and ACEP_1O. The inclusion of RV Tau stars in the PL/PW relations of Type-II Cepheids is still a matter of debate (see, for instance Bhardwaj, 2020, and references therein), we decided to fit two relations for Type-II Cepheids: one with RV,Tau stars and one without. Unfortunately, the paucity of RV Tau stars precludes the fitting of a high-confidence PW relation on their sample alone (the 'high-confidence' regime is defined in the next section).

All in all, we could ideally fit 25 (Windices) $\times$ 2 (calibrators) $\times$ 8 (subtypes) = 400 PW relations. However, based on the availability of the data, it was only possible to compute 220 PW relations out of these. The remaining 180 combinations either cannot be computed as certain surveys like those based in the Northern hemisphere did not observe the LMC, or the data was available but only for fewer than ten stars of a certain subtype.

192 relations out of these 220 PW relations were classified as high-confidence and are presented (and validated) next.

V.5   Model Fitting

For each of the PW relations, we used the absolute Wesenheit index, W^{abs}, as the predicted (dependent) variable, and the period as the predictor (independent) variable in the model presented in Fig. V.2. Here, we expound on the intricacies of this Bayesian model.

A period-Wesenheit relation is formulated as:

$$W^{\text{abs}} = \alpha + \beta \cdot \log_{10}(P), \qquad\qquad (\text{V.5})$$

where P is the period (in days), and α and β are the intercept and the slope of the relation, respectively. To fold in the measurement uncertainties on the W^{abs} and the period, and to be immune to outliers, we assumed

$$W^{\text{abs}} \sim \boldsymbol{t}(\alpha + \beta \cdot \log_{10}(P), \sigma, \nu). \qquad\qquad (\text{V.6})$$

Here, $\boldsymbol{t}$ is a Student's t-distribution centered on the measured value of the absolute Wesenheit. σ is the standard deviation of this distribution and it quantifies the intrinsic scatter (dispersion) of a given PW relation. The parameter ν is the degrees of freedom of the distribution and governs its shape. A ν value $\gtrsim 30$ results in an approximation of the Gaussian distribution. However, in the presence of outliers, ν can be ~ 1. In this case, the tails of a Student's t-distribution are quite heavy. Therefore, for a given mean value, more weight can be allotted to the outliers just by lowering the value of the ν parameter. This ensures that the model is not biased by the contaminants.

This property of our model is pivotal in ensuring bias-free PW relations as both our calibrators carried varying degrees of imperfections. For instance, the LMC sample contained line-of-sight interlopers of almost all subtypes, but especially so of RRab- and RRc-subtypes. More importantly, the *Gaia* parallax sample was riddled with large uncertainties and many outliers. As proven below (in Sect. V.6), our deployed model overcame these challenges and derived derive a large number of highly precise and accurate PW relations.

Equation V.6 constitutes the likelihood distribution of our Bayesian model. As illustrated in Fig. V.2, the priors for both α and β were set to uninformative normal distributions with a mean of 0 and standard deviation = 30. The prior for ν was set to follow a $\Gamma(2, 0.1)$ distribution following Juárez & Steel (2010), while the same for the intrinsic scatter followed a half-Cauchy distribution with scale parameter set to 10 (Gelman, 2006).

The posterior distributions were sampled using an extension of the Hamiltonian Monte Carlo algorithm (HMC; Duane et al. 1987), i.e., the No-U-Turn Sampler (NUTS; Hoffman & Gelman 2011). The commonly-used Markov Chain Monte Carlo (MCMC) sampler relies on the Metropolis-Hastings algorithm to sample the posterior phase space. The Hamiltonian Monte Carlo sampler sidesteps the time-consuming random walk behavior of the Metropolis-Hastings algorithm by computing the gradients of the distribution. These gradients are equivalent to the momenta of a physical system that constitute its time-invariant Hamiltonian (hence the name). These first-order gradients allow the HMC sampler to quickly converge on the target distribution, even in the presence of many dimensions and multi-modality. The HMC algorithm is parameterized by two quantities: (i) the size of the

sampling step (jump-size between one sample point and the next) and (ii) the total number of such steps. The NUTS extension of the HMC auto-tunes these parameters at practically negligible overhead cost. For an in-depth explanation of this sampler, refer Hoffman & Gelman (2011).

We sampled the posterior space for long enough to ensure a high effective sample size (> 1000 generally, c.f. Sorensen et al. 1995; Gelman et al. 2015) in both the central and the tail regions of the distributions (Vehtari et al., 2019), thus ensuring a high-confidence uncertainty determination of model parameters.

While we visually checked the convergence of the MCMC chains only for a subset of the models, we did set a pretty restrictive limit on the rank-normalized Gelman-Rubin statistic $\hat{R}$ (Gelman, 2006; Vehtari et al., 2019). We resampled the posteriors with different initial values, and with longer burn-in, until the model achieved $\hat{R} \leq 1.005$ for every parameter. This ensured that the MCMC chains were mixed well and the posterior phase space was explored efficiently. Finally, we included only those relations with the Monte Carlo standard error (MCSE, Kruschke 2015) ≤ 0.005 for both α and β, thereby retaining PL/PW laws from models with reliable posterior estimates. MCSE quantifies the uncertainty induced due to a discrete sampling of the continuous posterior distribution and this threshold ensures that the tails of posterior distribution were sampled effectively.

From the combinations discussed earlier, we present 192 PW relations and their model parameters here and in Appendices B and C. These relations satisfied all the aforementioned criteria and were therefore, classified as 'high-confidence' relations. The PW relation of the LMC DCEP_F stars based on $W_{G,\text{BP,RP}}^{\text{abs}}$ is plotted in Fig. V.3. The black solid line denotes the relation and the gray shaded region represented the 96% confidence interval. The narrowness of this interval underlines the precision of the PW relation as it is quantified by the uncertainty of the model parameters. These parameters, along with the number of stars used in the fit, and the Wesenheit coefficient used to construct the corresponding W^{abs}, are displayed on each PW plot. This annotated information facilitates an easy application of a particular law. The similar LMC-based P-$W_{G,\text{BP,RP}}^{\text{abs}}$ plots for the other subtypes are presented in Figures V.4–V.10. For a systematic comparison of different PW relations of the same subtype, their model parameters are listed in Table V.2 for the LMC DCEP_F sample (and in Table V.3 for the Milky Way DCEP_F sample). Similarly, the LMC-based relations of the other subtypes are plotted in Appendix B: Sect. 2.1, and their model parameters are listed in Appendix C.

The PW relations of the Milky Way stars (based on *Gaia* parallax) are presented in Appendix B: Sect. 2.2 and their parameters are aggregated in Appendix C.
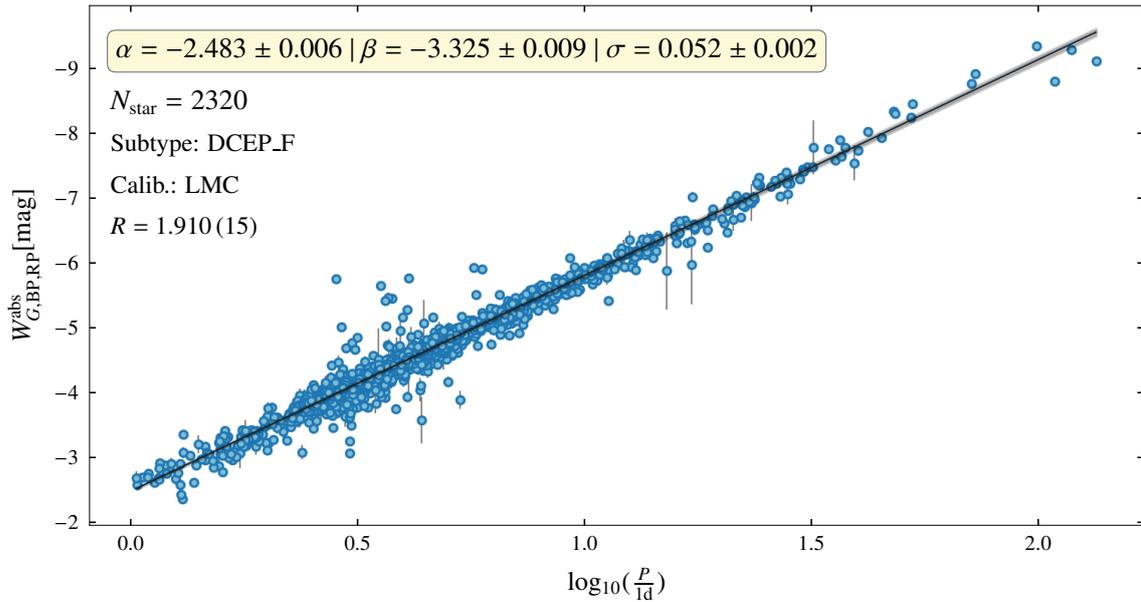
FIGURE V.3: Period-Wesenheit relation of the LMC classical Cepheids pulsating in the fundamental mode (DCEP_F). The *x*-axis is the decadic logarithm of the variability period (in days). The *y*-axis represents the absolute Wesenheit index constructed using the Fourier-fit intensity-averaged *Gaia* magnitudes computed by the SOS Cep&RRL pipeline (Ripepi et al., 2023; Clementini et al., 2023). The distance modulus of the LMC was adopted as the one computed by Pietrzyński et al. (2019): 18.477 ± 0.026 mag. The model parameters (and their uncertainties) and the Wesenheit coefficient, *R*, used to construct the Wesenheit index are displayed on the top. The uncertainty ($\times 1000$) on the Wesenheit coefficient is given in parentheses.
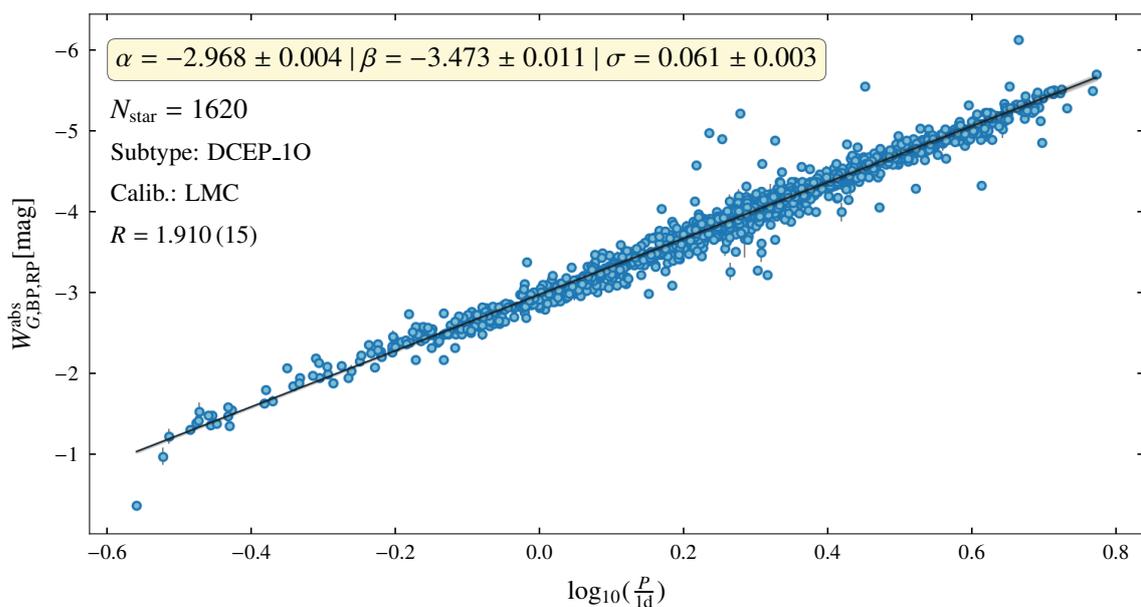
FIGURE V.4: Same as Fig. V.4 but DCEP_1O stars.
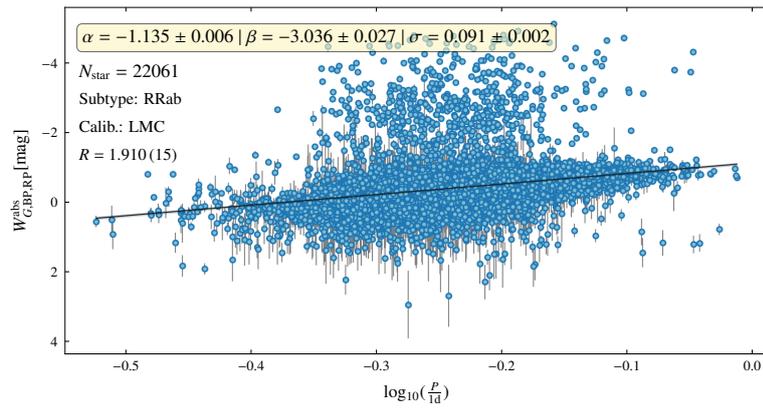
FIGURE V.5: Same as Fig. V.3 but for RRab-type stars.
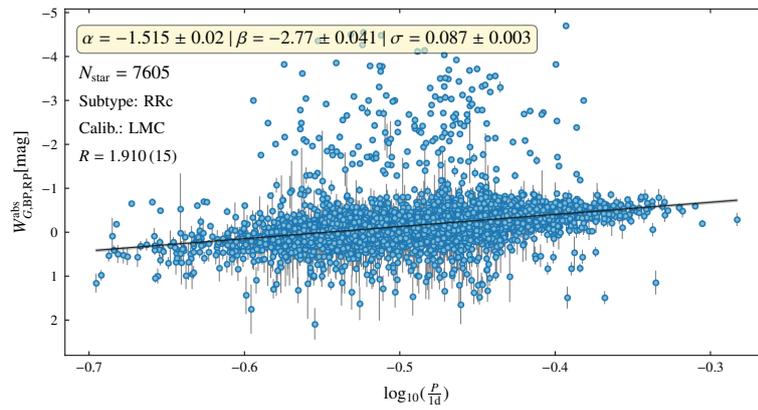
FIGURE V.6: Same as Fig. V.3 but for RRc-type stars.
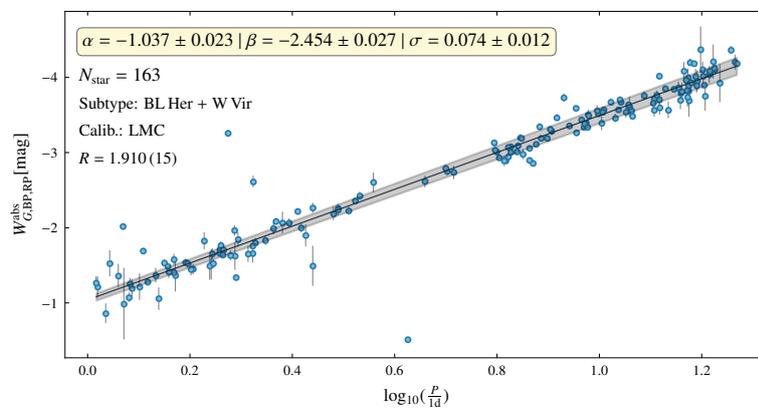
FIGURE V.7: Same as Fig. V.3 but for BL Her and W Vir stars.
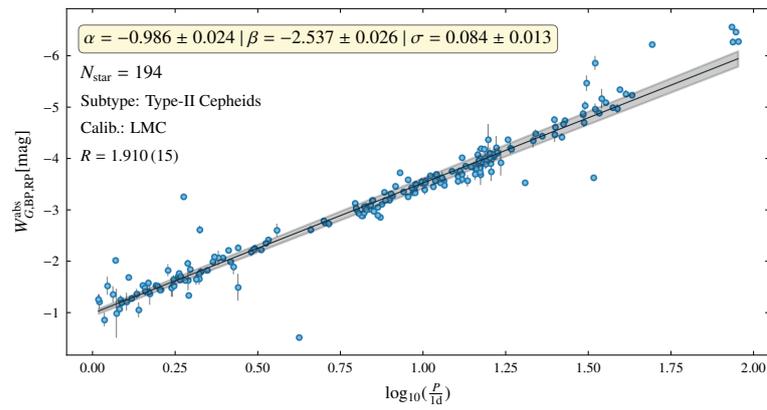
FIGURE V.8: Same as Fig. V.3 but for Type-II Cepheids.
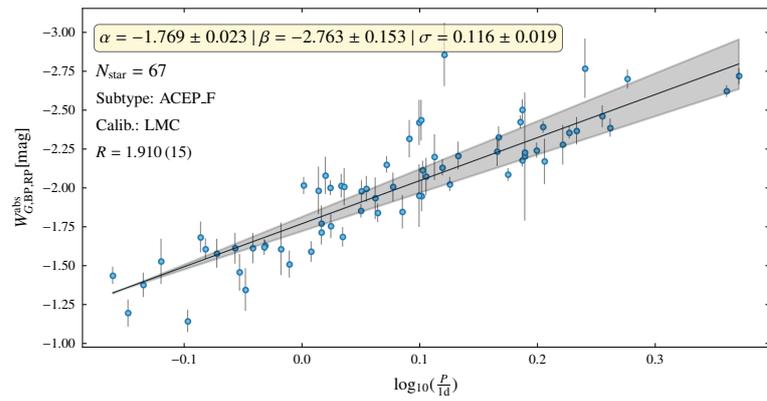
FIGURE V.9: Same as Fig. V.3 but for ACEP_F stars.
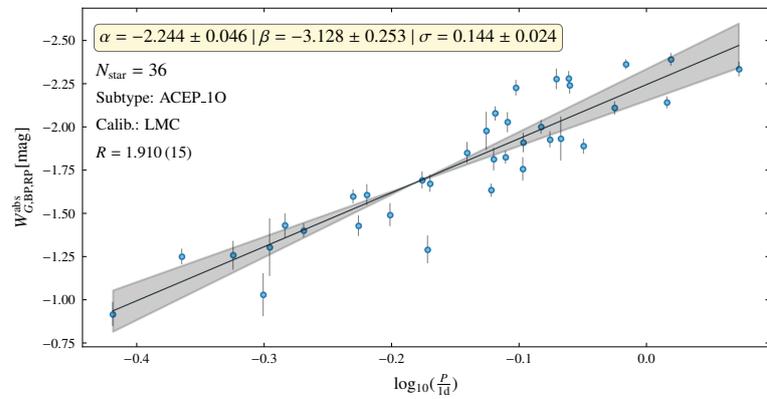
FIGURE V.10: Same as Fig. V.3 but for ACEP_1O stars.

TABLE V.2: Model parameters of the LMC DCEP_F PW relations.

| W	α	β	σ	N_{star}
$W_{G,\text{BP,RP}}$	-2.483 ± 0.006	-3.325 ± 0.009	0.052 ± 0.002	2320
$W_{G,G,\text{RP}}$	-2.938 ± 0.007	-3.305 ± 0.011	0.06 ± 0.003	2322
$W_{G,\text{BP,RP}}$ (mean_x_mag)	-2.506 ± 0.006	-3.321 ± 0.009	0.069 ± 0.002	2443
$W_{G,G,\text{RP}}$ (mean_x_mag)	-2.989 ± 0.008	-3.274 ± 0.012	0.093 ± 0.003	2443
$W_{G,\text{BP,RP}}$ (phot_x_mean_mag)	-2.497 ± 0.009	-3.343 ± 0.013	0.004 ± 0.003	2471
$W_{G,G,\text{RP}}$ (phot_x_mean_mag)	-3.009 ± 0.016	-3.273 ± 0.023	0.006 ± 0.005	2472
$W_{I,V,I}$ (OGLE)	-2.565 ± 0.006	-3.299 ± 0.009	0.055 ± 0.002	2312
$W_{Ks,J,Ks}$ (2MASS)	-2.626 ± 0.012	-3.223 ± 0.018	0.146 ± 0.004	1973
$W_{H,J,Ks}$ (2MASS)	-2.647 ± 0.012	-3.262 ± 0.017	0.123 ± 0.003	1956
$W_{Ks,J,Ks}$ (VISTA)	-2.599 ± 0.006	-3.268 ± 0.01	0.065 ± 0.002	2035
$W_{J,Y,Ks}$ (VISTA)	-2.623 ± 0.008	-3.239 ± 0.013	0.09 ± 0.003	2023
$W_{Ks,J,Ks}$ (VISTA; fit)	-2.597 ± 0.006	-3.283 ± 0.009	0.059 ± 0.002	2290
$W_{J,Y,Ks}$ (VISTA; fit)	-2.611 ± 0.008	-3.276 ± 0.012	0.078 ± 0.002	2263
$W_{\text{W2,W1,W2}}$ (unWISE)	-2.433 ± 0.014	-3.2 ± 0.02	0.159 ± 0.006	2107

TABLE V.3: Model parameters of the MW DCEP_F PW relations.

W	α	β	σ	N_{star}
$W_{G,\text{BP,RP}}$	-2.307 ± 0.044	-3.696 ± 0.058	0.009 ± 0.001	1139
$W_{G,G,\text{RP}}$	-2.744 ± 0.06	-3.011 ± 0.083	0.017 ± 0.001	1139
$W_{G,\text{BP,RP}}$ (mean_x_mag)	-2.341 ± 0.043	-3.673 ± 0.056	0.009 ± 0.001	1166
$W_{G,G,\text{RP}}$ (mean_x_mag)	-2.788 ± 0.058	-2.981 ± 0.08	0.016 ± 0.001	1168
$W_{G,\text{BP,RP}}$ (phot_x_mean_mag)	-2.366 ± 0.038	-3.642 ± 0.05	0.009 ± 0.001	1175
$W_{G,G,\text{RP}}$ (phot_x_mean_mag)	-2.76 ± 0.056	-3.031 ± 0.077	0.016 ± 0.001	1177
$W_{I,V,I}$ (OGLE)	-2.364 ± 0.088	-3.553 ± 0.135	0.016 ± 0.001	357
$W_{r,g,r}$ (ZTF)	-1.248 ± 0.094	-3.508 ± 0.14	0.026 ± 0.003	310
$W_{Ks,J,Ks}$ (2MASS)	-2.347 ± 0.064	-3.308 ± 0.097	0.012 ± 0.001	638
$W_{H,J,Ks}$ (2MASS)	-2.372 ± 0.068	-3.277 ± 0.106	0.012 ± 0.001	591
$W_{Ks,J,Ks}$ (VISTA)	-3.781 ± 0.216	0.598 ± 0.254	0.066 ± 0.009	146
$W_{\text{W2,W1,W2}}$ (unWISE)	-2.291 ± 0.109	-2.904 ± 0.191	0.017 ± 0.003	317

V.6   Validation of Distances

V.6.1   Computing the distance modulus

By combining Equations V.2, V.3, and V.5, we computed the distance modulus to our stars as:

$$\mu = W^{\text{abs}} - W = \alpha + \beta \cdot \log_{10}(P) - W. \tag{V.7}$$

The uncertainties on each of these parameters along with the intrinsic scatter of the PW relation (σ) were used to construct a common covariance matrix. These sources of uncertainties (and their nature, i.e., statistical or systematic) can be summarized as:

1. Uncertainty on α (syst), β (stat)

2. $\sigma / \sqrt{N_{\text{star}}}$ (syst; Riess, A. G., in priv comm.)

3. Uncertainty on the period (stat; estimated in Chapter II)

4. Uncertainty on the mean LMC distance modulus from Pietrzyński et al. (2019, syst)

5. Uncertainty on the apparent Wesenheit W, which included:

   (i) Uncertainty on the mean magnitudes (stat; see V.1)

   (ii) Uncertainty on the Wesenheit coefficient and the reddening law (stat; see Sect. V.2)

Using this covariance matrix and the mean vector of Eqn. V.7 as the parameters of a multivariate normal distribution, we generated 10^5 realizations of μ per star. The mean and standard deviation of these Monte Carlo samples were assumed as the μ and σ_μ for a given star. This process resulted in (say) 50 different distance modulus estimates per star. The true number for many subtypes is less than 50 due to the reasons discussed earlier.

V.6.2   Validation method

In the following validation, for the sake of clarity and brevity, these distance modulus estimates are referred to as distance estimates. Nevertheless, we performed the validation on the distance modulus itself and not the distance, as the uncertainties are Gaussian in nature only for the former.

We undertook a preliminary internal validation of these distance estimates, by computing the mean distance of the stars in our LMC sample and comparing it with that from Pietrzyński et al. (2019). To compute the mean distance of the LMC sample, we employed a Gaussian Mixture Model (GMM) based on the extreme deconvolution algorithm (XD, Bovy et al.

2011). This XD-GMM model estimates the underlying 'true' density of a distribution, even in the presence of data generated from one primary data source and multiple other noise sources. It affords a complete treatment to data with heteroskedastic uncertainty covariance matrices, which is the case for our distance estimates.

The number of components (Gaussian distributions) needed to be fit is the main parameter of such an XD-GMM model. For a subset of our PW relations, we performed 3-fold cross-validation to determine the optimal number of components between 1 and 5. We also examined the Bayesian Information Criterion (BIC) for each run of the cross-validation. Invariably, the best cross-validation score (or the lowest BIC) was attained by the model with 2 or 3 components, with most of the weight ($\gtrsim 85\%$) attributed to a single component. Moreover, the secondary distributions in all cases were flat, or in other words, had large standard deviations. This was unsurprising as the selection of our LMC sample was quite thorough. Furthermore, the origin of the additional distributions can be explained by taking into account the following noise sources: (i) even after weeding out possible interlopers using *Gaia* astrometry, some stars in the line-of-sight of the LMC (and not true members of the LMC) persisted. (ii) The photometry of a sample of stars towards the crowded central regions of the LMC was saturated or affected by blending (e.g, Jacyszyn-Dobrzeniecka et al., 2016, 2017). Hence, these stars could also contribute to the secondary 'noisy' distributions.

Regardless, for every set of distance estimates (i.e, every PW relation), we fit 5 identical XD GMM models each, with the number of components equal to 2 and 3. The final distance estimate of the LMC was computed as the median of the means of the primary distributions of each XD-GMM run. The uncertainty on this final determination was computed as the median of the uncertainty on the distance of each star in the respective sample. This mean estimate and its corresponding uncertainty are compared with the 'true' LMC distance in Figures V.11–V.18 for each subtype and for all their LMC-based PW relations.

V.6.3 Discussion

As expected, the PW relations based on Fourier-/template-fit mean magnitudes led to most accurate estimations of the LMC distance. Nonetheless, even the PW relations based on unphased mean magnitudes (computed as a simple mean of randomly-timed observations), we were able to derive them with high accuracy. For instance, the PW relations based on the OGLE and *Gaia* SOS Fourier-fit magnitudes resulted in a $\sim 99\%$ accurate and better than 2% precise distance estimate to the LMC.

However, even when using the biased `gaia_source` magnitudes (see Sect. II.2), we were able to achieve a distance that was $\sim 97\%$ accurate, though less precise. The loss of precision is by design and necessary. To account for the bias in these magnitudes (as they
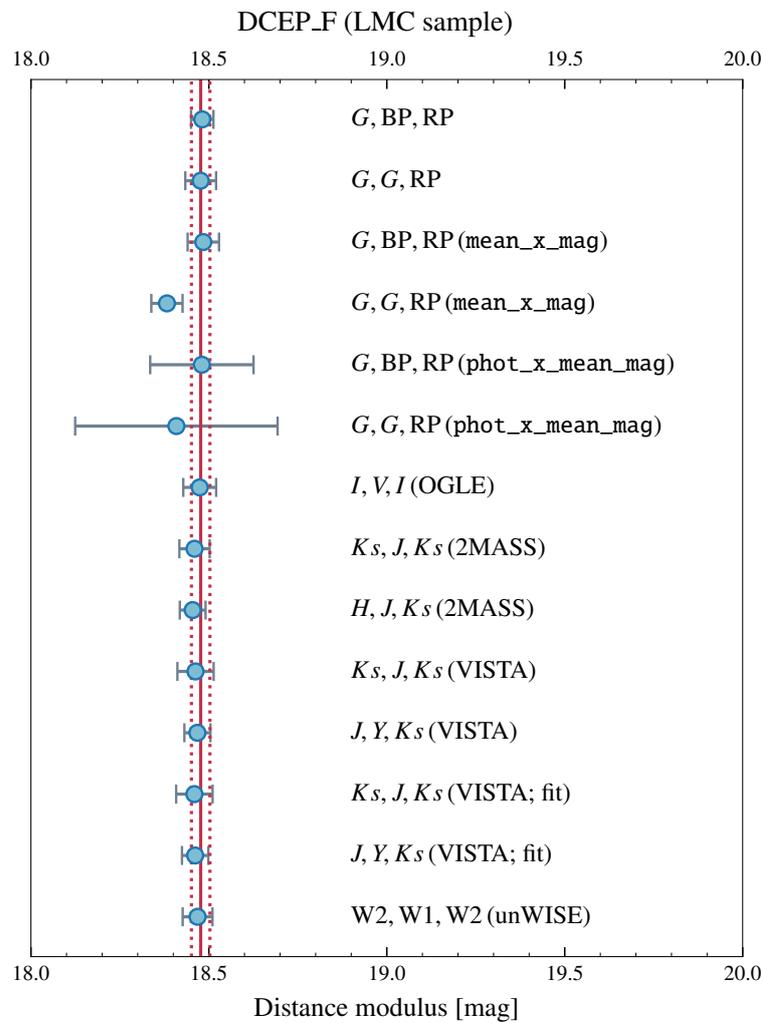
FIGURE V.11: Internal validation of the LMC DCEP_F PW relations. The three magnitudes listed for each comparison were used to construct the Wesenheit index for the respective relation. The solid and dotted red lines represent the "true" distance modulus (18.477 mag) and its uncertainty (0.026 mag) as computed by Pietrzyński et al. (2019). respectively.
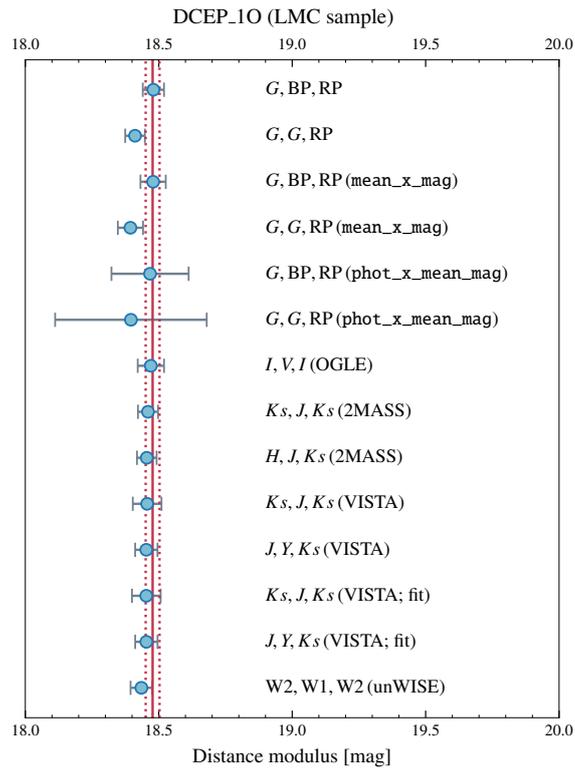
FIGURE V.12: Same as Fig. V.11 but for DCEP_1O stars.
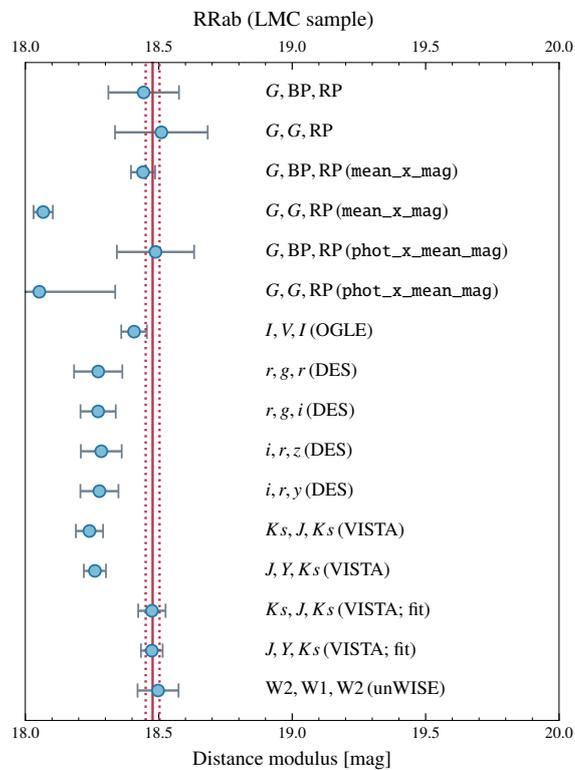
FIGURE V.13: Same as Fig. V.11 but for RRab stars.
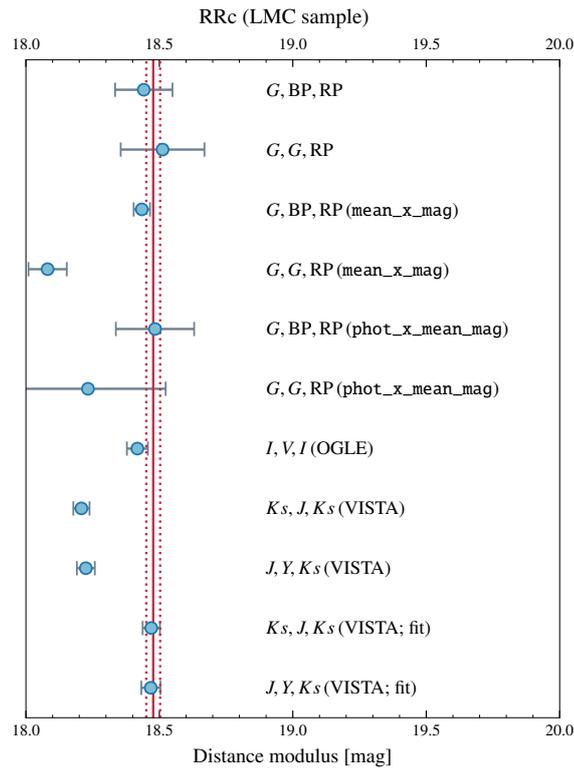
FIGURE V.14: Same as Fig. V.11 but for RRc stars.
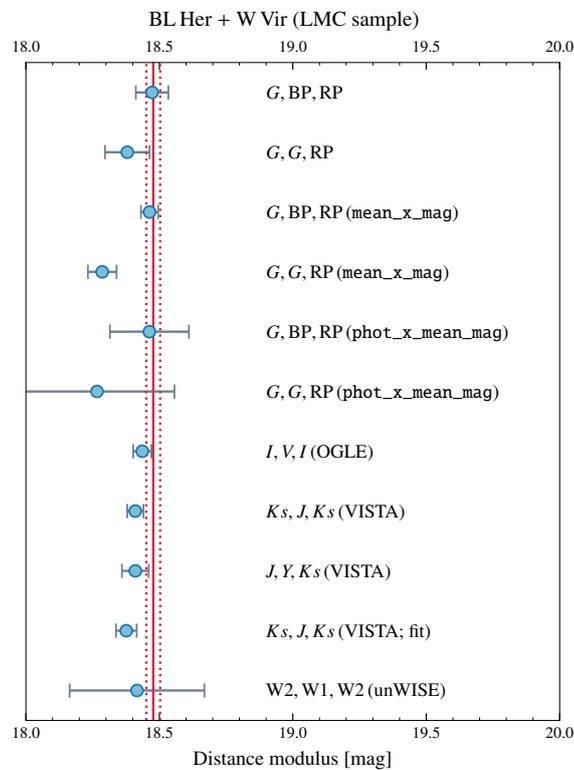
FIGURE V.15: Same as Fig. V.11 but for BL Her and W Vir stars.

Type-II Cepheids (LMC sample)
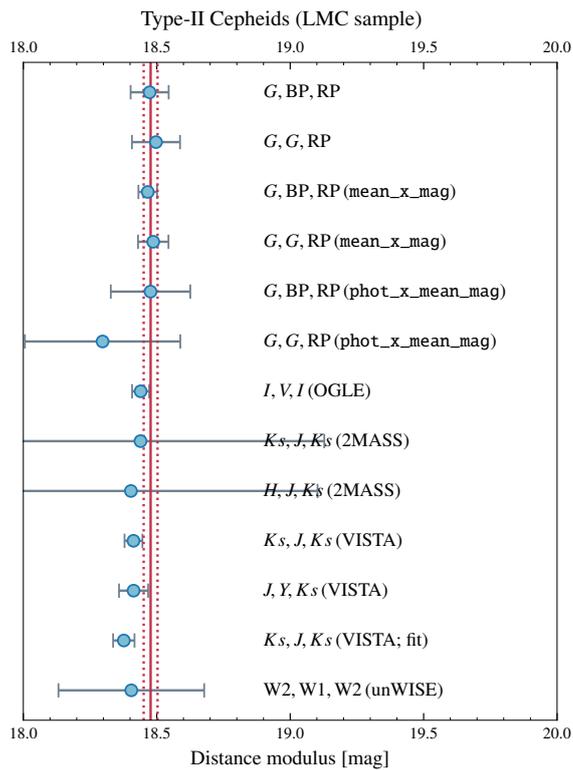
FIGURE V.16: Same as Fig. V.11 but for all Type-II Cepheids, i.e.,: BL Her, W Vir, and RV Tau stars.

ACEP_F (LMC sample)
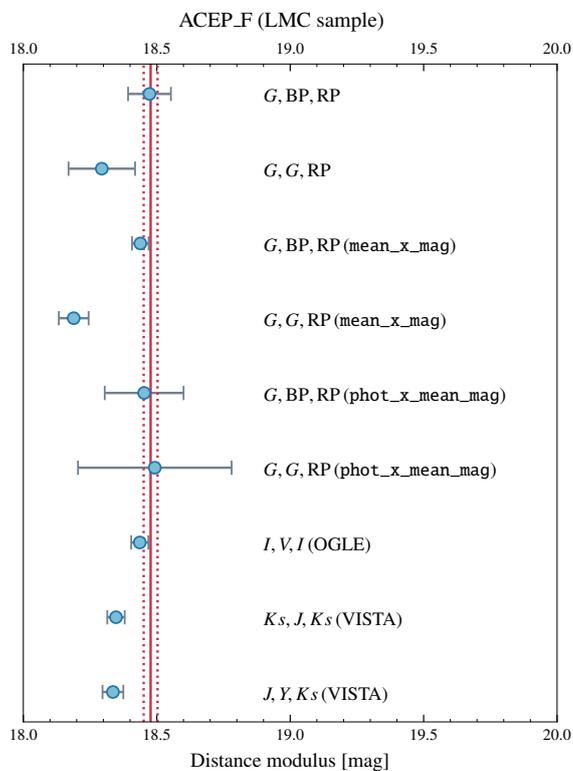
FIGURE V.17: Same as Fig. V.11 but for ACEP_F stars.
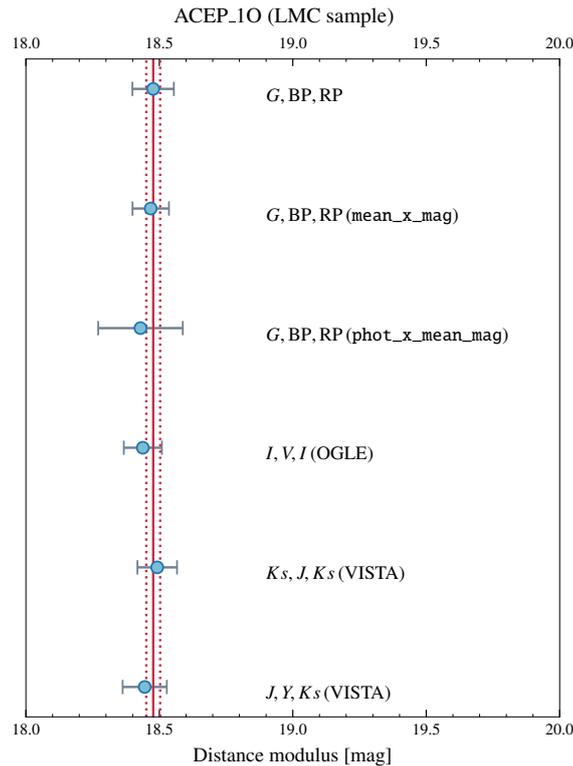
FIGURE V.18: Same as Fig. V.11 but for ACEP_1O stars.

were weighted with their uncertainties), we had added a systematic error of 0.05 mag for each star. Our Bayesian model was able to find an accurate fit by taking this inflated uncertainty into account.

We also highlight the importance of constructing a Wesenheit index using the correct formalism. The performance of the PW relations based on the Wesenheit defined as $W = G - R \cdot (G - RP)$ is systematically worse than that of the correctly-defined $W_{G,\mathrm{BP,RP}}$. This inaccuracy is also directly proportional to the intrinsic dispersion of the fits. For instance, for the DCEP_F relations (see Fig. V.11), the LMC distance modulus based on $W_{G,G,\mathrm{RP}}$ index is quite accurate. However, for the Type-II (or anomalous) Cepheids, this estimates is more than 1σ away from the expected 'true' value. Moreover, the performance of $W_{G,G,\mathrm{RP}}$ is progressively worse for the `mean_x_mag` and `phot_x_mean_mag` photometry, even more so than that for $W_{G,\mathrm{BP,RP}}$ index.

Due to their high intrinsic brightness and their concentrated location in the disk of the LMC, the DCEP_F PW relations resulted in the most accurate and precise estimates of the LMC distance. We have computed independent and almost just as accurate relations for the DCEP_1O sample (see Fig. V.12 and Tables C.1 and C.2). These relations can form the basis for a more active inclusion of DCEP_1O stars in the extragalactic distance scale.

The LMC RRab PW relations based on the DES Wesenheit indices grossly underestimate the distance to the LMC (see Fig. V.13 and V.14). This shortcoming can be explained by

considering the fact that DES only observes the exterior regions of the LMC. The DES LMC sample only consists ∼ 700 RRab stars, a good number of them might belong to the Milky Way anyway. Nevertheless, the Milky Way RRab $grizy$ PW relations based on the ZTF and PS1 data agree quite well with each other. Moreover, our relations represent the first-ever computation of RR Lyrae PW relations that also include the y-band (see Tables C.3– C.6).

The RRab PW relations based on the VISTA photometry in the LMC also represent a breaking point of our Bayesian model. As the fraction of outlier in this relation (see for e.g., Fig. B.49) is ≥ 30%, the contaminants bias the relation to lower distances. As a result, the VISTA-based LMC distance modulus estimates are underestimated (see Fig. V.13).

We detect a marginal difference in both the slope and the intercept of the two sets of Type-II Cepheids PW relations that we computed. The ones without RV Tau stars are validated in Fig. V.15 and listed in Tables C.7 and C.8. And the relations including RV Tau stars are validated in Fig. V.16 and their model parameters are given in Tables C.9 and C.10. These differences are usually smaller than the average intrinsic scatter of their PW relations.

Due to the paltry number of the anomalous Cepheids and the lack of their high S/N *Gaia* parallax, no Milky Way ACEP_F or ACEP_1O PW relation was classified as high-confidence by our criteria. Furthermore, owing to their evolutionary stage, the anomalous Cepheids are not bound by a well-defined mass-luminosity relation (Caputo et al., 2004; Ripepi et al., 2014). As a result, their PW relations are generally more dispersed. Nevertheless, their 'high-confidence' PW relations are listed in Tables C.11 and C.12 ad a validation of their LMC relations is presented in Figures V.17 and V.18.

Overall, many of the $grizy$, NIR, and MIR PW relations presented here have been computed for the very first time.

The remarkable level of agreement between our LMC distance estimates and the geometrically-determined estimate from Pietrzyński et al. (2019) underlines the accomplishment of the primary goal of this dissertation.

VI

Summary and Future Outlook

And the day came when the risk to remain tight in a bud was more painful than the risk it took to blossom.

– ANAIS NIN

Here, I present a qualitative overview of the results achieved over the course of this dissertation. I also elaborate the applications of the catalog that will be pursued in due course.

VI.1 Summary

– We combined catalogs of RR Lyrae stars and Cepheids from eight optical photometric surveys resulting in a collection of $\sim 311\,000$ RR Lyrae stars and $\sim 18\,000$ Cepheids in the Milky Way and the Magellanic clouds. The number of stars of various subtypes are listed in Table II.4.

– We developed a new cross-match algorithm based on Bhattacharyya distance that allowed us to incorporate the astrometric covariance matrix as well as the variability periods in finding the best match of a star across surveys.

– Based on $\sim 5.6 \times 10^6$ comparisons of period estimates from all surveys against each other, we showed that for $\sim 87\%$ of stars with $P \leq 1\,\mathrm{d}$, their periods are measured

accurately within 2 s. Similarly, periods are measured within 2 min (or 2 h) for $\sim 86\%$ (66%) of stars with $1 < P \leq 20\,\mathrm{d}$ (or $P > 20\,\mathrm{d}$). We also discussed how aliased periods due to low N_{obs}, and the inconsistency in reporting the "single" vs formal periods for RV Tau stars, lead to a lower overall accuracy of longer-period variables (DCEP_F, RV Tau). Overall, we showed how sparse, multi-band light curves contribute the highest fraction of relatively lower accuracy periods.

– We conducted similar comparisons to validate the classification of stars across surveys. After using the OGLE catalog as a control sample, we visually inspected light curves of stars with inconsistent classification. By using $F_{0.5}$ as the classification accuracy metric, we estimated a classification accuracy of over 96% for all surveys. Based on the contamination rate of each survey, we anticipate that a minimum of ~ 1800 stars out of $\sim 130\,000$ unique detections across surveys may have the wrong subtype.

– We validated these estimates by comparing our catalog against literature catalogs of similar subtypes and periods from *TESS* and *Kepler* light curves. We found that the classification accuracy is $\sim 98\%$ for Galactic classical Cepheids and $> 99\%$ for RR Lyrae stars in the SDSS Stripe 82 region. While this collection is essentially complete with respect to the literature, we did not estimate its absolute completeness in the Milky Way with respect to simulations, nor did we attempt to model the convolved selection function of these surveys.

– We established quality control criteria for *Gaia* DR3 astrometry and photometry tailored to RR Lyrae stars and Cepheids and flagged $\sim 40\,000$ stars with high astrometric noise and contaminated BP/RP photometry. We also flagged $\sim 10\,000$ stars that are detected as variable stars of other subtypes in literature catalogs. Additionally, we removed ~ 1000 stars that are highly likely to be one of ellipsoidal binaries, spotted stars, or non-stellar objects (QSOs, variable AGNs, etc.). Specifically, we found that a small number of RR Lyrae stars classified using sparse, multi-band light curves are likely mistaken non-stellar objects.

– In light of applications to LSST and JWST, we cross-matched our catalog against optical and IR surveys to obtain photometry in $ugrizy$, $JHKs$, and W1, W2 bands. We retrieved epoch line-of-sight velocity estimates from nine spectroscopic surveys and discussed biases in the average velocities by comparing them against barycentric velocities. Additionally, we assimilated high-resolution spectroscopic [Fe/H] estimates for ~ 8300 stars in our catalog.

– We outlined an extensive procedure to select the LMC and SMC population. As part of this algorithm, we introduced the modifications we made to the MCD estimator to

make it more robust by incorporating measurement uncertainties into it.

– We presented a detailed report on the computation of the uncertainties on mean magnitudes, in the limit of low number of random-phase observations.

– We discussed the construction of a Bayesian regression model that is immune to outliers and elaborated the criteria we employed to ensure a meaningful determination of model parameters.

– We computed 192 new period-Wesenheit relations for these stars under this probabilistic framework, resulting in ~ 50 distance estimates per star. Many of these relations have been computed for the very first time.

– We used the distance to the LMC sample to validate these PW relations, wherein the mean distance modulus of the LMC was computed using a suite of cross-validated extreme deconvolution-based Gaussian mixture models.

VI.2 Future Outlook

This collection can be used as a high-quality training sample for supervised classification of RR Lyrae stars and Cepheids. The catalog has also been selected to be used during the commissioning of the LSST data (Ivezić, Ž., in priv. comm.).

Moreover, we envision numerous use cases of this catalog, a subset of which we have already undertaken and will be presented in upcoming works of the series RRLCep: A catalog of RR Lyrae stars and Cepheids. These are:

– In "Paper III. The effect of metallicity on PL/PW laws of Cepheids and RR Lyrae stars", we study the metallicity dependence of the PL/PW relations for our target subtypes. The context for this analysis is: Groenewegen & Lub (2023), based on spectral energy distributions and the bolometric luminosities of classical Cepheids, predict zero to marginal dependence of the DCEP PL laws on metallicity (i.e., the slope of the metallicity term $\gamma \sim 0$). However, Breuval et al. (2021) predicted a wavelength-dependent non-zero metallicity dependence. Furthermore, Breuval et al. (2022) predict a wavelength-independent non-zero γ. To systematically test this in numerous bands, we have collected Cepheid iron-abundances from the Hocdé et al. (2023) compilation of literature high-resolution spectroscopic [Fe/H] estimates. Similarly, Crestani et al. (2021) released high-resolution spectroscopic metallicity estimates of over 10 000 RRab and RRc stars in the MW and the Magellanic Clouds. The metallicity dependence of RRL PL/PW is well-known (Catelan & Smith, 2015) but

calibrated in only a few bands. We will present new calibrations of this effect in over 20 bands and 10 unique Wesenheit indices.

– For Type-II and Anomalous Cepheids, the magnitude of the γ term is usually accepted to be quite small if not zero. However, no such result is based on high-resolution spectroscopic estimates of individual stellar metallicities as these have been measured only for a few subtypes of these Cepheids. We have collected the first-ever high-resolution spectra of ~ 100 Milky Way Type-II and anomalous Cepheids. The iron-abundances of this sample will be used to robustly estimate γ for these subtypes.

– Under the same Bayesian formalism presented here, we have computed new photometric metallicity relations for RR Lyrae stars using the comprehensive collection of [Fe/H] estimates from Crestani et al. (2021) and the globular cluster sample that was described in Chapter IV. The distances from the current version of the catalog have been used to identify members stars of Milky Way stellar streams, while their new photometric [Fe/H] estimates have been used to characterize the substructures. These results will be presented as the third installment of our other series "Milky Way archaeology using RR Lyrae stars and Type-II Cepheids".

– With the advent of spectroscopic surveys like WEAVE (Dalton, 2016), 4MOST (de Jong et al., 2019; Ibata et al., 2023), MOONS (Cirasuolo et al., 2020), thousands of these stream RR Lyrae stars in this catalog will garner a v_{los} estimate for the first time. The new measurements, combined with the ones we have already collected (along with precise distances of these stars), will enable dynamical studies of the Galaxy covering even larger distances (Prudil et al., 2022; Han et al., 2022). Furthermore, upcoming high-resolution spectra of Type-II Cepheids and RR Lyrae stars will also facilitate chemical tagging of the many intricate substructures that constitute the Milky Way halo (Naidu et al., 2020; Bonaca et al., 2021; Prudil et al., 2021), thereby determining their origins.

– The sample of Galactic classical Cepheids in this catalog is more complete than the one we used in Lemasle et al. (2022) to trace the Milky Way spiral arms. Additionally, in the quoted analysis, we had used photometry from 5-year unWISE Coadds. However, Meisner et al. (2022) have released more precise W1, W2 magnitudes using 9-year unWISE coadds. These magnitudes will help in determining distances to a larger sample of DCEP with better precision, thereby reducing the uncertainties in our spiral arms model.

– In Pipwala et al. (in prep.), we empirically proved that Type-II Cepheids can be used as accurate distance indicators ($\geq 99\%$) in the Local Group by employing them to

compute distances to the Andromeda galaxy. These distances were determined using PL/PW relations that were trained on the LMC sample of the current catalog and followed the same methodology that was detailed in Chapter V. We will build on this affirmative result by combining Type-II Cepheids and RR Lyrae stars and computing their joint IR PL/PW relations.

– We have combined literature catalogs of RR Lyrae stars and Cepheids identified in Galactic globular clusters and dwarf spheroidal galaxies from dedicated photometric surveys (not included here). These stars will be appended to the current catalog in the future and will improve the completeness within these subsystems even further. A homogeneous analysis of cluster members will be presented in Lemasle et al. (in prep.).

– The zero-point offset in *Gaia* DR3 parallaxes has been a matter of hot debate (see discussion in e.g., Groenewegen 2023). The zero-point correction computed by Lindegren et al. (2021a) tends to over-correct the parallax for Cepheids (e.g., Riess et al., 2022) and RR Lyrae stars (see, e.g., Bhardwaj, 2022). The Milky Way PW relations presented in Chapter V enable a systematic estimation of the *Gaia* parallax zero-point offset. We will pursue this in "Paper IV. Estimating the zero-point offset in *Gaia* DR3 parallaxes."

– In my collaboration with B. Madore and W. Freedman, the contents of this catalog will be used to compute period-luminosity-color relations that provide robust individual stellar distances.

– The Bhattacharyya distance-based cross-match algorithm that was developed here will be used to cross-match LSST multi-band catalogs of the Galactic bulge regions in my post-doctoral research.

I am certain that beyond these endeavors, the catalog holds significant potential that I am yet to foresee. Hence, the entire catalog will be released publicly along with a data model to facilitate its use by the community. Over time, I anticipate that this catalog and the methods outlined here will contribute significantly to the fields of Milky Way archaeology and near-field cosmology.

A

Supplementary Information I

In this appendix, we include the supporting material for the Chapters II and III.

1.1 Light curves phased with altered periods

OGLE *I*-band light curves phased using altered periods ($P_{\text{altered}} = P - 0.0005P$).

FIGURE A.1: OGLE *I*-band light curves of the same stars included in Fig. I.3 but phased with an altered period. The periods were altered by 0.05% and they are listed in each panel as P_{altered}.

1.2   Bhattacharyya Distance

While the commonly used methods either employ the Euclidean distance or the Mahalanobis distance (like in Pineau et al. 2017; Marrese et al. 2017, 2019), we used the Bhattacharyya distance as the metric for our cross-match algorithm. In an n-dimensional space, the Euclidean distance measures the separation between two points and the Mahalanobis distance measures the separation between a point and a distribution. Analogously, the Bhattacharyya distance measures the separation between two n-d distributions characterized using their means and covariance matrices. As shown already (and repeated here for continuity), for two multivariate normal distributions $p_i = \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i); i \in [1, 2]$, the Bhattacharyya distance between them can be calculated as:

$$
\begin{aligned}
D_B(p_1, p_2) = \; & \frac{1}{8}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\
& + \frac{1}{2} \ln\left(\frac{\det(\boldsymbol{\Sigma})}{\sqrt{\det(\boldsymbol{\Sigma}_1)\,\det(\boldsymbol{\Sigma}_2)}}\right),
\end{aligned}
\tag{A.1}
$$

where det(**M**) denotes the determinant of a matrix **M**, and the joint covariance $\boldsymbol{\Sigma} = \frac{1}{2} \times (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)$.

In case the two distributions are normally distributed but with identical covariance matrices, the Bhattacharyya distance is proportional to the squared Mahalanobis distance. The squared Mahalanobis distance between a point $p_i = (x_i, y_i)$ and a bivariate distribution $q = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in 2-d space is computed as:

$$
D_M^2(p_i, q) = ([x_i, y_i]^T - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} ([x_i, y_i]^T - \boldsymbol{\mu}).
\tag{A.2}
$$

In some studies, the Mahalanobis distance is sometimes used to measure the distance between two distributions. The way to include the covariance matrix of the second distribution in Eq. A.2 is by using the joint covariance matrix. The latter is computed by taking the average of the two covariance matrices ($\boldsymbol{\Sigma} = \frac{\Sigma_1 + \Sigma_2}{2}$). However, for cases in which the systematic error between the coordinates is larger than the statistical uncertainty, D_B is better than D_M at encapsulating the 'true' distance between such a coordinate pair. As a result, in cross-matches where the propagation of coordinates to the same epoch using proper motions is not possible or when this propagation is inadequate when astrometric uncertainties are underestimated, D_B can effectively account for these inadequacies even without synthetically broadening coordinate uncertainties (see Appendix A1 of Marrese et al. 2017). Therefore, we selected D_B as the metric around which to build our algorithm.

Moreover, since Bhattacharyya distance enables the comparison of the shape of the astro-photometric ellipse, it can be a employed to match galaxy catalogs. We will discuss these

TABLE A.1: Parameters of the ellipse for each region.

Notes.

(*1*) Galactic co-ordinates of the center in J2016 epoch. (*2*) Semi-major axis of the ellipse (along *l*). (*3*) Semi-minor axis of the ellipse (along *b*).

Region	(l_0, b_0)*1* (deg)	a'*2* (deg)	b'*3* (deg)
LMC	280.46, -32.75	12	10
SMC	302.96, -43.98	8.5	9
Bulge	2, 1	22	33

applications and the difference in its performance against that of the Mahalanobis distance in a forthcoming publication.

1.3   On-sky demarcations of substructures

We distributed stars into samples of the following regions: LMC, SMC, Galactic bulge, disk, and halo. Three ellipses (with parameters listed in Table A.1) were used to categorize stars into LMC, SMC, and Galactic bulge regions. Outside of these regions, stars with $|b| \leq 20°$ were classified as disk stars, with the halo sample comprising of the remaining stars. The halo sample contains stars from many Milky Way dwarf galaxies, globular clusters, and M31 and M33 galaxies.

These categories were assigned solely on the basis of the boundaries listed here and do not imply full membership. This categorization was undertaken only to account for varying line-of-sight number density of stars in the these regions.

1.4   Confusion matrices against OGLE classification
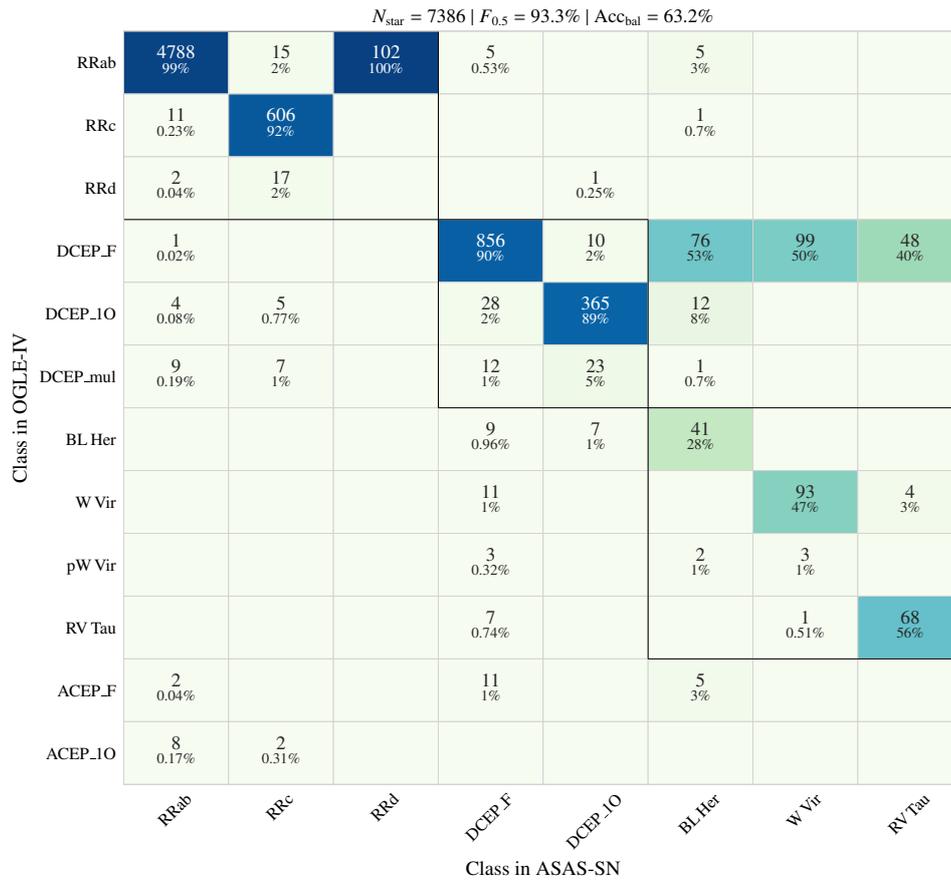
FIGURE A.2: Like Fig. III.6 but for ASAS-SN catalog.
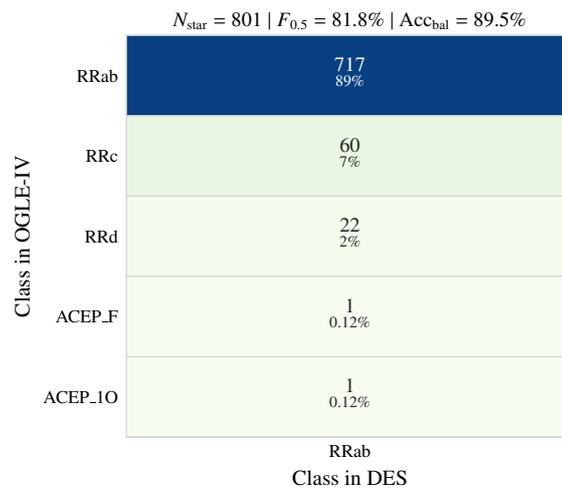
FIGURE A.3: Like Fig. III.6 but for DES.

$N_{star} = 6784 \mid F_{0.5} = 99.0\% \mid Acc_{bal} = 99.2\%$

Class in OGLE-IV	RRab	RRc
RRab	4790 / 99%	9 / 0.46%
RRc	12 / 0.25%	1946 / 99%
RRd	4 / 0.08%	7 / 0.36%
DCEP_1O		2 / 0.1%
BL Her	8 / 0.17%	
ACEP_F	3 / 0.06%	
ACEP_1O	2 / 0.04%	1 / 0.05%

Class in PS1

FIGURE A.4: Like Fig. III.6 but for PS1.

$N_{star} = 5504 \mid F_{0.5} = 98.5\% \mid Acc_{bal} = 98.1\%$

Class in OGLE-IV	RRab	RRc	DCEP_F	DCEP_1O	DCEP_mul	BL Her	W Vir	RV Tau
RRab	3686 / 99%	6 / 0.43%				1 / 3%	1 / 4%	
RRc	2 / 0.05%	1366 / 97%						
RRd	9 / 0.24%	26 / 1%						
aRRd		2 / 0.14%						
DCEP_F			245 / 100%					
DCEP_1O		2 / 0.14%		86 / 100%				
DCEP_mul					26 / 96%			
BL Her						22 / 91%		
W Vir						1 / 4%	15 / 100%	
RV Tau								6 / 100%
ACEP_F	1 / 0.03%							
ACEP_1O	1 / 0.03%							

Class in ZTF

FIGURE A.5: Like Fig. III.6 but for ZTF catalog.
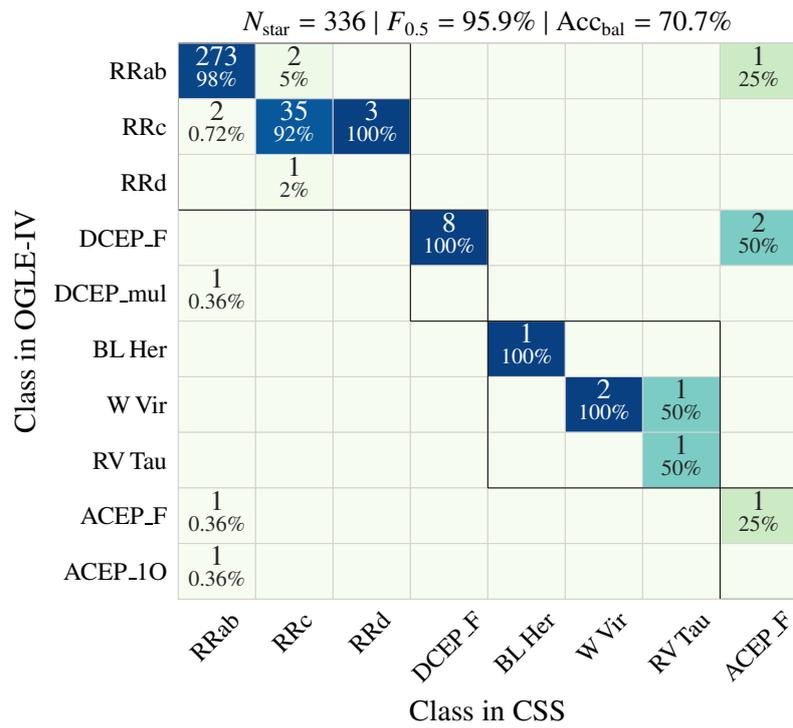
FIGURE A.6: Like Fig. III.6 but for CSS catalog.
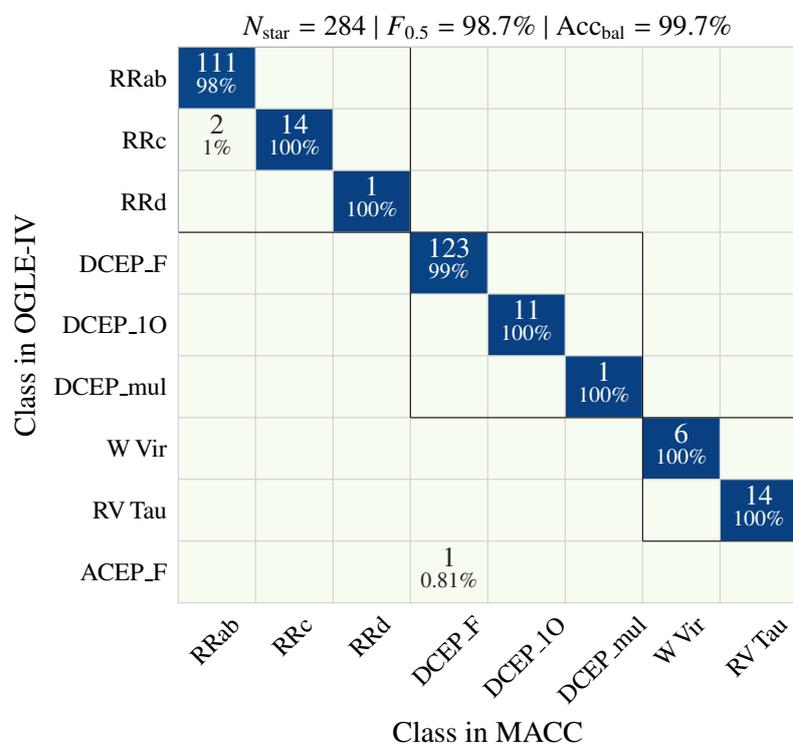
FIGURE A.7: Like Fig. III.6 but for MACC.

1.5 Confusion matrices against final classification in our catalog

FIGURE A.8: Like Fig. III.7 but for ASAS-SN catalog.

$$N_{\text{star}} = 4226 \mid F_{0.5}^{\text{RRLCep}} = 96.1\% \mid \text{Acc}_{\text{bal}} = 97.8\%$$

Final class in our catalog	
RRab	4134 97%
RRc	71 1%
RRd	18 0.43%
ACEP_F	2 0.05%
ACEP_1O	1 0.02%
	RRab

Class in DES

FIGURE A.9: Like Fig. III.7 but for DES.

$$N_{\text{star}} = 56349 \mid F_{0.5}^{\text{RRLCep}} = 98.9\% \mid \text{Acc}_{\text{bal}} = 99.1\%$$

Final class in our catalog	RRab	RRc
RRab	41735 99%	41 0.28%
RRc	74 0.18%	14219 98%
RRd	71 0.17%	131 0.91%
DCEP_F	1	
DCEP_1O		3 0.02%
BL Her	15 0.04%	
ACEP_F	47 0.11%	
ACEP_1O	10 0.02%	2 0.01%
	RRab	RRc

Class in PS1

FIGURE A.10: Like Fig. III.7 but for PS1.

FIGURE A.11: Like Fig. III.7 but for ZTF catalog.
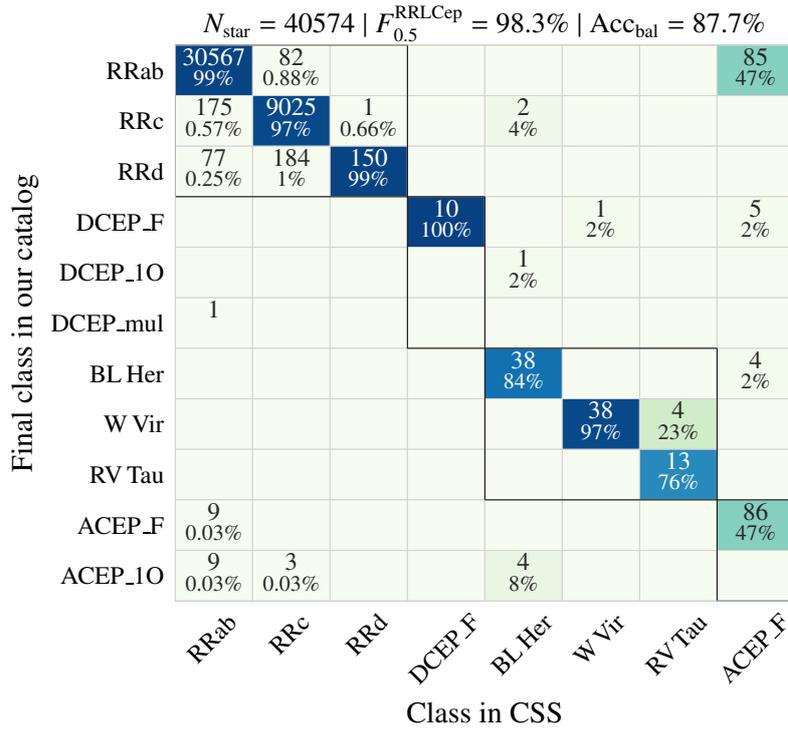
FIGURE A.12: Like Fig. III.7 but for CSS catalog.
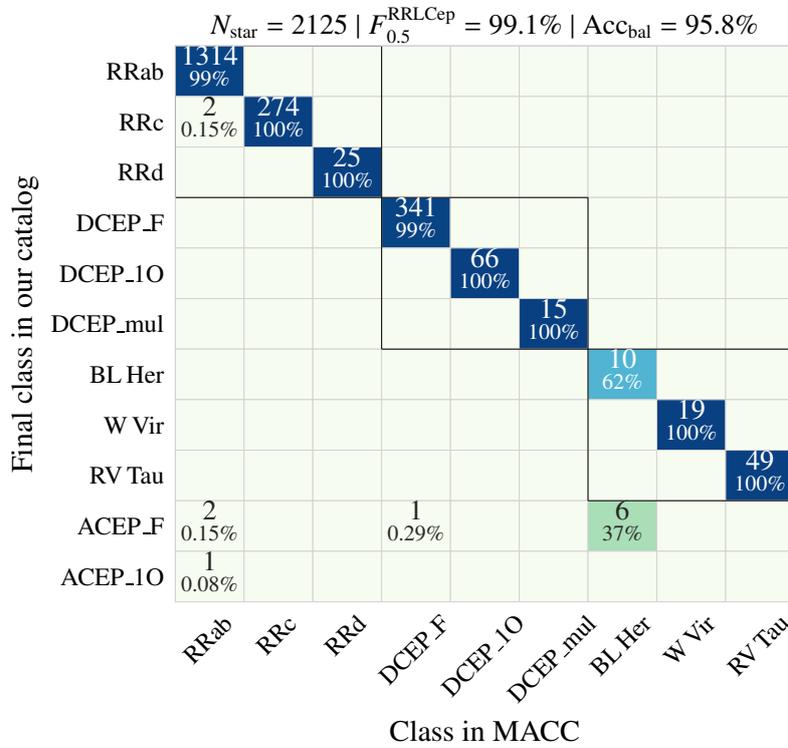
FIGURE A.13: Like Fig. III.7 but for MACC.

B

Supplementary information II

In this appendix, we present the PW plots of all the subtypes based on the LMC (Sect. 2.1) and the *Gaia* parallax sample (in Sect. 2.2).

2.1 Period-Wesenheit relations trained on the LMC sample

2.1.1 *Gaia* DR3 `gaia_source` uncertainty-weighted magnitudes
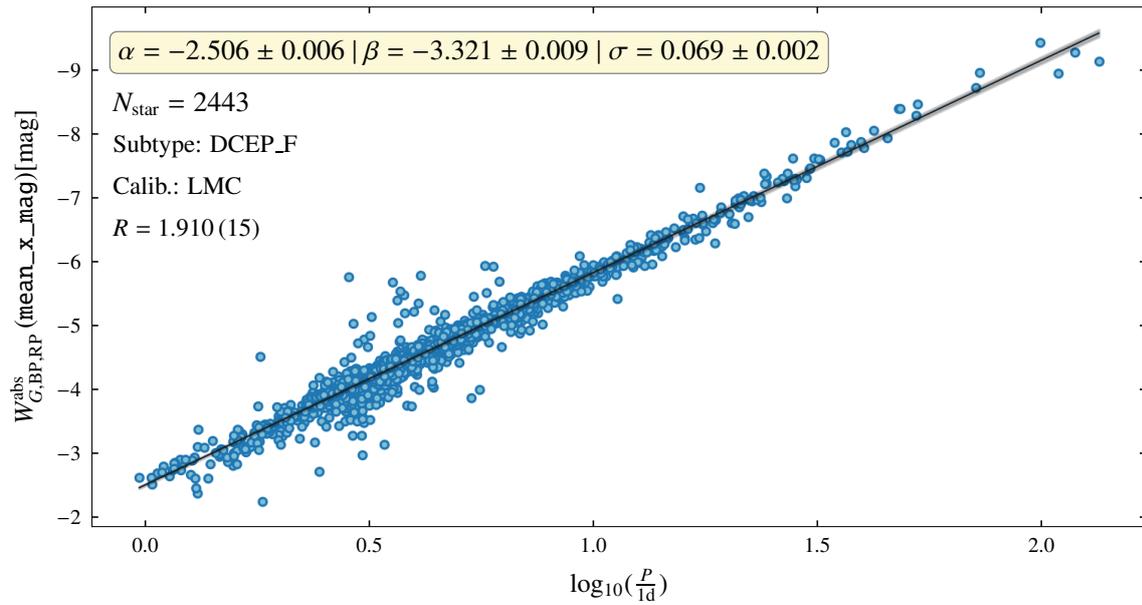
FIGURE B.1: Same as Fig. V.3 but with a Wesenheit index constructed using unweighted *Gaia* magnitudes as given in the *Gaia* DR3 `vari_summary` table, namely, `mean_g_mag_fov`, `mean_bp_mag`, and `mean_rp_mag`.
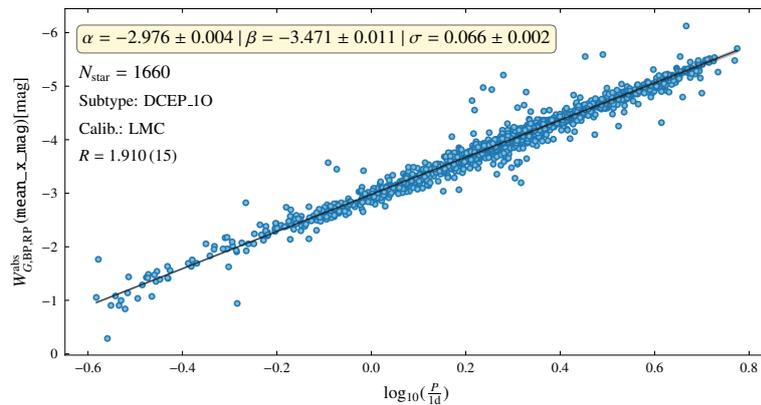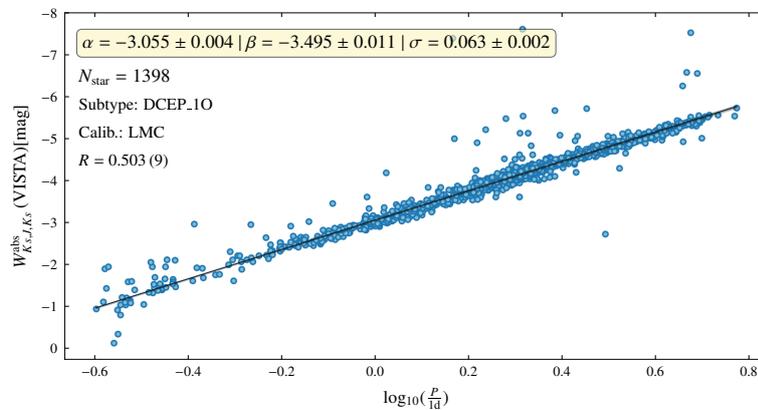

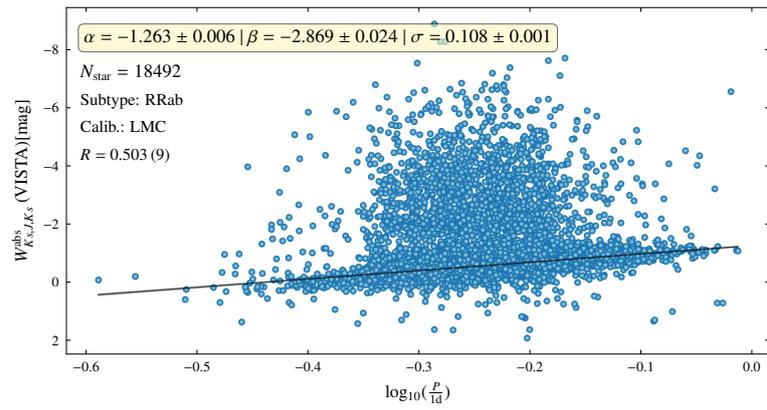**FIGURE B.2:** Same as Fig. B.1 but DCEP_1O stars.
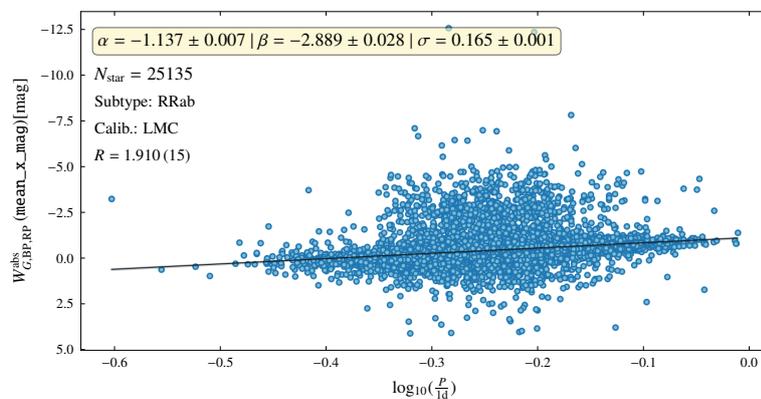
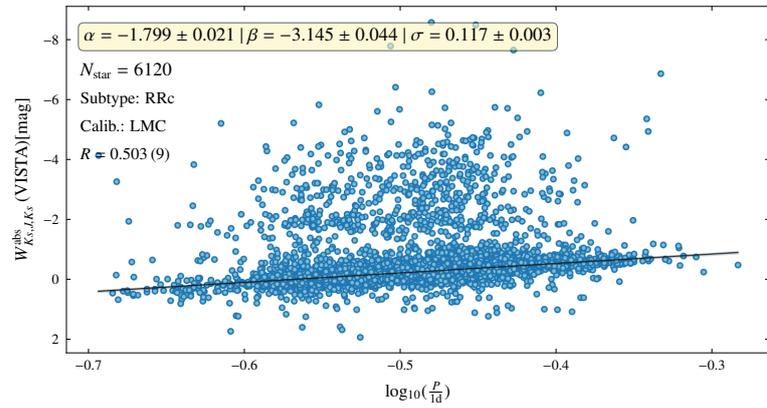**FIGURE B.3:** Same as Fig. B.1 but for RRab-type stars.
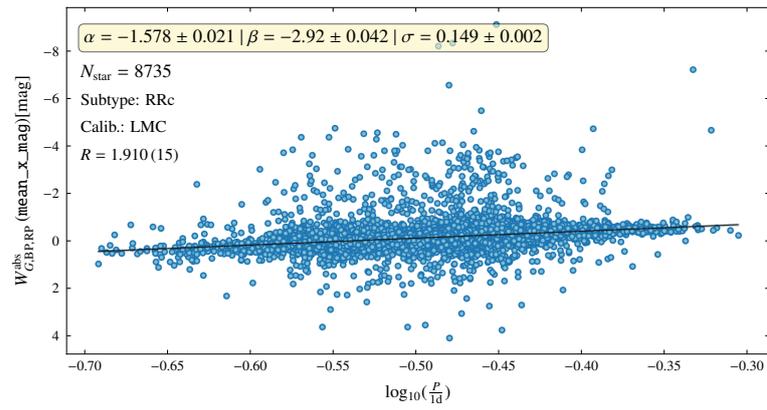
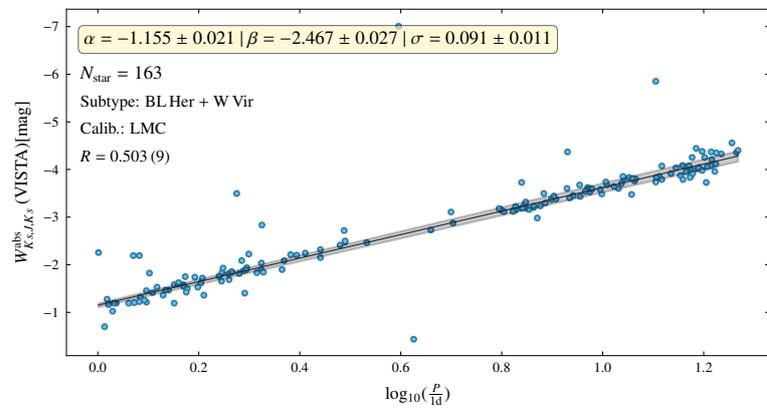**FIGURE B.4:** Same as Fig. B.1 but for RRc-type stars.
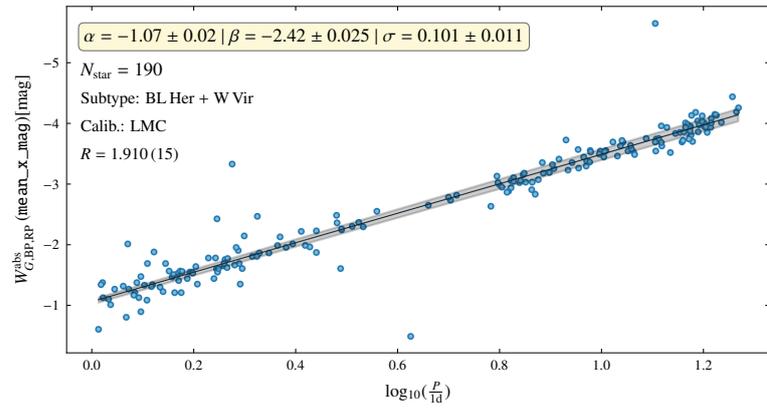
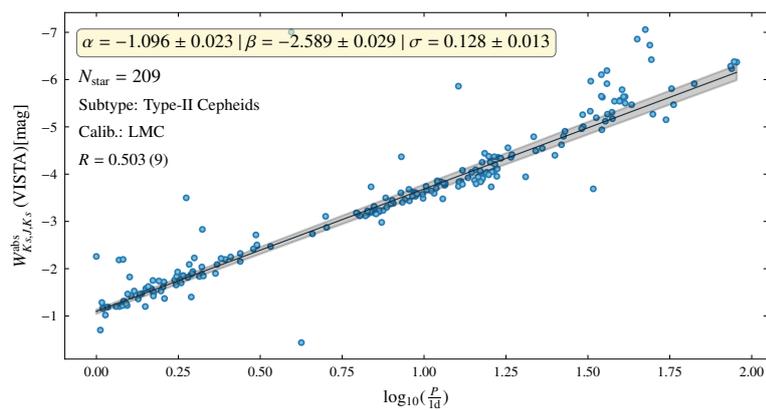**FIGURE B.5:** Same as Fig. B.1 but for BL Her and W Vir stars.
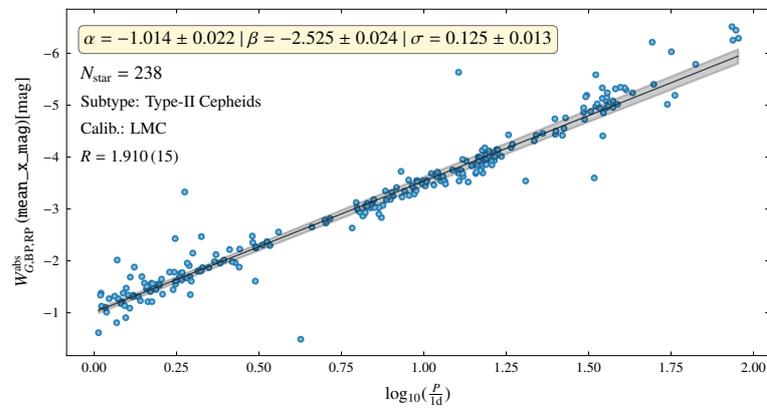
FIGURE B.6: Same as Fig. B.1 but for Type-II Cepheids.
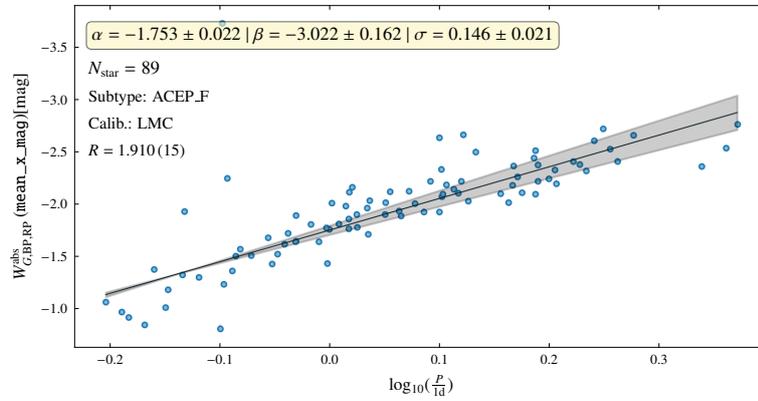
FIGURE B.7: Same as Fig. B.1 but for ACEP_F stars.
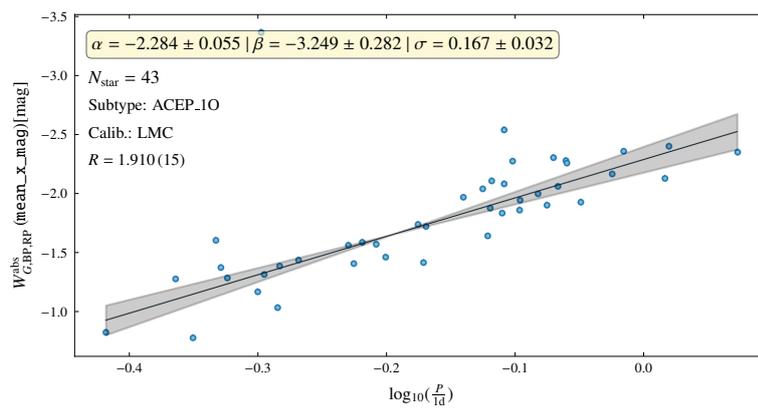
FIGURE B.8: Same as Fig. B.1 but for ACEP_1O stars.

2.1.2 *Gaia* DR3 `vari_summary` mean magnitudes
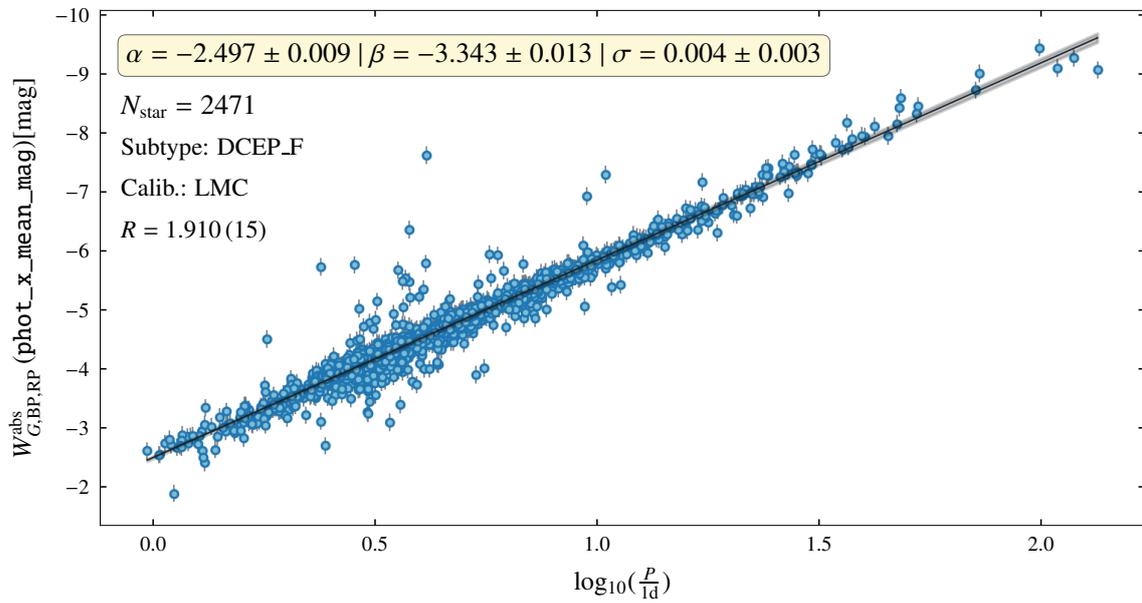
FIGURE B.9: Same as Fig. V.3 but with uncertainty-weighted *Gaia* magnitudes as given in the *Gaia* DR3 `source_table`, namely, `phot_g_mean_mag`, `phot_bp_mean_mag`, and `phot_rp_mean_mag`. An additional uncertainty of 0.05 mag was assumed in these magnitudes to account for the bias found by Eyer et al. (2023) and also illustrated in Fig. II.14.
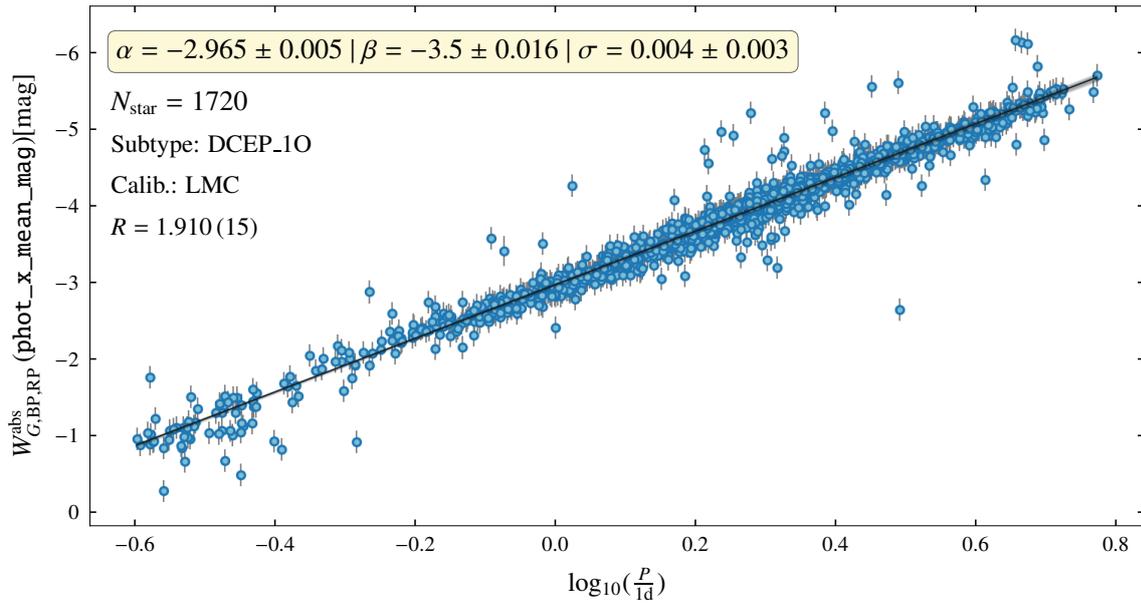
FIGURE B.10: Same as Fig. B.9 but DCEP_1O stars.
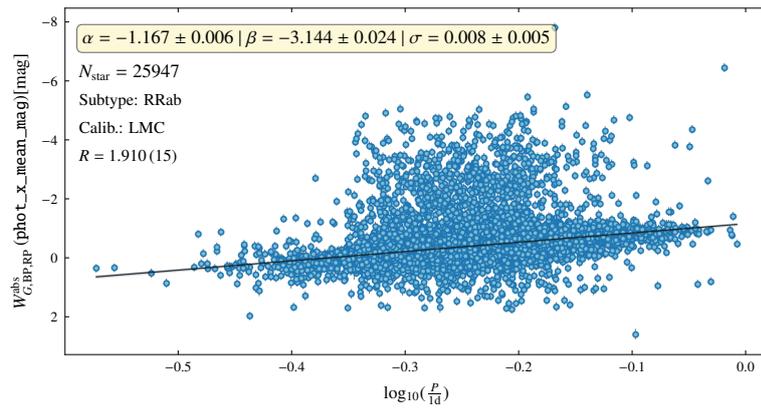
FIGURE B.11: Same as Fig. B.9 but for RRab-type stars.
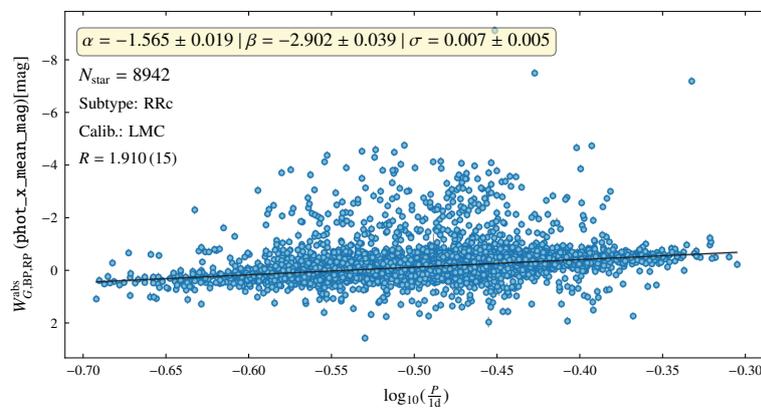
FIGURE B.12: Same as Fig. B.9 but for RRc-type stars.
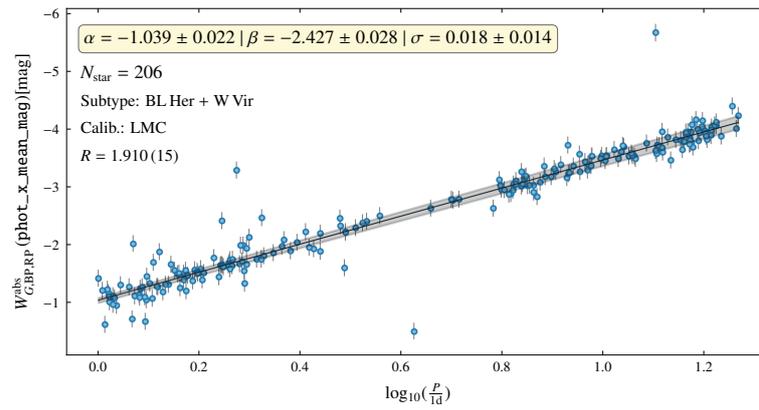
FIGURE B.13: Same as Fig. B.9 but for BL Her and W Vir stars.
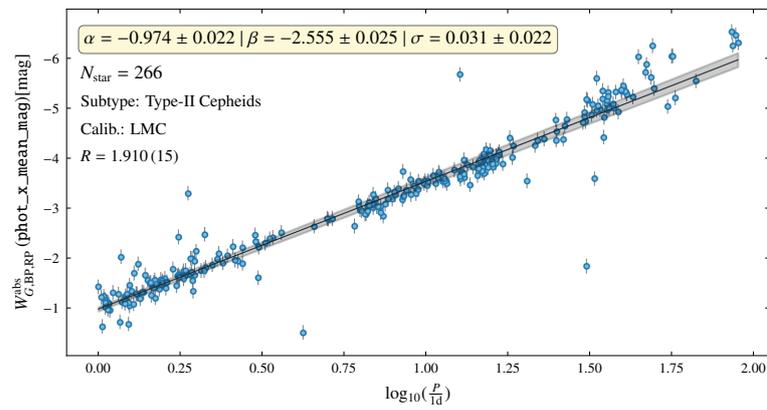
**FIGURE B.14:** Same as Fig. B.9 but for Type-II Cepheids.
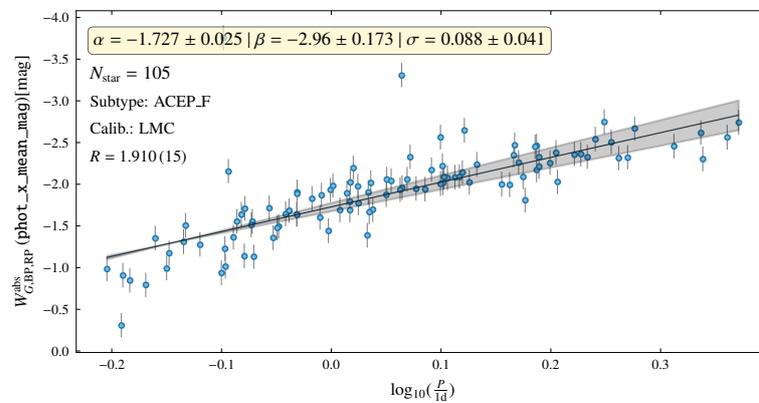
FIGURE B.15: Same as Fig. B.9 but for ACEP_F stars.
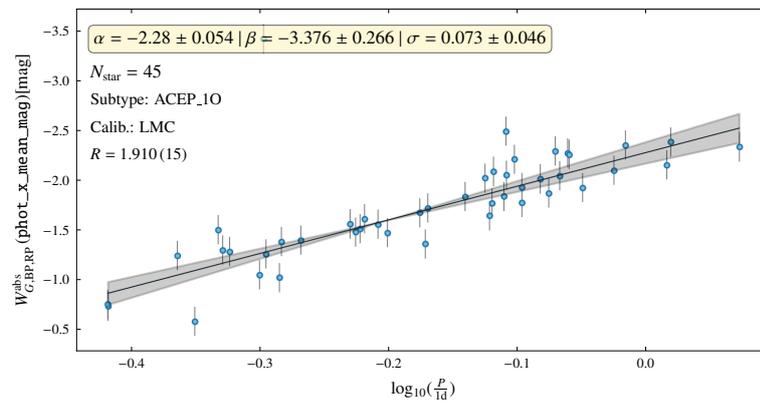
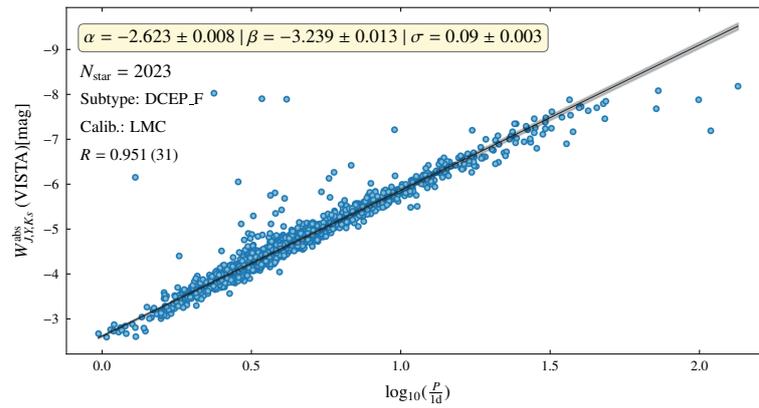**FIGURE B.16:** Same as Fig. B.9 but for ACEP_1O stars.

2.1.3 OGLE photometry
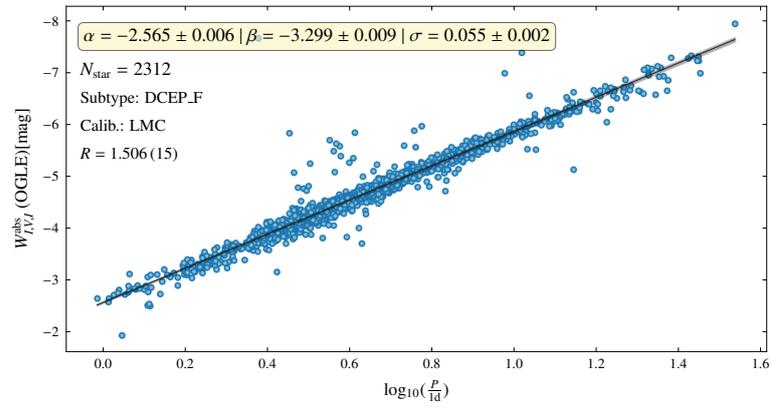
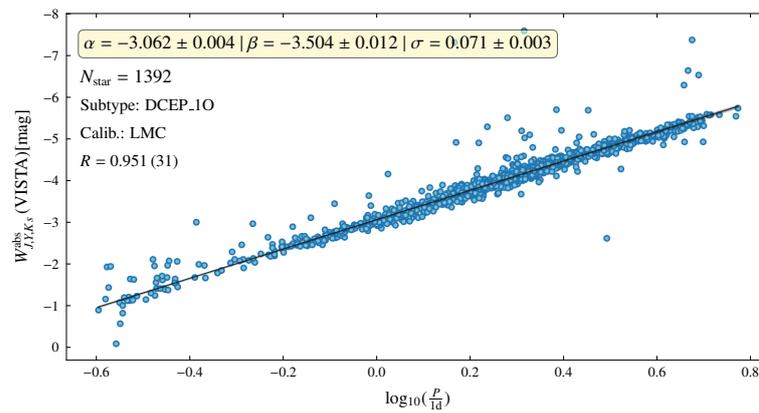**FIGURE B.17:** OGLE PW relation of the LMC DCEP_F stars
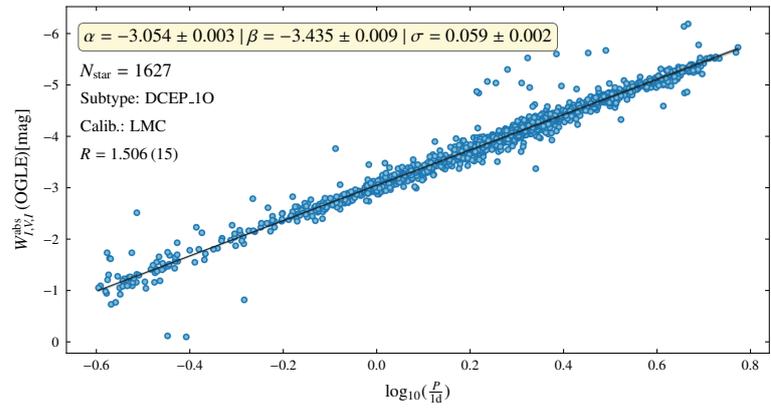
**FIGURE B.18:** Same as Fig. B.17 but DCEP_1O stars.
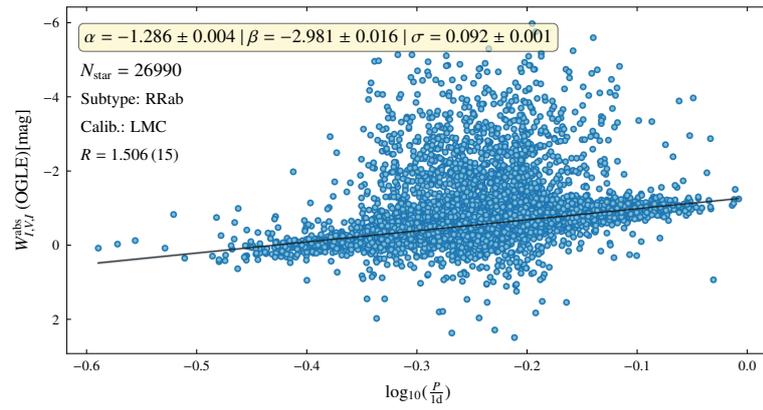
FIGURE B.19: Same as Fig. B.17 but for RRab-type stars.
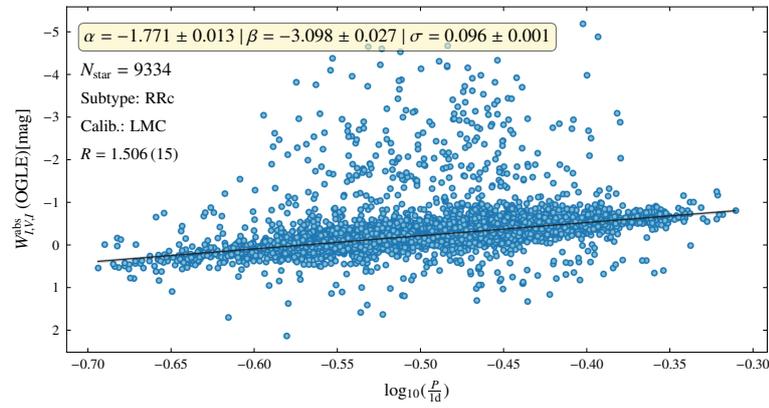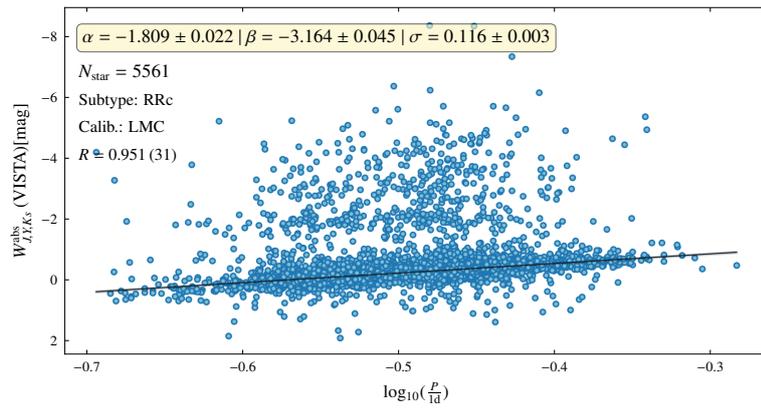

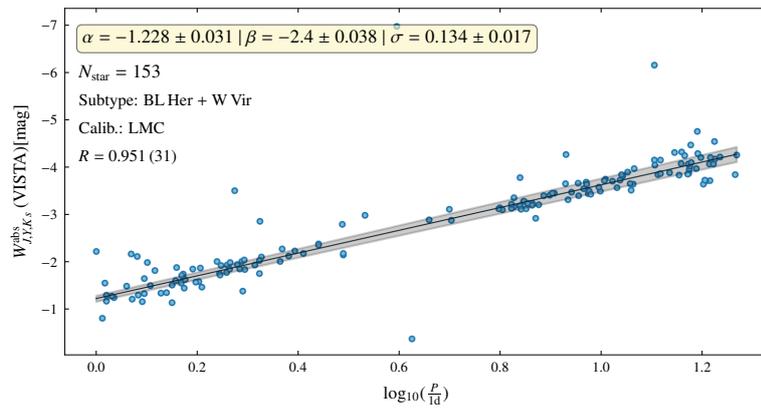**FIGURE B.20:** Same as Fig. B.17 but for RRc-type stars.
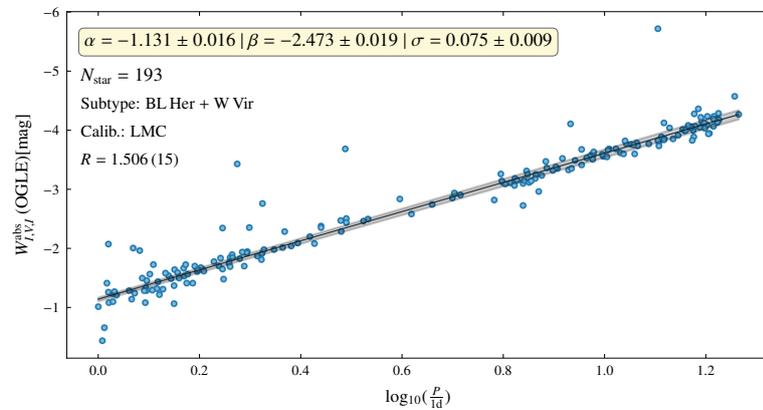
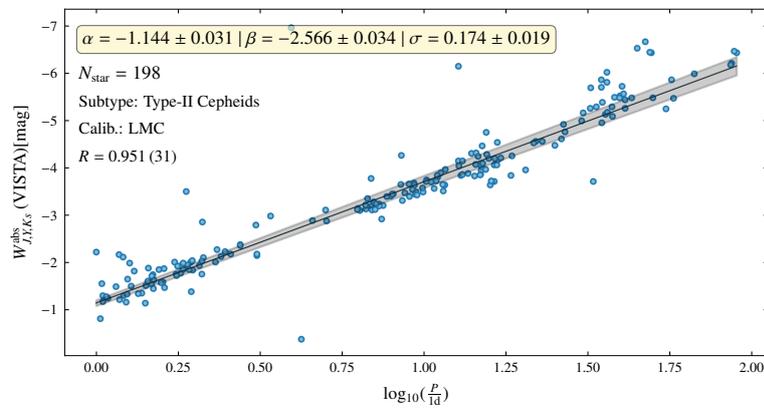**FIGURE B.21:** Same as Fig. B.17 but for BL Her and W Vir stars.
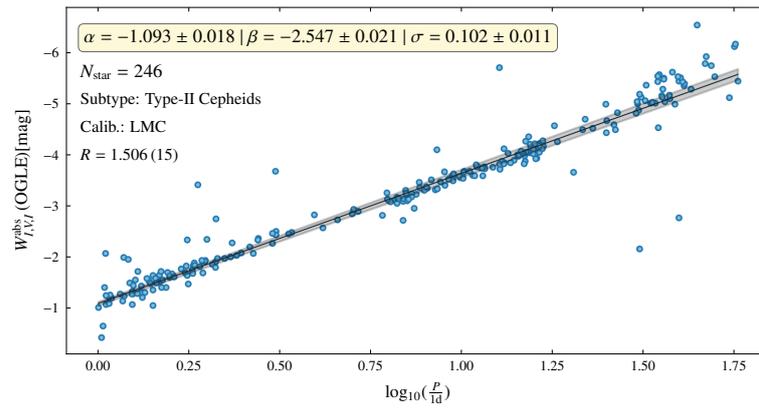
FIGURE B.22: Same as Fig. B.17 but for Type-II Cepheids.
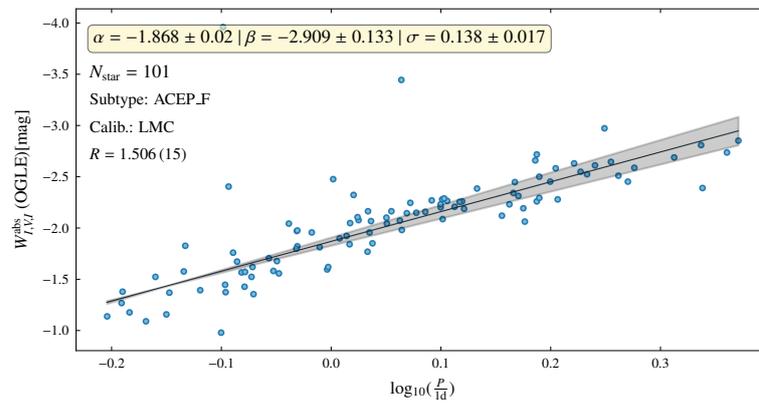
FIGURE B.23: Same as Fig. B.17 but for ACEP_F stars.
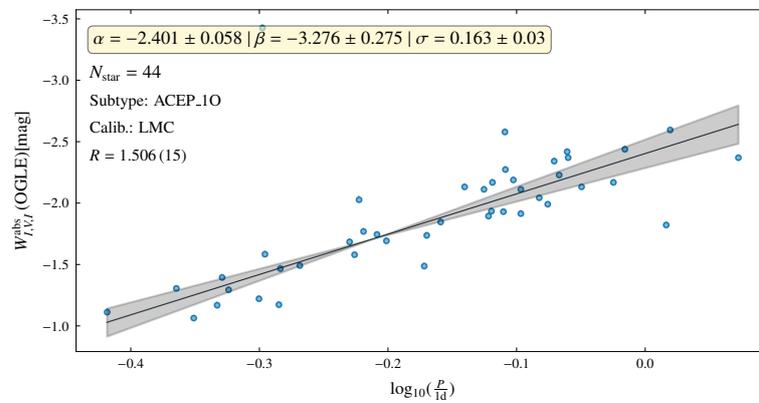
FIGURE B.24: Same as Fig. B.17 but for ACEP_1O stars.

2.1.4 2MASS and VISTA relations

Here, we present the LMC PW relations trained using NIR photometry from the 2MASS and VISTA surveys.
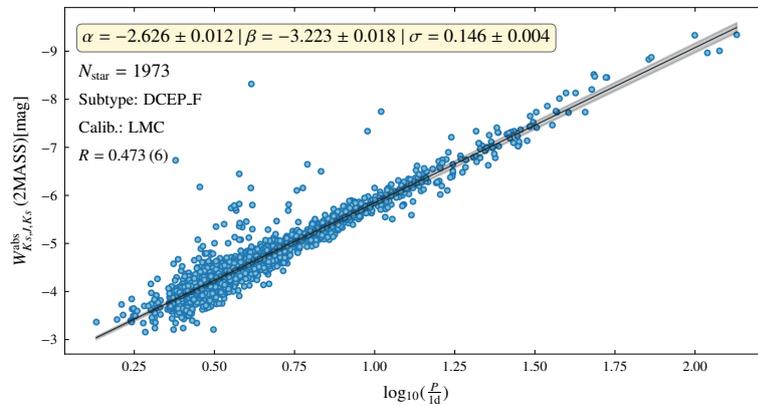
FIGURE B.25: 2MASS PW relation of the LMC DCEP_F stars with the Wesenheit index constructed using the J, Ks bands.
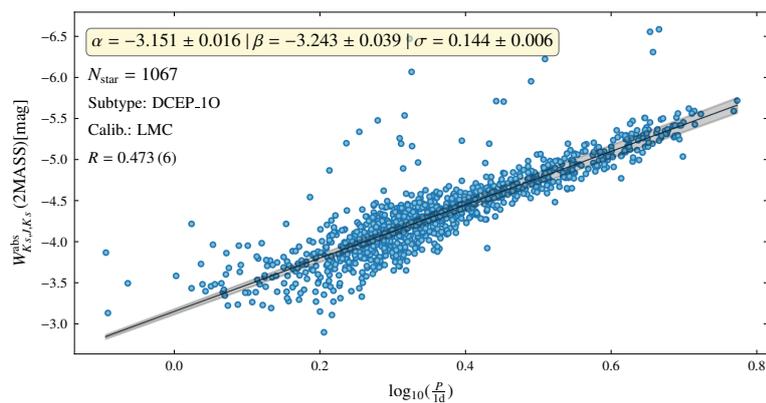
FIGURE B.26: Same as Fig. B.25 but DCEP_1O stars.
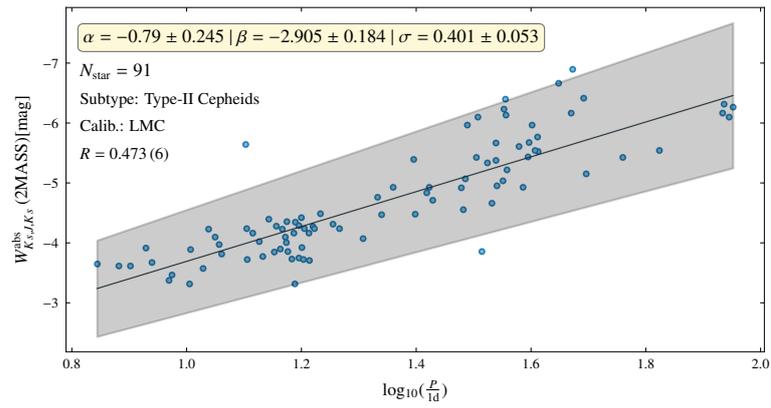
FIGURE B.27: Same as Fig. B.25 but for Type-II Cepheids.
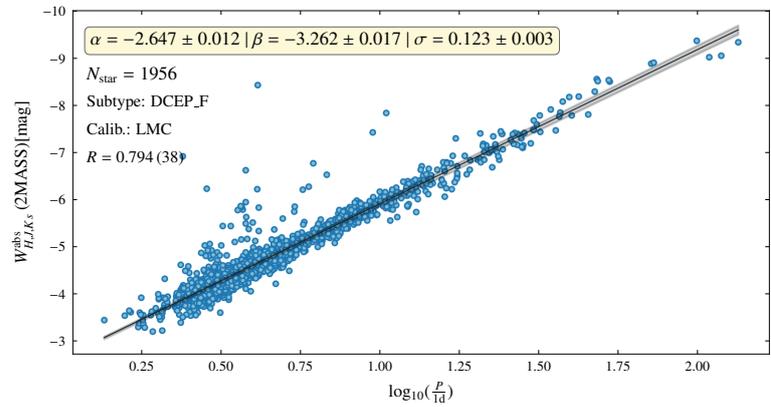
FIGURE B.28: 2MASS PW relation of the LMC DCEP_F stars with the Wesenheit index constructed using all $J, H,$ and Ks bands.
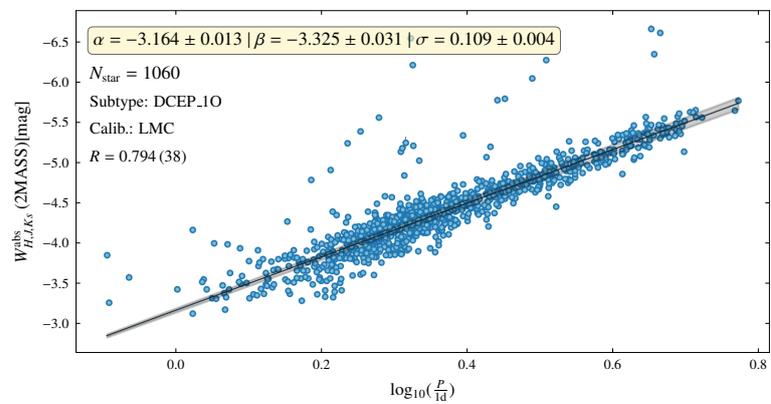
FIGURE B.29: Same as Fig. B.28 but DCEP_1O stars.
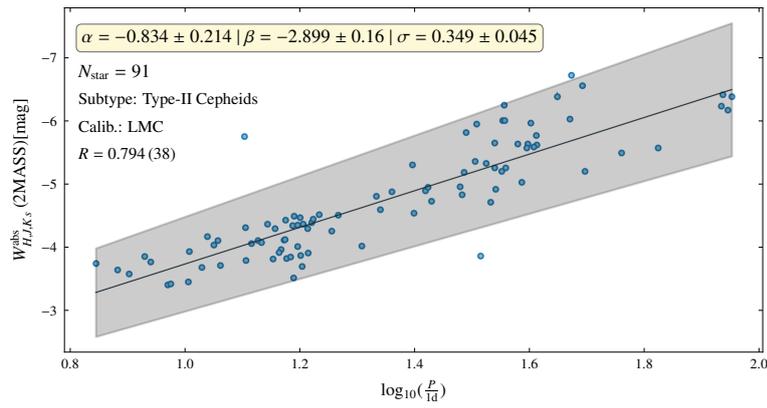
FIGURE B.30: Same as Fig. B.28 but for Type-II Cepheids.
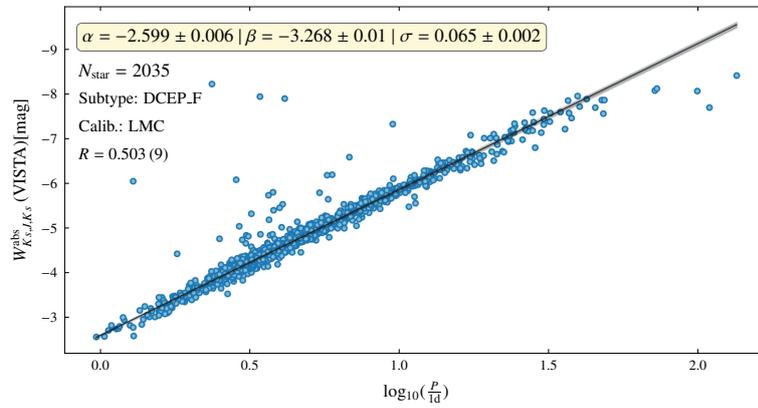
FIGURE B.31: VISTA PW relation of the LMC DCEP_F stars with the Wesenheit index constructed using J and Ks magnitudes that were not corrected for phase.

FIGURE B.32: Same as Fig. B.31 but DCEP_1O stars.

FIGURE B.33: Same as Fig. B.31 but for RRab-type stars.

FIGURE B.34: Same as Fig. B.31 but for RRc-type stars.

FIGURE B.35: Same as Fig. B.31 but for BL Her and W Vir stars.

FIGURE B.36: Same as Fig. B.31 but for Type-II Cepheids.
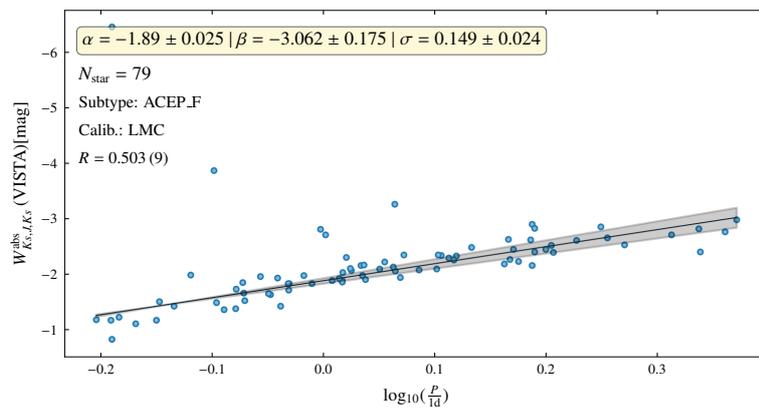
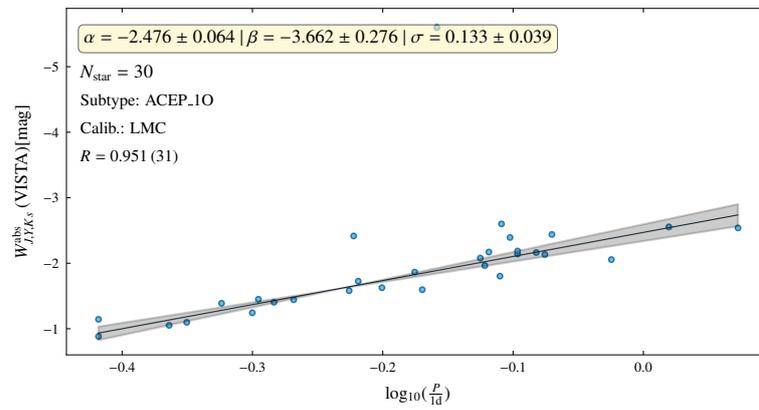**FIGURE B.37:** Same as Fig. B.31 but for ACEP_F stars.
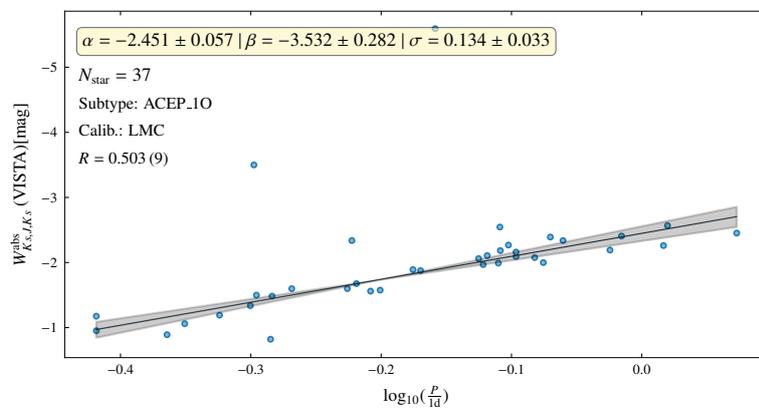
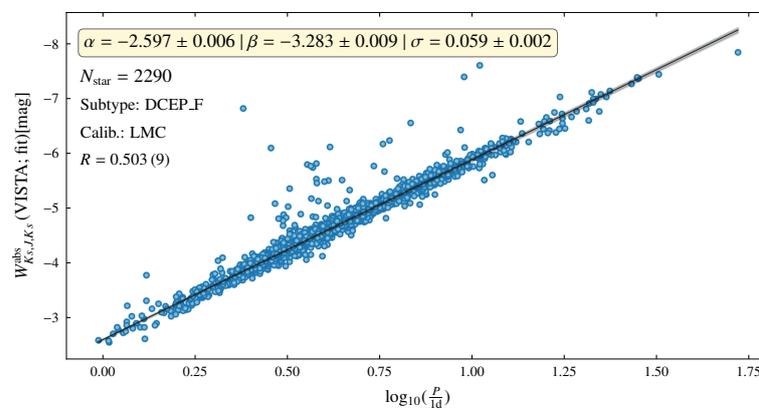**FIGURE B.38:** Same as Fig. B.31 but for ACEP_1O stars.

FIGURE B.39: VISTA PW relation of the LMC DCEP_F stars with the Wesenheit index constructed using Y, J and Ks magnitudes that were not corrected for phase.
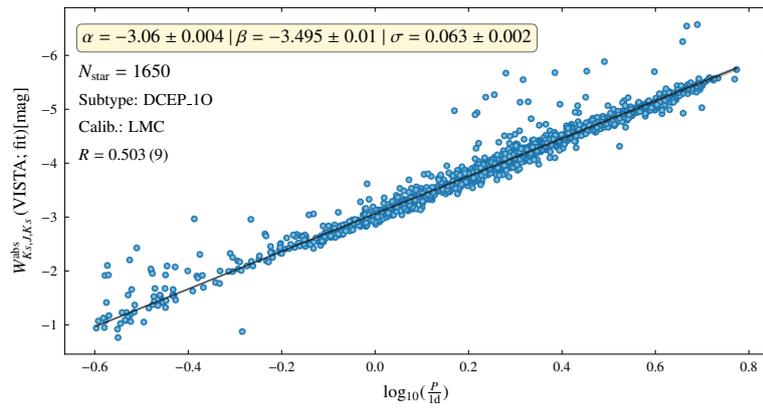

FIGURE B.40: Same as Fig. B.39 but DCEP_1O stars.
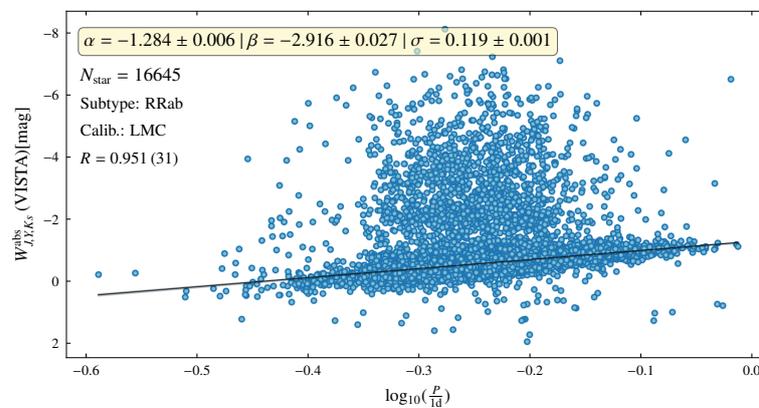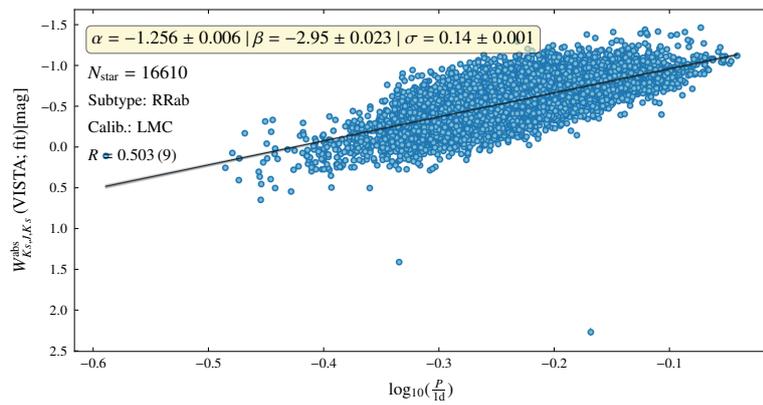

FIGURE B.41: Same as Fig. B.39 but for RRab-type stars.

FIGURE B.42: Same as Fig. B.39 but for RRc-type stars.

FIGURE B.43: Same as Fig. B.39 but for BL Her and W Vir stars.

FIGURE B.44: Same as Fig. B.39 but for Type-II Cepheids.
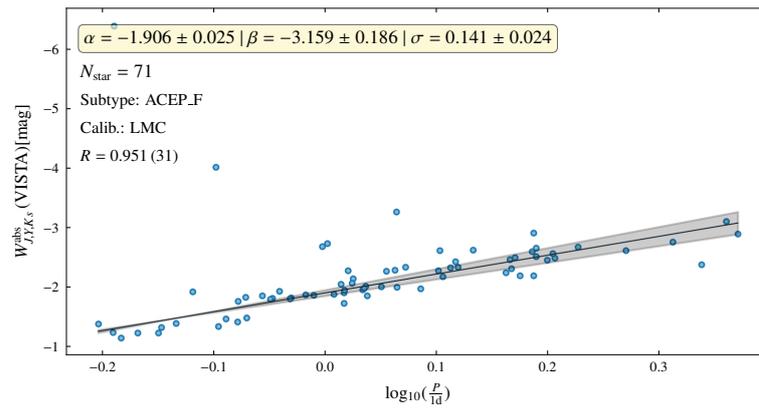
FIGURE B.45: Same as Fig. B.39 but for ACEP_F stars.

FIGURE B.46: Same as Fig. B.39 but for ACEP_1O stars.

FIGURE B.47: VISTA PW relation of the LMC DCEP_F stars with the Wesenheit index constructed using J and Ks magnitudes that were Fourier/template fit.

FIGURE B.48: Same as Fig. B.47 but DCEP_1O stars.

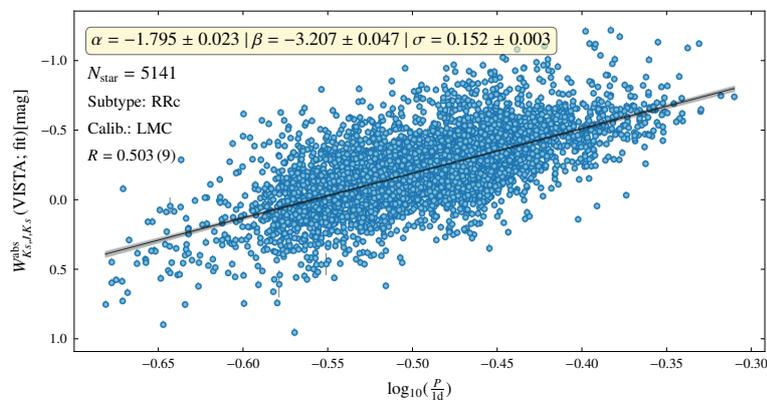**FIGURE B.49:** Same as Fig. B.47 but for RRab-type stars.

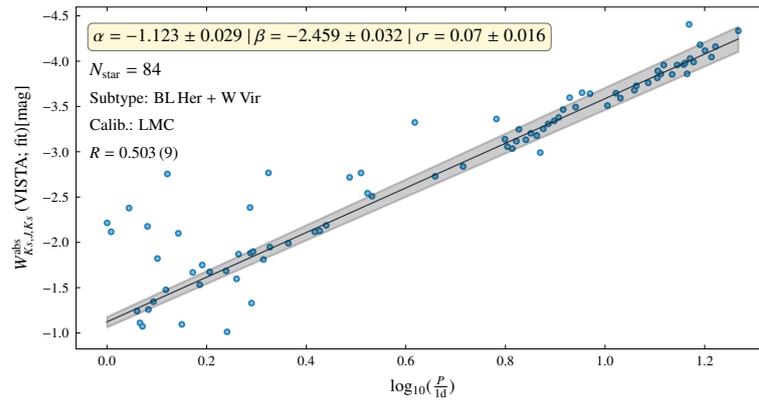**FIGURE B.50:** Same as Fig. B.47 but for RRc-type stars.

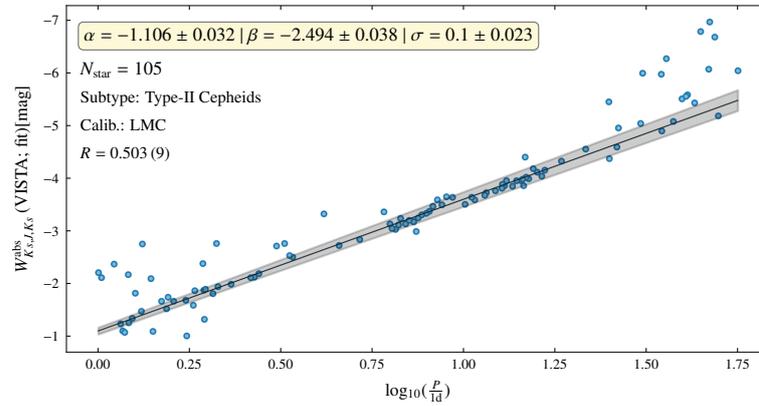**FIGURE B.51:** Same as Fig. B.47 but for BL Her and W Vir stars.

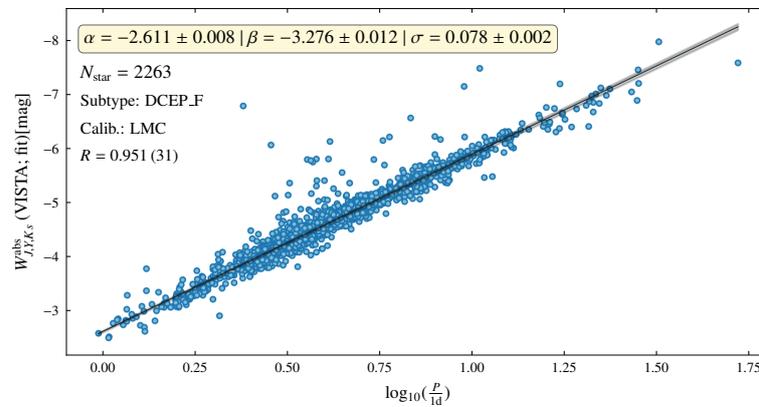**FIGURE B.52:** Same as Fig. B.47 but for Type-II Cepheids.

FIGURE B.53: VISTA PW relation of the LMC DCEP_F stars with the Wesenheit index constructed using Y, J and Ks magnitudes that were Fourier/template fit.
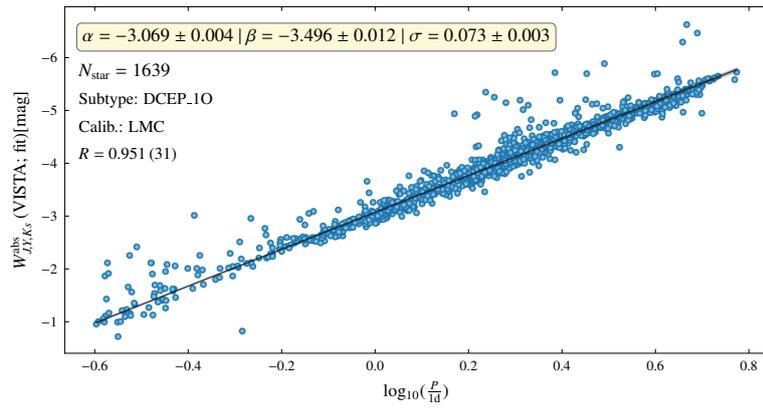
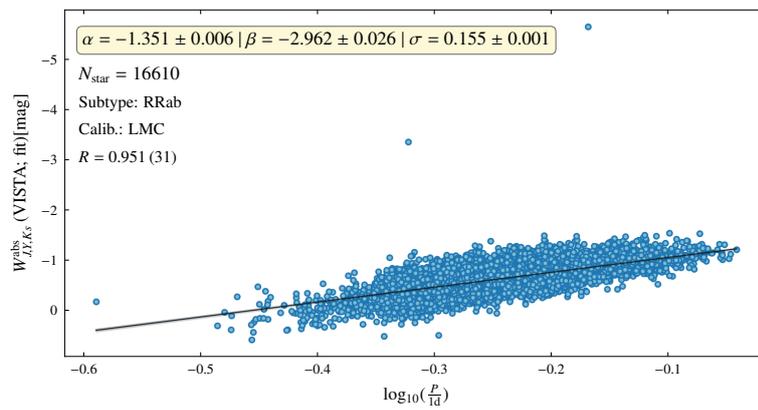**FIGURE B.54:** Same as Fig. B.53 but DCEP_1O stars.

FIGURE B.55: Same as Fig. B.53 but for RRab-type stars.
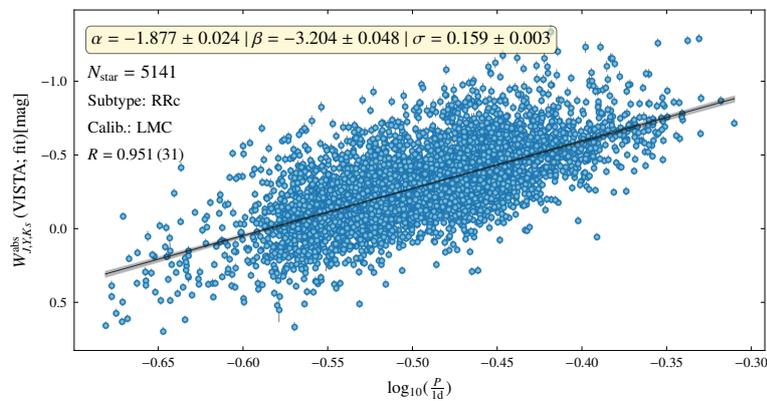

FIGURE B.56: Same as Fig. B.53 but for RRc-type stars.
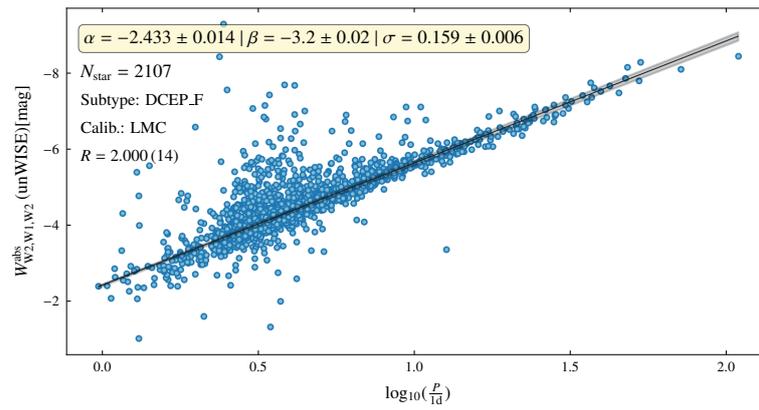
FIGURE B.57: unWISE PW relation of the LMC DCEP_F stars.
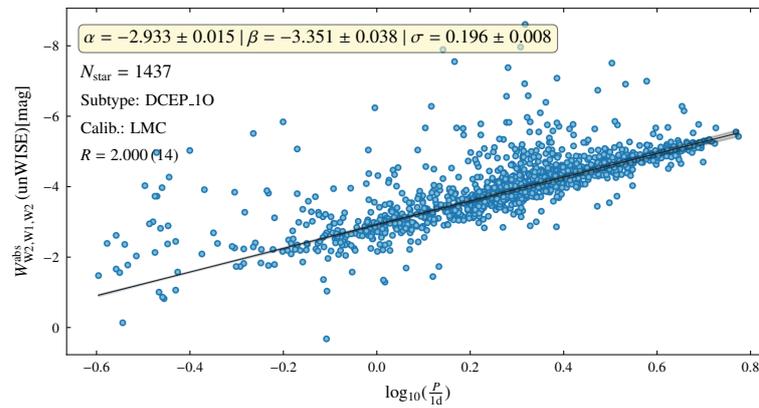

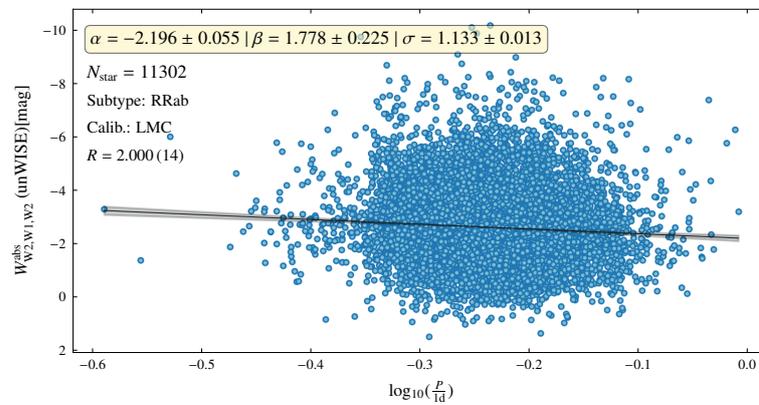**FIGURE B.58:** Same as Fig. B.57 but DCEP_1O stars.

FIGURE B.59: Same as Fig. B.57 but for RRab-type stars.
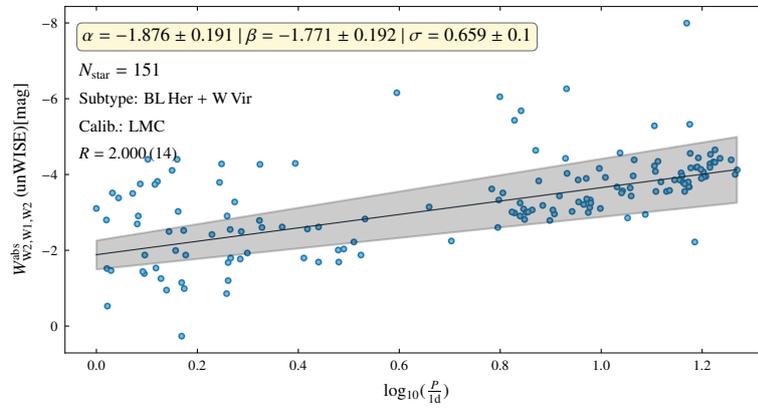
FIGURE B.60: Same as Fig. B.57 but for BL Her and W Vir stars.
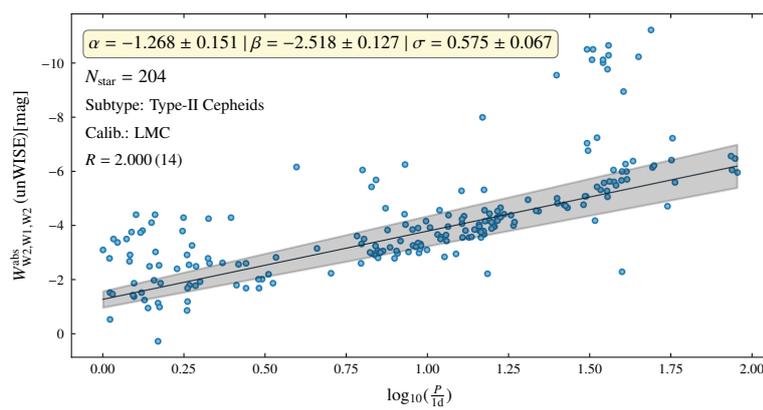
FIGURE B.61: Same as Fig. B.57 but for Type-II Cepheids.

2.2 Period-Wesenheit relations trained on the *Gaia* DR3 parallax sample

Here, we present the PW relations trained using the astrometry-based luminosity based on the *Gaia* DR3 parallax.

2.2.1 *Gaia* relations based on Fourier-fit, intensity-averaged magnitudes
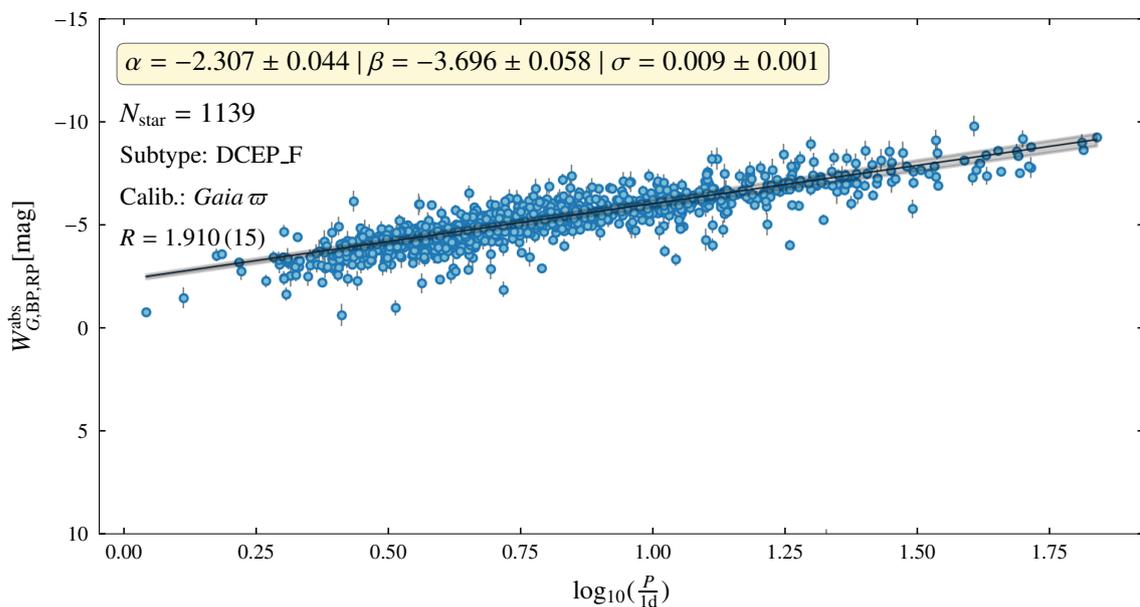
FIGURE B.62: Period-Wesenheit relation of the Milky Way classical Cepheids pulsating in the fundamental mode (DCEP_F). The *x*-axis is the decadic logarithm of the variability period (in days). The *y*-axis represents the absolute Wesenheit index constructed using the Fourier-fit intensity-averaged *Gaia* magnitudes computed by the SOS Cep&RRL pipeline (Ripepi et al., 2023; Clementini et al., 2023). Astrometry-based luminosity of the star was computed using their *Gaia* DR3 parallax. The model parameters (and their uncertainties) and the Wesenheit coefficient, R, used to construct the Wesenheit index are displayed on the top. The uncertainty on the Wesenheit coefficient ($\times 1000$) is given in in parentheses.

FIGURE B.63: Same as Fig. B.62 but DCEP_1O stars.

FIGURE B.64: Same as Fig. B.62 but for RRab-type stars.
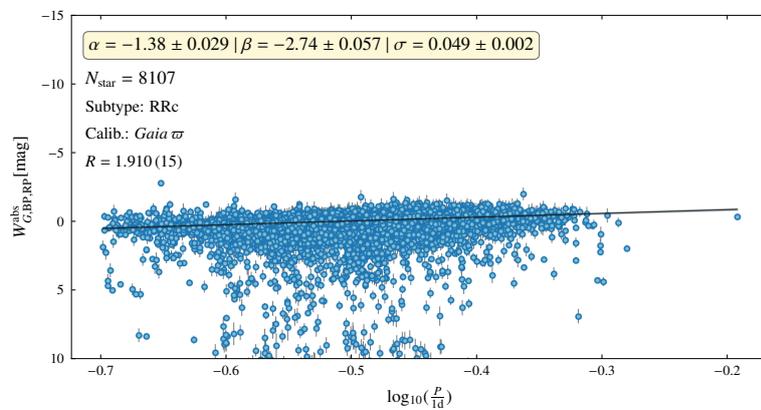
FIGURE B.65: Same as Fig. B.62 but for RRc-type stars.

FIGURE B.66: Same as Fig. B.62 but for Type-II Cepheids.

2.2.2   OGLE relations

FIGURE B.67: OGLE PW relation of the MW DCEP_F stars.
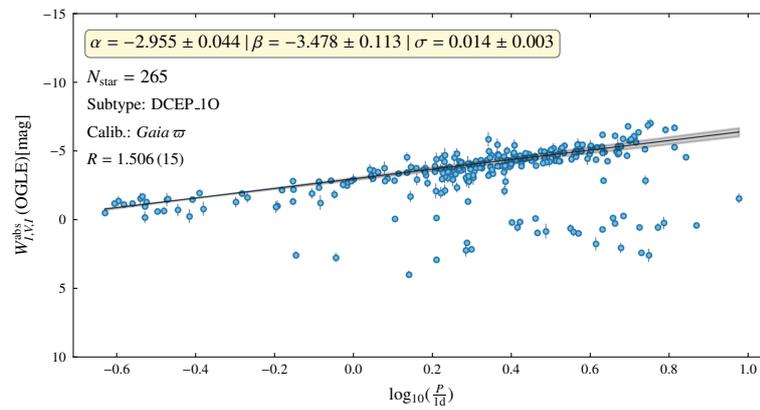
FIGURE B.68: Same as Fig. B.67 but DCEP_1O stars.

2.2.3   ZTF relations

FIGURE B.69: Same as Fig. B.67 but for RRab-type stars.

FIGURE B.70: Same as Fig. B.67 but for RRc-type stars.

FIGURE B.71: Same as Fig. B.67 but for Type-II Cepheids.

FIGURE B.72: ZTF PW relation of the MW DCEP_F stars.
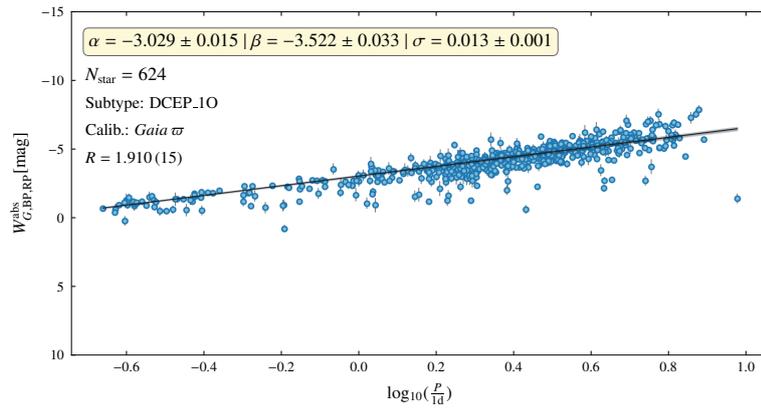

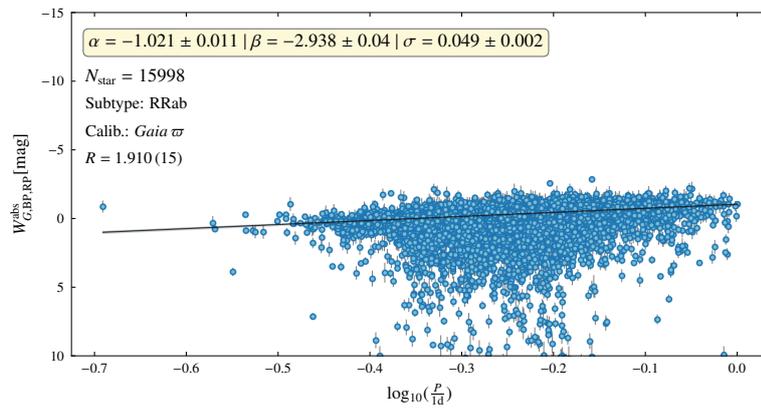**FIGURE B.73:** Same as Fig. B.72 but DCEP_1O stars.

FIGURE B.74: Same as Fig. B.72 but for RRab-type stars.

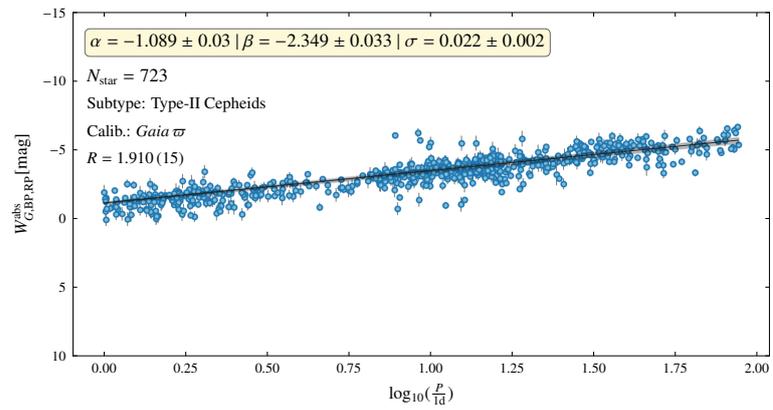**FIGURE B.75:** Same as Fig. B.72 but for RRc-type stars.

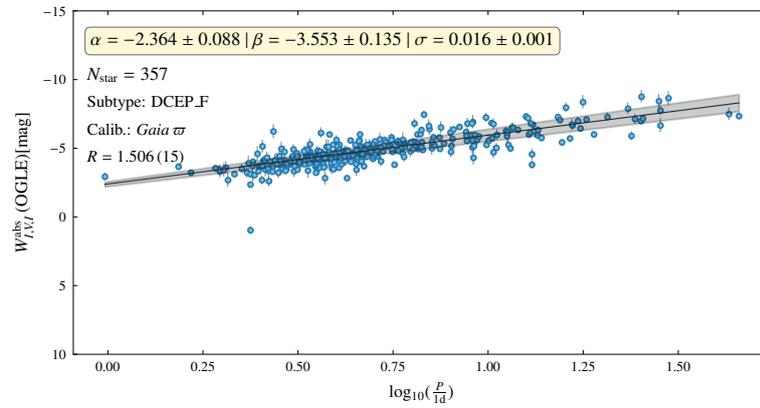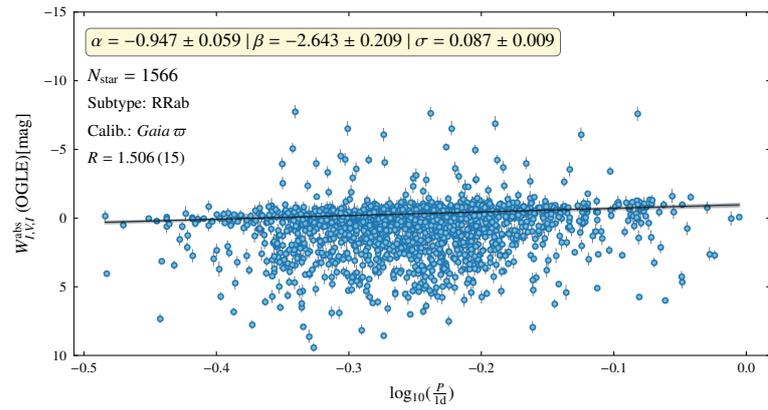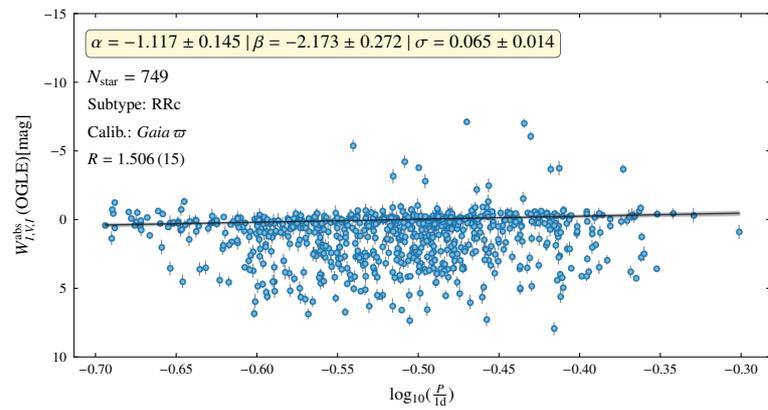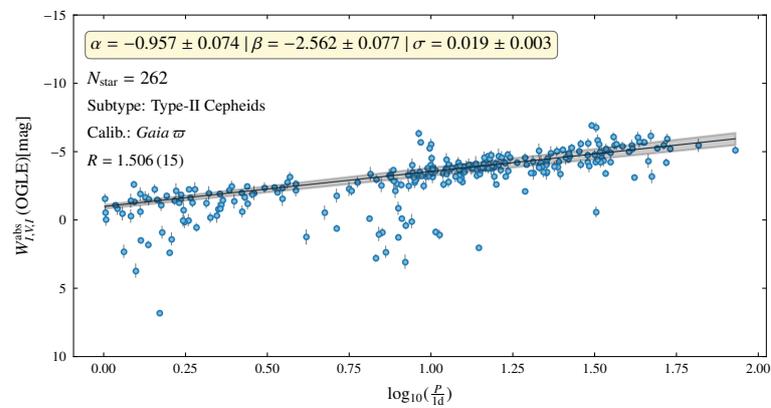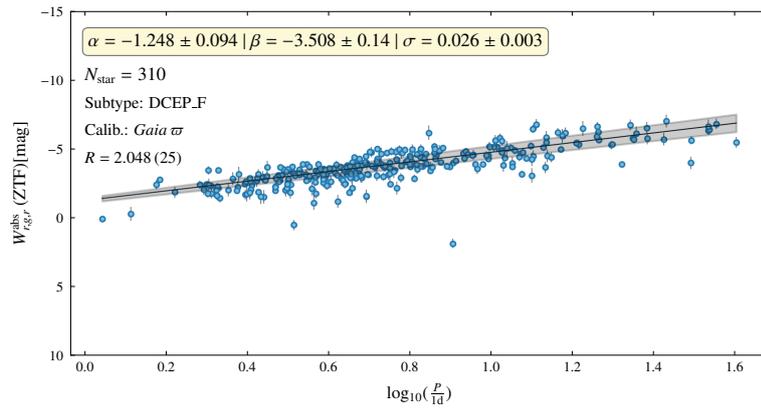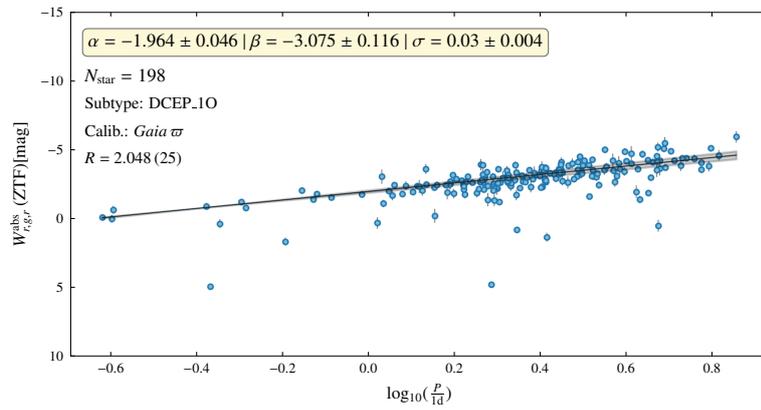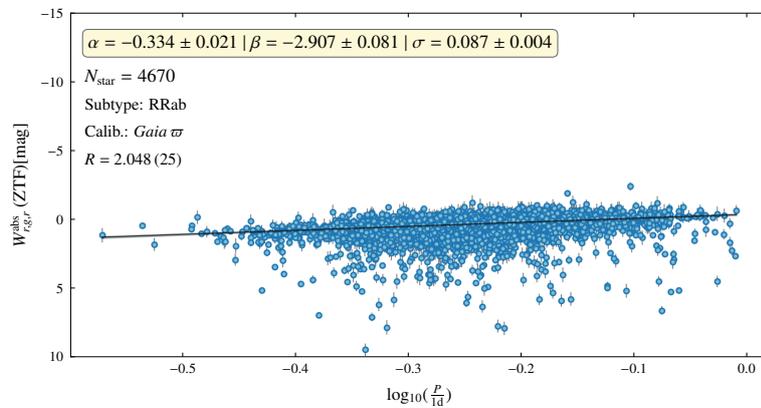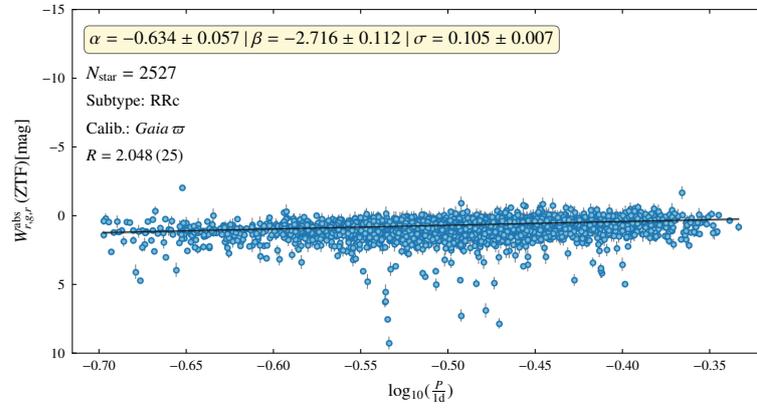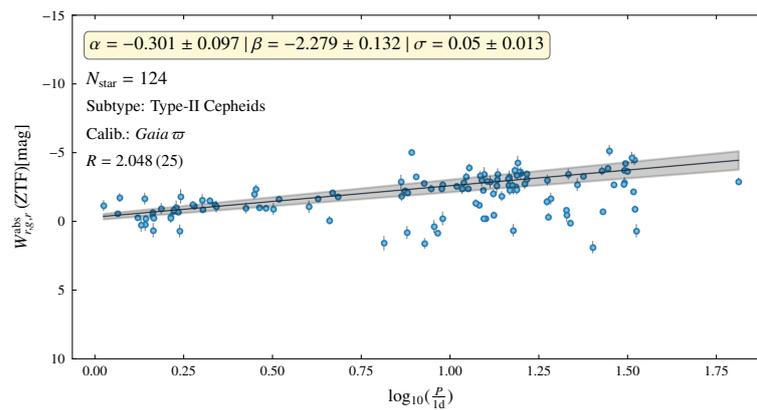**FIGURE B.76:** Same as Fig. B.72 but for Type-II Cepheids.

C

Supplementary information III

In this appendix, we list the model parameters of the 192 PW relations.

TABLE C.1: Model parameters of the LMC DCEP_1O PW relations.

W	α	β	σ	N_{star}
$W_{G,\text{BP,RP}}$	-2.968 ± 0.004	-3.473 ± 0.011	0.061 ± 0.003	1620
$W_{G,G,\text{RP}}$	-3.364 ± 0.005	-3.525 ± 0.013	0.064 ± 0.003	1620
$W_{G,\text{BP,RP}}$ (mean_x_mag)	-2.976 ± 0.004	-3.471 ± 0.011	0.066 ± 0.002	1660
$W_{G,G,\text{RP}}$ (mean_x_mag)	-3.385 ± 0.006	-3.5 ± 0.015	0.077 ± 0.003	1660
$W_{G,\text{BP,RP}}$ (phot_x_mean_mag)	-2.965 ± 0.005	-3.5 ± 0.016	0.004 ± 0.003	1720
$W_{G,G,\text{RP}}$ (phot_x_mean_mag)	-3.434 ± 0.011	-3.426 ± 0.032	0.008 ± 0.006	1721
$W_{I,V,I}$ (OGLE)	-3.054 ± 0.003	-3.435 ± 0.009	0.059 ± 0.002	1627
$W_{Ks,J,Ks}$ (2MASS)	-3.151 ± 0.016	-3.243 ± 0.039	0.144 ± 0.006	1067
$W_{H,J,Ks}$ (2MASS)	-3.164 ± 0.013	-3.325 ± 0.031	0.109 ± 0.004	1060
$W_{Ks,J,Ks}$ (VISTA)	-3.055 ± 0.004	-3.495 ± 0.011	0.063 ± 0.002	1398
$W_{J,Y,Ks}$ (VISTA)	-3.062 ± 0.004	-3.504 ± 0.012	0.071 ± 0.003	1392
$W_{Ks,J,Ks}$ (VISTA; fit)	-3.06 ± 0.004	-3.495 ± 0.01	0.063 ± 0.002	1650
$W_{J,Y,Ks}$ (VISTA; fit)	-3.069 ± 0.004	-3.496 ± 0.012	0.073 ± 0.003	1639
$W_{\text{W2,W1,W2}}$ (unWISE)	-2.933 ± 0.015	-3.351 ± 0.038	0.196 ± 0.008	1437

TABLE C.2: Model parameters of the MW DCEP_1O PW relations.

W	α	β	σ	N_{star}
$W_{G,\text{BP,RP}}$	-3.029 ± 0.015	-3.522 ± 0.033	0.013 ± 0.001	624
$W_{G,G,\text{RP}}$	-3.199 ± 0.02	-3.218 ± 0.042	0.018 ± 0.001	624
$W_{G,\text{BP,RP}}$ (mean_x_mag)	-3.038 ± 0.015	-3.548 ± 0.032	0.01 ± 0.001	649
$W_{G,G,\text{RP}}$ (mean_x_mag)	-3.221 ± 0.021	-3.255 ± 0.043	0.015 ± 0.001	649
$W_{G,\text{BP,RP}}$ (phot_x_mean_mag)	-3.043 ± 0.015	-3.553 ± 0.033	0.009 ± 0.001	680
$W_{G,G,\text{RP}}$ (phot_x_mean_mag)	-3.224 ± 0.022	-3.301 ± 0.046	0.013 ± 0.001	680
$W_{I,V,I}$ (OGLE)	-2.955 ± 0.044	-3.478 ± 0.113	0.014 ± 0.003	265
$W_{r,g,r}$ (ZTF)	-1.964 ± 0.046	-3.075 ± 0.116	0.03 ± 0.004	198
$W_{Ks,J,Ks}$ (2MASS)	-2.9 ± 0.02	-3.364 ± 0.044	0.012 ± 0.002	468
$W_{H,J,Ks}$ (2MASS)	-2.913 ± 0.019	-3.325 ± 0.045	0.011 ± 0.002	448
$W_{Ks,J,Ks}$ (VISTA)	-2.254 ± 0.118	-2.365 ± 0.221	0.074 ± 0.02	120
$W_{J,Y,Ks}$ (VISTA)	-1.464 ± 0.131	-0.906 ± 0.224	0.15 ± 0.034	68
$W_{H,J,Ks}$ (VISTA)	-0.965 ± 0.103	-0.588 ± 0.192	0.145 ± 0.041	63
$W_{\text{W2,W1,W2}}$ (unWISE)	-2.675 ± 0.026	-3.261 ± 0.069	0.014 ± 0.003	319

TABLE C.3: Model parameters of the LMC RRab PW relations.

W	α	β	σ	N_{star}
$W_{G,\mathrm{BP,RP}}$	-1.135 ± 0.006	-3.036 ± 0.027	0.091 ± 0.002	22061
$W_{G,G,\mathrm{RP}}$	-1.507 ± 0.012	-2.433 ± 0.049	0.219 ± 0.003	22132
$W_{G,\mathrm{BP,RP}}$ (`mean_x_mag`)	-1.137 ± 0.007	-2.889 ± 0.028	0.165 ± 0.001	25135
$W_{G,G,\mathrm{RP}}$ (`mean_x_mag`)	-1.422 ± 0.015	-2.426 ± 0.062	0.378 ± 0.004	25693
$W_{G,\mathrm{BP,RP}}$ (`phot_x_mean_mag`)	-1.167 ± 0.006	-3.144 ± 0.024	0.008 ± 0.005	25947
$W_{G,G,\mathrm{RP}}$ (`phot_x_mean_mag`)	-1.406 ± 0.014	-2.201 ± 0.058	0.205 ± 0.007	26906
$W_{I,V,I}$ (OGLE)	-1.286 ± 0.004	-2.981 ± 0.016	0.092 ± 0.001	26990
$W_{r,g,r}$ (DES)	-0.453 ± 0.064	-2.715 ± 0.265	0.247 ± 0.013	777
$W_{r,g,i}$ (DES)	-0.335 ± 0.047	-2.644 ± 0.192	0.172 ± 0.009	776
$W_{i,r,z}$ (DES)	-0.261 ± 0.054	-3.104 ± 0.225	0.189 ± 0.01	775
$W_{i,r,y}$ (DES)	-0.14 ± 0.049	-2.848 ± 0.204	0.174 ± 0.009	775
$W_{Ks,J,Ks}$ (VISTA)	-1.263 ± 0.006	-2.869 ± 0.024	0.108 ± 0.001	18492
$W_{J,Y,Ks}$ (VISTA)	-1.284 ± 0.006	-2.916 ± 0.027	0.119 ± 0.001	16645
$W_{Ks,J,Ks}$ (VISTA; fit)	-1.256 ± 0.006	-2.95 ± 0.023	0.14 ± 0.001	16610
$W_{J,Y,Ks}$ (VISTA; fit)	-1.351 ± 0.006	-2.962 ± 0.026	0.155 ± 0.001	16610
$W_{\mathrm{W2,W1,W2}}$ (unWISE)	-2.196 ± 0.055	1.778 ± 0.225	1.133 ± 0.013	11302

TABLE C.4: Model parameters of the MW RRab PW relations.

W	α	β	σ	N_{star}
$W_{G,\text{BP,RP}}$	-1.021 ± 0.011	-2.938 ± 0.04	0.049 ± 0.002	15998
$W_{G,G,\text{RP}}$	-1.425 ± 0.011	-2.996 ± 0.04	0.041 ± 0.001	16010
$W_{G,\text{BP,RP}}$ (mean_x_mag)	-1.0 ± 0.011	-2.826 ± 0.04	0.054 ± 0.002	17097
$W_{G,G,\text{RP}}$ (mean_x_mag)	-1.413 ± 0.011	-2.868 ± 0.042	0.045 ± 0.001	17281
$W_{G,\text{BP,RP}}$ (phot_x_mean_mag)	-0.987 ± 0.011	-2.661 ± 0.041	0.051 ± 0.002	18148
$W_{G,G,\text{RP}}$ (phot_x_mean_mag)	-1.392 ± 0.011	-2.703 ± 0.043	0.04 ± 0.002	18381
$W_{I,V,I}$ (OGLE)	-0.947 ± 0.059	-2.643 ± 0.209	0.087 ± 0.009	1566
$W_{r,g,r}$ (ZTF)	-0.334 ± 0.021	-2.907 ± 0.081	0.087 ± 0.004	4670
$W_{r,g,r}$ (PS1)	-0.405 ± 0.023	-2.902 ± 0.089	0.089 ± 0.004	6399
$W_{r,g,i}$ (PS1)	-0.267 ± 0.022	-2.883 ± 0.086	0.093 ± 0.004	6399
$W_{i,r,z}$ (PS1)	-0.003 ± 0.023	-2.965 ± 0.089	0.106 ± 0.005	6399
$W_{i,r,y}$ (PS1)	0.006 ± 0.024	-3.001 ± 0.096	0.128 ± 0.005	6395
$W_{Ks,J,Ks}$ (2MASS)	-1.156 ± 0.013	-3.106 ± 0.048	0.062 ± 0.002	15234
$W_{H,J,Ks}$ (2MASS)	-1.218 ± 0.012	-3.191 ± 0.047	0.06 ± 0.002	15121
$W_{Ks,J,Ks}$ (VISTA)	-1.024 ± 0.018	-3.041 ± 0.07	0.073 ± 0.003	10391
$W_{J,Y,Ks}$ (VISTA)	-0.991 ± 0.042	-3.27 ± 0.164	0.091 ± 0.006	2858
$W_{H,J,Ks}$ (VISTA)	-1.005 ± 0.048	-3.491 ± 0.185	0.096 ± 0.008	2387
$W_{Ks,J,Ks}$ (VISTA; fit)	-0.682 ± 0.063	-3.145 ± 0.232	0.115 ± 0.013	1750
$W_{J,Y,Ks}$ (VISTA; fit)	-0.588 ± 0.07	-2.916 ± 0.265	0.137 ± 0.014	1739
$W_{H,J,Ks}$ (VISTA; fit)	-0.743 ± 0.063	-3.242 ± 0.238	0.109 ± 0.013	1738
$W_{\text{W2,W1,W2}}$ (unWISE)	-0.976 ± 0.012	-3.022 ± 0.046	0.062 ± 0.002	15903

TABLE C.5: Model parameters of the LMC RRc PW relations.

W	α	β	σ	N_{star}
$W_{G,\text{BP,RP}}$	-1.515 ± 0.02	-2.77 ± 0.041	0.087 ± 0.003	7605
$W_{G,G,\text{RP}}$	-1.939 ± 0.04	-2.804 ± 0.082	0.215 ± 0.006	7611
$W_{G,\text{BP,RP}}$ (mean_x_mag)	-1.578 ± 0.021	-2.92 ± 0.042	0.149 ± 0.002	8735
$W_{G,G,\text{RP}}$ (mean_x_mag)	-1.975 ± 0.05	-3.235 ± 0.104	0.349 ± 0.006	8940
$W_{G,\text{BP,RP}}$ (phot_x_mean_mag)	-1.565 ± 0.019	-2.902 ± 0.039	0.007 ± 0.005	8942
$W_{G,G,\text{RP}}$ (phot_x_mean_mag)	-2.007 ± 0.047	-3.19 ± 0.096	0.187 ± 0.012	9273
$W_{I,V,I}$ (OGLE)	-1.771 ± 0.013	-3.098 ± 0.027	0.096 ± 0.001	9334
$W_{Ks,J,Ks}$ (VISTA)	-1.799 ± 0.021	-3.145 ± 0.044	0.117 ± 0.003	6120
$W_{J,Y,Ks}$ (VISTA)	-1.809 ± 0.022	-3.164 ± 0.045	0.116 ± 0.003	5561
$W_{Ks,J,Ks}$ (VISTA; fit)	-1.795 ± 0.023	-3.207 ± 0.047	0.152 ± 0.003	5141
$W_{J,Y,Ks}$ (VISTA; fit)	-1.877 ± 0.024	-3.204 ± 0.048	0.159 ± 0.003	5141

TABLE C.6: Model parameters of the MW RRc PW relations.

W	α	β	σ	N_{star}
$W_{G,\text{BP,RP}}$	-1.38 ± 0.029	-2.74 ± 0.057	0.049 ± 0.002	8107
$W_{G,G,\text{RP}}$	-1.76 ± 0.031	-2.814 ± 0.059	0.045 ± 0.002	8119
$W_{G,\text{BP,RP}}$ (mean_x_mag)	-1.247 ± 0.052	-2.887 ± 0.104	0.206 ± 0.005	10945
$W_{G,G,\text{RP}}$ (mean_x_mag)	-1.645 ± 0.05	-2.952 ± 0.101	0.166 ± 0.004	11132
$W_{G,\text{BP,RP}}$ (phot_x_mean_mag)	-1.188 ± 0.053	-2.833 ± 0.108	0.232 ± 0.005	11436
$W_{G,G,\text{RP}}$ (phot_x_mean_mag)	-1.587 ± 0.051	-2.867 ± 0.102	0.176 ± 0.004	11640
$W_{I,V,I}$ (OGLE)	-1.117 ± 0.145	-2.173 ± 0.272	0.065 ± 0.014	749
$W_{r,g,r}$ (ZTF)	-0.634 ± 0.057	-2.716 ± 0.112	0.105 ± 0.007	2527
$W_{r,g,r}$ (PS1)	-0.394 ± 0.086	-2.343 ± 0.169	0.181 ± 0.011	3120
$W_{r,g,i}$ (PS1)	-0.271 ± 0.084	-2.274 ± 0.164	0.184 ± 0.011	3120
$W_{i,r,z}$ (PS1)	-0.042 ± 0.083	-2.177 ± 0.163	0.195 ± 0.012	3120
$W_{i,r,y}$ (PS1)	0.043 ± 0.088	-2.061 ± 0.173	0.222 ± 0.013	3119
$W_{Ks,J,Ks}$ (2MASS)	-1.349 ± 0.055	-2.918 ± 0.111	0.216 ± 0.006	9527
$W_{H,J,Ks}$ (2MASS)	-1.424 ± 0.054	-3.0 ± 0.108	0.21 ± 0.007	9456
$W_{Ks,J,Ks}$ (VISTA)	-1.333 ± 0.054	-2.762 ± 0.106	0.117 ± 0.005	5872
$W_{J,Y,Ks}$ (VISTA)	-1.372 ± 0.13	-3.033 ± 0.262	0.142 ± 0.011	1645
$W_{\text{W2,W1,W2}}$ (unWISE)	-1.112 ± 0.053	-2.727 ± 0.105	0.22 ± 0.006	10058

TABLE C.7: Model parameters of the LMC BL Her and W Vir PW relations.

W	α	β	σ	N_{star}
$W_{G,\text{BP,RP}}$	-1.037 ± 0.023	-2.454 ± 0.027	0.074 ± 0.012	163
$W_{G,G,\text{RP}}$	-1.5 ± 0.04	-2.474 ± 0.045	0.137 ± 0.019	163
$W_{G,\text{BP,RP}}$ (mean_x_mag)	-1.07 ± 0.02	-2.42 ± 0.025	0.101 ± 0.011	190
$W_{G,G,\text{RP}}$ (mean_x_mag)	-1.534 ± 0.038	-2.438 ± 0.043	0.167 ± 0.018	190
$W_{G,\text{BP,RP}}$ (phot_x_mean_mag)	-1.039 ± 0.022	-2.427 ± 0.028	0.018 ± 0.014	206
$W_{G,G,\text{RP}}$ (phot_x_mean_mag)	-1.549 ± 0.044	-2.448 ± 0.052	0.032 ± 0.025	206
$W_{I,V,I}$ (OGLE)	-1.131 ± 0.016	-2.473 ± 0.019	0.075 ± 0.009	193
$W_{Ks,J,Ks}$ (VISTA)	-1.155 ± 0.021	-2.467 ± 0.027	0.091 ± 0.011	163
$W_{J,Y,Ks}$ (VISTA)	-1.228 ± 0.031	-2.4 ± 0.038	0.134 ± 0.017	153
$W_{Ks,J,Ks}$ (VISTA; fit)	-1.123 ± 0.029	-2.459 ± 0.032	0.07 ± 0.016	84
$W_{\text{W2,W1,W2}}$ (unWISE)	-1.876 ± 0.191	-1.771 ± 0.192	0.659 ± 0.1	151

TABLE C.8: Model parameters of the MW BL Her and W Vir PW relations.

W	α	β	σ	N_{star}
$W_{G,\text{BP,RP}}$	-1.093 ± 0.034	-2.315 ± 0.044	0.027 ± 0.003	551
$W_{G,G,\text{RP}}$	-1.493 ± 0.034	-2.184 ± 0.046	0.026 ± 0.003	551
$W_{G,\text{BP,RP}}$ (`mean_x_mag`)	-1.108 ± 0.036	-2.317 ± 0.045	0.023 ± 0.003	584
$W_{G,G,\text{RP}}$ (`mean_x_mag`)	-1.492 ± 0.034	-2.199 ± 0.045	0.023 ± 0.002	584
$W_{G,\text{BP,RP}}$ (`phot_x_mean_mag`)	-1.112 ± 0.039	-2.307 ± 0.047	0.021 ± 0.003	627
$W_{G,G,\text{RP}}$ (`phot_x_mean_mag`)	-1.497 ± 0.037	-2.179 ± 0.048	0.021 ± 0.003	630
$W_{I,V,I}$ (OGLE)	-0.988 ± 0.078	-2.508 ± 0.095	0.02 ± 0.005	189
$W_{r,g,r}$ (ZTF)	-0.282 ± 0.102	-2.323 ± 0.151	0.049 ± 0.014	101
$W_{Ks,J,Ks}$ (2MASS)	-1.073 ± 0.041	-2.352 ± 0.056	0.025 ± 0.003	479
$W_{H,J,Ks}$ (2MASS)	-1.116 ± 0.042	-2.36 ± 0.056	0.024 ± 0.003	473
$W_{Ks,J,Ks}$ (VISTA)	-0.966 ± 0.057	-2.029 ± 0.074	0.042 ± 0.005	371
$W_{J,Y,Ks}$ (VISTA)	-0.927 ± 0.101	-1.958 ± 0.116	0.029 ± 0.008	181
$W_{H,J,Ks}$ (VISTA)	-0.813 ± 0.095	-1.972 ± 0.109	0.02 ± 0.01	180
$W_{Ks,J,Ks}$ (VISTA; fit)	-0.797 ± 0.144	-1.831 ± 0.226	0.027 ± 0.018	84
$W_{J,Y,Ks}$ (VISTA; fit)	-0.729 ± 0.139	-1.887 ± 0.222	0.025 ± 0.018	84
$W_{H,J,Ks}$ (VISTA; fit)	-0.766 ± 0.15	-1.934 ± 0.242	0.031 ± 0.02	78
$W_{\text{W2,W1,W2}}$ (unWISE)	-0.943 ± 0.04	-2.252 ± 0.056	0.025 ± 0.004	420

TABLE C.9: Model parameters of the LMC Type-II Cepheids PW relations.

W	α	β	σ	N_{star}
$W_{G,\text{BP,RP}}$	-0.986 ± 0.024	-2.537 ± 0.026	0.084 ± 0.013	194
$W_{G,G,\text{RP}}$	-1.428 ± 0.04	-2.591 ± 0.04	0.153 ± 0.02	194
$W_{G,\text{BP,RP}}$ (`mean_x_mag`)	-1.014 ± 0.022	-2.525 ± 0.024	0.125 ± 0.013	238
$W_{G,G,\text{RP}}$ (`mean_x_mag`)	-1.46 ± 0.038	-2.563 ± 0.037	0.185 ± 0.017	238
$W_{G,\text{BP,RP}}$ (`phot_x_mean_mag`)	-0.974 ± 0.022	-2.555 ± 0.025	0.031 ± 0.022	266
$W_{G,G,\text{RP}}$ (`phot_x_mean_mag`)	-1.48 ± 0.041	-2.577 ± 0.04	0.031 ± 0.024	266
$W_{I,V,I}$ (OGLE)	-1.093 ± 0.018	-2.547 ± 0.021	0.102 ± 0.011	246
$W_{Ks,J,Ks}$ (2MASS)	-0.79 ± 0.245	-2.905 ± 0.184	0.401 ± 0.053	91
$W_{H,J,Ks}$ (2MASS)	-0.834 ± 0.214	-2.899 ± 0.16	0.349 ± 0.045	91
$W_{Ks,J,Ks}$ (VISTA)	-1.096 ± 0.023	-2.589 ± 0.029	0.128 ± 0.013	209
$W_{J,Y,Ks}$ (VISTA)	-1.144 ± 0.031	-2.566 ± 0.034	0.174 ± 0.019	198
$W_{Ks,J,Ks}$ (VISTA; fit)	-1.106 ± 0.032	-2.494 ± 0.038	0.1 ± 0.023	105
$W_{\text{W2,W1,W2}}$ (unWISE)	-1.268 ± 0.151	-2.518 ± 0.127	0.575 ± 0.067	204

TABLE C.10: Model parameters of the MW Type-II Cepheids PW relations.

W	α	β	σ	N_{star}
$W_{G,BP,RP}$	-1.089 ± 0.03	-2.349 ± 0.033	0.022 ± 0.002	723
$W_{G,G,RP}$	-1.489 ± 0.031	-2.21 ± 0.036	0.023 ± 0.002	723
$W_{G,BP,RP}$ (mean_x_mag)	-1.094 ± 0.033	-2.363 ± 0.034	0.02 ± 0.002	810
$W_{G,G,RP}$ (mean_x_mag)	-1.502 ± 0.032	-2.185 ± 0.035	0.022 ± 0.002	810
$W_{G,BP,RP}$ (phot_x_mean_mag)	-1.093 ± 0.036	-2.349 ± 0.036	0.02 ± 0.002	896
$W_{G,G,RP}$ (phot_x_mean_mag)	-1.493 ± 0.034	-2.187 ± 0.038	0.022 ± 0.002	904
$W_{I,V,I}$ (OGLE)	-0.957 ± 0.074	-2.562 ± 0.077	0.019 ± 0.003	262
$W_{r,g,r}$ (ZTF)	-0.301 ± 0.097	-2.279 ± 0.132	0.05 ± 0.013	124
$W_{Ks,J,Ks}$ (2MASS)	-1.065 ± 0.036	-2.382 ± 0.041	0.022 ± 0.002	661
$W_{H,J,Ks}$ (2MASS)	-1.119 ± 0.039	-2.369 ± 0.042	0.021 ± 0.002	641
$W_{Ks,J,Ks}$ (VISTA)	-1.038 ± 0.054	-1.865 ± 0.059	0.045 ± 0.005	531
$W_{J,Y,Ks}$ (VISTA)	-1.016 ± 0.092	-1.802 ± 0.088	0.034 ± 0.005	267
$W_{H,J,Ks}$ (VISTA)	-0.922 ± 0.084	-1.774 ± 0.08	0.03 ± 0.007	263
$W_{Ks,J,Ks}$ (VISTA; fit)	-0.799 ± 0.14	-1.824 ± 0.222	0.027 ± 0.018	85
$W_{J,Y,Ks}$ (VISTA; fit)	-0.743 ± 0.14	-1.863 ± 0.224	0.025 ± 0.017	85
$W_{H,J,Ks}$ (VISTA; fit)	-0.772 ± 0.153	-1.918 ± 0.245	0.031 ± 0.021	79
$W_{W2,W1,W2}$ (unWISE)	-0.931 ± 0.04	-2.276 ± 0.052	0.028 ± 0.004	476

TABLE C.11: Model parameters of the LMC ACEP_F PW relations.

W	α	β	σ	N_{star}
$W_{G,BP,RP}$	-1.769 ± 0.023	-2.763 ± 0.153	0.116 ± 0.019	67
$W_{G,G,RP}$	-2.223 ± 0.045	-2.626 ± 0.283	0.184 ± 0.053	67
$W_{G,BP,RP}$ (mean_x_mag)	-1.753 ± 0.022	-3.022 ± 0.162	0.146 ± 0.021	89
$W_{G,G,RP}$ (mean_x_mag)	-2.263 ± 0.043	-2.78 ± 0.291	0.256 ± 0.045	89
$W_{G,BP,RP}$ (phot_x_mean_mag)	-1.727 ± 0.025	-2.96 ± 0.173	0.088 ± 0.041	105
$W_{G,G,RP}$ (phot_x_mean_mag)	-2.275 ± 0.041	-2.731 ± 0.261	0.081 ± 0.058	105
$W_{I,V,I}$ (OGLE)	-1.868 ± 0.02	-2.909 ± 0.133	0.138 ± 0.017	101
$W_{Ks,J,Ks}$ (VISTA)	-1.89 ± 0.025	-3.062 ± 0.175	0.149 ± 0.024	79
$W_{J,Y,Ks}$ (VISTA)	-1.906 ± 0.025	-3.159 ± 0.186	0.141 ± 0.024	71

TABLE C.12: Model parameters of the LMC ACEP_1O PW relations.

W	α	β	σ	N_{star}
$W_{G,\text{BP,RP}}$	-2.244 ± 0.046	-3.128 ± 0.253	0.144 ± 0.024	36
$W_{G,\text{BP,RP}}$ (`mean_x_mag`)	-2.284 ± 0.055	-3.249 ± 0.282	0.167 ± 0.032	43
$W_{G,\text{BP,RP}}$ (`phot_x_mean_mag`)	-2.28 ± 0.054	-3.376 ± 0.266	0.073 ± 0.046	45
$W_{I,V,I}$ (OGLE)	-2.401 ± 0.058	-3.276 ± 0.275	0.163 ± 0.03	44
$W_{Ks,J,Ks}$ (VISTA)	-2.451 ± 0.057	-3.532 ± 0.282	0.134 ± 0.033	37
$W_{J,Y,Ks}$ (VISTA)	-2.476 ± 0.064	-3.662 ± 0.276	0.133 ± 0.039	30

Bibliography

Abbott, T. M. C., Adamów, M., Aguena, M., et al. 2021, The Dark Energy Survey Data Release 2, ApJS, 255, 20 21, 25, 43

Ablimit, I., Zhao, G., Flynn, C., & Bird, S. A. 2020, The Rotation Curve, Mass Distribution, and Dark Matter Content of the Milky Way from Classical Cepheids, ApJ, 895, L12 1

Aherne, F. J., Thacker, N. A., & Rockett, P. I. 1998, The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data, Kybernetika, 34, [363] 45

Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra, ApJS, 249, 3 107

Anders, F., Khalatyan, A., Queiroz, A. B. A., et al. 2022, Photo-Astrometric Distances, Extinctions, and Astrophysical Parameters for Gaia EDR3 Stars Brighter than G = 18.5, A&A, 658, A91 3

Andrae, R. 2010, Error Estimation in Astronomy: A Guide 122

Arenou, F. & Luri, X. 1999, Distances and Absolute Magnitudes from Trigonometric Parallaxes, 167, 13 129

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package, ApJ, 935, 167 240

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package, AJ, 156, 123 240

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, Astropy: A Community Python Package for Astronomy, A&A, 558, A33 240

Baade, W. 1946, A Search For the Nucleus of Our Galaxy, PASP, 58, 249 11

Baade, W. 1952, Extragalactic nebulae, Trs. IAU, 8, 397 7, 10

Babusiaux, C., Fabricius, C., Khanna, S., et al. 2023, Gaia Data Release 3. Catalogue Validation, A&A, 674, A32 39, 106, 119

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, Estimating Distances from Parallaxes. V. Geometric and Photogeometric Distances to 1.47 Billion Stars in Gaia Early Data Release 3, AJ, 161, 147 72, 129

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., & Andrae, R. 2018, Estimating Distance from Parallaxes. IV. Distances to 1.33 Billion Stars in Gaia Data Release 2, AJ, 156, 58 72

Bailey, S. I. 1902, A Discussion of Variable Stars in the Cluster ω Centauri, Annals of Harvard College Observatory, 38, 1 7, 10, 111

Bailey, S. I., Leland, E. F., Woods, I. E., & Pickering, E. C. 1919, Variable Stars in the Cluster Messier 15, Annals of Harvard College Observatory, 78, 195 7

Balick, B. & Brown, R. L. 1974, Intense Sub-Arcsecond Structure in the Galactic Center., ApJ, 194, 265 11

Baumgardt, H. & Vasiliev, E. 2021, Accurate Distances to Galactic Globular Clusters through a Combination of Gaia EDR3, HST, and Literature Data, MNRAS, 505, 5957 111, 113

Beaton, R. L., Bono, G., Braga, V. F., et al. 2018, Old-Aged Primary Distance Indicators, ß, 214, 113 19

Bellinger, E. P., Kanbur, S. M., Bhardwaj, A., & Marconi, M. 2020, When a Period Is Not a Full Stop: Light-curve Structure Reveals Fundamental Parameters of Cepheid and RR Lyrae Stars, MNRAS, 491, 4752 15

Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, The Zwicky Transient Facility: System Overview, Performance, and First Results, PASP, 131, 018002 21, 25, 34

Benkő, J. M., Plachy, E., Netzel, H., et al. 2023, Time Series Analysis of Bright TESS RRc Stars: Additional Modes, Phase Variations, and More, MNRAS, 521, 443 7, 87, 93

Bentley, J. L. 1975, Multidimensional Binary Search Trees Used for Associative Searching, Communications of the ACM, 18, 509 45

Bhardwaj, A. 2020, High-Precision Distance Measurements with Classical Pulsating Stars, Journal of Astrophysics and Astronomy, 41, 23 1, 4, 19, 130

Bhardwaj, A. 2022, RR Lyrae and Type II Cepheid Variables in Globular Clusters: Optical and Infrared Properties, Universe, 8, 122 111, 149

Bhardwaj, A., Macri, L. M., Rejkuba, M., et al. 2017, Large Magellanic Cloud Near-infrared Synoptic Survey. IV. Leavitt Laws for Type II Cepheid Variables, AJ, 153, 154 98

Bhardwaj, A., Marconi, M., Rejkuba, M., et al. 2023, Precise Empirical Determination of Metallicity Dependence of Near-infrared Period-Luminosity Relations for RR Lyrae Variables, ApJ, 944, L51 98

Bhardwaj, A., Rejkuba, M., de Grijs, R., et al. 2020, Near-Infrared Census of RR Lyrae Variables in the Messier 3 Globular Cluster and the Period-Luminosity Relations, AJ, 160, 220 125

Bhattacharyya, A. 1943, On a Measure of Divergence between Two Statistical Populations Defined by Their Probability Distributions., Bulletin of the Calcutta Mathematical Society, 35, 99 45

Bhattacharyya, A. 1946, On a Measure of Divergence between Two Multinomial Populations, Sankhyā: The Indian Journal of Statistics, 7, 401 45

Bland-Hawthorn, J. & Gerhard, O. 2016, The Galaxy in Context: Structural, Kinematic, and Integrated Properties, Annual Review of Astronomy and Astrophysics, 54, 529 2

Bland-Hawthorn, J., Sharma, S., Tepper-Garcia, T., et al. 2019, The GALAH Survey and Gaia DR2: Dissecting the Stellar Disc's Phase Space by Age, Action, Chemistry, and Location, MNRAS, 486, 1167 12

Blažko, S. 1907, Mitteilung Über Veränderliche Sterne, Astronomische Nachrichten, 175, 325 7

Boch, T., Pineau, F., & Derriere, S. 2012, The CDS Cross-Match Service, 461, 291 50

Bódi, A. & Kiss, L. L. 2019, Physical Properties of Galactic RV Tauri Stars from Gaia DR2 Data, ApJ, 872, 60 11

Bonaca, A., Naidu, R. P., Conroy, C., et al. 2021, Orbital Clustering Identifies the Origins of Galactic Stellar Streams, ApJ, 909, L26 12, 15, 17, 148

Bono, G., Braga, V. F., Fiorentino, G., et al. 2020, Evolutionary and Pulsation Properties of Type II Cepheids, A&A, 644, A96 1, 10

Bono, G., Caputo, F., & Di Criscienzo, M. 2007, RR Lyrae Stars in Galactic Globular Clusters. VI. The Period-Amplitude Relation, A&A, 476, 779 111

Bono, G., Caputo, F., Marconi, M., & Musella, I. 2010, Insights into the Cepheid Distance Scale, ApJ, 715, 277 7, 98

Bono, G., Caputo, F., Santolamazza, P., Cassisi, S., & Piersimoni, A. 1997, Evolutionary Scenario for Metal-Poor Pulsating Stars.II.Anomalous Cepheids, AJ, 113, 2209 1, 10

Bovy, J. 2015, Galpy: A Python Library for Galactic Dynamics, ApJS, 216, 29 240

Bovy, J., Hogg, D. W., & Roweis, S. T. 2011, Extreme Deconvolution: Inferring Complete Distribution Functions from Noisy, Heterogeneous and Incomplete Observations, Annals of Applied Statistics, 5, 1657 137

Bovy, J., Rix, H.-W., Green, G. M., Schlafly, E. F., & Finkbeiner, D. P. 2016, On Galactic Density Modeling in the Presence of Dust Extinction, ApJ, 818, 130 119

Braga, V. F., Bono, G., Fiorentino, G., et al. 2020, Separation between RR Lyrae and Type II Cepheids and Their Importance for a Distance Determination: The Case of Omega Cen, A&A, 644, A95 10, 20

Braga, V. F., Crestani, J., Fabrizio, M., et al. 2021, On the Use of Field RR Lyrae as Galactic Probes. V. Optical and Radial Velocity Curve Templates, ApJ, 919, 85 105

Braga, V. F., Stetson, P. B., Bono, G., et al. 2016, On the RR Lyrae Stars in Globulars. IV. ω Centauri Optical UBVRI Photometry, AJ, 152, 170 113

Braga, V. F., Stetson, P. B., Bono, G., et al. 2019, New Near-Infrared JHKs Light-Curve Templates for RR Lyrae Variables, A&A, 625, A1 7, 21, 98

Braga, V. F., Stetson, P. B., Bono, G., et al. 2018, On the RR Lyrae Stars in Globulars. V. The Complete Near-infrared (JHK s) Census of ω Centauri RR Lyrae Variables, AJ, 155, 137 125

Breuval, L., Kervella, P., Wielgórski, P., et al. 2021, The Influence of Metallicity on the Leavitt Law from Geometrical Distances of Milky Way and Magellanic Cloud Cepheids, ApJ, 913, 38 147

Breuval, L., Riess, A. G., Kervella, P., Anderson, R. I., & Romaniello, M. 2022, An Improved Calibration of the Wavelength Dependence of Metallicity on the Cepheid Leavitt Law, ApJ, 939, 89 147

Buder, S., Sharma, S., Kos, J., et al. 2021, The GALAH+ Survey: Third Data Release, MNRAS, 506, 150 107

Cantat-Gaudin, T. & Anders, F. 2020, Clusters and Mirages: Cataloguing Stellar Aggregates in the Milky Way, A&A, 633, A99 112

Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, Painting a Portrait of the Galactic Disc with Its Stellar Clusters, A&A, 640, A1 111, 112, 113

Caputo, F., Bono, G., Fiorentino, G., Marconi, M., & Musella, I. 2005, Pulsation and Evolutionary Masses of Classical Cepheids. I. Milky Way Variables, ApJ, 629, 1021 1

Caputo, F., Castellani, V., Degl'Innocenti, S., Fiorentino, G., & Marconi, M. 2004, Bright Metal-Poor Variables: Why "Anomalous" Cepheids?, A&A, 424, 927 7, 10, 144

Carnerero, M. I., Raiteri, C. M., Rimoldini, L., et al. 2023, Gaia Data Release 3. The First Gaia Catalogue of Variable AGN, A&A, 674, A24 91, 93

Carroll, B. W. & Ostlie, D. A. 1996, An Introduction to Modern Astrophysics (Cambridge University Press) 2

Carroll, B. W. & Ostlie, D. A. 2017, An Introduction to Modern Astrophysics, Second Edition (Cambridge University Press) 2

Cassisi, S. & Salaris, M. 2013, Old Stellar Populations: How to Study the Fossil Record of Galaxy Formation (Wiley-VCH) 1

Castro-Ginard, A., Jordi, C., Luri, X., et al. 2020, Hunting for Open Clusters in Gaia DR2: 582 New Open Clusters in the Galactic Disc, A&A, 635, A45 112

Castro-Ginard, A., Jordi, C., Luri, X., et al. 2022, Hunting for Open Clusters in Gaia EDR3: 628 New Open Clusters Found with OCfinder, A&A, 661, A118 112

Castro-Ginard, A., Jordi, C., Luri, X., et al. 2018, A New Method for Unveiling Open Clusters in Gaia. New Nearby Open Clusters Confirmed by DR2, A&A, 618, A59 112

Catelan, M. & Smith, H. A. 2015, Pulsating Stars (Wiley-VCH) 1, 147

Ceraski, W. 1905, Trois Nouvelles Variables, Astronomische Nachrichten, 168, 29 111

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, The Pan-STARRS1 Surveys 21, 25, 41

Chen, T. & Guestrin, C. 2016, XGBoost: A Scalable Tree Boosting System 42

Chen, X., Wang, S., Deng, L., de Grijs, R., & Yang, M. 2018, Wide-Field Infrared Survey Explorer (WISE) Catalog of Periodic Variable Stars, ApJS, 237, 28 99, 101, 125

Chen, X., Wang, S., Deng, L., et al. 2020, The Zwicky Transient Facility Catalog of Periodic Variable Stars, ApJS, 249, 18 xv, 25, 34, 36, 71, 89, 93, 103

Christy, C. T., Jayasinghe, T., Stanek, K. Z., et al. 2023, The ASAS-SN Catalogue of Variable Stars X: Discovery of 116 000 New Variable Stars Using G-band Photometry, MNRAS, 519, 5271 25, 37, 89, 93

Christy, C. T., Jayasinghe, T., Stanek, K. Z., et al. 2022, Citizen ASAS-SN Data Release. I. Variable Star Classification Using Citizen Science, PASP, 134, 024201 37, 72

Cioni, M. R. L., Clementini, G., Girardi, L., et al. 2011, The VMC Survey. I. Strategy and First Data, A&A, 527, A116 100

Cirasuolo, M., Fairley, A., Rees, P., et al. 2020, MOONS: The New Multi-Object Spectrograph for the VLT, The Messenger, 180, 10 148

Clement, C. M., Muzzin, A., Dufton, Q., et al. 2001, Variable Stars in Galactic Globular Clusters, AJ, 122, 2587 111

Clementini, G., Ripepi, V., Garofalo, A., et al. 2023, Gaia Data Release 3. Specific Processing and Validation of All-Sky RR Lyrae and Cepheid Stars: The RR Lyrae Sample, A&A, 674, A18 20, 25, 39, 106, 122, 128, 133, 187

Clementini, G., Ripepi, V., Molinaro, R., et al. 2019, Gaia Data Release 2. Specific Characterisation and Validation of All-Sky Cepheids and RR Lyrae Stars, A&A, 622, A60 53, 74

Cox, J. P. 1980, Theory of Stellar Pulsation. (PSA-2), Volume 2, Vol. 2 (Princeton University Press) 1, 3

Crestani, J., Fabrizio, M., Braga, V. F., et al. 2021, On the Use of Field RR Lyrae as Galactic Probes. II. A New ΔS Calibration to Estimate Their Metallicity, ApJ, 908, 20 147, 148

Cross, N. J. G., Collins, R. S., Mann, R. G., et al. 2012, The VISTA Science Archive, A&A, 548, A119 239

Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST), Research in Astronomy and Astrophysics, 12, 1197 106

Cusano, F., Moretti, M. I., Clementini, G., et al. 2021, The VMC Survey - XLII. Near-infrared Period-Luminosity Relations for RR Lyrae Stars and the Structure of the Large Magellanic Cloud, MNRAS, 504, 1 100, 102

Cutri, R. M., Wright, E. L., Conrow, T., et al. 2021, VizieR Online Data Catalog: AllWISE Data Release (Cutri+ 2013), VizieR Online Data Catalog, II/328 98

Dalton, G. 2016, WEAVE: The Next Generation Spectroscopy Facility for the WHT, 507, 97 148

Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, The Dark Energy Survey: More than Dark Energy - an Overview, MNRAS, 460, 1270 25, 43

de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, 4MOST: Project Overview and Information for the First Call for Proposals, The Messenger, 175, 3 148

Dékány, I. & Grebel, E. K. 2022, Photometric Metallicity Prediction of Fundamental-mode RR Lyrae Stars in the Gaia Optical and K s Infrared Wave Bands by Deep Learning, ApJS, 261, 33 15

Dékány, I., Hajdu, G., Grebel, E. K., & Catelan, M. 2019, Into the Darkness: Classical and Type II Cepheids in the Zona Galactica Incognita, ApJ, 883, 58 11

Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, LAMOST Experiment for Galactic Understanding and Exploration (LEGUE) — The Survey's Science Plan, Research in Astronomy and Astrophysics, 12, 735 106

Dias, W. S., Alessi, B. S., Moitinho, A., & Lépine, J. R. D. 2002, New Catalogue of Optically Visible Open Clusters and Candidates, A&A, 389, 871 111

Distefano, E., Lanzafame, A. C., Brugaletta, E., et al. 2023, Gaia Data Release 3. Rotational Modulation and Patterns of Colour Variation in Solar-like Variables, A&A, 674, A20 88, 93

Drake, A. J., Beshore, E., Djorgovski, S. G., et al. 2012, The First Data Release of the Catalina Surveys, 219, 428.20 21, 25, 30, 31

Drake, A. J., Catelan, M., Djorgovski, S. G., et al. 2013a, Probing the Outer Galactic Halo with RR Lyrae from the Catalina Surveys, ApJ, 763, 32 25, 31, 34

Drake, A. J., Catelan, M., Djorgovski, S. G., et al. 2013b, Evidence for a Milky Way Tidal Stream Reaching Beyond 100 Kpc, ApJ, 765, 154 25, 31

Drake, A. J., Djorgovski, S. G., Catelan, M., et al. 2017, The Catalina Surveys Southern Periodic Variable Star Catalogue, MNRAS, 469, 3688 25, 31, 32, 34, 71, 89, 93

Drake, A. J., Graham, M. J., Djorgovski, S. G., et al. 2014, The Catalina Surveys Periodic Variable Star Catalog, ApJS, 213, 9 25, 31, 32, 89, 93

Drimmel, R., Cabrera-Lavers, A., & López-Corredoira, M. 2003, A Three-Dimensional Galactic Extinction Model, A&A, 409, 205 119

Drlica-Wagner, A., Carlin, J. L., Nidever, D. L., et al. 2021, The DECam Local Volume Exploration Survey: Overview and First Data Release, ApJS, 256, 2 103

Drlica-Wagner, A., Ferguson, P. S., Adamów, M., et al. 2022, The DECam Local Volume Exploration Survey Data Release 2, ApJS, 261, 38 103, 104

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. 1987, Hybrid Monte Carlo, Physics Letters B, 195, 216 131

ESA. 1997, The HIPPARCOS and TYCHO Catalogues. Astrometric and Photometric Star Catalogues Derived from the ESA HIPPARCOS Space Astrometry Mission, ESA Special Publication, 1200 47

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2022, First Sagittarius A* Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole in the Center of the Milky Way, ApJ, 930, L12 11

Eyer, L., Audard, M., Holl, B., et al. 2023, Gaia Data Release 3. Summary of the Variability Processing and Analysis, A&A, 674, A13 20, 52, 88, 128, 167

Fabricius, C., Luri, X., Arenou, F., et al. 2021, Gaia Early Data Release 3. Catalogue Validation, A&A, 649, A5 54

Fabrizio, M., Braga, V. F., Crestani, J., et al. 2021, On the Use of Field RR Lyrae As Galactic Probes: IV. New Insights Into and Around the Oosterhoff Dichotomy, ApJ, 919, 118 43

Fiorentino, G., Bono, G., Monelli, M., et al. 2015, Weak Galactic Halo-Dwarf Spheroidal Connection from RR Lyrae Stars, ApJ, 798, L12 42

Fiorentino, G. & Monelli, M. 2012, Anomalous Cepheids in the Large Magellanic Cloud. Insight into Their Origin and Connection with the Star Formation History, A&A, 540, A102 10

Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, The Dark Energy Camera, AJ, 150, 150 103

Flewelling, H. A., Magnier, E. A., Chambers, K. C., et al. 2020, The Pan-STARRS1 Database and Data Products, ApJS, 251, 7 25, 41

Freedman, W. L. & Madore, B. F. 2010, A Physically Based Method for Scaling Cepheid Light Curves for Future Distance Determinations, ApJ, 719, 335 98

Freedman, W. L. & Madore, B. F. 2023, The Cepheid Extragalactic Distance Scale: Past, Present and Future 20

Freedman, W. L., Madore, B. F., Gibson, B. K., et al. 2001, Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant, ApJ, 553, 47 7, 19

Gaia Collaboration, Bailer-Jones, C. A. L., Teyssier, D., et al. 2023a, Gaia Data Release 3. The Extragalactic Content, A&A, 674, A41 90, 93

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018a, Gaia Data Release 2. Summary of the Contents and Survey Properties, A&A, 616, A1 12

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021a, Gaia Early Data Release 3. Summary of the Contents and Survey Properties, A&A, 649, A1 54

Gaia Collaboration, De Ridder, J., Ripepi, V., et al. 2023b, Gaia Data Release 3. Pulsations in Main Sequence OBAF-type Stars, A&A, 674, A36 88, 93

Gaia Collaboration, Helmi, A., van Leeuwen, F., et al. 2018b, Gaia Data Release 2. Kinematics of Globular Clusters and Dwarf Galaxies around the Milky Way, A&A, 616, A12 15, 113, 115

Gaia Collaboration, Klioner, S. A., Lindegren, L., et al. 2022, Gaia Early Data Release 3. The Celestial Reference Frame (Gaia-CRF3), A&A, 667, A148 91

Gaia Collaboration, Luri, X., Chemin, L., et al. 2021b, Gaia Early Data Release 3. Structure and Properties of the Magellanic Clouds, A&A, 649, A7 12, 14, 15, 18, 114, 115, 118

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, The Gaia Mission, A&A, 595, A1 12, 20, 25

Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023c, Gaia Data Release 3. Summary of the Content and Survey Properties, A&A, 674, A1 12, 15, 21, 38, 128, 129

Gallagher, III, J. S. & Wyse, R. F. G. 1994, Dwarf Spheroidal Galaxies: Keystones of Galaxy Evolution, PASP, 106, 1225 113

Gallenne, A., Kervella, P., Borgniet, S., et al. 2019, Multiplicity of Galactic Cepheids from Long-Baseline Interferometry. IV. New Detected Companions from MIRC and PIONIER Observations, A&A, 622, A164 105

Gavras, P., Rimoldini, L., Nienartowicz, K., et al. 2023, Gaia Data Release 3. Cross-match of Gaia Sources with Variable Objects from the Literature, A&A, 674, A22 38

Gavrilchenko, T., Klein, C. R., Bloom, J. S., & Richards, J. W. 2014, A Mid-Infrared Study of RR Lyrae Stars with the Wide-field Infrared Survey Explorer All-Sky Data Release, MNRAS, 441, 715 98

Gelman, A. 2006, Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper), Bayesian Analysis, 1, 515 131, 132

Gelman, A., Carlin, J. B., Stern, H. S., et al. 2015, Bayesian Data Analysis, 3rd edn. (New York: Chapman and Hall/CRC) 132

Gilmore, G., Randich, S., Asplund, M., et al. 2012, The Gaia-ESO Public Spectroscopic Survey, The Messenger, 147, 25 107

Gilmore, G., Randich, S., Worley, C. C., et al. 2022, The Gaia-ESO Public Spectroscopic Survey: Motivation, Implementation, GIRAFFE Data Processing, Analysis, and Final Data Products, A&A, 666, A120 109

Gilmore, G. & Reid, N. 1983, New Light on Faint Stars - III. Galactic Structure towards the South Pole and the Galactic Thick Disc., MNRAS, 202, 1025 11

Gomel, R., Mazeh, T., Faigler, S., et al. 2023, Gaia Data Release 3. Ellipsoidal Variables with Possible Black Hole or Neutron Star Secondaries, A&A, 674, A19 88, 93

González-Fernández, C., Hodgkin, S. T., Irwin, M. J., et al. 2018, The VISTA ZYJHKs Photometric System: Calibration from 2MASS, MNRAS, 474, 5459 102

Graczyk, D., Pietrzyński, G., Thompson, I. B., et al. 2020, A Distance Determination to the Small Magellanic Cloud with an Accuracy of Better than Two Percent Based on Late-type Eclipsing Binary Stars, ApJ, 904, 13 111

Graham, M. J., Kulkarni, S. R., Bellm, E. C., et al. 2019, The Zwicky Transient Facility: Science Objectives, PASP, 131, 078001 25

Grebel, E. K., Gallagher, III, J. S., & Harbeck, D. 2003, The Progenitors of Dwarf Spheroidal Galaxies, AJ, 125, 1926 113

Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, A 3D Dust Map Based on Gaia, Pan-STARRS 1, and 2MASS, ApJ, 887, 93 119

Green, M. J., Maoz, D., Mazeh, T., et al. 2023, 15 000 Ellipsoidal Binary Candidates in TESS: Orbital Periods, Binary Fraction, and Tertiary Companions, MNRAS, 522, 29 xvi, 89, 90, 93

Groenewegen, M. 2023, Primary Period-Luminosity-Relation Calibrators in the Milky Way: Cepheids and RR Lyrae Physical Basis, Calibration, and Applications 121, 149

Groenewegen, M. & Lub, J. 2023, Spectral Energy Distributions of Classical Cepheids in the Magellanic Clouds 147

Groenewegen, M. a. T. & Jurkovic, M. I. 2017, The Period-Luminosity and Period-Radius Relations of Type II and Anomalous Cepheids in the Large and Small Magellanic Clouds, A&A, 604, A29 10

Hajdu, G., Pietrzyński, G., Jurcsik, J., et al. 2021, Studies of RR Lyrae Variables in Binary Systems. I. Evidence of a Trimodal Companion Mass Distribution, ApJ, 915, 50 92, 93, 95

Hambly, N., Andrae, R., De Angeli, F., et al. 2022, Gaia DR3 Documentation Chapter 20: Datamodel Description, Tech. rep. 52, 109

Han, J. J., Conroy, C., Johnson, B. D., et al. 2022, The Stellar Halo of the Galaxy Is Tilted and Doubly Broken, AJ, 164, 249 148

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Array Programming with NumPy, Nature, 585, 357 240

Haschke, R., Grebel, E. K., & Duffau, S. 2011, New Optical Reddening Maps of the Large and Small Magellanic Clouds, AJ, 141, 158 5

Hattori, K., Valluri, M., & Vasiliev, E. 2021, Action-Based Distribution Function Modelling for Constraining the Shape of the Galactic Dark Matter Halo, MNRAS, 508, 5468 1

Heinze, A. N., Tonry, J. L., Denneau, L., et al. 2018, A First Catalog of Variable Stars Measured by the Asteroid Terrestrial-impact Last Alert System (ATLAS), AJ, 156, 241 105

Hobbs, D., Lindegren, L., Bastian, U., et al. 2018, Gaia DR2 Documentation Chapter 3: Astrometry, Tech. rep. 47

Hocdé, V., Smolec, R., Moskalik, P., Ziółkowska, O., & Singh Rathour, R. 2023, Metallicity Estimations of MW, SMC, and LMC Classical Cepheids from the Shape of the V- and I-band Light Curves, A&A, 671, A157 147

Hoffman, M. D. & Gelman, A. 2011, The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo 131, 132

Huang, K.-W. & Koposov, S. E. 2021, Search for Globular Clusters Associated with the Milky Way Dwarf Galaxies Using Gaia DR2, MNRAS, 500, 986 113

Hubble, E. 1929, A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae, Proceedings of the National Academy of Science, 15, 168 1, 19

Hunt, E. L. & Reffert, S. 2021, Improving the Open Cluster Census. I. Comparison of Clustering Algorithms Applied to Gaia DR2 Data, A&A, 646, A104 112

Hunt, E. L. & Reffert, S. 2023, Improving the Open Cluster Census. II. An All-Sky Cluster Catalogue with Gaia DR3, A&A, 673, A114 111, 112

Hunter, J. D. 2007, Matplotlib: A 2D Graphics Environment, Computing in Science and Engineering, 9, 90 240

Ibata, R., Battaglia, G., Bellazzini, M., et al. 2023, 4MOST Gaia RR Lyrae Survey (4GRoundS), The Messenger, 190, 10 148

Ibata, R., Bellazzini, M., Thomas, G., et al. 2020, A Panoramic Landscape of the Sagittarius Stream in Gaia DR2 Revealed with the STREAMFINDER Spyglass, ApJ, 891, L19 1, 7

Inno, L., Matsunaga, N., Romaniello, M., et al. 2015, New NIR Light-Curve Templates for Classical Cepheids, A&A, 576, A30 98

Iorio, G. & Belokurov, V. 2019, The Shape of the Galactic Halo with Gaia DR2 RR Lyrae. Anatomy of an Ancient Major Merger, MNRAS, 482, 3868 7

Iorio, G. & Belokurov, V. 2021, Chemo-Kinematics of the Gaia RR Lyrae: The Halo and the Disc, MNRAS, 502, 5686 7

Iorio, G., Belokurov, V., Erkal, D., et al. 2018, The First All-Sky View of the Milky Way Stellar Halo with Gaia+2MASS RR Lyrae, MNRAS, 474, 2142 7

Irwin, J. B. 1955, Cepheid Variables and Galactic Structure, Monthly Notes of the Astronomical Society of South Africa, 14, 38 111

Irwin, M. J., Lewis, J., Hodgkin, S., et al. 2004, VISTA Data Flow System: Pipeline Processing for WFCAM and VISTA, 5493, 411 126, 239

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, LSST: From Science Drivers to Reference Design and Anticipated Data Products, ApJ, 873, 111 102

Iwanek, P., Soszyński, I., Kozłowski, S., et al. 2022, The OGLE Collection of Variable Stars: Nearly 66,000 Mira Stars in the Milky Way, ApJS, 260, 46 89, 93

Iwanek, P., Soszyński, I., Skowron, D., et al. 2018, Three-Dimensional Distributions of Type II Cepheids and Anomalous Cepheids in the Magellanic Clouds. Do These Stars Belong to the Old, Young or Intermediate-Age Population?, Acta Astron., 68, 213 115

Jacyszyn-Dobrzeniecka, A. M., Mróz, P., Kruszyńska, K., et al. 2020a, OGLE-ing the Magellanic System: RR Lyrae Stars in the Bridge, ApJ, 889, 26 7

Jacyszyn-Dobrzeniecka, A. M., Skowron, D. M., Mróz, P., et al. 2016, OGLE-ing the Magellanic System: Three-Dimensional Structure of the Clouds and the Bridge Using Classical Cepheids, Acta Astron., 66, 149 7, 115, 138

Jacyszyn-Dobrzeniecka, A. M., Skowron, D. M., Mróz, P., et al. 2017, OGLE-ing the Magellanic System: Three-Dimensional Structure of the Clouds and the Bridge Using RR Lyrae Stars, Acta Astron., 67, 1 129, 138

Jacyszyn-Dobrzeniecka, A. M., Soszyński, I., Udalski, A., et al. 2020b, OGLE-ing the Magellanic System: Cepheids in the Bridge, ApJ, 889, 25 12

Jayasinghe, T., Kochanek, C. S., Stanek, K. Z., et al. 2018, The ASAS-SN Catalogue of Variable Stars I: The Serendipitous Survey, MNRAS, 477, 3145 25, 37, 89, 93

Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019a, The ASAS-SN Catalogue of Variable Stars - II. Uniform Classification of 412 000 Known Variables, MNRAS, 486, 1907 37

Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019b, The ASAS-SN Catalogue of Variable Stars III: Variables in the Southern TESS Continuous Viewing Zone, MNRAS, 485, 961 37

Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2020, The ASAS-SN Catalogue of Variable Stars - V. Variables in the Southern Hemisphere, MNRAS, 491, 13 25, 37, 93

Ji, A. P., Li, T. S., Hansen, T. T., et al. 2020, The Southern Stellar Stream Spectroscopic Survey (S5): Chemical Abundances of Seven Stellar Streams, AJ, 160, 181 15

Jönsson, H., Holtzman, J. A., Allende Prieto, C., et al. 2020, APOGEE Data and Spectral Analysis from SDSS Data Release 16: Seven Years of Observations Including First Results from APOGEE-South, AJ, 160, 120 107

Joy, A. H. 1937, Radial Velocities of Cepheid Variable Stars, ApJ, 86, 363 10

Juárez, M. A. & Steel, M. F. J. 2010, Model-Based Clustering of Non-Gaussian Panel Data Based on Skew-t Distributions, Journal of Business & Economic Statistics, 28, 52 131

Jurcsik, J., Hajdu, G., Szeidl, B., et al. 2012, Long-Term Photometric Monitoring of RR Lyrae Stars in Messier 3, MNRAS, 419, 2173 7

Jurkovic, M. I. 2018, Anomalous Cepheids Discovered in a Sample of Galactic Short Period Type II Cepheids, Serbian Astronomical Journal, 197, 13 86

Jurkovic, M. I., Plachy, E., Molnár, L., et al. 2023, Type II and Anomalous Cepheids in the Kepler K2 Mission, MNRAS, 518, 642 86

Kailath, T. 1967, The Divergence and Bhattacharyya Distance Measures in Signal Selection, IEEE Transactions on Communication Technology, 15, 52 45

Katz, D., Sartoretti, P., Guerrier, A., et al. 2023, Gaia Data Release 3. Properties and Validation of the Radial Velocities, A&A, 674, A5 106, 110

Kim, D.-W. & Bailer-Jones, C. A. L. 2016, A Package for the Automated Classification of Periodic Variable Stars, A&A, 587, A18 37

Kiss, L. L. & Bódi, A. 2017, Amplitude Variations of Modulated RV Tauri Stars Support the Dust Obscuration Model of the RVb Phenomenon, A&A, 608, A99 11

Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, Jupyter Notebooks—a Publishing Format for Reproducible Computational Workflows, 87–90 240

Kochanek, C. S., Shappee, B. J., Stanek, K. Z., et al. 2017, The All-Sky Automated Survey for Supernovae (ASAS-SN) Light Curve Server v1.0, PASP, 129, 104502 21, 25, 37

Kodric, M., Riffeser, A., Hopp, U., et al. 2018, Cepheids in M31: The PAndromeda Cepheid Sample, AJ, 156, 130 11

Koposov, S. E., Erkal, D., Li, T. S., et al. 2023, S 5: Probing the Milky Way and Magellanic Clouds Potentials with the 6-D Map of the Orphan-Chenab Stream, MNRAS 7

Kordopatis, G., Gilmore, G., Steinmetz, M., et al. 2013, The Radial Velocity Experiment (RAVE): Fourth Data Release, AJ, 146, 134 107

Kovacs, G., Kisvarsanyi, E. G., & Buchler, J. R. 1990, Cepheid Radial Velocity Curves Revisited, ApJ, 351, 606 105

Krone-Martins, A. & Moitinho, A. 2014, UPMASK: Unsupervised Photometric Membership Assignment in Stellar Clusters, A&A, 561, A57 112

Kruschke, J. 2015, Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan, Academic Press (Academic Press) 132

Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. 2019, ArviZ a unified library for exploratory analysis of Bayesian models in Python, Journal of Open Source Software, 4, 1143 240

Kunder, A., Koch, A., Rich, R. M., et al. 2012, The Bulge Radial Velocity Assay (BRAVA). II. Complete Sample and Data Release, AJ, 143, 57 109

Kunder, A., Pérez-Villegas, A., Rich, R. M., et al. 2020, The Bulge Radial Velocity Assay for RR Lyrae Stars (BRAVA-RR) DR2: A Bimodal Bulge?, AJ, 159, 270 7, 109

Lala, H. N., Lemasle, B., Grebel, E. K., et al. 2023, RRLCep: A Catalog of RR Lyrae Stars and Cepheids. I. A Comprehensive Compilation of Literature Surveys, submitted (Paper I) 125

Lallement, R., Babusiaux, C., Vergely, J. L., et al. 2019, Gaia-2MASS 3D Maps of Galactic Interstellar Dust within 3 Kpc, A&A, 625, A135 119

Lallement, R., Vergely, J. L., Babusiaux, C., & Cox, N. L. J. 2022, Updated Gaia-2MASS 3D Maps of Galactic Interstellar Dust, A&A, 661, A147 119

Lampens, P. 2021, Eclipsing Systems with Pulsating Components (Types β Cep, δ Sct, γ Dor or Red Giant) in the Era of High-Accuracy Space Data, Galaxies, 9, 28 95

Leavitt, H. S. 1908, 1777 Variables in the Magellanic Clouds, Annals of Harvard College Observatory, 60, 87 7, 111, 121

Leavitt, H. S. & Pickering, E. C. 1912, Periods of 25 Variable Stars in the Small Magellanic Cloud., Harvard College Observatory Circular, 173, 1 7, 111, 121

Lebzelter, T., Mowlavi, N., Lecoeur-Taibi, I., et al. 2023, Gaia Data Release 3. The Second Gaia Catalogue of Long-Period Variable Candidates, A&A, 674, A15 88, 93

Lee, M. G., Freedman, W. L., & Madore, B. F. 1993, The Tip of the Red Giant Branch as a Distance Indicator for Resolved Galaxies, ApJ, 417, 553 19

Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008, The SEGUE Stellar Parameter Pipeline. I. Description and Comparison of Individual Methods, AJ, 136, 2022 107

Lemasle, B., François, P., Genovali, K., et al. 2013, Galactic Abundance Gradients from Cepheids. α and Heavy Elements in the Outer Disk, A&A, 558, A31 1

Lemasle, B., Lala, H. N., Kovtyukh, V., et al. 2022, Tracing the Milky Way Warp and Spiral Arms with Classical Cepheids, A&A, 668, A40 1, 15, 18, 98, 100, 125, 148

Lewis, J. R., Irwin, M., & Bunclark, P. 2010, Pipeline Processing for VISTA, 434, 91 239

Li, T. S., Koposov, S. E., Zucker, D. B., et al. 2019, The Southern Stellar Stream Spectroscopic Survey (S5): Overview, Target Selection, Data Reduction, Validation, and Early Science, MNRAS, 490, 3508 15, 107

Lindegren, L., Bastian, U., Biermann, M., et al. 2021a, Gaia Early Data Release 3. Parallax Bias versus Magnitude, Colour, and Position, A&A, 649, A4 149

Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021b, Gaia Early Data Release 3. The Astrometric Solution, A&A, 649, A2 15, 129

Lindegren, L., Lammers, U., Hobbs, D., et al. 2012, The Astrometric Core Solution for the Gaia Mission. Overview of Models, Algorithms, and Software Implementation, A&A, 538, A78 54

Liu, C., Fu, J., Shi, J., et al. 2020, LAMOST Medium-Resolution Spectroscopic Survey (LAMOST-MRS): Scientific Goals and Survey Plan 106

Luri, X., Brown, A. G. A., Sarro, L. M., et al. 2018, Gaia Data Release 2. Using Gaia Parallaxes, A&A, 616, A9 129

Lyke, B. W., Higley, A. N., McLane, J. N., et al. 2020, The Sloan Digital Sky Survey Quasar Catalog: Sixteenth Data Release, ApJS, 250, 8 91, 93

Madore, B. F. 1976, A Reddening-independent Formulation of the Period-Luminosity Relation: The Wesenheit Function, Royal Greenwich Observatory Bulletins, 182, 153 5, 126

Madore, B. F. 1982, The Period-Luminosity Relation. IV. Intrinsic Relations and Reddenings for the Large Magellanic Cloud Cepheids., ApJ, 253, 575 5, 126

Madore, B. F. & Freedman, W. L. 1991, The Cepheid Distance Scale, PASP, 103, 933 1, 2, 3, 5, 7, 126

Madore, B. F. & Freedman, W. L. 2005, Nonuniform Sampling and Periodic Signal Detection, ApJ, 630, 1054 122

Magnier, E. A., Chambers, K. C., Flewelling, H. A., et al. 2020a, The Pan-STARRS Data-processing System, ApJS, 251, 3 25

Magnier, Eugene. A., Schlafly, Edward. F., Finkbeiner, D. P., et al. 2020b, Pan-STARRS Photometric and Astrometric Calibration, ApJS, 251, 6 25, 42

Majaess, D., Turner, D., & Lane, D. 2009, Type II Cepheids as Extragalactic Distance Candles, Acta Astron., 59, 403 11

Malhan, K. & Ibata, R. A. 2018, STREAMFINDER - I. A New Algorithm for Detecting Stellar Streams, MNRAS, 477, 4063 15

Malhan, K., Ibata, R. A., Sharma, S., et al. 2022, The Global Dynamical Atlas of the Milky Way Mergers: Constraints from Gaia EDR3-based Orbits of Globular Clusters, Stellar Streams, and Satellite Galaxies, ApJ, 926, 107 12, 15

Maneewongvatana, S. & Mount, D. M. 1999, Analysis of Approximate Nearest Neighbor Searching with Clustered Point Sets 45

Marconi, M., Molinaro, R., Ripepi, V., et al. 2017, The VMC Survey - XXIII. Model Fitting of Light and Radial Velocity Curves of Small Magellanic Cloud Classical Cepheids, MNRAS, 466, 3206 105

Marocco, F., Eisenhardt, P. R. M., Fowler, J. W., et al. 2021, The CatWISE2020 Catalog, ApJS, 253, 8 99, 125

Marrese, P. M., Marinoni, S., Fabrizio, M., & Altavilla, G. 2019, Gaia Data Release 2. Cross-match with External Catalogues: Algorithms and Results, A&A, 621, A144 153

Marrese, P. M., Marinoni, S., Fabrizio, M., & Giuffrida, G. 2017, Gaia Data Release 1. Cross-match with External Catalogues. Algorithm and Results, A&A, 607, A105 153

Marshall, D. J., Robin, A. C., Reylé, C., Schultheis, M., & Picaud, S. 2006, Modelling the Galactic Interstellar Extinction Distribution in Three Dimensions, A&A, 453, 635 119

Martínez-Vázquez, C. E., Monelli, M., Bernard, E. J., et al. 2017, The ISLAnds Project. III. Variable Stars in Six Andromeda Dwarf Spheroidal Galaxies, ApJ, 850, 137 111, 113

Matsunaga, N., Feast, M. W., & Menzies, J. W. 2009, Period-Luminosity Relations for Type II Cepheids and Their Application, MNRAS, 397, 933 11

Matsunaga, N., Feast, M. W., & Soszyński, I. 2011, Period-Luminosity Relations of Type II Cepheids in the Magellanic Clouds, MNRAS, 413, 223 98

Matsunaga, N., Fukushi, H., Nakada, Y., et al. 2006, The Period-Luminosity Relation for Type II Cepheids in Globular Clusters, MNRAS, 370, 1979 11, 98

McConnachie, A. W. 2012, The Observed Properties of Dwarf Galaxies in and around the Local Group, AJ, 144, 4 113

McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2013, First Scientific Results from the VISTA Hemisphere Survey (VHS), The Messenger, 154, 35 100, 239

Medina, G. E., Lemasle, B., & Grebel, E. K. 2021, A Revisited Study of Cepheids in Open Clusters in the Gaia Era, MNRAS, 505, 1342 112

Meisner, A. M., Lang, D., Schlafly, E. F., & Schlegel, D. J. 2022, 9-Yr Deep Sky unWISE Coadds, Research Notes of the American Astronomical Society, 6, 188 148

Minniti, D., Lucas, P. W., Emerson, J. P., et al. 2010, VISTA Variables in the Via Lactea (VVV): The Public ESO near-IR Variability Survey of the Milky Way, New Astronomy, 15, 433 100

Molnar, T. A., Sanders, J. L., Smith, L. C., et al. 2022, Variable Star Classification across the Galactic Bulge and Disc with the VISTA Variables in the Vía Láctea Survey, MNRAS, 509, 2566 102

Morganson, E., Gruendl, R. A., Menanteau, F., et al. 2018, The Dark Energy Survey Image Processing Pipeline, PASP, 130, 074501 104

Mowlavi, N., Holl, B., Lecoeur-Taïbi, I., et al. 2023, Gaia Data Release 3. The First Gaia Catalogue of Eclipsing-Binary Candidates, A&A, 674, A16 88, 93

Murphy, S. J. 2018, Pulsating Stars in Binary Systems: A Review 95

Naidu, R. P., Conroy, C., Bonaca, A., et al. 2020, Evidence from the H3 Survey That the Stellar Halo Is Entirely Comprised of Substructure, ApJ, 901, 48 2, 12, 15, 148

Naidu, R. P., Conroy, C., Bonaca, A., et al. 2021, Reconstructing the Last Major Merger of the Milky Way with the H3 Survey, arXiv e-prints, 2103, arXiv:2103.03251 12, 15

Navarrete, C., Catelan, M., Contreras Ramos, R., et al. 2017, Near-IR Period-Luminosity Relations for Pulsating Stars in ω Centauri (NGC 5139), A&A, 604, A120 113

Neeley, J. R., Marengo, M., Bono, G., et al. 2015, On the Distance of the Globular Cluster M4 (NGC 6121) Using RR Lyrae Stars. II. Mid-infrared Period-luminosity Relations., ApJ, 808, 11 122

Neeser, M., Lewis, J., Madsen, G., et al. 2016, Science-Grade Imaging Data for HAWK-I, VIMOS, and VIRCAM: The ESO-UK Pipeline Collaboration, The Messenger, 166, 36 126

Nemec, J. M. 1984, Studies of Extragalactic RR Lyrae Stars, Studies of Extragalactic RR Lyrae Stars, PhD thesis 7

Netzel, H. & Smolec, R. 2019, The Census of Non-Radial Pulsation in First-Overtone RR Lyrae Stars of the OGLE Galactic Bulge Collection, MNRAS, 487, 5584 87

Netzel, H. & Smolec, R. 2022, Asteroseismology of RR Lyrae Stars with Non-Radial Modes, MNRAS, 515, 3439 93

Netzel, H., Smolec, R., Soszyński, I., & Udalski, A. 2018, Blazhko Effect in the First Overtone RR Lyrae Stars of the OGLE Galactic Bulge Collection, MNRAS, 480, 1229 7, 87, 93

Ngeow, C.-C., Bhardwaj, A., Dekany, R., et al. 2022a, Zwicky Transient Facility and Globular Clusters: The RR Lyrae Gri-Band Period-Luminosity-Metallicity and Period-Wesenheit-Metallicity Relations, AJ, 163, 239 111, 114

Ngeow, C.-C., Bhardwaj, A., Graham, M. J., et al. 2022b, Zwicky Transient Facility and Globular Clusters: The Period-Luminosity and Period-Wesenheit Relations for Anomalous Cepheids Supplemented with Large Magellanic Cloud Sample, AJ, 164, 191 111, 114

Ngeow, C.-C., Bhardwaj, A., Henderson, J.-Y., et al. 2022c, Zwicky Transient Facility and Globular Clusters: The Period-Luminosity and Period-Wesenheit Relations for Type II Cepheids, AJ, 164, 154 19, 111, 114

Nidever, D. L., Dey, A., Fasbender, K., et al. 2021, Second Data Release of the All-sky NOIRLab Source Catalog, AJ, 161, 192 104

Nidever, D. L., Dey, A., Olsen, K., et al. 2018, First Data Release of the All-sky NOAO Source Catalog, AJ, 156, 131 104

Nidever, D. L., Olsen, K., Walker, A. R., et al. 2017, SMASH: Survey of the MAgellanic Stellar History, AJ, 154, 199 104

Oosterhoff, P. T. 1939, Some Remarks on the Variable Stars in Globular Clusters, The Observatory, 62, 104 43

Oriol, A.-P., Virgile, A., Colin, C., et al. 2023, PyMC: A Modern and Comprehensive Probabilistic Programming Framework in Python, PeerJ Computer Science, 9, e1516 240

Paczynski, B. 1986, Gravitational Microlensing by the Galactic Halo, ApJ, 304, 1 27

pandas development team, T. 2020, pandas-dev/pandas: Pandas, Zenodo 240

Pawlak, M., Soszyński, I., Udalski, A., et al. 2016, The OGLE Collection of Variable Stars. Eclipsing Binaries in the Magellanic System, Acta Astron., 66, 421 89, 93

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Scikit-Learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825 116

Pérez, F. & Granger, B. E. 2007, IPython: a System for Interactive Scientific Computing, Computing in Science and Engineering, 9, 21 240

Pietrukowicz, P., Mróz, P., Soszyński, I., et al. 2013, Eclipsing Binary Stars in the OGLE-III Galactic Disk Fields, Acta Astron., 63, 115 89, 93

Pietrukowicz, P., Ratajczak, M., Soszyński, I., et al. 2022, New Short-Period δ Scuti Stars in OGLE-IV Fields toward the Galactic Bulge, Acta Astron., 72, 161 89, 93

Pietrukowicz, P., Soszyński, I., Netzel, H., et al. 2020, Over 10000 δ Scuti Stars toward the Galactic Bulge from OGLE-IV, Acta Astron., 70, 241 89, 93

Pietrukowicz, P., Soszyński, I., & Udalski, A. 2021, Classical Cepheids in the Milky Way, Acta Astron., 71, 205 7, 34, 71, 80, 82, 83

Pietrzyński, G., Graczyk, D., Gallenne, A., et al. 2019, A Distance to the Large Magellanic Cloud That Is Precise to One per Cent, Nature, 567, 200 111, 129, 133, 137, 139, 144

Pineau, F.-X., Boch, T., Derrière, S., & Schaaff, A. 2020, The CDS Cross-match Service: Key Figures, Internals and Future Plans, 522, 125 50

Pineau, F.-X., Derriere, S., Motch, C., et al. 2017, Probabilistic Multi-Catalogue Positional Cross-Match, A&A, 597, A89 42, 153

Plachy, E., Pál, A., Bódi, A., et al. 2021, TESS Observations of Cepheid Stars: First Light Results, ApJS, 253, 11 11, 15, 86

Planck, M. 1901, Ueber Das Gesetz Der Energieverteilung Im Normalspectrum, Annalen der Physik, 309, 553 2

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, Planck 2018 Results. VI. Cosmological Parameters, A&A, 641, A6 19

Pojmanski, G. 1997, The All Sky Automated Survey, Acta Astron., 47, 467 21, 25, 40

Pojmanski, G. 2002, The All Sky Automated Survey. Catalog of Variable Stars. I. 0 h - 6 hQuarter of the Southern Hemisphere, Acta Astron., 52, 397 25, 40

Pojmanski, G., Pilecki, B., & Szczygiel, D. 2005, The All Sky Automated Survey. Catalog of Variable Stars. V. Declinations 0 Arcd - +28 Arcd of the Northern Hemisphere, Acta Astron., 55, 275 40

Poleski, R., Soszyński, I., Udalski, A., et al. 2010, The Optical Gravitational Lensing Experiment. The OGLE-III Catalog of Variable Stars. VI. Delta Scuti Stars in the Large Magellanic Cloud, Acta Astron., 60, 1 89, 93

Price-Whelan, A., Sipőcz, B., Wagg, T., et al. 2022, Adrn/Gala: V1.6.1, Zenodo 240

Price-Whelan, A. M. 2017, Gala: A Python Package for Galactic Dynamics, The Journal of Open Source Software, 2, 388 240

Price-Whelan, A. M., Hogg, D. W., Rix, H.-W., et al. 2018, Binary Companions of Evolved Stars in APOGEE DR14: Search Method and Catalog of ~5000 Companions, AJ, 156, 18 107

Prudil, Z., Hanke, M., Lemasle, B., et al. 2021, Milky Way Archaeology Using RR Lyrae and Type II Cepheids. I. The Orphan Stream in 7D Using RR Lyrae Stars, A&A, 648, A78 15, 148

Prudil, Z., Koch-Hansen, A. J., Lemasle, B., et al. 2022, Milky Way Archaeology Using RR Lyrae and Type II Cepheids. II. High-velocity RR Lyrae Stars and Milky Way Mass, A&A, 664, A148 109, 148

Prudil, Z. & Skarka, M. 2017, Blazhko Effect in the Galactic Bulge Fundamental Mode RR Lyrae Stars - I. Incidence Rate and Differences between Modulated and Non-Modulated Stars, MNRAS, 466, 2602 95

Ramos, P., Mateu, C., Antoja, T., et al. 2020, Full 5D Characterisation of the Sagittarius Stream with Gaia DR2 RR Lyrae, A&A, 638, A104 7

Randich, S., Gilmore, G., & Gaia-ESO Consortium. 2013, The Gaia-ESO Large Public Spectroscopic Survey, The Messenger, 154, 47 107

Rathour, R. S., Smolec, R., & Netzel, H. 2021, Frequency Analysis of OGLE-IV Photometry for Classical Cepheids in Galactic Fields: Non-Radial Modes and Modulations, MNRAS, 505, 5412 87, 93

Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2019, Trigonometric Parallaxes of High-mass Star-forming Regions: Our View of the Milky Way, ApJ, 885, 131 12, 15

Reyes, M. C. & Anderson, R. I. 2023, A 0.9% Calibration of the Galactic Cepheid Luminosity Scale Based on Gaia DR3 Data of Open Clusters and Cepheids, A&A, 672, A85 112, 113

Richards, J. W., Starr, D. L., Butler, N. R., et al. 2011, On Machine-learned Classification of Variable Stars with Sparse and Noisy Time-series Data, ApJ, 733, 10 40, 81

Richards, J. W., Starr, D. L., Miller, A. A., et al. 2012, Construction of a Calibrated Probabilistic Classification Catalog: Application to 50k Variable Sources in the All-Sky Automated Survey, ApJS, 203, 32 25, 40, 41

Riello, M., De Angeli, F., Evans, D. W., et al. 2021, Gaia Early Data Release 3. Photometric Content and Validation, A&A, 649, A3 53, 54

Riess, A. G., Anand, G. S., Yuan, W., et al. 2023, Crowded No More: The Accuracy of the Hubble Constant Tested with High Resolution Observations of Cepheids by JWST 19

Riess, A. G. & Breuval, L. 2023, The Local Value of H$\_0$ 19

Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019a, Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM, ApJ, 876, 85 1, 111, 125

Riess, A. G., Narayan, G., & Calamida, A. 2019b, Calibration of the WFC3-IR Count-rate Nonlinearity, Sub-percent Accuracy for a Factor of a Million in Flux, Tech. rep. 125

Riess, A. G., Yuan, W., Casertano, S., Macri, L. M., & Scolnic, D. 2020, The Accuracy of the Hubble Constant Measurement Verified through Cepheid Amplitudes, ApJ, 896, L43 19, 125

Riess, A. G., Yuan, W., Macri, L. M., et al. 2022, A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 Km S-1 Mpc-1 Uncertainty from the Hubble Space Telescope and the SH0ES Team, ApJ, 934, L7 2, 7, 19, 126, 149

Rijsbergen, C. J. V. 1979, Information Retrieval (Butterworths) 76

Rimoldini, L., Holl, B., Gavras, P., et al. 2023, Gaia Data Release 3. All-sky Classification of 12.4 Million Variable Sources into 25 Classes, A&A, 674, A14 25, 38

Ripepi, V., Chemin, L., Molinaro, R., et al. 2022, The VMC Survey - XLVIII. Classical Cepheids Unveil the 3D Geometry of the LMC, MNRAS, 512, 563 100, 129

Ripepi, V., Clementini, G., Molinaro, R., et al. 2023, Gaia Data Release 3. Specific Processing and Validation of All Sky RR Lyrae and Cepheid Stars: The Cepheid Sample, A&A, 674, A17 11, 19, 20, 25, 39, 75, 80, 83, 106, 122, 128, 133, 187

Ripepi, V., Marconi, M., Moretti, M. I., et al. 2014, The VMC Survey - VIII. First Results for Anomalous Cepheids, MNRAS, 437, 2307 144

Ritter, A. 1879, Untersuchungen Über Die Höhe Der Atmosphäre Und Die Constitution Gasförmiger Weltkörper, Annalen der Physik, 244, 157 2

Rousseeuw, P. J. 1984, Least Median of Squares Regression, Journal of the American Statistical Association, 79, 871 115

Rousseeuw, P. J. & Driessen, K. V. 1999, A Fast Algorithm for the Minimum Covariance Determinant Estimator, Technometrics, 41, 212 115, 116

Sandage, A. 1958, Current Problems in the Extragalactic Distance Scale., ApJ, 127, 513 3, 7

Sandage, A. & Tammann, G. A. 2006, Absolute Magnitude Calibrations of Population I and II Cepheids and Other Pulsating Variables in the Instability Strip of the Hertzsprung-Russell Diagram, Annual Review of Astronomy and Astrophysics, 44, 93 1

Sartoretti, P., Katz, D., Cropper, M., et al. 2018, Gaia Data Release 2. Processing the Spectroscopic Data, A&A, 616, A6 109

Savino, A., Weisz, D. R., Skillman, E. D., et al. 2022, The Hubble Space Telescope Survey of M31 Satellite Galaxies. I. RR Lyrae-based Distances and Refined 3D Geometric Structure, ApJ, 938, 101 19

Saydjari, A. K., Schlafly, E. F., Lang, D., et al. 2023, The Dark Energy Camera Plane Survey 2 (DECaPS2): More Sky, Less Bias, and Better Uncertainties, ApJS, 264, 28 103

Schlafly, E. F. & Finkbeiner, D. P. 2011, Measuring Reddening with Sloan Digital Sky Survey Stellar Spectra and Recalibrating SFD, ApJ, 737, 103 29

Schlafly, E. F., Meisner, A. M., & Green, G. M. 2019, The unWISE Catalog: Two Billion Infrared Sources from Five Years of WISE Imaging, ApJS, 240, 30 90, 98, 99, 125

Schlafly, E. F., Meisner, A. M., Stutz, A. M., et al. 2016, The Optical-infrared Extinction Curve and Its Variation in the Milky Way, ApJ, 821, 78 119

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds, ApJ, 500, 525 29, 104, 119

Schmidt, E. G. 2013, Type II Cepheid Candidates. IV. Objects from the Northern Sky Variability Survey, AJ, 146, 61 41

Schmidt, E. G., Hemen, B., Rogalla, D., & Thacker-Lynn, L. 2009, Photometry of Type II Cepheid Candidates from the Northern Part of the All Sky Automated Survey, AJ, 137, 4598 40

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. 2017, DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN, ACM Transactions on Database Systems (TODS), 42, 1 34

Schwarzschild, M. 1940, On the Variables in Messier 3, Harvard College Observatory Circular, 437, 1 7

Scowcroft, V., Freedman, W. L., Madore, B. F., et al. 2011, The Carnegie Hubble Program: The Leavitt Law at 3.6 Mm and 4.5 Mm in the Large Magellanic Cloud, ApJ, 743, 76 122

Sesar, B. 2012, Template RR Lyrae Hα, Hβ, and Hγ Velocity Curves, AJ, 144, 114 105

Sesar, B., Hernitschek, N., Mitrović, S., et al. 2017, Machine-Learned Identification of RR Lyrae Stars from Sparse, Multi-band Data: The PS1 Sample, AJ, 153, 204 25, 42, 75, 103, 105

Sesar, B., Ivezić, Ž., Grammer, S. H., et al. 2010, Light Curve Templates and Galactic Distribution of RR Lyrae Stars from Sloan Digital Sky Survey Stripe 82, ApJ, 708, 717 xvi, 42, 43, 84, 85, 86

Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, The Man behind the Curtain: X-Rays Drive the UV through NIR Variability in the 2013 Active Galactic Nucleus Outburst in NGC 2617, ApJ, 788, 48 21, 25, 37

Sharon, A., Kushnir, D., Yuan, W., Macri, L., & Riess, A. 2023, Reassessing the Constraints from SH0ES Extragalactic Cepheid Amplitudes on Systematic Blending Bias 125

Shen, J., Rich, R. M., Kormendy, J., et al. 2010, Our Milky Way as a Pure-disk Galaxy—A Challenge for Galaxy Formation, ApJ, 720, L72 11

Skarka, M. 2013, Known Galactic Field Blazhko Stars, A&A, 549, A101 88, 93

Skowron, D. M., Skowron, J., Mróz, P., et al. 2019, A Three-Dimensional Map of the Milky Way Using Classical Cepheid Variable Stars, Science, 365, 478 1, 68

Skowron, D. M., Skowron, J., Udalski, A., et al. 2021, OGLE-ing the Magellanic System: Optical Reddening Maps of the Large and Small Magellanic Clouds from Red Clump Stars, ApJS, 252, 23 5, 119

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, The Two Micron All Sky Survey (2MASS), AJ, 131, 1163 30, 100

Smolec, R., Moskalik, P., Plachy, E., Soszyński, I., & Udalski, A. 2018, Diversity of Dynamical Phenomena in Type II Cepheids of the OGLE Collection, MNRAS, 481, 3724 10

Smolec, R., Ziółkowska, O., Ochalik, M., & Śniegowska, M. 2023, First Overtone Cepheids of the OGLE Magellanic Cloud Collection - beyond Radial Modes, MNRAS, 519, 4010 87, 93

Sorensen, DA., Andersen, S., Gianola, D., & Korsgaard, I. 1995, Bayesian Inference in Threshold Models Using Gibbs Sampling, Genetics, Selection, Evolution : GSE, 27, 229 132

Soszyński, I., Pawlak, M., Pietrukowicz, P., et al. 2016a, The OGLE Collection of Variable Stars. Over 450 000 Eclipsing and Ellipsoidal Binary Systems Toward the Galactic Bulge, Acta Astron., 66, 405 89, 93

Soszyński, I., Pietrukowicz, P., Skowron, J., et al. 2021, Over 24 000 δ Scuti Stars in the Galactic Bulge and Disk from the OGLE Survey, Acta Astron., 71, 189 89, 93

Soszyński, I., Smolec, R., Udalski, A., & Pietrukowicz, P. 2019a, Type II Cepheids Pulsating in the First Overtone from the OGLE Survey, ApJ, 873, 43 10

Soszyński, I., Smolec, R., Udalski, A., et al. 2020a, OGLE-GAL-ACEP-091: The First Known Multi-mode Anomalous Cepheid, ApJ, 901, L25 10

Soszyński, I., Udalski, A., Skowron, J., et al. 2022, The OGLE Collection of Variable Stars. Over 2600 δ Scuti Stars in the Small Magellanic Cloud, Acta Astron., 72, 245 89, 93

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2008, The Optical Gravitational Lensing Experiment. The OGLE-III Catalog of Variable Stars. II.Type II Cepheids and Anomalous Cepheids in the Large Magellanic Cloud, Acta Astron., 58, 293 10

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2009, The Optical Gravitational Lensing Experiment. The OGLE-III Catalog of Variable Stars. IV. Long-Period Variables in the Large Magellanic Cloud, Acta Astron., 59, 239 11, 89, 93

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2011, The Optical Gravitational Lensing Experiment. The OGLE-III Catalog of Variable Stars. XIII. Long-Period Variables in the Small Magellanic Cloud, Acta Astron., 61, 217 89, 93

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2013, The Optical Gravitational Lensing Experiment. The OGLE-III Catalog of Variable Stars. XV. Long-Period Variables in the Galactic Bulge, Acta Astron., 63, 21 89, 93

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2019b, Final Release of the OGLE Collection of Cepheids and RR Lyrae Stars in the Magellanic System. The Outer Regions, Acta Astron., 69, 87 27, 87

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2020b, Additional Galactic Cepheids from the OGLE Survey, Acta Astron., 70, 101 27

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2015a, The OGLE Collection of Variable Stars. Anomalous Cepheids in the Magellanic Clouds, Acta Astron., 65, 233 27

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2015b, The OGLE Collection of Variable Stars. Classical Cepheids in the Magellanic System, Acta Astron., 65, 297 27, 30

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2018, The OGLE Collection of Variable Stars. Type II Cepheids in the Magellanic System, Acta Astron., 68, 89 11

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2016b, The OGLE Collection of Variable Stars. Over 45 000 RR Lyrae Stars in the Magellanic System, Acta Astron., 66, 131 7, 27

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2017a, Concluding Henrietta Leavitt's Work on Classical Cepheids in the Magellanic System and Other Updates of the OGLE Collection of Variable Stars, Acta Astron., 67, 103 7, 25, 27

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2017b, The OGLE Collection of Variable Stars. Classical, Type II, and Anomalous Cepheids toward the Galactic Center, Acta Astron., 67, 297 7, 27

Soszyński, I., Udalski, A., Wrona, M., et al. 2019c, Over 78 000 RR Lyrae Stars in the Galactic Bulge and Disk from the OGLE Survey, Acta Astron., 69, 321 1, 27, 30

Steinmetz, M., Matijevič, G., Enke, H., et al. 2020, The Sixth Data Release of the Radial Velocity Experiment (RAVE). I. Survey Description, Spectra, and Radial Velocities, AJ, 160, 82 107

Storey-Fisher, K., Hogg, D. W., Rix, H.-W., et al. 2023, Quaia, the Gaia-unWISE Quasar Catalog: An All-Sky Spectroscopic Quasar Sample 90, 93

Stringer, K. M., Drlica-Wagner, A., Macri, L., et al. 2021, Identifying RR Lyrae Variable Stars in Six Years of the Dark Energy Survey, ApJ, 911, 109 25, 43, 44, 75, 103

Stringer, K. M., Long, J. P., Macri, L. M., et al. 2019, Identification of RR Lyrae Stars in Multiband, Sparsely Sampled Data from the Dark Energy Survey Using Template Fitting and Random Forest Classification, AJ, 158, 16 25, 43

Tonry, J. L., Denneau, L., Heinze, A. N., et al. 2018, ATLAS: A High-cadence All-sky Survey System, PASP, 130, 064505 105

Torra, F., Castañeda, J., Fabricius, C., et al. 2021, Gaia Early Data Release 3. Building the Gaia DR3 Source List - Cross-match of Gaia Observations, A&A, 649, A10 54

Torrealba, G., Catelan, M., Drake, A. J., et al. 2015, Discovery of ∼9000 New RR Lyrae in the Southern Catalina Surveys, MNRAS, 446, 2251 25, 31, 32

Turner, D. G., Majaess, D. J., Lane, D. J., et al. 2009, HDE 344787, A Cepheid Analogue To Polaris, 213, 433.06 81

Udalski, A., Soszyński, I., Pietrukowicz, P., et al. 2018, OGLE Collection of Galactic Cepheids, Acta Astron., 68, 315 25, 27, 68

Udalski, A., Szymanski, M. K., Soszynski, I., & Poleski, R. 2008, The Optical Gravitational Lensing Experiment. Final Reductions of the OGLE-III Data, Acta Astron., 58, 69 20, 27, 89

Udalski, A., Szymański, M. K., & Szymański, G. 2015, OGLE-IV: Fourth Phase of the Optical Gravitational Lensing Experiment, Acta Astron., 65, 1 20, 25, 27, 89

Ulla, A., Creevey, O. L., Álvarez, M. A., et al. 2022, Gaia DR3 Documentation Chapter 11: Astrophysical Parameters, Tech. rep. 119

van der Marel, R. P. & Cioni, M.-R. L. 2001, Magellanic Cloud Structure from Near-Infrared Surveys. I. The Viewing Angles of the Large Magellanic Cloud, AJ, 122, 1807 129

Van Rossum, G. & Drake Jr, F. L. 1995, Python tutorial (Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands) 240

VanderPlas, J. T. & Ivezić, Ž. 2015, Periodograms for Multiband Astronomical Time Series, ApJ, 812, 18 64

Vasiliev, E. 2023, Dear Magellanic Clouds, Welcome Back! 12, 113

Vasiliev, E. & Baumgardt, H. 2021, Gaia EDR3 View on Galactic Globular Clusters, MNRAS, 505, 5978 7, 15, 111, 113

Vasiliev, E. & Belokurov, V. 2020, The Last Breath of the Sagittarius dSph, MNRAS, 497, 4162 12

Vasiliev, E., Belokurov, V., & Erkal, D. 2021, Tango for Three: Sagittarius, LMC, and the Milky Way, MNRAS, 501, 2279 12

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. 2019, Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC 132

Vincenty, T. 1975, Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations, Survey Review, 23, 88 47

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods, 17, 261 240

Wang, S. & Chen, X. 2019, The Optical to Mid-infrared Extinction Law Based on the APOGEE, Gaia DR2, Pan-STARRS1, SDSS, APASS, 2MASS, and WISE Surveys, ApJ, 877, 116 119, 126, 127

Waskom, M. L. 2021, seaborn: statistical data visualization, Journal of Open Source Software, 6, 3021 240

Weiland, J. L., Arendt, R. G., Berriman, G. B., et al. 1994, COBE Diffuse Background Experiment Observations of the Galactic Bulge, ApJ, 425, L81 11

Wiecki, T., Salvatier, J., Vieira, R., et al. 2023, Pymc-Devs/Pymc: V5.7.2, Zenodo 240

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance, AJ, 140, 1868 98

Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, LAMOST Spectral Survey — An Overview, Research in Astronomy and Astrophysics, 12, 723 106

Zinn, R. & Dahn, C. C. 1976, Variable 19 in NGC 5466: An Anomalous Cepheid in a Globular Cluster., AJ, 81, 527 111

Zinn, R. & Searle, L. 1976, The Masses of the Anomalous Cepheids in the Draco System., ApJ, 209, 734 7

Zwitter, T., Kos, J., Buder, S., et al. 2021, The GALAH+ Survey: A New Library of Observed Stellar Spectra Improves Radial Velocities and Hints at Motions within M67, MN-RAS, 508, 4202 107

Publications of Hitesh Lala

> *Yesterday I was clever, so I wanted to change the*
> *world. Today I am wise, so I am changing myself.*
> – Rumi

First- and second-authored publications

➤ **Lala, H.N.**, Lemasle, B., Pipwala, V., Grebel, E.K., Bono, G., Braga, V.F., Crestani, J., Fabrizio, M., Fiorentino, G.
"RRLCep: A Catalog of RR Lyrae stars and Cepheids. II. Distance Determination"
in preparation. Based on the Chapter V of this thesis.

➤ Pipwala, V., **Lala, H.N.**, Lemasle, B., Grebel, E.K., Bono, G., Fiorentino, G.
"Using Type-II Cepheids as Extragalactic Standard Candles: I. Distances to M31"
in preparation.

➤ **Lala, H.N.**, Lemasle, B., Grebel, E.K., Bono, G., Braga, V.F., Crestani, J., Fabrizio, M., Fiorentino, G
"RRLCep: A Catalog of RR Lyrae stars and Cepheids. I. A Comprehensive Compilation of Literature Surveys"
in preparation. Based on the Chapters II–IV of this thesis.

➤ Lemasle, B., **Lala, H.N.**, Kovtyukh, V., Hanke, M., Prudil, Z., Bono, G., Braga, V. F., da Silva, R., Fabrizio, M., Fiorentino, G., François, P., Grebel, E. K., Kniazev, A
"Tracing the Milky Way warp and spiral arms with classical Cepheids"
2022, Astronomy & Astrophysics, Volume 668, A40

Co-authored publications

➤ Fabrizio, M., Braga, V. F., Crestani, J., Bono, G., Ferraro, I., Fiorentino, G., Iannicola, G., Preston, G. W., Sneden, C., Thévenin, F., Altavilla, G., Chaboyer, B., Dall'Ora, M., da Silva, R., Grebel, E. K., Gilligan, C. K., **Lala, H.N.**, Lemasle, B., Magurno, D., Marengo, M., Marinoni, S., Marrese, P. M., Martínez-Vázquez, C. E., Matsunaga, N., Monelli, M., Mullen, J. P., Neeley, J., Nonino, M., Prudil, Z., Salaris, M., Stetson, P. B., Valenti, E., Zoccali, M.
"On the Use of Field RR Lyrae As Galactic Probes: IV. New Insights Into and Around the Oosterhoff Dichotomy"
2021, The Astrophysical Journal, Volume 919, Issue 2, 118

➤ Braga, V. F., Crestani, J., Fabrizio, M., Bono, G., Sneden, C., Preston, G. W., Storm, J., Kamann, S., Latour, M., **Lala, H.N.**, Lemasle, B., Prudil, Z., Altavilla, G., Chaboyer, B., Dall'Ora, M., Ferraro, I., Gilligan, C. K., Fiorentino, G., Iannicola, G., Inno, L., Kwak, S., Marengo, M., Marinoni, S., Marrese, P. M., Martínez-Vázquez, C. E., Monelli, M., Mullen, J. P., Matsunaga, N., Neeley, J., Stetson, P. B., Valenti, E., Zoccali, M.
"On the Use of Field RR Lyrae as Galactic Probes. V. Optical and Radial Velocity Curve Templates"
2021, The Astrophysical Journal, Volume 919, Issue 2, 85

➤ Crestani, J., Braga, V. F., Fabrizio, M., Bono, G., Sneden, C., Preston, G., Ferraro, I., Iannicola, G., Nonino, M., Fiorentino, G., Thévenin, F., Lemasle, B., Prudil, Z., Alves-Brito, A., Altavilla, G., Chaboyer, B., Dall'Ora, M., D'Orazi, V., Gilligan, C., Grebel, E. K., Koch-Hansen, A. J., **Lala, H.N.**, Marengo, M., Marinoni, S., Marrese, P. M., Martínez-Vázquez, C., Matsunaga, N., Monelli, M., Mullen, J. P., Neeley, J., da Silva, R., Stetson, P. B., Salaris, M., Storm, J., Valenti, E., Zoccali, M.
"On the Use of Field RR Lyrae as Galactic Probes. III. The α-element Abundances"
2021, The Astrophysical Journal, Volume 914, Issue 1, 10

➤ Prudil, Z., Hanke, M., Lemasle, B., Crestani, J., Braga, V. F., Fabrizio, M., Koch-Hansen, A. J., Bono, G., Grebel, E. K., Matsunaga, N., Marengo, M., da Silva, R., Dall'Ora, M., Martínez-Vázquez, C. E., Altavilla, G., **Lala, H.N.**, Chaboyer, B., Ferraro, I., Fiorentino, G., Gilligan, C., Nonino, M., Thévenin, F.
"Milky Way archaeology using RR Lyrae and type II Cepheids. I. The Orphan stream in 7D using RR Lyrae stars"
2021, Astronomy & Astrophysics, Volume 648, A78

➤ Crestani, J., Fabrizio, M., Braga, V. F., Sneden, C., Preston, G., Ferraro, I., Iannicola, G., Bono, G., Alves-Brito, A., Nonino, M., D'Orazi, V., Inno, L., Monelli, M., Storm, J., Altavilla, G., Chaboyer, B., Dall'Ora, M., Fiorentino, G., Gilligan, C., Grebel, E. K., **Lala, H.N.**, Lemasle, B., Marengo, M., Marinoni, S., Marrese, P. M., Martínez-Vázquez, C. E., Matsunaga, N., Mullen, J. P., Neeley, J., Prudil, Z., da Silva, R., Stetson, P. B., Thévenin, F., Valenti, E., Walker, A., Zoccali, M.
"On the Use of Field RR Lyrae as Galactic Probes. II. A New ΔS Calibration to Estimate Their Metallicity"
2021, The Astrophysical Journal, Volume 908, Issue 1, 20

➤ Momany, Y., Zaggia, S., Montalto, M., Jones, D., Boffin, H. M. J., Cassisi, S., Moni Bidin, C., Gullieuszik, M., Saviane, I., Monaco, L., Mason, E., Girardi, L., D'Orazi, V., Piotto, G., Milone, A. P., **Lala, H.N.**, Stetson, P. B., Beletsky, Y.
"A plague of magnetic spots among the hot stars of globular clusters"
2020, Nature Astronomy, Volume 4, 1092-1101

Acknowledgments

An unshared life is not living.

– Samuel S. Wise

The past four years have changed me as a person. They have altered my goals. They have changed how I see the world. They have changed how I look at life itself. So many lovely people have stood by me throughout this transformation. I would like to thank a non-exhaustive subset of them here.

To **Eva**, thank you so much for taking me on as a Ph.D. student, it has been a wonderful experience. You are an exemplary scientist, an erudite professor, and an adept director. In spite of your frantic schedule, you were always generous with your time. I will always be grateful to you for finding the funding for my contract extensions. Lastly, thank you for rechristening me as *Bonifacio*. I hope to live up to this name.

To **Bertrand**, I write this with equal parts satisfaction and disappointment. I am disappointed that I will not get to see you more often. It pains me to write this as a significant part of our journey together is drawing to a close. But I am so happy that you offered me this position and, thankfully, I was wise enough to accept it. Your style of supervision is unrivaled and it fitted me like a glove. One might even describe this style as an *iron fist in a velvet glove*. Thankfully, I never had to deal with the iron fist. You hold this uncanny ability of absorbing my stress and imparting comfort. After our every interaction, I felt more confident of my abilities. You always treated me as an equal and it has been an honor working with you and not for you. I'll always have a friend in you and you'll always have a friend in me. You have cracked the secret to happiness: never let the child within us die. I wish the child in you never grows up.

To **Mom**, *tujhe sab hai pata..., hai na Maa?* Your clock-like precision in calling me every morning after waking up and every night before going to bed, it timestamped my days as well. Throughout the past six years, whenever I have felt lonely, you have called me unfailingly. I miss being cared by you, I miss being hugged by you, I miss being pampered by you. I just love and miss you a lot, *Maren*. No words will ever be able to fully surmise what you mean to me.

To **Dad**, you haven't taught me a single thing. You have never lectured me on anything. You have just been. By just being there, you have molded every second of my life. You exemplify patience, embody virtuosity, and personify kindness. By being this paragon of a human being, you have set me up for success. Yet you have never said *bada naam karega, beta humara aisa kaam karega*. You just supported my every choice, fulfilled my every demand without any questions or expectations. I love you loads, *Papa*.

To **Shaku** *Didi*, you are the darling of my heart. You are so young at heart even at this age when you are practically older than *besamjh swami*. You are a true fighter and the lifeblood of our family. To **Saru** *Didi*, thank you for taking care of my parents and your best friend. I wish to adhere to a schedule as well as you have done all your life. To **Sillu** *Didi*, thank you for Vicky, Maggi, and all the *faloodas*. Let's share one again very soon.

To **Tilak**, *kya hi bol du idar jo roz nahi bolta*. Pune, Potsdam, Poland, I am so glad I get to witness your evolution first-hand. Your daily calls invariably include my perpetual boasting and your constant repackaging of said boasting as modesty. "The frailty of genius, John; it always needs an audience". I am sure that Arthur Conan Doyle (and Mark Gatiss) did not intend for this sentence to be applied to a Sherlock with my limited genius. But you have played the role of John (and the audience) to aplomb. My heart fills with pride every time I refer to you as my best friend. On behalf of all Sindhis, I declare you an honorary Sindhi and a certified *Charya*.

To **Divyanshu**, my personal TeXGPT, my culinary mentor, my sports buddy, and a thousand other monikers that I can assign you, thank you so much. You fed me food with one hand while dealing with a broken tooth with the other; how I can ever thank you, *Bandar*? As you reminded me a few days ago, I had asked you to continue to be weird in the Acknowledgments of my Master's thesis. Here, I ask you to continue being this gentle-but-firm, free-spirited-but-determined person that you have become since June 28, 2017. Retain the weirdness, though. Never not be weird. I am so very proud of you, Divya.

To **Nilima**, thank you so much for being the Rosa to my Jake, the El to my Moz, and the Lily to my Ted. From pandas and HAMILTON, to my firstborn itself, your friendship exacts a very heavy toll. But your support, your brilliance, and your (dare I say Divya-like) cheerfulness that you bring to my life far far outweigh what I could ever give you. I hereby promise to never stop discussing numbers and budgets with you. I know you don't hate it when I do that. From customary acquaintances of Fergusson and infrequent interactions of Belgrade, to constant conversations of Bonn and regular stay-overs of Heidelberg, neither of us imagined our friendship to blossom this wonderfully. I am delighted that it did. *Bas? Baki call pe.*

To **Aish**, it was so great to have your company during our Master's degree that I never realized how excruciating AstroMundus was. Your absence was deeply felt throughout my

experiences (and the LaTeX templates) that you shared with me put me in good stead for the final stretch of my Ph.D. Thanks a lot for inviting me to your defense, it was an honor to meet your mom. It is down to her that my mom will get to attend my defense.

To **Michael Hanke** and **Zdenek Prudil**, thanks to both of you for being anally pedantic about the plots in your (and everyone else's) papers; it helped me learn a lot. You were an important factor in me choosing Heidelberg and I frequently wish that our time at ARI would have overlapped a bit more. To **Camilla Hansen-Koch**, thanks for being a part of my thesis committee and ensuring that everything ran smoothly. To **Anna Pasquali**, *grazie mille* for adding an Italian touch to our group meetings. I always appreciated your child-like, infectious enthusiasm.

To **Simone** and **Yazan**, I owe so much to the two of you. Thank you for being an indispensable part of my scientific career. I hope to see you again very soon.

To **Sabine Reffert**, thank you so much for agreeing to be a referee of this thesis. Your patience and cooperation mean a lot to me. To **Coryn Bailer-Jones** and **Luca Amendola**, I am grateful that the two of you agreed to be a part of my examination panel. To **Karina Schoenfeld**, **Eleonara Grauer-Gamet**, **Saskia Mayer**, **Martina Buchhaupt**, and many other administrators at ARI and the HGSFP, thank you so much for your dedication. Your work positively affects the lives of so many people around the world.

This dissertation has been possible only because Tilak, Nilima, and Divya took a significant amount of time off their own schedules and focused on babysitting me. I will forever be grateful to the three of you for your magnanimity. I formally invite you to write your theses from the comforts of my next home, wherever it might be.

Data and Software

Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration.

This work used the following software: Python (Van Rossum & Drake Jr, 1995), IPython (Pérez & Granger, 2007; Kluyver et al., 2016), NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), Matplotlib (Hunter, 2007), Seaborn (Waskom, 2021), pandas (pandas development team, 2020), Astropy (Astropy Collaboration et al., 2013, 2018, 2022), gala (Price-Whelan, 2017; Price-Whelan et al., 2022), Galpy (Bovy, 2015), Pymc (Wiecki et al., 2023; Oriol et al., 2023), ArviZ (Kumar et al., 2019), TeXstudio, and Ubuntu.