Dissertation

submitted to the
Instituto de Astrofísica, Facultad de Física
Pontificia Universidad Católica de Chile, Chile
for the degree of
Doctor in Astrophysics

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Put forward by

M.Sc. Gergely Hajdu
Born in Miskolc, Hungary

Oral examination: 06 August 2019

Structure of the obscured Galactic disk with pulsating variables

Gergely Hajdu

Referees: Prof. Dr. Márcio Catelan
Prof. Dr. Eva Grebel

ABSTRACT

Bright pulsating variables, such as Cepheids and RR Lyrae, are prime probes of the structure of both the young and old stellar components of the Milky Way. However, the far side of the Galactic disk has not yet been mapped using such variables as tracers, due to the severe extinction caused by foreground interstellar dust.

In this thesis, the near-infrared light curves from the VISTA Variables in the Vía Láctea survey are utilized to penetrate these regions of high extinction and thus discover thousands of previously "hidden" Cepheid and RR Lyrae variables. The analysis of the light curves of RR Lyrae variables, was performed with a newly developed fitting algorithm, and their metallicities determined from their near-infrared light-curve shapes, using a newly developed method. These photometric metal abundances, combined with their positions within the Galactic disk, lend support to theories of an early, inside-out formation of the Galactic disk.

The newly discovered Cepheids were classified into the old (Type II) and young (Classical) subtypes. A new near-infrared extinction law was determined using the Type II Cepheids, taking advantage of their concentration around the Galactic center. The distribution of the Classical Cepheids in the Galactic disk follows both the Galactic warp and the flare of the Galactic disk at large Galactocentric radii. A first attempt has been made to connect the current locations of the Classical Cepheids to the spiral arm structure of the Milky Way.

ZUSAMMENFASSUNG

Helle pulsierende variable Sterne, sogenannte Variablen wie z.B. Cepheiden und RR-Lyrae Sterne, stellen ein Hauptwerkzeug für die Untersuchung der Struktur, sowohl der jungen als auch der alten Sternkomponenten, der Milchstraße dar. Die uns gegenüberliegende Seite der galaktischen Scheibe wurde jedoch noch nicht mithilfe von Variablen kartiert, aufgrund von starker Vordergrundextinktion hervorgerufen durch interstellaren Staub.

In dieser Dissertation werden die Nahinfrarot-Lichtkurven der Beobachtungskampagne VISTA Variables in the Vía Láctea verwendet, um diese Regionen mit hoher Extinktion zu durchdringen und so Tausende von zuvor "verborgenen" Cepheid- und RR Lyrae-Variablen in diesen Bereichen zu untersuchen. Die Analyse der Lichtkurven von RR-Lyrae-Variablen wurde mit einem neu entwickelten Anpassungsalgorithmus durchgeführt, und ihre Metallizitäten wurden anhand ihrer Lichtkurvenformen im Nahinfrarot unter Verwendung einer neu entwickelten Methode bestimmt. Diese photometrischen Metallhäufigkeiten stützen, zusammen mit den Positionen der Variablen innerhalb der Galaktischen Scheibe, die Theorie über eine frühe, von innen nach außen gerichtete Bildung der Galaktischen Scheibe.

Die neu entdeckten Cepheiden wurden in alte (Typ II) und junge (klassischer Typ) Untertypen eingeteilt. Ein neues Nahinfrarot-Extinktionsgesetz wurde unter Verwendung der Typ-II-Cepheiden unter Ausnutzung ihrer Eigenschaften der Konzentration um das galaktische Zentrum bestimmt. Die Verteilung der klassischen Cepheiden in der galaktischen Scheibe folgt sowohl ihrer Verzerrung als auch ihrer unregelmäßigen Verdickung bei großen galaktozentrischen Radien. Schlussendlich wurde ein erster Versuch unternommen, die derzeitigen Positionen der klassischen Cepheiden mit der Spiralarmstruktur der Milchstraße zu verbinden.

RESUMEN

Las estrellas variables pulsantes brillantes, tales como Cefeidas y RR-Lyras, son sondas fundamentales de la estructura de las componentes vieja y joven de la Vía Láctea. Sin embargo, el lado más alejado del disco de la Galaxia aún no ha sido mapeado usando tales variables debido a la severa extinción causada por polvo interestelar en frente de ellas.

En esta tesis, las curvas de luz en infrarrojo cercano de la muestra VISTA Variables in the Vía Láctea son utilizadas para penetrar esas regiones y descubrir miles de Cefeidas o RR Lyras previamente "escondidas". El análisis de las curvas de luz de las variables RR Lyras fue ejecutado con un algoritmo de ajuste, y sus metalicidades determinadas de siluetas de sus curvas de luz en el infrarrojo cercano usando métodos recientemente desarrollados. Estas abundancias fotométricas de metales, con sus posiciones en el disco Galáctico, apoyan las teorílas de la formación de adentro hacia fuera del disco de la Galaxia.

Las Cefeidas descubiertas fueron clasificadas en los subtipos viejo (Tipo II) y nuevo (Clásicas). Una nueva ley de extinción en el infrarrojo cercano fue determinada usando Cefeidas Tipo II, utilizando su concentración en el centro Galáctico. La distribución de las Cefeidas Clásicas en el disco Galáctico sigue el pandeo Galáctico y el ensanchamiento del disco de la Galaxia en grandes radios Galactocéntricos. Un primer intento se efectuó con el fin de conectar las ubicaciones actuales de las Cefeidas Clásicas con la estructura de brazos espirales de la Vía Láctea.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1

Introduction

1.1 The Milky Way disk

The possibility that our home, the Milky Way galaxy, might show a spiral structure was probably first proposed by Alexander (1852), but it has taken a century until this idea was finally proven by observations of the distribution of stars of spectral types O and B in the Solar neighborhood (Morgan et al. 1952), as well as by mapping the distribution of neutral hydrogen (Kerr 1962). These methods, useing young stellar tracers, as well as radio observations combined with kinematic distance estimates, remain the two main ways for studying the Galactic spiral arm structure of the Milky Way.

Both methods face serious limitations: for example, radial velocity mapping is blind towards the central region of the Galaxy, as gas (and stars) move mostly perprendicularly to the line of sight, resulting in zero radial velocities. Furthermore, in these studies, the distance is calculated by assuming circular motion of the gas in the Galactic disk, combining radial velocities measured from radio emission lines (such as neutral hydrogen 21 cm emission) with a Galactic rotation curve. Deviations from circular rotation however, are expected, as spiral arms are known to excite non-circular streaming motions both in the case of Milky Way (Fresneau et al. 2005) as well as in that of other galaxies (Kim and Kim 2014; Erroz-Ferrer et al. 2015). Figure 1.1 shows a modern map of the Milky Way disk, as traced by neutral hydrogen emission, highlighting the limitations of this approach.

In contrast, young stellar tracers, such as young open clusters, OB stars, and Classical Cepheids (introduced in Section 1.3), have consistent, individually determined distances, and they have been used before to delineate the spiral arm structure in the Solar neighborhood (see, e.g., Majaess et al. 2009). However, very high extinction values in the Galactic plane have so far prevented the detection of these objects at large distances in the Galactic disk. The perfect stellar tracers of the spiral structure in these regions are the high

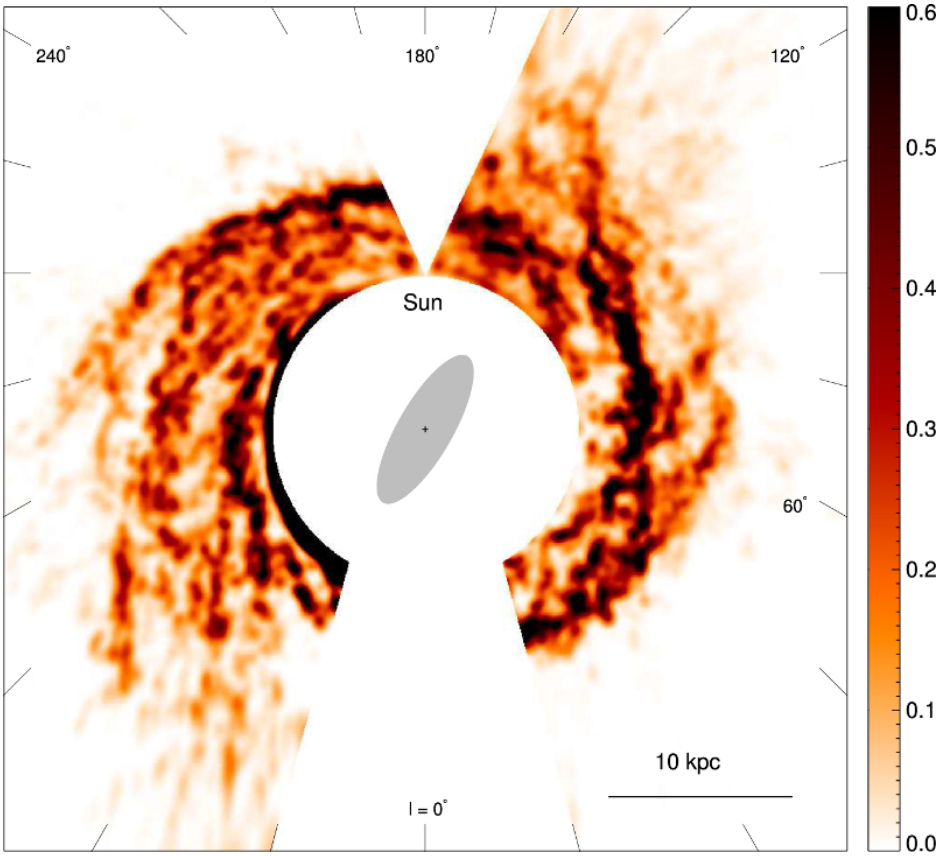

Figure 1.1: Face-on map of atomic hydrogen distribution in the Milky Way disk. The color scale is proportional to the calculated surface density of the hydrogen (figure from Koo et al. 2017).

mass star forming regions (HMSFR), as their water and methanol maser emissions (Elitzur 1992) allow radio-interferometric observations to directly measure their parallaxes up to ~ 20 kpc away, i.e., to the distance of the farthest star forming regions on the other side of the Milky Way (Sanna et al. 2017). Figure 1.2 shows the distribution of the maser sources with measured parallaxes in the Galactic disk. As can be seen, these observations are not constrained by kinematical uncertainties, in contrast to the kinematically-constrained radial velocity mapping of the gas in the Galactic disk (compare to Figure 1.1). However, this figure also highlights some of the problems related to these tracers: they are relatively rare, when compared to other stellar tracers of the spiral structure. Furthermore, the low number of such objects with measured parallaxes is, at least, partially caused by the high number of interferometric radio observations needed to obtain the parallax of a single HMSFR.

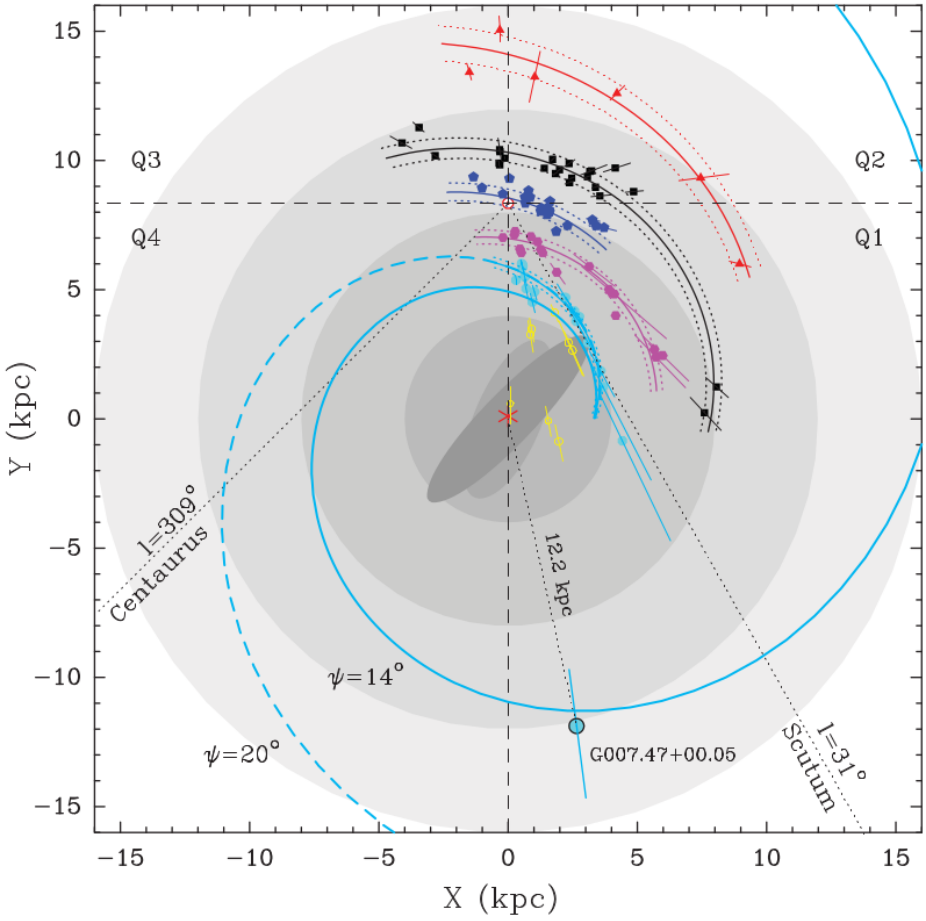

Figure 1.2: Galactic map of high mass star forming regions with measured maser parallaxes (figure from Sanna et al. 2017).

Besides the spiral structure, the disk of the Milky Way exhibits two additional major large-scale features: the Galactic warp and the flaring in its outer parts. Both effects modify the flat shape of the disk, which is otherwise followed by the inner regions. In the outer regions, the scale height of the Galactic disk increases, leading to a thicker outer disk (see, e.g., López-Corredoira and Molgó 2014, as well as references there). Furthermore, in the outer regions of the Milky Way, the disk is warped upwards in the north and downward in the south, as shown by both the distribution of atomic gas (Kerr 1957; Oort et al. 1958), as well as that of the stars (Reylé et al. 2009; Li et al. 2019). Figure 1.3 shows a selection of disk galaxies exhibiting the warping feature present in the disk of the Milky Way.

It was proposed by Gilmore and Reid (1983) that the older stars belonging to the

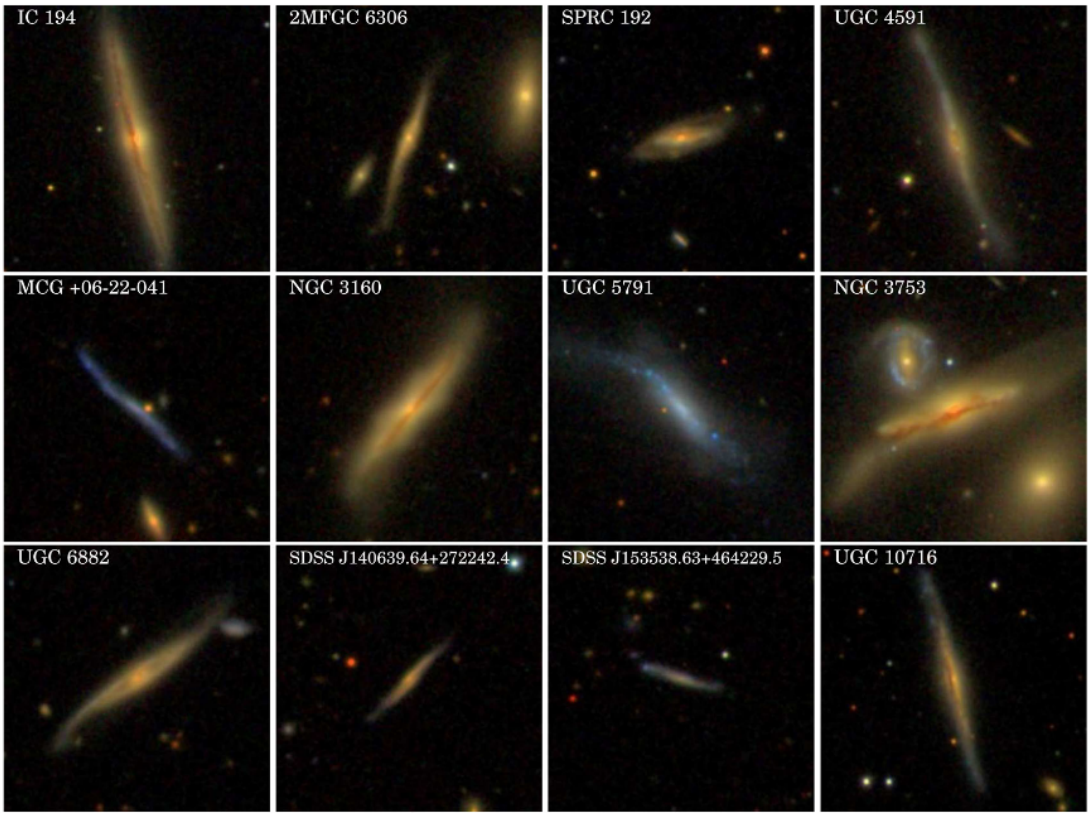

Figure 1.3: SDSS thumbnail images of galaxies with warping disks (figure from Reshetnikov et al. 2016).

Galactic disk are kinematically separated from the younger thin-disk stars, calling this component the thick disk. The thick disk is characterized by older ages, lower metal abundances, as well as larger scale height. The formation of these two components in a cosmological framework was investigated by Brook et al. (2012), by performing hydrodynamical simulations of a Milky Way-like forming disk galaxy, resulting in an inside-out formation scenario, where the thick disk formed earlier, followed by the thin disk, with both components being observable in present times. Spectroscopic observations (Bovy et al. 2012) lend support to this scenario. The recent identification of stars originating from a massive dwarf galaxy that has been accreted onto the Milky Way about 8 – 11 Gyr ago (Belokurov et al. 2018; Helmi et al. 2018), called the "Gaia Sausage" or Enceladus, suggests that a large fraction of the stars found in the thick disk might have originated from such mergers, similarly to the simulations of Villalobos and Helmi (2008).

Due to their diagnostic value, RR Lyrae variables (Section 1.2) hold crucial information

about the old stellar populations in which they reside. Therefore, finding and characterizing the RR Lyrae variables throughout the Galactic disk can provide the missing pieces to our understanding of its formation and evolution, particularly in establishing the relationship between the thick disk, halo, and bulge components, all of which are known to host old stellar populations. Similarly to RR Lyrae stars, Classical Cepheids (Section 1.3) can also contribute vastly to our understanding of the Galactic disk. In contrast to the RR Lyrae variables, however, these young stars can reveal the spiral structure of the Galactic disk (Majaess et al. 2009).

1.2 RR Lyrae variables

RR Lyrae variables are classical radial pulsators with pulsation periods between 0.2 and 1.0 day, found in the intersection of the horizontal branch and the Cepheid instability strip (IS, see Figure 1.4), with absolute magnitudes of $M_V \sim 0.6$. They are one of the most useful types of variable stars, as their intrinsic brightness and characteristic light-curve shapes allow their discovery out to large distances (Smith 2004; Catelan and Smith 2015). They are often found in old stellar populations, as the canonical formation scenario of RR Lyrae stars requires ages greater than 10 Gyr. As such, they serve as one of the most useful tracers of old stellar populations. They can be often found in large quantities in globular clusters (Clement et al. 2001).

The first discovered member of this variable star type was U Lep (Kapteyn 1890), followed by the discovery of a seventh-magnitude star in the constellation Lyra by W.F. Flemming (Pickering et al. 1901), which later received the name RR Lyrae and became the prototype of this class of variable stars. Within a few short years of these initial discoveries, until 1913 more than 500 variables of the class had been found in globular clusters (Bailey and Pickering 1913) by the photographic monitoring program led by S.I. Bailey.

RR Lyrae variables are traditionally categorized into three subtypes: the RRab variables pulsate in the radial fundamental mode, the RRc stars pulsate in the first radial overtone, while the members of the RRd subtype pulsate simultaneously in both modes. The original RRa, RRb, RRc subtypes (Bailey 1902) were revised by Schwarzschild (1940), who suggested that the larger-amplitude RRa and the lower-amplitude RRb classes are

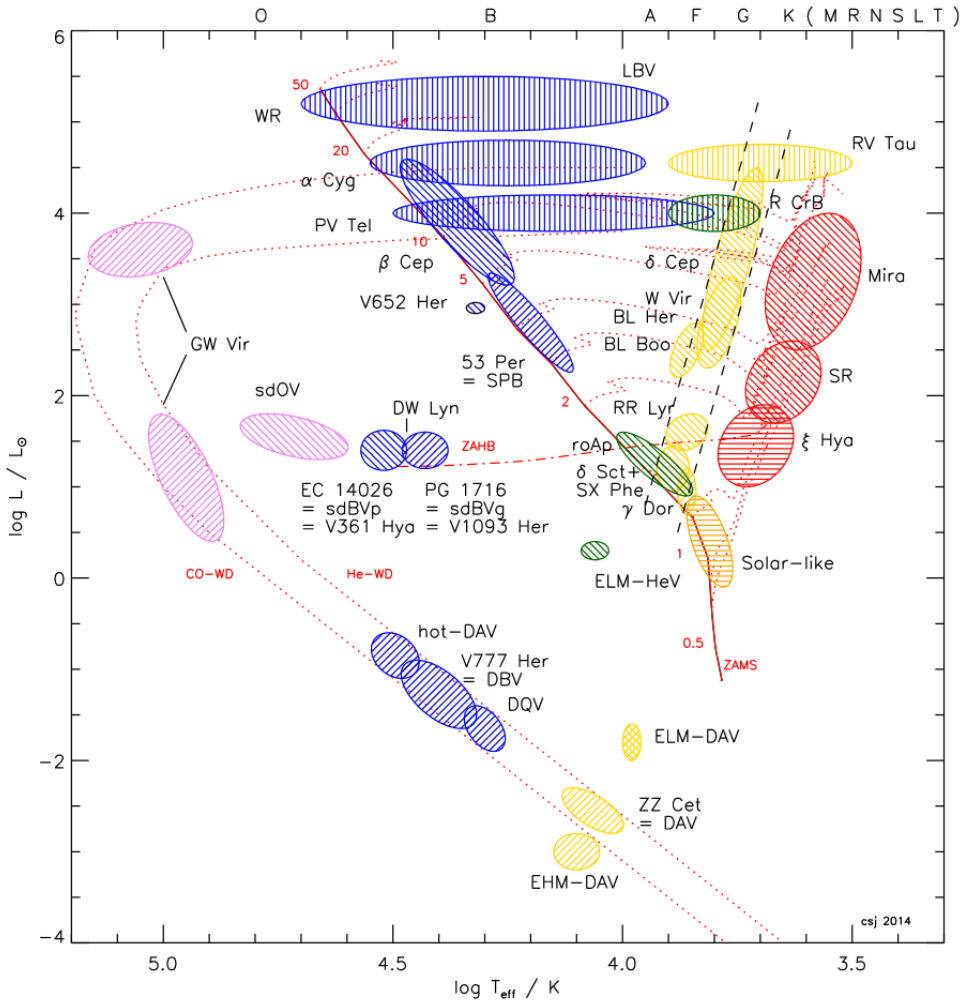

Figure 1.4: Schematic Hertzsprung-Russel diagram showing the positions of major classes of pulsating variables. The zero-age main sequence is shown by the red continuous line, with the numbers next to it showing the corresponding masses in solar units. The zero-age horizontal branch is shown by the dash-dotted line, while the dashed lines denote the edges of the Cepheid instability strip. Horizontal shading (≡) represents acoustically driven modes, perpendicular shading (\\\) opacity-driven p modes, inverse perpendicular shading (///) buoyancy-driven g modes, and vertical shading (|||) strange modes (see, e.g., Catelan and Smith 2015, for the description of these mechanisms). Individual post-MS stellar evolutionary tracks are shown by the dotted lines. (Figure from Jeffery and Saio 2016.)

both pulsating in the fundamental mode, while the RRc variables are pulsating in the radial first overtone mode. Their pulsations are caused by the κ (Baker and Kippenhahn 1962) and γ (Cox et al. 1966) mechanisms, mainly driven by the H and He partial ion-

ization zones in the stellar interior (for a succinct description, see Section 5.9.2 of Catelan and Smith 2015).

During their pulsation cycles, RR Lyrae stars experience changes in both surface temperature and radius. The combination of these two effects lead to the observed light-curve shapes, which will also depend on the observed wavelength. Figure 1.5 shows the change in the light-curve shape of a typical RRab variable. In the U band, the change in the surface temperature dominates, while towards progressively longer wavelengths, the contribution of the change in the radius of the variable becomes commensurable, as more and more the Rayleigh-Jeans tail of the Planck function becomes dominant (see also Das et al. 2018). This also results in lowered amplitudes towards the infrared bands, posing a challenge to their detection in infrared surveys, as they become more easily confused with other types of variables, such as eclipsing binaries, especially at the lower photometric precisions achievable for ground-based near-infrared observations.

Besides their characteristic shapes, and hence easy identification, what makes RR Lyrae variables excellent distance indicators is the existence of the period-luminosity-metallicity (PLZ) relations: the intrinsic magnitudes and colors of the stars can be approximated as a function of their pulsation periods and their metallicities (Catelan et al. 2004; Marconi et al. 2015). In case no spectroscopically measured metallicity estimates are available, in the optical bands it can be estimated as a function of the period and the light-curve shape of the variables (Jurcsik and Kovács 1996; Smolec 2005; Morgan et al. 2007).

Even before 1990, thousands of RR Lyrae variables were known to exist within the Galactic disk, bulge, halo and the globular clusters of the Milky Way. With the advent of large-scale optical detectors, microlensing projects like EROS-2 (Kim et al. 2014), MACHO (Alcock et al. 2003) and OGLE (Udalski et al. 2015), as well as other transient and asteroid-hunting projects like LINEAR (Sesar et al. 2013), the Catalina Sky Survey (Drake et al. 2013; Torrealba et al. 2015), and Pan-STARRS1 (Sesar et al. 2017), thousands more were discovered. The current number of known Galactic RR Lyrae variables is over 100 000, with the two largest contributors being the OGLE project (Soszyński et al. 2014) and the *Gaia* astrometric satellite (Clementini et al. 2019).

Nothwithstanding, the fact that these surveys have greatly increased our undestanding of the old stellar population of the Galactic bulge and halo, the large amount of extinction

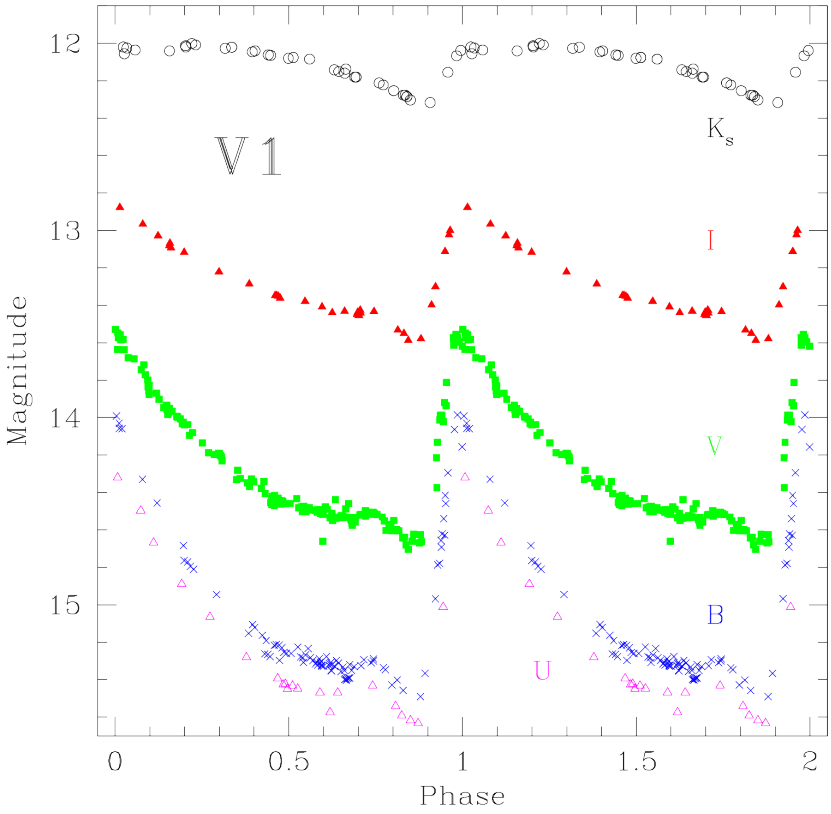

Figure 1.5: Typical folded $UBVIK_S$ light curves of an RRab variable (figure from Catelan et al. 2013).

found in the Galactic plane have so far prevented the systematic investigation of the RR Lyrae population of the Galactic disk (see e.g., Figure 26 of Clementini et al. 2019). As an example, at the distance of the Galactic bulge, the extinction in the plane of the Galaxy plane can reach $A(K_S) \sim 3.5$ mag (Gonzalez et al. 2012), corresponding to $A(V) > 30$ mag, making this region inaccessible to any optical survey.

1.3 Cepheid variables

Cepheid variables are similar to RR Lyrae stars in the sense that they are also located within the IS, however, their periods are found between 1.0 and 100 days (Catelan and Smith 2015). Their pulsations are supported by the κ and γ mechanisms, just as in the case of RR Lyrae variables.

They are categorized into two main groups: Classical (also known as Type I or δ) Cepheids and Type II Cepheids. Classical Cepheids are young (< 200 Myr), evolved pop-

ulation I stars of masses $2 - 20 M_\odot$ (Turner 1996), crossing the instability strip multiple times during their post-main sequence evolution (see Figure 1.4 for their position on the Hertzsprung-Russel diagram). The brightest Classical Cepheids reach absolute magnitudes of about $M_V \sim -6$ mag, while the faintest ones at $M_V \sim -2$ mag are still considerably brighter than the brightest RR Lyrae variables.

In contrast to the Classical Cepheids, Type II Cepheids are old, evolved low-mass stars. At the beginning of the 20$^{\text{th}}$ century, the distinction between the two classes of Cepheids was unknown. In the spectroscopic study of Joy (1937), hydrogen emission lines were noticed in the spectra of the variable W Virginis. Later, Joy (1940) reported similar features during the pulsation cycle of a Cepheid variable in the globular cluster Messier 3 to those of W Vir. Finally, after the indentification of Population I and II stars by Baade (1944), it has been firmly established that the Cepheid variable and W Vir belong to a separate subgroup of Cepheid variables, with the latter becoming the prototype of its class.

Similarly to the RR Lyrae stars, the light curves of Classical Cepheids exhibit characteristic shapes in different bands, as shown in Figure 1.7. However, around $P \sim 8$ days, a small bump appears on the the light curves of Classical Cepheids near the end of the descending branch. With increasing periods, this bump moves closer to the maxima of the light curves, eventually becoming the main maxima for Cepheids with $P \gtrsim 11$ days. This phenomenon is called the Hertzsprung progression (Hertzsprung 1926; also see Figure 7 of Soszyński et al. 2010), and is thought to be caused by a resonance between the fundamental and second radial overtone modes (Bono et al. 2000; Gastine and Dintrans 2008).

The period-luminosity (PL) relation of Classical Cepheids was discovered by Henrietta Leavitt (Leavitt and Pickering 1912). The existence of this relation allowed Hubble (1926) to determine that the Andromeda "cloud" was indeed a galaxy, separated from, and located outside of, the Milky Way. In the years since, Classical Cepheids have kept their status as one of the most powerful tools to measure extragalactic distances, and hence in the calibration of the extragalactic distance scale, which resulted in continued interest in the ever more accurate calibration of the PL relation (Macri et al. 2015; Riess et al. 2018; Groenewegen 2018).

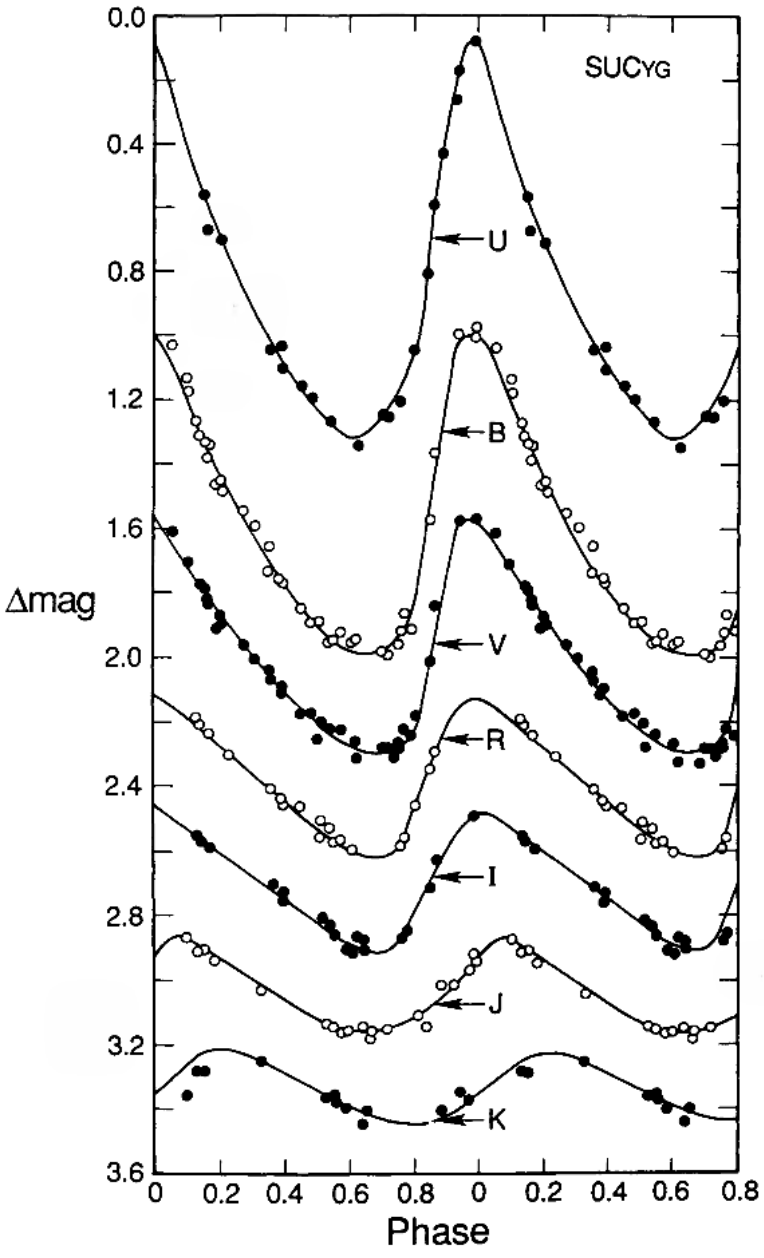

Figure 1.6: Folded *UBVRIJK* light curves of a typical Classical Cepheid (figure from Madore and Freedman 1991).

Another relation followed by Classical Cepheids is the period-age relation (Bono et al. 2005; Anderson et al. 2016): longer period Classical Cepheids are on average younger. As younger post-main sequence stars are also more massive, they also cross the IS at higher luminosities, becoming brighter, and therefore longer-period Classical Cepheids. However, there are substantial uncertainties in this relation related to our incomplete

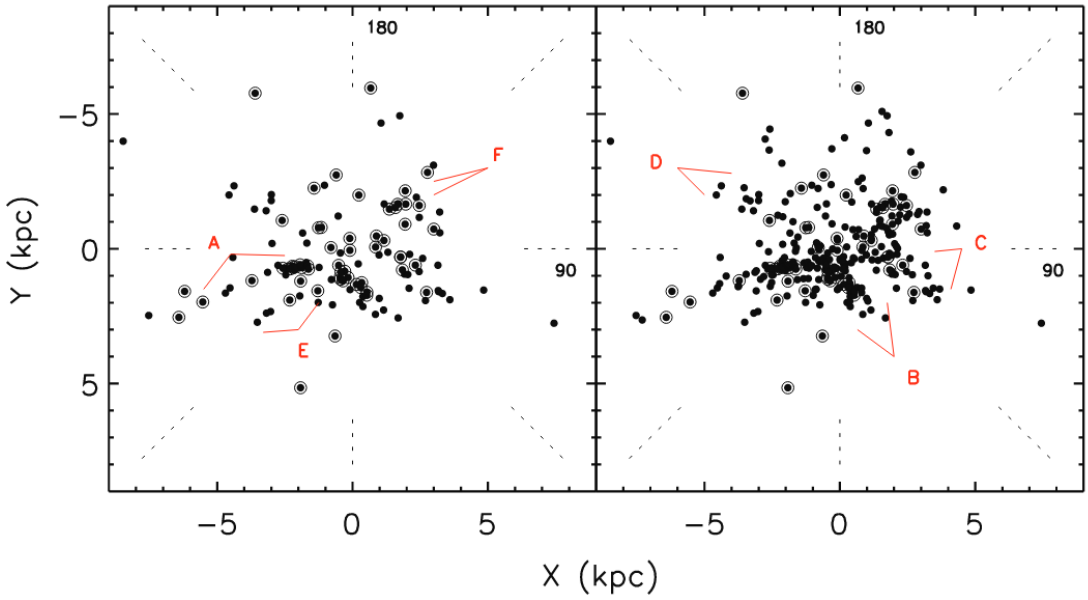

Figure 1.7: The Galactic distribution of Classical Cepheids (solid circles) and young open clusters (circled dots) in the Solar neighborhood, showing the local spiral structure. *Left:* Only long period (young) Classical Cepheids are shown ($P \geq 13$ d). *Right:* As on the right panel, but short period (older) Classical Cepheids are also included ($P \geq 5$ d). (Figure from Majaess et al. 2009.)

knowledge of the initial rotation rates of stars on the upper main sequence (Anderson et al. 2016). Furthermore, ideally, the crossing number of the Classical Cepheids (second or third crossing) should be known beforehand to chose the correct relation for the determination of the age; however, this is only known for stars with a long history of repeated observations, and hence observed period changes (Fadeyev 2013).

Recently, Udalski et al. (2018) announced the discovery of more than 1000 new Classical Cepheids, more than doubling the number of previous known objects of this type in the Milky Way. Using this data, Skowron et al. (2018) derived key properties of the Galactic disk, while also showing that the Classical Cepheids in the southern Galactic disk do indeed follow the Galactic warp (Kerr 1957; Oort et al. 1958; Reylé et al. 2009). Furthermore, they have identified three dominant star formation episodes which have taken place in the Galactic disk in the last 100 million years. As young stars, the distribution of Classical Cepheid variables partially still follow the distribution of their original birth places, which for most of them, were the star-forming regions found within the

spiral arms. Skowron et al. (2018) showed how Cepheids formed in the three major star-formation episodes have reached their current positions, migrating out of their respective spiral arms. However, similarly to the case of RR Lyrae variables, only a handful of studies have explored the Classical Cepheids in the most obscured regions of the Milky Way photometrically (see, e.g, Dékány et al. 2015a,b; Matsunaga et al. 2016), and even fewer spectroscopically (Inno et al. 2019), making our view of the structure of the disk of the Milky Way incomplete, as there are at least 15 000 Classical Cepheids expected to reside within the Milky Way (Majaess et al. 2009).

1.4 The VISTA Variables in the Vía Láctea survey

The VISTA Variables in the Vía Láctea ESO Public Survey (VVV; Minniti et al. 2010) started in 2010 and concluded in 2016, and utilized the 4m-class Visible and Infrared Survey Telescope for Astronomy (VISTA; McPherson et al. 2006) of the European Southern Observatory (ESO). The observations taken by VISTA using the near-infrared imager VIRCAM are reduced by the VISTA Data Flow System (Emerson et al. 2004) at the Cambridge Astronomical Survey Unit (CASU).

As one of the largest ESO Public Surveys, the $YZJHK_S$ single-epoch observations, as well as the K_S-band time series photometry, provided by VVV have given us a unique insight into the inner Milky Way, which is otherwise obscured by the extinction caused by the dust. Therefore VVV complements the optical variability surveys, such as that carried out by the OGLE project (Udalski et al. 2015)

RR Lyrae variables towards the VVV bulge fields have already been utilized in a variety of ways to study the 3D distribution of the old bulge stellar population (Dékány et al. 2013), their properties in the outer bulge regions (Gran et al. 2016), the distance to the Galactic center (Majaess et al. 2018), as well as the old stellar population residing therein (Contreras Ramos et al. 2018b). In contrast, there have been fewer studies concerning RRL stars in the disk regions of the VVV survey. Minniti et al. (2017a) conducted a preliminary study of about 25% of the VVV disk area, in the region defined by the rectangular area $-2.^{\circ}24 \lesssim b \lesssim -1.^{\circ}05$, $295^{\circ} \lesssim l \lesssim 350^{\circ}$ in Galactic coordinates, identifying 404 RRL stars. Furthermore, RR Lyrae variables have been utilized to confirm a globular cluster

candidate (Minniti et al. 2017b) in the Galactic disk. The recent development of a machine learned RR Lyrae classifier (Elorrieta et al. 2016) allows the distinction of RRab stars from other types of variables in the K_S-band time series observations obtained by the VVV survey.

Similarly to the RR Lyrae variables, investigations of Classical and Type II Cepheids in the VVV survey have been most constrained to the Galactic bulge area. As up to this point, no machine learned classifier has been published that can identify Cepheid variables, and classifying them into Classical and Type II subclasses, utilizing only K_S-band observations, other criteria must be utilized to distinguish the two subclasses of variables. Dékány et al. (2015a) identified two long-period variable stars, which have almost exactly the same periods, color excesses, and apparent brightnesses, while being separated only by an angular distance of 18.″3 on the sky. Due to these peculiar properties, it has been concluded that these two variables are Classical Cepheids, still found in the open cluster where they have formed, but the other, dimmer stars are invisible in the VVV observations due to the high amounts of extinction. Similarly, Dékány et al. (2015b) identified a sample of Classical Cepheids located towards the Galactic bulge, which showed higher extinction values than expected from reddening maps (Gonzalez et al. 2012). Meanwhile, Bhardwaj et al. (2017a) used Type II Cepheids identified by the OGLE survey (Soszyński et al. 2017) to constrain their period-luminosity relationships, while Braga et al. (2019) identified Type II Cepheids in the direction of the Galactic center.

1.5 The layout of the thesis

This thesis presents a summary of the efforts undertaken during my PhD studies towards the discovery and characterization of RR Lyrae and Cepheid variables in the near-infrared observations of the VVV survey. This work has been conducted in close collaboration with researchers both at the Instituto de Astrofísica de la Pontificia Universidad Católica de Chile, as well as the Astronomisches Rechen-Institut der Universät Heidelberg.

The individual chapters each closely track a specific publication (either already published or in preparation), presenting the results of a particular aspect of the work. These

chapters, highlighting my contributions, are:

- Chapter 2 presents my results on the characterization of the near-infrared light-curve shapes of RRab stars. I developed a novel, robust fitting method of K_S-band RRab light-curves, based on their light-curve representations as a sum of principal components. This fit also estimates the J-band light-curve shape, allowing the determination of accurate mean J magnitudes even if only one observation is available in that band. I have found that the amplitudes of the principal component fits, together with the pulsation period, allow the estimation of the metallicities of RRab stars based only on K_S-band light curves. The results presented in this chapter were published in Hajdu et al. (2018).

- Chapter 3 presents the RRab variable search undertaken in the VVV disk regions. I have participated in every step of the analysis process, and in particular, I fit the K_S-band light-curves of RRab stars with the method developed in Chapter 2, after which I also determined their individual metallicities, which were crucial for allowing a detailed analysis of the metallicity distribution function of the RRab stars across the Galactic disk. The results presented in this chapter were published in Dékány et al. (2018).

- Motivated by some light-curve inconsistencies found in the H-band observations of RR Lyrae stars analyzed during the disk RRab search, I performed a revision of the photometric calibration of the VISTA telescope. In particular, in Chapter 4, I present the serious photometric zero-point calibration issues I have discovered, resulting from blending of stars in the Two Micron All Sky Survey (2MASS; Skrutskie et al. 2006) catalog, which results in biased the zero points of VISTA observations. These zero-point issues seriously diminish both the photometric accuracy as well as the precision of the VISTA telescope photometry in dense stellar fields. I give suggestions for the improvement of the VISTA pipelines, and recalibrate all pawprint measurements of the VVV survey. The results presented in this chapter will be published in Hajdu et al. (in prep.).

- The recalibrated pawprint photometric measurements, in combination with the

Cepheid light-curve classifier developed by Dr. István Dékány were used to search the Galactic plane area covered by the VVV survey for Classical and Type II Cepheids, as presented in Chapter 5. A new extinction law is derived using the Type II Cepheids, which is used to calculate the distances of the VVV Classical Cepheids in the Galactic disk. I show that the radial age distribution of Classical Cepheids in the far side of the Galactic disk follows that of the Cepheids in the near side. Furthermore, I perform a simple azimuthal rotation of the current positions of VVV Classical Cepheids in order to connect them to the spiral arms where they were originally born. The results presented in this chapter will be published in Dékány et al. (in prep.).

- Chapter 6 provides a brief summary or the presented work and gives an outlook for future studies building upon them.

2

RR Lyrae light curves in the near-infrared

2.1 Light-curve properties of RR Lyrae variables

RR Lyrae variables are useful tracers of old stellar populations of galaxies, as well as serving as standard candles for distance determinations (Chapter 1). While their optical light-curve properties are very well studied, the same cannot be said about the near-infrared (near-IR) light-curves. This is mainly due to their diminished amplitudes and technical challenges associated with near-IR observations, in general. Furthermore, it is known that their near-IR light curves contain fewer features than at optical wavelengths, making their discovery challenging due to the possible confusion with binary variables of the W UMa type. This is especially true of RR Lyrae showing first overtone (RRc) or double-mode (RRd) pulsations. Nevertheless, the lower absorption in the infrared bands, as well as the lowered effect of metallicity on the absolute magnitudes, makes near-IR observations desirable for the determination of the properties of old stellar populations. In this Chapter, the near-IR properties of fundamental-mode RR Lyrae variables (RRab subtype, from here on RRLs) are examined, in order to develop the necessary methods for the characterization of RRL variables discovered in the VVV survey.

Generally, to derive photometrical distances to these variables, their mean apparent magnitudes in at least two bands have to be determined, allowing the estimation of the line-of-sight extinction towards each star, after adopting an absolute-to-selective extinction ratio (extinction "law"). In studies of classical radial pulsators, such as Cepheids and RRLs, it is customary to fit the light curves with a truncated Fourier series (see, e.g., Simon and Lee 1981), and to use the intercept of this fit as a measure of the mean apparent magnitude. The lower amplitudes and relatively high scatter of near-IR photometry present a

challenge for this technique: in time series with a limited amount of measurements, there are not enough data to accurately determine the coefficients of the high-order truncated Fourier series needed to describe sharp features, such as the region of the minimum light of the light curves. Alternatively, light curve templates, such as those of Jones et al. (1996), could be used as model representations of the time-series. However, this approach has drawbacks: the Jones et al. (1996) templates only cover the *K* band, and the *J*-band light curves of fundamental-mode RR Lyrae are markedly different; with only four RRL templates, not all possible light-curve shapes are represented.

Although the effect of metallicity on the absolute magnitudes is slightly lessened in the near-IR compared to the optical (e.g., Bono et al. 2003; Catelan et al. 2004), knowledge of individual RRL metallicities can still provide valuable insight into the star formation histories of the oldest populations of the Milky Way in parts not accessible by optical spectroscopy and/or photometry. Relationships between the light-curve shape, period, and the iron abundance [Fe/H], such as the widely used relation of Jurcsik and Kovács (1996), provide a convenient estimate in the optical regime. Despite their obvious usefulness, no such relation has been established so far in the near-IR bands.

Table 2.1: Collection of RR Lyrae near-IR photometric observations

ID[a]	Period[b]	J[c]	Ref.	ID[a]	Period[b]	J[c]	Ref.
AV Peg	0.390375	+	10	ωCen NV458	0.620326	+	11
V445 Oph	0.397020	+	4	ωCen V032	0.620347	-	11
W Crt	0.412012	+	12	ωCen V018	0.621689	+	11
AR Per	0.425549	+	10	ωCen V096	0.624527	+	11
SW And	0.442262	+	1,9,10	SS Leo	0.626344	+	4
RR Leo	0.452390	+	10	ωCen V004	0.627320	+	11
ωCen V112	0.474359	+	11	ωCen V115	0.630469	+	11
BB Pup	0.480532	+	12	ωCen V146	0.633092	-	11
ωCen V130	0.493250	-	11	ωCen V040	0.634072	+	11
WVSC 054[d]	0.501267	+	5	ωCen V122	0.634929	+	11
ωCen V074	0.503209	+	11	ωCen V041	0.662942	+	11
ωCen NV457	0.508619	+	11	ωCen V013	0.669039	+	11
ωCen V023	0.510870	+	11	ωCen V114	0.675307	+	11
ωCen V107	0.514102	+	11	ωCen V149	0.682728	+	11

Continued on next page

Table 2.1 – *Continued from previous page*

ID[a]	Period[b]	J[c]	Ref.	ID[a]	Period[b]	J[c]	Ref.
WVSC 055[d]	0.514810	+	5	ω Cen V046	0.686971	+	11
ω Cen V005	0.515274	-	11	ω Cen V088	0.690211	+	11
WVSC 047[d]	0.519611	+	5	NSV 660	0.636985	+	13
ω Cen V008	0.521329	+	11	W Tuc	0.642230	+	2
WVSC 046[d]	0.529820	+	5	ω Cen V086	0.647844	+	11
ω Cen V120	0.548537	+	11	X Ari	0.651180	+	3,6
WVSC 050[d]	0.551535	+	5	ω Cen V134	0.652903	+	11
ω Cen V100	0.552745	+	11	ω Cen V069	0.653195	-	11
RR Cet	0.553038	+	10	ω Cen V132	0.655656	-	11
TU UMa	0.557648	+	1	ω Cen V102	0.691396	+	11
ω Cen V067	0.564446	-	11	ω Cen V097	0.691898	+	11
ω Cen V044	0.567545	+	11	ω Cen V007	0.713026	+	11
ω Cen V056	0.568023	-	11	VY Ser	0.714094	+	4,8
SW Dra	0.569670	+	7	ω Cen V116	0.720074	-	11
ω Cen V106	0.569903	-	11	ω Cen V034	0.733967	+	11
RV Oct	0.571163	+	12	ω Cen V172	0.738049	-	11
ω Cen V113	0.573375	+	11	ω Cen V085	0.742758	+	11
ω Cen V051	0.574152	+	11	ω Cen V109	0.744098	+	11
WY Ant	0.574341	+	12	ω Cen V111	0.762905	+	11
ω Cen V073	0.575215	+	11	ω Cen V099	0.766181	-	11
ω Cen V055	0.581724	+	11	ω Cen V054	0.772915	+	11
ω Cen V181	0.588370	+	11	ω Cen V038	0.779061	+	11
ω Cen V025	0.588500	-	11	ω Cen V026	0.784720	+	11
ω Cen V045	0.589116	+	11	ω Cen V057	0.794402	+	11
ω Cen V125	0.592888	+	11	ω Cen V015	0.810642	+	11
ω Cen V108	0.594458	-	11	ω Cen V268	0.812922	+	11
RV Phe	0.596400	+	2	ω Cen V063	0.825943	+	11
TT Lyn	0.597436	+	10	ω Cen V128	0.834988	+	11
ω Cen V033	0.602324	+	11	ω Cen V144	0.835320	+	11
ω Cen V090	0.603404	+	11	ω Cen V003	0.841258	+	11
ω Cen V049	0.604650	+	11	ω Cen V411	0.844880	+	11
UU Cet	0.606074	+	2	ω Cen V104	0.866308	+	11
ω Cen V079	0.608276	+	11	ω Cen V091	0.895225	+	11
ω Cen V118	0.611618	+	11	ω Cen V150	0.899367	+	11
ω Cen V020	0.615559	+	11	ω Cen NV455	0.932517	+	11

Continued on next page

Table 2.1 – *Continued from previous page*

ID[a]	Period[b]	J[c]	Ref.	ID[a]	Period[b]	J[c]	Ref.
ωCen V027	0.615680	+	11	ωCen V263	1.012158	+	11
ωCen V062	0.619770	+	11				

Notes:

[a] Unique identifier of the variable.

[b] Period in days.

[c] Flag whether the *J* band data are present and utilized in Sect 2.2.4.

[d] The variables WVSC 054, 055, 047, 046 and 050 are V83, V116, V108, V55 and V40, respectively, from the globular cluster Messier 3 (Clement et al. 2001; Ferreira Lopes et al. 2015).

References (photometric system): 1 – Barnes et al. (1992) (CIT), 2 – Cacciari et al. (1992) (ESO), 3 – Fernley et al. (1989) (AAO), 4 – Fernley et al. (1990) (SAAO), 5 – Ferreira Lopes et al. (2015) (WFCAM), 6 – Jones et al. (1987a) (CIT), 7 – Jones et al. (1987b) (CIT), 8 – Jones et al. (1988) (CIT), 9 – Jones et al. (1992) (CIT), 10 – Liu and Janes (1989) (CIT), 11 – Navarrete et al. (2015) (VISTA), 12 – Skillen et al. (1993) (SAAO), 13 – Szabó et al. (2014) (2MASS). The photometric systems are defined by González-Fernández et al. (2018, VISTA), Hodgkin et al. (2009, WFCAM), Carpenter (2001) and references therein (all other systems).

In this Chapter, special attention is paid to the particular properties of light curves produced by the VVV survey. Several new methods are introduced, driven by the desire of making maximum use of the RRLs in the VVV survey, for the analysis of variable stars in general, and for the near-IR light curves of RRLs in particular. Utilizing a high-quality RRL sample collected from the available literature (Section 2.2.1), principal component analysis (PCA) is applied to the K_S-band light curves with the aim of decreasing the number of parameters parameters required to accurately describe the various light-curve shapes of RRL stars in this band (Section 2.2). It is demonstrated that the *J*-band light curve shape can also be approximated, utilizing the K_S-band principal component amplitudes, allowing the determination of the *J*-band average magnitudes, even from a single observation

(Section 2.2.4). Utilizing the first few principal components as basis vectors, a robust non-linear fitting technique is developed (Section 2.3). Finally, it is demonstrated, on a selected sample of OGLE-IV RRLs, that the light-curve shapes of RRLs in the K_S band, together with the pulsation period, can be used to estimate their metallicities, similarly to their optical light curves (Section 2.4). The methods developed here are utilized in Chapter 3 for the characterization of the RRL population of the VVV disk fields.

2.2 Model representation of near-IR RRab light-curves

Due to its widespread use and intuitive nature, PCA has been elected as the analysis method for the near-IR RRL light curves. PCA is a widely used dimensionality reduction procedure to transform an original set of variables by an orthogonal transformation into a new set of linearly uncorrelated variables called principal components (PCs). Generally, the first few PCs contain most of the variation in the original data. The procedure was first described more than a century ago by Pearson (1901), and then rediscovered and named by Hotelling (1933). PCA has two main uses for data analysis: (1) reducing the number of dimensions of a data set by keeping only the most significant PCs and (2) identifying hidden trends in the data. As astronomical data sets are inherently multidimensional (e.g., images, spectra, individual element abundances of stars, etc.), PCA has been adopted for both purposes by the astronomical community. Among the myriad applications of PCA in astronomy are: spectral classification of galaxies (Galaz and de Lapparent 1998), stars (Singh et al. 1998), and quasars (Yip et al. 2004); modeling systematics in light curves (Jordán et al. 2013); removal of galaxies from images with the aim of finding gravitationally lensed background galaxies (Paraficz et al. 2016); analysis of galaxy velocity curves (Kalinova et al. 2017); as well as looking for correlations between diffuse interstellar band features (Ensor et al. 2017).

In the context of variable stars, the most important applications of PCA for the study of Cepheids and RR Lyrae stars were presented by Kanbur et al. (2002), Kanbur and Mariani (2004), Tanvir et al. (2005), and Bhardwaj et al. (2017b). Furthermore, a range of variable star classes were analyzed with PCA by Deb and Singh (2009), in order to evaluate PCs as a metric for light curves in large databases.

2.2.1 The light-curve training set

High-quality RRL K_S and J band photometry was collected from the available literature for 101 RRLs. Table 2.1 summarizes this (training) data set. The available data can be categorized into three main types: near-IR photometry taken with the aim of performing Baade-Wesselink analysis of RRLs (Barnes et al. 1992; Cacciari et al. 1992; Fernley et al. 1989; Fernley et al. 1990; Jones et al. 1987a, 1987b, 1988, 1992; Liu and Janes 1989; Skillen et al. 1993); serendipitous observations of RRLs in the 2MASS (Szabó et al. 2014) and WFCAM (Ferreira Lopes et al. 2015) calibration fields; and the extensive J and K_S variability study of ωCentauri by Navarrete et al. (2015). There are other sources of near-IR time-series photometry available in the literature for RRLs (see, e.g., Angeloni et al. 2014), but only photometry where the phase coverage and quality of the observations are adequate for the accurate determination of the near-IR light-curve shapes, at least in the K_S band, are utilized here.

Examining the data set presented in Table 2.1 reveals that the variables in the field of ωCentauri (Navarrete et al. 2015) dominate the sample. Although this could introduce a heavy bias towards a particular metallicity, this is minimized by the fact that ωCen contains multiple stellar populations (e.g., Valcarce and Catelan 2011; Gratton et al. 2012; and references therein), and at least two populations with metallicities [Fe/H] ~ -1.2 and ~ -1.7 contribute to the RRL sample (Sollima et al. 2006, and references therein). The variables utilized from the photometry of Ferreira Lopes et al. (2015) are all members of the globular cluster M3, with a metallicity of ~ -1.5 (Carretta et al. 2009). Furthermore, the field RRLs have a wide metallicity distribution, therefore this sample covers most of the possible metallicity range of RRL variables.

The period distribution of the sample also covers most of the possible period range of ab-type RRLs, from 0.39 to 1.01 days. While very metal rich RRLs can have periods as short as about 0.35 days, their optical light-curve shapes are not drastically different from those with periods around 0.4 days. Therefore, it can surmised that the present data set can be considered representative of RRLs and their light-curve shapes.

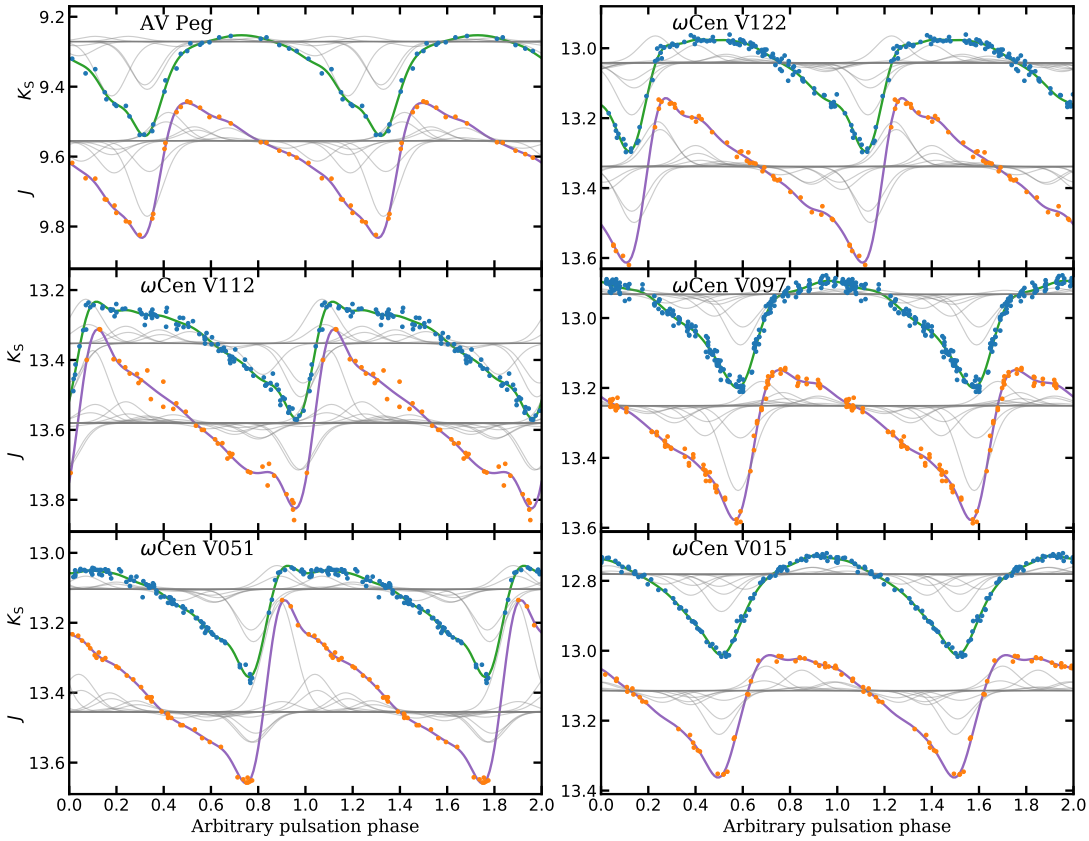

Figure 2.1: Typical folded K_S (top, blue points) and J-band (bottom, orange points) light curves of the RRL sample. Each light curve is modeled with a sum of periodic Gaussian (von Mises) basis functions (Eq. 2.1), following Eq. 2.2. LASSO regularization results in most periodic Gaussians having zero amplitudes. The individual periodic Gaussian functions with non-zero amplitudes are illustrated by the faint grey lines for each band and variable. The green and purple curves illustrate the light curve fit (sum of the periodic Gaussians) in the K_S and J bands, respectively.

2.2.2 Data preparation

Before PCA can be applied to the data set presented in Section 2.2.1, the phase-folded light-curve shapes must be described and resampled onto a regular grid. As mentioned before, it can be hard to accurately represent the near-IR RRL light curves as a Fourier series. Furthermore, the light curves have vastly different numbers of data points: NSV 660 has almost 3000, while for the RRL found in ω Centauri, Navarrete et al. (2015) only acquired a total of 100 and 42 epochs in the K_S and J bands, respectively.

As an alternative to a global Fourier-series based fit, folded light curves can also be

described utilizing techniques based on local fitting techniques. An example of this can be found in the work of Reyner et al. (2010), where bicubic splines were utilized to fit RR Lyrae light curves. Here an alternative method is presented, based on the linear sums of a series of basis functions. Such basis functions could be, for example, Gaussians, which can be aligned to the phased light-curve points with ordinary least squares (OLS) regression. For the problem of fitting RRL light curves accurately, the Gaussian sum fitting example of Fig. 8.4 of Ivezić et al. (2014) has been adopted, for both the K_S and J-band RRL light curves. As RRL light curves are strictly periodic, Gaussians have been replaced with their periodic analogue, the circular normal (also called Von Mises) distributions of the form:

$$f(x) = \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)}, \tag{2.1}$$

where μ is the measure of location (analogous to the mean of the Gaussian distribution), κ is the measure of concentration (where $1/\kappa$ is analogous to the variance, σ^2 of a Gaussian), and $I_0(\kappa)$ is the modified Bessel function of order 0. To model the light-curve shapes, the sum of 100 of these basis functions has been defined, distributed evenly between phases 0 and 1:

$$\mathrm{LC} = m + \sum_{i=0}^{99} A_i \frac{e^{\kappa \cos[2\pi(x-\frac{i}{100})]}}{2\pi I_0(\kappa)}, \tag{2.2}$$

where A_i are the individual amplitudes of the circular normal distributions, and m is the intercept of the fit. Although OLS could be used to find the amplitudes of Eq. 2.2, most of the light curves have less than 100 points available, leading to an underdetermined problem. In such a case, regularization can be introduced to penalize the magnitude of independent parameters (in this case the amplitudes A_i), by modifying the loss function. This is achieved by utilizing the least absolute shrinkage and selection operator (LASSO or L1 regularization, Tibshirani 1996). LASSO adds the sum of absolute values of the regression coefficients multiplied by a regularization parameter α to the loss function. While utilizing LASSO, generally most coefficients end up being zero[1].

[1]Another popular choice for these kinds of problems is the Tikhonov regularization, also known as L2 regularization or Ridge regression, where the sum of the squares of the fit coefficients times α is

In order to implement this fitting procedure, linear regression routines of scikit-learn (Pedregosa et al. 2011) were utilized to fit the folded K_S and J light curves of each RRL with the model presented by Eq. 2.2, utilizing LASSO regularization. This fit has two arbitrary hyperparameters: κ and α. Both cross-validation (leave-one-out for stars with few light-curve points, N-fold otherwise) and manual inspection of the resulting light curve fits were utilized to determine the optimal values of these parameters. High values for the concentration parameter κ (analogous to a small σ^2 in the case of a simple Gaussian) grants the model the ability to fit sharp features, such as the rising branches of certain variables. However, it can also cause an overfit in phase ranges with few points. Conversely, at low values of κ the model cannot fit sharp features. It has been found that a numerical value of $\kappa = 6$ is optimal for the model for both the K_S and J-band light curves of RRLs. Cross-validation and inspection of fits with different values of the regularization parameter α resulted in a range of different optimal values for different stars, typically in the range between 10^{-4} and 10^{-5}. As visual inspection did not reveal significant differences when changing the regularization parameter between these two values, an intermediate value of $10^{-4.5}$ is adopted for all of the stars.

Figure 2.1 illustrates the quality of the light curves and their fits. During the fitting process, some outlying points have been removed manually and the periods of some variables have been revised, when tension existed between different photometric sources. Table 2.1 contains the revised periods for all variables. Figure 2.2 compares the fits with Fourier series of different orders. As can be seen, this method provides a better representation for variables with light-curve gaps.

The utilized light curves had been obtained in a variety of photometric systems, as detailed in Table 2.1. Therefore, the resulting light curve fits were all transformed to the photometric system of VISTA, in every pulsation phase. For variables not in the VISTA, WFCAM, or 2MASS systems, first they were transformed to the system of 2MASS utilizing the updated transformation formulae[2] of Carpenter (2001). Then the 2MASS and WFCAM

added to the loss function. In contrast to LASSO, Ridge regression does not result in sparse solutions (i.e., most parameters do not end up being zero), making the interpretation of the results of the fit harder to interpret.

[2] http://www.astro.caltech.edu/~jmc/2mass/v3/transformations/

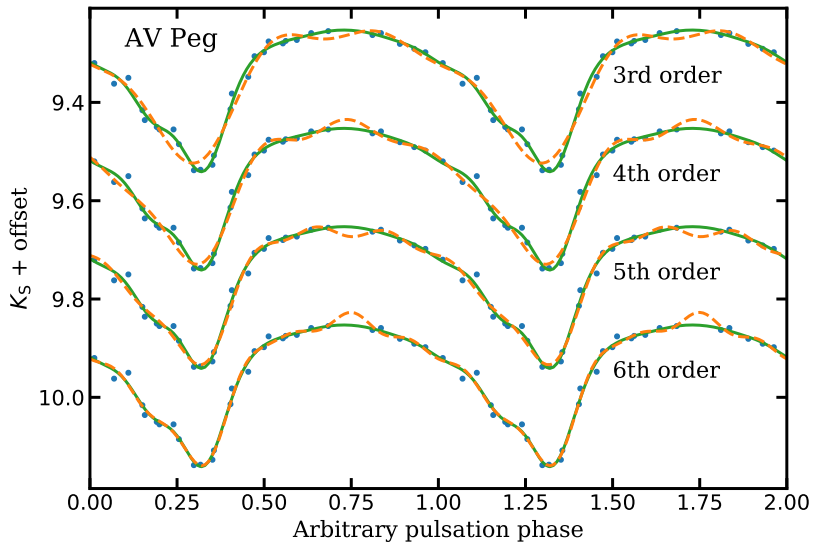

Figure 2.2: Comparison between the circular normal distribution basis fits (Eq. 2.2, continuous lines) and Fourier series of different orders (dashed lines).

photometry was transformed to the VISTA system with the help of the CASU version 1.4 transformations given by the Cambridge Astronomical Survey Unit[3].

2.2.3 Application of PCA on the K_S-band light curves

In order to apply PCA, the light curve fits are sampled on a grid of 100 phase points, evenly distributed from 0.0 to 0.99. In the analysis of pulsating variables, it is customary to set the light curve maxima at phase 0.0; however, inspecting the K_S-band light curve examples of Figure 2.1 reveals that the maxima of RRLs in the K_S band are ill-defined: the timing of the maxima of the light curve depends heavily on the strength of the bump on the rising branch. Therefore, as an alternative, the light curves were aligned by the much sharper feature of the light curve minima (similarly to the case of eclipsing binaries).

In most applications of PCA, the sample values (the magnitudes in this case) are usually

[3] `http://casu.ast.cam.ac.uk/surveys-projects/vista/technical/photometric-properties` The transformations between the VISTA and WFCAM systems were carried out using the relations updated on July 30, 2014. Updated transformations were provided by González-Fernández et al. (2018). The changes in the resulting transformed magnitudes were revised, and were found to be less than 0.005 mag. Since, in addition, this only affects the five stars from Ferreira Lopes et al. (2015), none of the results in this chapter are significantly affected by the choice of transformation equations between these two systems.

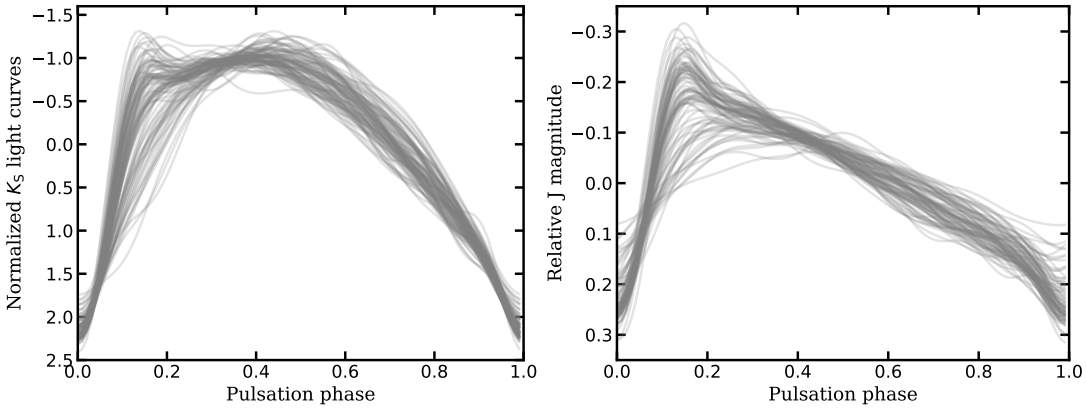

Figure 2.3: *Left:* The K_S-band normalized, folded and minima-aligned RR Lyrae light curves of the training set. *Right:* The J-band folded and minima aligned RR Lyrae light curves.

normalized to a mean of 0 and scatter of 1 along each dimension (phase) after subtracting the mean value along each dimension (phase). As the goal here is to describe the light-curve shapes of RRLs in the K_S band as a linear combination of PCs, it has been found that better results are obtained by not subtracting the mean in each dimension. Furthermore, in this case, each light curve is normalized independently to have a mean of 0 and scatter of 1. These aligned, normalized input light curves can be seen in the left panel of Figure 2.3.

PCA was carried out by utilizing a related transformation called singular value decomposition (see, e.g. Chapter 7.3.1 of Ivezić et al. 2014), adopted from the PCA module of scikit-learn (Pedregosa et al. 2011)[4]. Figure 2.4 shows the first six PCs, according to the PCA decomposition of the RRL K_S-band light curve sample. As the light curves were chosen *not* to be mean subtracted and normalized in each dimension (phase), the first PC contains the average light-curve shape of the normalized light curves. As the original light curves (presented in the left panel of Figure 2.3) can be reconstructed as a linear combination of the PCs, the role of further components in the description of individual light curves can be easily understood. For example, the second component modifies the average light curve (the first PC), so that the bump at the end of the rising branch (around phase 0.15) can become more or less pronounced, while the third component is important

[4]The PCA module present in scikit-learn automatically subtracts the mean along each data dimension, hence the need for the modification of the original algorithm.

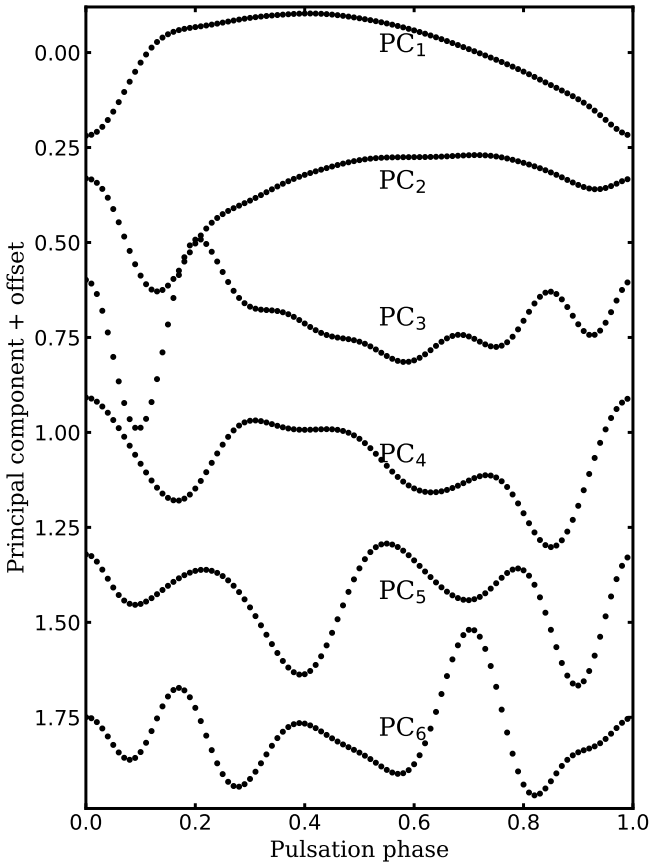

Figure 2.4: The first six principal components of the decomposition of the K_S-band RRL light curves.

to reproduce the double-peaked light curves displayed by some of the variables.

The power of the PCA lies in the fact that generally the first few PCs, together with their amplitudes, are sufficient to describe the original input data with high accuracy, and the rest of the PCs can be discarded, greatly reducing the number of parameters. The number of significant PCs can be decided by examining the fraction of the variance explained by the components. As each light curve was normalized individually instead of normalizing the magnitudes along each dimension (phase), obviously the first PC dominates, with 95.6 percent of the explained variance. The rest of the PCs explain 3.68, 0.14, 0.01, 0.007, etc. percent of the total variance, or 88.8, 3.4, 2.3, 1.7, etc. percent of the residual variance, when the variance explained by the first PC is subtracted from the total variance. Based on these values, as well as inspection of the effect that the omission of different PCs have on the reconstructed light curves, it has been deemed that the first

four PCs are sufficient for describing the K_S-band light-curve shapes of RRab stars.

The PCA method has been applied on the normalized light curves of the sample of variables, in order to emphasize the light-curve shape differences among RRLs in the K_S band, instead of the different pulsation amplitudes that each variable has. Consequently, the individual amplitudes (also called eigenvalues) of the first four PCs, u_{1j}, u_{2j}, u_{3j} and u_{4j}, where $j = 1..101$ is the index of the variables, are insufficient to reconstruct the original light curves, as they have lost the amplitude information due to the normalization of the individual light curve variances. However, by multiplying these amplitudes with the normalization constants applied to the light curves (or, alternatively, by utilizing OLS to directly align the PCs to the light curves), a modified set of PC amplitudes U_{1j}, U_{2j}, U_{3j} and U_{4j} are created for each star in the sample. The sum of the PCs multiplied with these amplitudes recover the original light-curve shapes (and amplitudes) with high accuracy.

The distribution of the amplitudes U_{ij} is illustrated in Figure 2.5. The shapes of the light curves of RRLs, represented by the amplitudes of these first few PCs, potentially contain information on the physical properties of the variables themselves, when the pulsation periods of the stars are taken into account. The first amplitude, U_1 (from now on, the j index is omitted for simplicity) can be viewed as an analogue to the total amplitude in amplitude-period (also called Bailey) diagrams. As described before, additional PCs with non-zero amplitudes modify the average light-curve shape. Therefore, by inspecting Figure 2.5, it is immediately obvious that, at a given period, the amplitudes U_1 and U_2 separate mainly the stars into two sequences, which can be associated with the Oosterhoff type I and II groups of variables (Oo; Oosterhoff 1939; see Catelan and Smith 2015, for a recent review and references). As these groups at least partly correlate with metallicity, such a clear separation hints at the effect metallicity has on the near-IR light curve shapes of RRL variables, similarly to their visual light curves. The possibility of estimating the metallicity of individual RRLs based on their K_S-band light curves, as described by their PC amplitudes U_i, is going to be explored in Section 2.4.

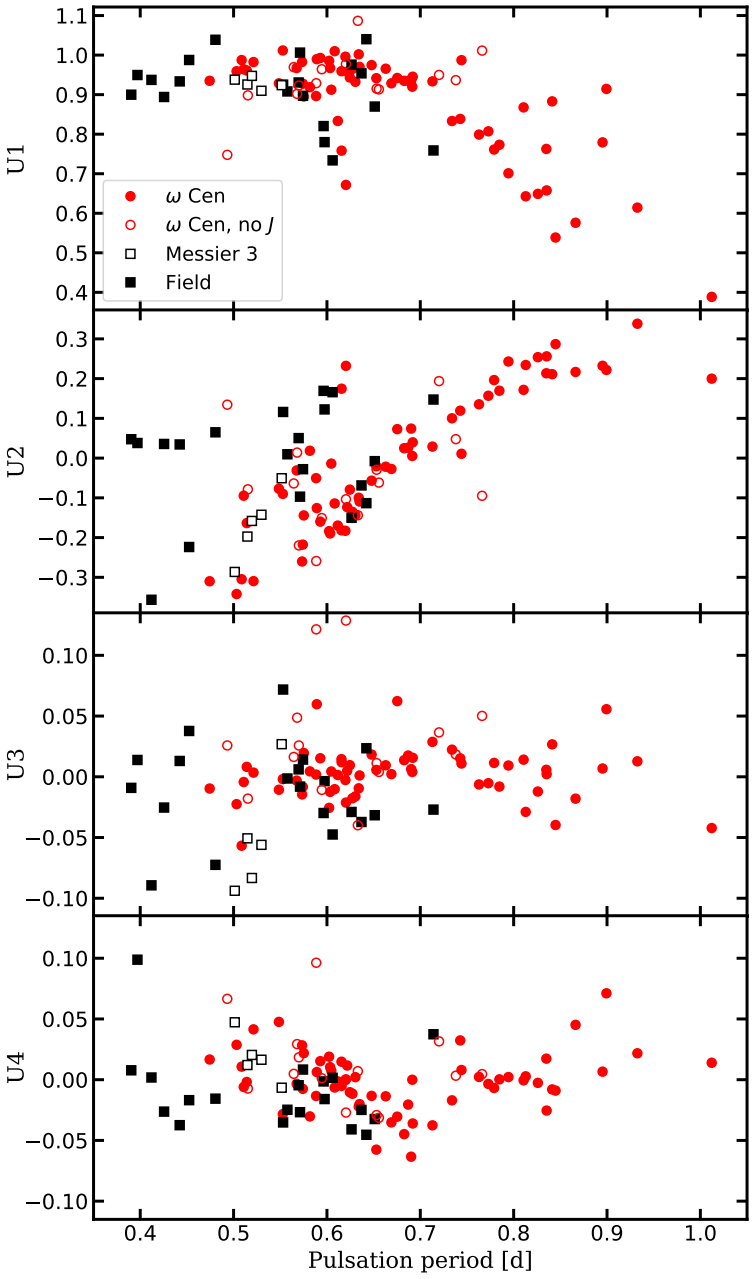

Figure 2.5: Amplitudes U_i of the PCs of the training sample. Circles and squares denote RRLs from the ω Centauri photometry of Navarrete et al. (2015) and other sources, respectively. Empty squares denote RRLs in Messier 3 from the photometry of Ferreira Lopes et al. (2015), while empty circles denote variables not considered for the approximation of the J-band light-curve shapes in Section 2.2.4.

2.2.4 Approximation of the J-band light curve shape

The light-curve shapes and amplitudes of pulsating variables vary among bandpasses, due to the difference in the relative contribution of the change of radius and photospheric temperature during the pulsation cycle to the emitted flux at different wavelengths (Catelan and Smith 2015, and references therein). Comparing the K_S and J-band light curves in Figure 2.1 clearly demonstrates this.

During the course of the VVV survey, each field has been observed only a few times in the J band. In order to determine accurate mean J magnitudes for the RRLs, it is necessary to describe the light-curve shapes of the stars in the J band, i.e., the deviation of the J-band magnitude from its average in each phase of the pulsation cycle. Usually the difference between the light-curve shapes is ignored, and the K_S-band light-curve shape is used to estimate the average J-band magnitude. However, this method, depending on the light-curve phases of the observations, introduces additional scatter in the derived magnitudes.

As the PC amplitudes U_i provide a concise description of the K_S-band light-curve shapes, it can be evaluated whether the J-band the light-curve shapes can be approximated with their use. Besides these amplitudes, the effect the period is also considered as a possible additional parameter, in order to assess its possible effect on the J-band light curves of variables that otherwise possess the same light-curve shapes in the K_S band.

Fourteen of the variables marked in Table 2.1, coming from the ω Centauri sample of Navarrete et al. (2015), have problematic J-band photometry, mostly due to inadequate phase coverage, therefore they are omitted, and the remaining 87 stars are analyzed. These J-band light curves are first sampled in the same 100 phase points as has been done for the K_S-band light curves in Section 2.2.3, where phase 0.0 is the phase of the K_S-band light-curve minimum for each of the variables[5]. These curves are aligned to have a mean of 0, but in contrast to the PCA analysis of the K_S-band light curves, they are not normalized in amplitude.

The training set of J-band light curves is shown in the right panel of Figure 2.3. In

[5]All of the stars have simultaneous J and K_S-band photometry, therefore the possibility of systematic differences between the real pulsation phases between these two bands in negligible.

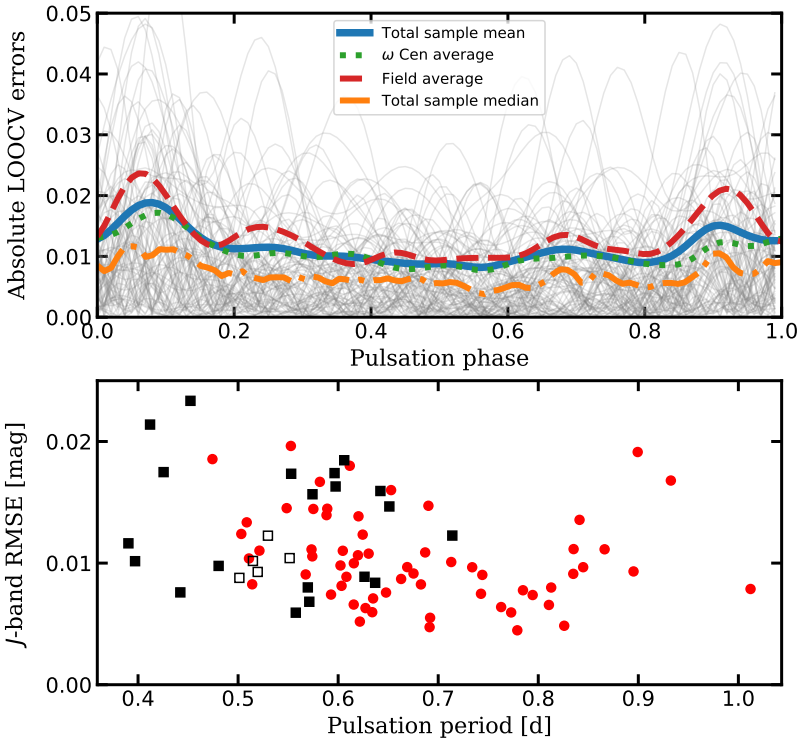

Figure 2.6: *Top:* Absolute J-band residuals of the J-band light-curve approximation shape using the PC amplitudes, as a function of the pulsation phase. Thin grey lines show the individual absolute residuals, according to the LOOCV. Continuous, dashed, and dotted lines mark the average errors for the complete sample, the stars of ω Cen, and the sample omitting the ω Cen stars, respectively. The dot-dashed line shows the median error for each phase bin. *Bottom:* J-band root-mean-squared errors of the LOOCV as a function of pulsation period. The symbols are the same as in Fig. 2.5.

each light-curve phase the approximation of the J-band magnitudes for all of the 87 stars is sought as a linear combination of a set of yet to be determined parameters as the basis vectors. Additionally, it is supposed that the connection between the J magnitudes in different light-curve phases should depend on the same parameters. The candidate parameters are the period P, the four PC amplitudes of the K_S-band light curves U_1, U_2, U_3 and U_4, as well as their polynomial combinations up to the third order (P^2, PU_1, PU_2, PU_3, PU_4, U_1^2, etc.).

An exhaustive search was performed for the best parameter combination, where various permutations were evaluated, containing up to six candidate parameters. As the sample size is fairly small, the usage of a hold-out set, as well as that of N-Fold Cross Val-

idation, was deemed inadequate for the validation of the results. For the former method, the exact (random) choice of the hold-out set can heavily bias the result for such a small data set. Similarly, utilizing N-Folds can also lead to biased results (for example, if all the short period variables are selected to be in one fold). Therefore, for each candidate parameter set, Leave-One-Out Cross Validation (LOOCV) was performed the following way:

1. 87 separate sub-cases are evaluated, in each one omitting the light curve of a different variable;

2. In each case, a linear solution is found utilizing OLS for the J-band magnitudes, using the candidate parameter set as basis;

3. The J-band light curves of the omitted variables are approximated with this parameter set and the OLS coefficients, and their residuals are calculated;

4. For each candidate parameters set, the total root mean square error (RMSE) was used as a performance metric.

It is found that, out of all the possible parameter combinations, by only considering the PC amplitudes of the K_S-band light curves U_1, U_2, U_3 and U_4, the total RMSE of the linear prediction is only 0.012 magnitudes. Although this value can still be lowered by including combinations with the period, such as PU_1, is is noted that this mainly decreases the errors for variables coming from ω Centauri, where multiple stars have very similar periods and K_S and J light-curve shapes, while the errors for other RRLs, especially the short period ones, increase drastically. Figure 2.6 illustrates the residuals both as a function of the pulsation phase (top panel) and for individual stars from the LOOCV (bottom panel).

Based on this analysis, it can be concluded that the J-band light-curve shapes of RRLs can be estimated using their K_S-band PCA amplitudes with high precision. Figure 2.7 illustrates the relation between the K_S-band PCs (left panel) and the coefficients found by OLS for the prediction of the J-band light-curve shape (right panel), when they are calculated using all of the 87 RRLs with good J light curves.

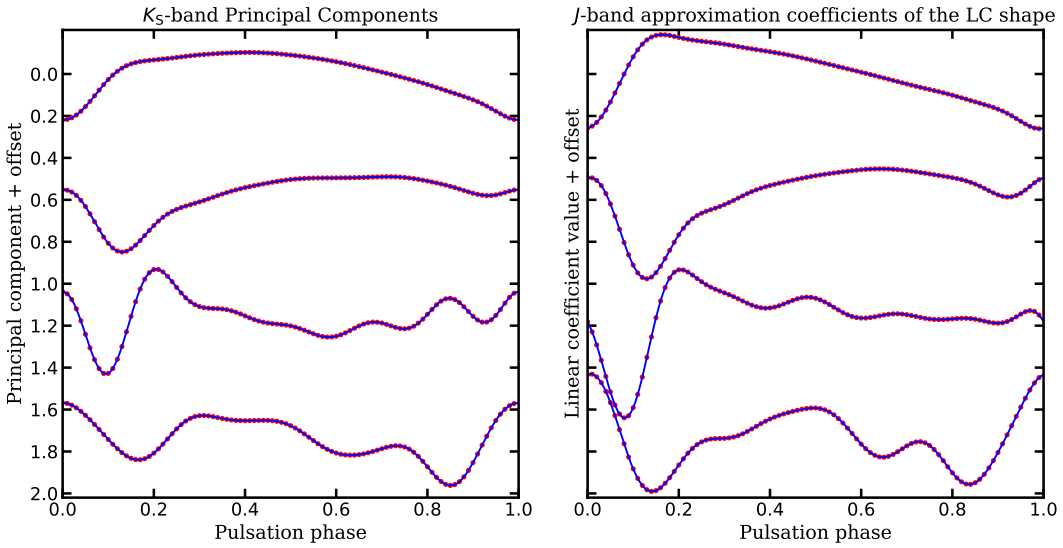

Figure 2.7: *Left:* The first four principal components (red circles) and their continuous Fourier representations (blue lines), utilizing Eq. 2.3. *Right:* The coefficients for the approximation of the J-band light-curve magnitude in each phase (red points), and their Fourier representations (blue lines). Note that the coefficients utilized to reconstruct the J-band light curves describe similar light-curve features as the K_S-band PCs, albeit with larger amplitudes, especially around the rising branch. This behaviour is in line with the fact that RRLs have larger amplitudes at shorter wavelengths, due to the larger stellar surface temperatures near the light-curve maximum.

2.3 Robust fitting of RRL K_S-band light curves

In Section 2.2 it has been shown that the PC amplitudes U_i provide a compact description of the K_S-band light-curve shape; furthermore, the J-band light-curve shapes can also be approximated using the same U_i coefficients as obtained from the K_S-band data (Section 2.2.4). Utilizing these results, it can be evaluated whether an accurate and convenient method for determining the light-curve shapes, periods and average magnitudes of RRLs can be developed, while also taking into account the specific properties of the VVV survey.

The PCs themselves have been utilized before by Tanvir et al. (2005), in the case of Cepheids, to produce a series of realistic template light curves in the optical, which could be used to determine specific parameters, such as the periods or the magnitude at maximum light, from relatively scarce photometry. In contrast to the study of Tanvir

et al. (2005), the method implemented here fits the PCs directly to the VVV K_S-band light curves. In Section 2.2.3, the PCs were determined on a fixed grid of light-curve phases. However, in order to perform a general fit, for example that of a light curve of a variable not in the training set, the PCs have to be expressed as continuous functions of the light-curve phase. To this effect, the individual PCs are transformed to a Fourier representation of the form:

$$PCF_i(x) = \sum_{k=1}^{12} \left[A_{ik} \sin(2\pi k x) + B_{ik} \sin(2\pi k x) \right], \tag{2.3}$$

where x is the light-curve phase, using OLS fitting of the individual PCs. The continuous lines on the left panel of Fig. 2.7 demonstrate the transformed PCs. Then, the sum of these new, continuous representations of the PCs can be used as basis functions to describe the light-curve shapes of RRLs as:

$$\mathrm{LC}(t) = m_0 + \sum_{i=1}^{4} \left[U_i \times PCF_i \left(\frac{t - E_0}{P} \right) \right], \tag{2.4}$$

where t is the Julian Date, and the seven free parameters are the PC amplitudes U_i, the magnitude average m_0, the period P and the zero epoch E_0.

The light-curve representation proposed by Eq. 2.4 is not a linear function, therefore to fit a real light curve, the usage of nonlinear regression is needed. A method has been devised to adjust these seven free parameters, and the model presented in Equation 2.4, to the light curves of RRLs in the VVV survey. Unfortunately, in the more crowded fields and near bright stars, VVV photometry tends to contain a fraction of outlying points. Traditionally, iterative threshold rejection (also known as *sigma clipping*) is utilized to omit outlying measurements in time series analysis of massive data sets. However, in cases of unfortunate data distribution and due to the fact that in the first iteration erroneous observations can have a large effect on the fit, this can result in the flagging of good data as outliers. To avoid this problem, the normally used squared error loss is replaced with the Huber loss function (**eq:huber**) of the form:

$$L_\delta(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{for } |y - f(x) < \delta| \\ \delta |y - f(x)| - \frac{1}{2}\delta^2 & \text{otherwise.} \end{cases} \tag{2.5}$$

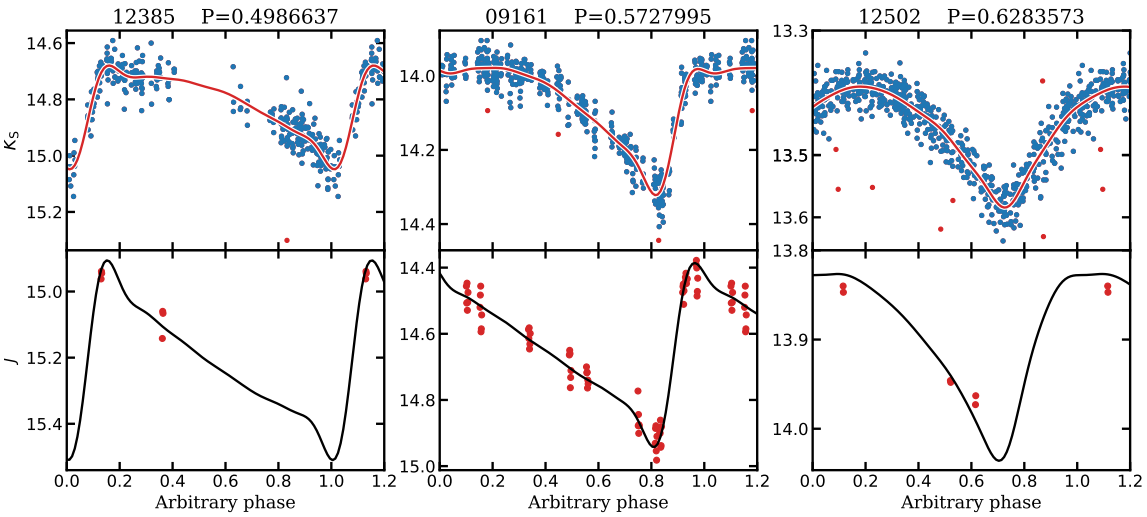

Figure 2.8: Examples of the near-IR light-curve fitting methods implemented in Sect. 2.3. OGLE IDs and periods (in days) are shown on the top of the panels. *Top:* K_S-band light curves and their fits using the first four PCs derived in Sect. 2.2. *Bottom:* J-band light-curve points and the approximations of the J-band light-curve shapes from the PC amplitudes, as described in Sect. 2.2.4. Variable no. 12385 (left panel) illustrates the case of a single large gap in the folded light curve, a challenge for traditional Fourier fitting. No. 09161 has an unusually large number of J-band light-curve points, allowing us to demonstrate the quality of the predicted J light curve. No. 12502 has a much more symmetric light-curve shape, but the method developed here is flexible enough to find a good fit for it.

This loss function behaves identically to the squared error loss for data points with residuals smaller than δ. For residuals bigger than δ, the loss grows linearly with increasing residuals. Therefore, outlying points weigh less than they would, if squared error loss was utilized, a convenient feature in the case of outliers.

The distribution of PC amplitudes of the PCA sample (Fig. 2.5) gives *a priori* information on the possible shapes of RRL light curves, if the sample analyzed in Sect. 2.2 is accepted as representative of RRLs in general. This information is utilized as follows: the first PC amplitude is limited to be in the range $U_1 \in [U_{1,\max} + \delta_{U_1}, U_{1,\min} - \delta_{U_1}]$, where $U_{1,\max}$ and $U_{1,\min}$ are the largest and smallest U_1 amplitudes of the PCA training sample (see Sect. 2.2.3), respectively, and $\delta_{U_1} = (U_{1,\max} - U_{1,\min})/10$. In other words, the U_1 amplitudes are required to lie between the largest and smallest U_1 values, but they are also allowed to be larger or smaller up to ±10% of the total range of U_1 amplitudes. As

for the other amplitudes, their relative values, U_2/U_1, U_3/U_1 and U_4/U_1, need to adhere to analogous requirements on their ranges. These requirements ensure that the fit light curves have shapes that are sufficiently similar to those in the PCA sample.

Besides using the Huber loss function, and limiting the possible range of PC amplitudes during the fitting process, a few additional steps have been implemented during the fitting process, in order to ensure that the light-curve shapes and JHK_S average magnitudes are determined with the highest possible accuracy for VVV RRL light curves:

1. All K_S-band light-curve points brighter than 9 mag, fainter than 20 mag, as well as points farther than ± 0.5 mag from the median are discarded;

2. The remaining light-curve points are fit with the light-curve model of Eq. 2.4, while utilizing the Huber loss function in the form of Eq. 2.5, with $\delta = 0.05$ mag, and the constraints on the ranges of U_1 and $U_{2,3,4}/U_1$, as described above;

3. Points with a residual larger than 3.5σ are omitted, where σ is determined from the absolute median deviation of the fit from step 2;

4. The fit is repeated on the remaining data in the same way as described in step 2;

5. As the light-curve shapes of RRLs are very similar in the H and K_S bands (e.g., they are the same with a scatter of ~ 0.02 mag, depending on the pulsation phase; see Figs. 1 and 2 of Barnes et al. 1992), the K_S light-curve shapes are also used to determine the average magnitudes by fitting them directly to the H-band measurements;

6. The J-band light-curve shapes are predicted using the U_i magnitudes, as described in Sect. 2.2.4, and these are fit to the J-band measurements to determine the mean J magnitude.

The principal outputs of this procedure are the K_S-band light-curve parameters, as well as the robust magnitude estimates of the target RRLs in the JHK_S bands. Figure 2.8 illustrates the VVV K_S-band fits (*top*), as well as the predicted J-band light curves of three

RRLs from the OGLE bulge RRL sample of Soszyński et al. (2014). All these fitting steps are implemented in the program PyFiNeR, which is available on GitHub[6].

2.4 Metallicity estimation from K_S-band photometry

The idea of estimating the metallicity of RRLs from their Fourier light-curve parameters originates from Kovács and Zsoldos (1995). Their original method was substantially improved by Jurcsik and Kovács (1996), who found a simple linear relationship between the iron abundances of RRLs, their periods and the epoch-independent Fourier phase differences ϕ_{31} ($= \phi_3 - 3\phi_1$, where ϕ_3 and ϕ_1 are the Fourier phases of the third and first harmonics of the light curve, respectively; Simon and Lee 1981) of their V-band light curves. Their formula is based on a set of RRL variables with low spectral resolution ΔS measurements (originally defined by Preston 1959; measurements were from Layden 1994 and Suntzeff et al. 1994), calibrated to a unified metallicity scale based on high-dispersion spectroscopic observations of individual RRL variables (Jurcsik 1995).

Following the example set by Jurcsik and Kovács (1996), Smolec (2005) developed similar formulae for the Cousins I band, notable as the band in which most observations of the OGLE surveys are being carried out (Udalski et al. 2015). Smolec (2005) gave two alternative formulae for the derivation of the iron abundance, a 2- and a 3-term formula, and he argued that the latter, which also includes the amplitude of the second Fourier harmonic, provided better results. It has to be emphasized that both the V- and I-band formulae of Jurcsik and Kovács (1996) and Smolec (2005), respectively, have a residual scatter of only ~0.14 dex, which is similar to the accuracy of spectrophotometric methods, and are on the same metallicity scale defined by Jurcsik (1995).

A similar calibration between [Fe/H] and the near-IR light-curve parameters has long been lacking, despite the fact that with the advent of large time-domain near-IR photometric surveys, such as the VVV and the VISTA Magellanic Survey (VMC; Cioni et al. 2011), among others, such an empirical calibration is of key importance, since a large fraction of the newly discovered distant RRL stars along the Galactic plane are beyond the faint magnitude limit of optical surveys. In the following, a method is developed for

[6] github.com/gerhajdu/pyfiner

the determination of metallicity from the light-curve shapes of RRLs in the K_S band, utilizing the overlapping RRLs found by the OGLE project (Soszyński et al. 2014) and also observed by the VVV survey.

2.4.1 VVV photometry of bulge RR Lyrae variables

In order to determine the light-curve shapes of the OGLE+VVV RRLs, the VVV data processed by the VISTA Data Flow System (VDFS, Emerson et al. 2004), provided by the Cambridge Astronomical Survey Unit (CASU), was retrieved. The aperture photometry of detector frame stacks (*pawprints*) was extracted using the Starlink Tables Infrastructure Library (STIL, **stil**) at the coordinates of the OGLE RRLs.

The light-curve fitting algorithm PyFiNeR, described in Section 2.3 was utilized to determine the K_S-band light-curve parameters of RRab variables listed in the OGLE-IV bulge RR Lyrae catalog (Soszyński et al. 2014).

The VDFS provides photometry in circular apertures of different radii, but due to the source crowding in the Galactic bulge fields, the adopted radii have to be chosen carefully. Generally, smaller apertures provide better photometry for dimmer objects, but in relatively uncrowded cases, this does not always holds true. Furthermore, variables found in overlapping regions of tiles and pawprints can have different offsets between them, but generally the best aperture also minimizes this offset. Therefore, for each star the optimal aperture has to be chosen on a case-by-case basis.

The fitting procedure described in Sect. 2.3 was utilized on the five smallest apertures provided by the VDFS for each OGLE RRL. It has been found that the final value of the Huber cost function divided by the number of data points after the 3.5σ clip is a good indicator of the quality of the photometry obtained with a given aperture with respect to the others. Hence, in this section the utilized PC amplitudes correspond to those determined with the aperture for which this value is the smallest.

2.4.2 Revision of *I*-band RR Lyrae photometric metallicity estimates

As large numbers of RRLs are available with both OGLE and VVV photometry, correlations between the photometric abundances determined on the former, and the light-

curve shapes of the latter, can be searched for in a statistical manner. As the main bulge population of RRLs has a mean spectroscopic iron abundance of [Fe/H]~ -1 (Walker and Terndrup 1991), this component is expected to be dominant in the calculated photometric metallicities as well.

The OGLE-IV (or OGLE-III, if the former was not available) I-band light curves of RRLs were fit with a sixth-order Fourier series, in order to calculate the Fourier parameters necessary for the metal abundance formulae of Smolec (2005). Blazhko modulation is common among RRLs (for a study based on a subsample of the bulge-field RRLs, see Prudil and Skarka 2017), and the light-curve shapes of RRLs affected by the Blazhko effect are never equivalent to those of normal RRLs (Jurcsik et al. 2002). Therefore, every RRL which presented a hints of the Blazhko effect in their individually inspected OGLE light curves were eliminated. This inspection also revealed some RRL light curves, where the classification of the variable is dubious. These stars were also removed from the sample, in order to preserve its purity. Furthermore, variables with low-quality K_S-band light curves (Sect. 2.4.1) were also removed.

Eqs. 2 and 3 of Smolec (2005) were utilized to calculate the photometric metal abundance for the remaining variables as:

$$[\mathrm{Fe/H}]_{\mathrm{Smolec\,Eq\,2.}} = -3.142 - 4.902\,P + 0.824\,\phi_{31} \tag{2.6}$$

$$[\mathrm{Fe/H}]_{\mathrm{Smolec\,Eq\,3.}} = -6.125 - 4.795\,P + 1.181\,\phi_{31} + 7.876\,A_2, \tag{2.7}$$

where P is the pulsation period, ϕ_{31} is the epoch independent phase difference between the first and third Fourier harmonics (Simon and Lee 1981), and A_2 is the amplitude of the second Fourier harmonic.

The left panel of Figure 2.9 reveals a striking systematic on top of the overall linear trend in the photometric metallicity calculated with Eq. 2 of Smolec (2005) as a function of the pulsation period. In the main locus of stars, longer-period RRLs seem to have systematically higher metallicities than their shorter-period counterparts. Moreover, a similar trend can be seen on the right panel: the lower-amplitude stars have systematically higher metallicities.

In globular clusters, for both Oosterhoff groups of fundamental-mode variables, the amplitude decreases with increasing periods (see, e.g., Figure 1 of Jurcsik et al. 2017). It is

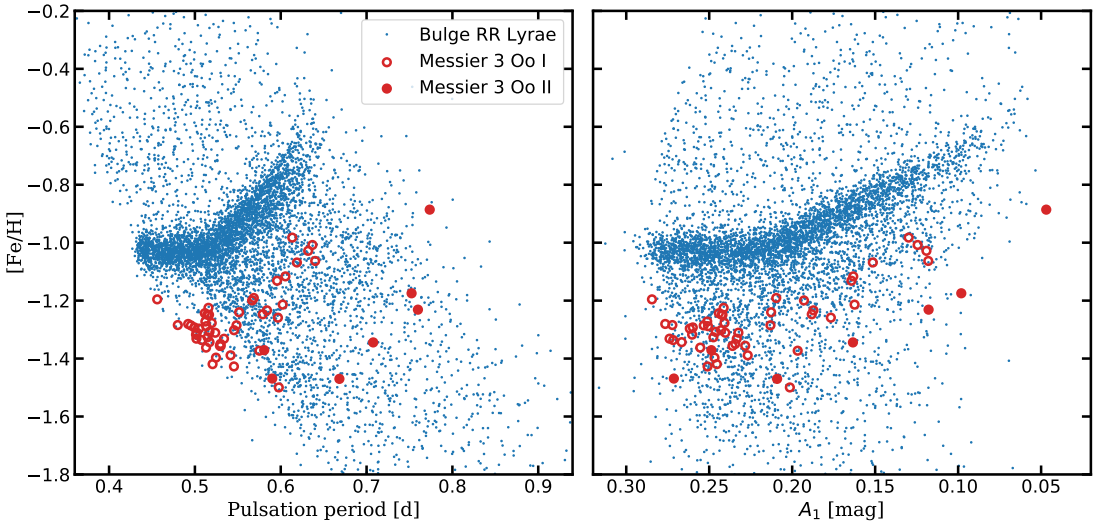

Figure 2.9: *Left:* Photometric iron abundances of 6193 RR Lyrae stars, calculated from the *I*-band light-curve parameters with Eq. 2 of Smolec (2005), as a function of the pulsation period. Oo I and II variables from M3 are overplotted as large empty and filled circles, respectively. *Right:* Same as the left panel, but as a function of the amplitude of the first harmonic of the Fourier series fit. Both the bulge field and the M3 RRLs display a systematic bias towards higher abundances, as the pulsation period increases and the pulsation amplitude decreases. Note that at a given period or amplitude, the metallicities derived for the Oo I and II RRLs in the monometallic globular cluster Messier 3 (calculated from the photometry of Jurcsik et al. 2017) at similar amplitudes are systematically offset.

hard to imagine a scenario where systematically higher-metallicity RRLs would only populate the lower-amplitude, longer-period part of the diagram, while the lower-metallicity variables occupy the higher-amplitude, shorter-period part. Therefore, it has to be concluded that the metal abundance formula described by Eq. 2 of Smolec (2005) suffers from a systematic bias as a function of the amplitude/period distribution of the bulge RRL variables. This finding is confirmed by a demonstrably monometallic dataset: on both panels of Fig. 2.9, empty and filled circles illustrate the *I*-band photometric metallicities of the Oo I and II variables in the globular cluster Messier 3, which is well known to be monometallic (Kraft et al. 1992; Cohen and Meléndez 2005), highlighting the systematic offset between these two groups of objects as well.

The second formula given by Smolec (2005) in the form of Eq. 3 gives much more

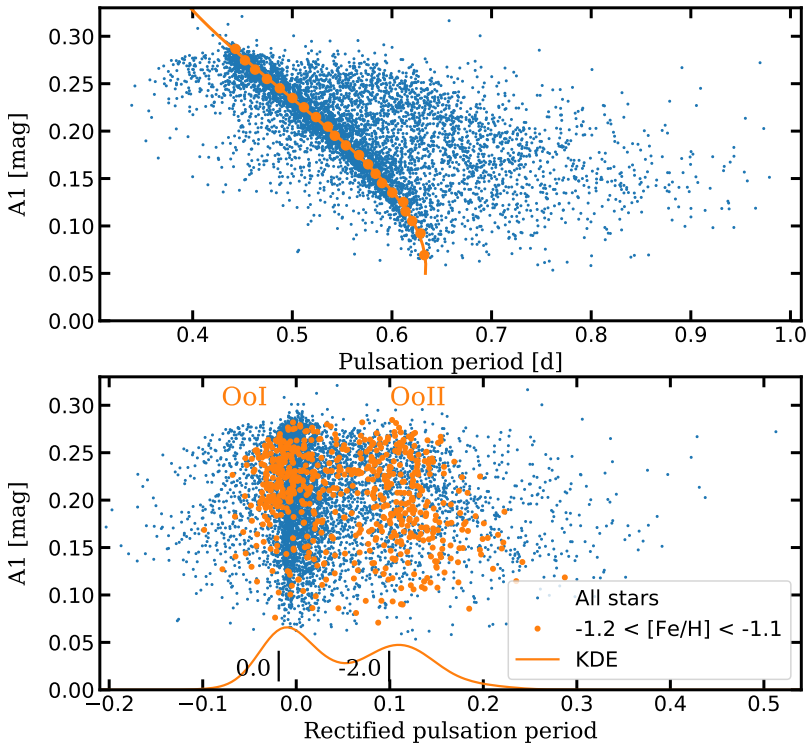

Figure 2.10: Iron abundance-dependent separation of Oosterhoff classes. *Top:* The distribution of the amplitudes of the first harmonic of the *I*-band Fourier fits of bulge RRLs as a function of the pulsation periods. The main ridge of Oo I stars is localized using the KDE maxima in different amplitude bins (circles). These are fit with a third order polynomial, illustrated by the continuous line. *Bottom:* Same as above, but as a function of the rectified pulsation periods, calculated as the difference between the real pulsation period and the position of the Oo I ridge of the top panel, as marked by the continuous line. The RRLs in the abundance bin $-1.2 < [\mathrm{Fe/H}] < -1.1$ are marked, and the continuous line at the bottom shows their KDE distribution. The local minima in the middle of the KDEs of different abundance bins change, reflecting the metallicity dependence of the Oosterhoff phenomenon. RRLs are classified as Oo I or II based on their position in this diagram, using a criterion that is a linear function of their calculated iron abundance. As examples, the separation limits for [Fe/H] = 0.0 and −2.0 are marked by the vertical black lines.

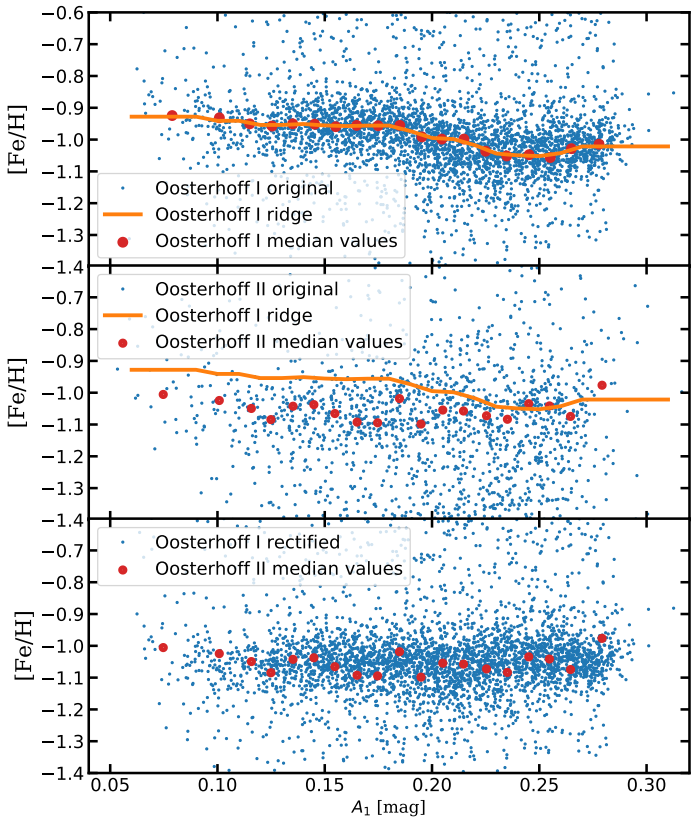

Figure 2.11: Photometric iron abundances (small blue points) calculated from the *I*-band light-curve parameters with Eq. 3 of Smolec (2005). *Top:* The photometric metallicity estimates of Oo I variables. The distribution still displays a dependency on the amplitude of the first Fourier harmonic, albeit to a much smaller degree than found using Eq. 2 (Figure 2.9; Smolec 2005). Red filled circles denote the average metallicities, as calculated from the maximum value of a KDE of data points in different bins amplitude. K-nearest regression (with K=2, marked as a continuous orange line) was utilized to generalize this effect, and this function was used to rectify the individual abundance estimates. *Middle:* Photometric metallicity estimates of Oo II variables. The KDE estimates (red filled circles) of the amplitude bins reveal no significant dependency on the estimated metallicities. The contiguous orange line from the top panel is drawn for comparison purposes. *Bottom:* The rectified metallicity estimates of the Oo I RRLs. The KDE estimates from the middle panel (filled red circles) are repeated for comparison purposes.

consistent abundance estimates as a function of the amplitude. However, in order to compare the behaviour of the calculated abundances of the two Oo groups of variables, they have to be separated. This was done using the period-amplitude diagram, utilizing a cut that is dependent on the calculated iron abundance, as illustrated by Figure 2.10. Comparing the top and middle panels of Figure 2.11 indicates that there is still a slight trend for the Oos I variables as a function of the amplitude, as well as an offset with respect to the Oo II stars, which have a calculated average [Fe/H] of −1.05 dex across the whole amplitude range. This offset has been corrected for, using the ridge line of Oo I RRLs, derived from kernel density estimates (KDEs) of the estimated abundances of stars in amplitude bins. The resulting rectified Oo I metallicity estimates are consistent with those of the Oo II variables, as well as across the whole range of RRL amplitudes (bottom panel of Figure 2.11).

A recent calibration of the RRL light curve – metallicity relation was performed by Nemec et al. (2013) by combining spectroscopic metallicities and *Kepler* photometry, and could serve as an alternative to the corrected metallicities calculated in this chapter. The relation given by Eq. 2 of Nemec et al. (2013) is:

$$\begin{aligned}[\mathrm{Fe/H}]_\mathrm{N} = &-8.65 && -40.12P && +5.96\phi_{31}^{Kp} && +6.27\phi_{31}^{Kp}P && -0.72\left(\phi_{31}^{Kp}\right)^2, \\ &\pm 4.64 && \pm 5.18 && \pm 1.72 && \pm 0.96 && \pm 0.17\end{aligned} \tag{2.8}$$

where ϕ_{31}^{Kp} is the epoch-independent phase difference in the *Kepler* photometric band. The relation between this quantitiy in the *Kepler* RRL observations and the V band has been determined by Jeon et al. (2014) as:

$$\phi_{31}^{Kp} = \phi_{31}^{V} - 0.174, \tag{2.9}$$

while Skowron et al. (2016) estimates the relation between the I_C and V bands in their Eq. 6 as:

$$\phi_{31}^{V} = 5.331 - 0.75\phi_{31}^{I_\mathrm{C}} + 0.122\left(\phi_{31}^{I_\mathrm{C}}\right)^2. \tag{2.10}$$

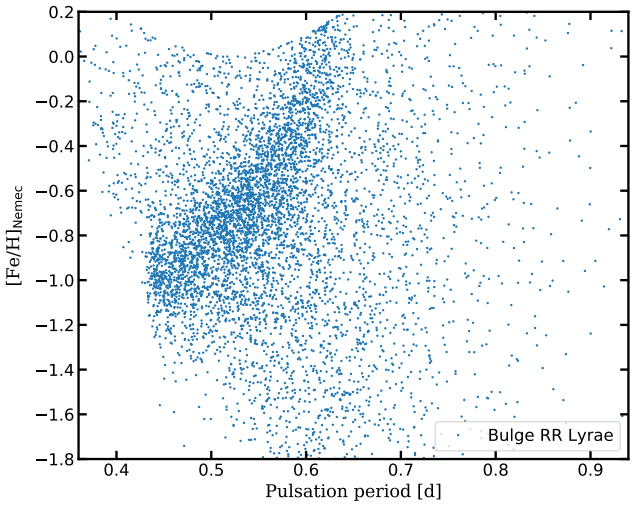

Figure 2.12: Photometric iron abundances calculated from the *I*-band light-curve parameters with Equations 2.8 – 2.10, following the RRL light curve shape – metallicity calibration of Nemec et al. (2013). Note that, according to their calibration, a star belonging to the main locus of bulge RRL stars can have a metallicity between −1 and 0, depending on the period of the pulsation.

The distribution of metallicities for the bulge RRL sample, when calculated using Equations 2.8 – 2.10, is shown in Figure 2.12. The main locus of bulge RRL stars shows even more severe period-dependent bias than that on Figure 2.9, when the Nemec et al. (2013) relation is used to calculate the photometric metallicities. It can be surmised that the main source of this metallicity bias lies in the limited sample (26 stars) and fairly complex model (5 parameters) employed by Nemec et al. (2013). Furthermore, the quoted large errors in the individual parameters in Equation 2.8 hint at strong degeneracies between the parameters, which can also be interpreted as a sign of severe data overfitting. Studies based on RRL metallicity distributions, for example that of Skowron et al. (2016) for the Magellanic Clouds, will have to be revised once less biased, carefully evaluated RRL light-curve shape – metallicity relations become available. Furthermore, quite alarmingly, this relation was used for the calculation of *Gaia* DR2 photometric metallicities of RRL stars (Clementini et al. 2019). Due to the biases discussed here, it is strongly recommended that any future studies avoid using metallicities calculated with this formula, and, as an extension, trusting those given by the *Gaia* DR2 catalog, as this bias is evident in the extremely wide ranges of metallicities calculated for Messier 3 and 62 variables (Figures 21

and 22 in Clementini et al. 2019, respectively).

2.4.3 Iron abundance estimation from the K_S-band light-curve shapes

In order to assess whether it is possible to determine the iron abundance of RRLs from the K_S band light-curve shapes, the K_S-band light-curve parameters of OGLE RRLs in the VVV fields determined in Sect. 2.4.1 were combined with the photometric metallicities estimated in Sect. 2.4.2. The data were subjected to the following quality cuts: all variables were discarded where either the first PC amplitude, U_1, or the PC amplitude ratios U_2/U_1, U_3/U_1 or U_4/U_1, exceeded the limit described in Sect. 2.3, as these indicate problematic photometry. Furthermore, the RRL sample on which Smolec (2005) calibrated the relation used to determine the initial [Fe/H] of the RRLs in question only extends down to [Fe/H] = -1.7. Therefore even the rectified abundance estimates are uncertain below this threshold. Nevertheless, because the typical abundance of Oo II globular clusters bearing RRLs is about [Fe/H]~ -2.2 dex, RRLs are only discarded below this limit. These cuts result in a final sample of 6193 RRLs.

The optical photometric metallicity as a function of the period and the first two PC amplitudes is shown in Figure 2.13. Some trends are readily apparent: as expected from the form of Eq. 3 of Smolec (2005) (Eq. 2.6), longer-period RRLs generally have higher iron abundances. However, even at the same period and amplitude U_1, stars of different U_2 have systematically different iron abundances.

Fourier parameters were also considered as additional parameters for this regression problem. During the light-curve fitting, the individual PCs are represented as Fourier series (Eq. 2.3), and the total light curve is given as the sum of these series multiplied by the PC amplitudes (Eq. 2.4). Therefore, traditional Fourier parameters, such as the Fourier amplitudes and epoch-independent phase differences (Simon and Lee 1981), can be calculated in a straightforward way. Consequently, in this analysis, the low-order Fourier parameters, such as the amplitude of the first three Fourier harmonics A_1, A_2, A_3, as well as the epoch-independent phase differences ϕ_{21} and ϕ_{31}, are also considered as additional possible parameters. Together with the pulsation period, the total number of possible independent variables is therefore 10.

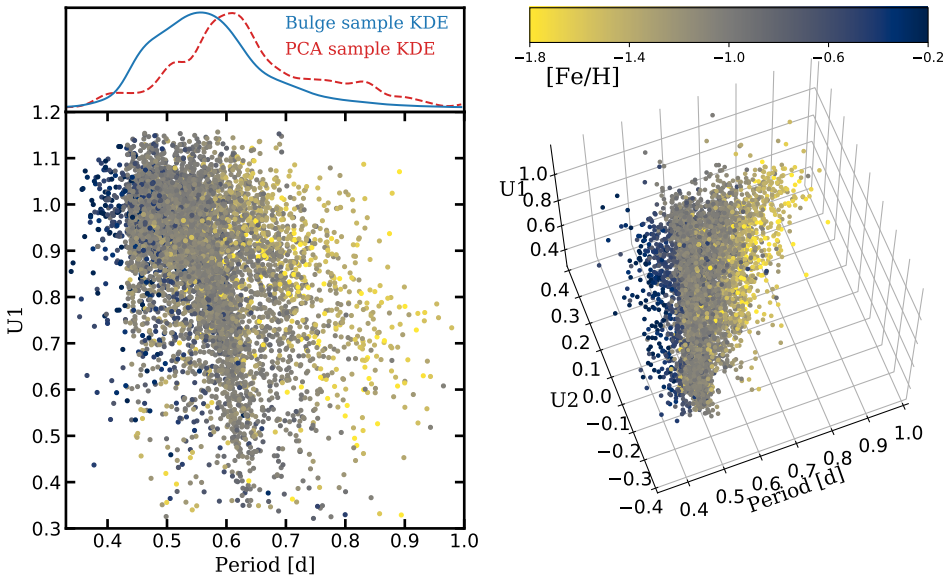

Figure 2.13: The K_S-band light-curve parameter – metallicity training sample. *Left:* The iron abundances show a significant overlap, when considering only the period and the first PC amplitudes. The upper panel illustrates the period distributions of the RRL sample used to to determine the K_S-band principal components (Section 2.2.1; dashed line), as well as that of the bulge RRL sample (Section 2.4.1; continuous line). *Right:* Considering U_2 as an additional parameter, stars with different I-band photometric [Fe/H] values clearly start separating.

In order to determine the iron abundance from different combinations of the considered parameters, the regression routines implemented in scikit-learn (Pedregosa et al. 2011) were explored. It has been found that regardless of the parameters, the underlying relation is nonlinear and multimodal due to the systematic offset between the Oo I and II variables. Additionally, the heavy excess of bulge RRLs with metallicities around [Fe/H] ~ -1 dex biases any kind of regression without resampling or weighting the input data.

The best results were achieved using the MLPRegressor routine, implementing a Multi-Layer Perceptron regressor, a type of neural network. Artificial neural networks are inspired by and modeled after biological neural networks, and have been used in many branches of science as well as in commercial applications for their ability to learn highly non-linear relations inherent in many types of data (see Haykin 2009 for a general description of neural networks).

Neural networks have many different parameters that have to be decided before the training of the network (hyperparameters). As these can vastly influence their performance,

Table 2.2: Parameters of the hyperparameter grid search

Hyperparameter	Candidate parameter value
Hidden layers[a]	(20), (50), (100) (20, 20), (20, 50), (20, 100) (50, 20), (50, 50), (50, 100) (100, 20), (100, 50), (100, 100)
α[b]	$10^{1.0}$, $10^{0.5}$, $10^{0.0}$, $10^{-0.5}$, $10^{-1.0}$ $10^{-1.5}$, $10^{-2.0}$, $10^{-2.5}$, $10^{-3.0}$ $10^{-3.5}$, $10^{-4.0}$, $10^{-4.5}$, $10^{-5.0}$
Activation function[c]	logistic, tanh, relu
Solver[c]	lbfgs, sgd, adam
Independent variables	P, $U_{1..4}$ P, $U_{1..3}$, ϕ_{31} P, $U_{1..3}$, ϕ_{21}, ϕ_{31} P, $A_{1..3}$, ϕ_{21}, ϕ_{31} P, $U_{1..4}$, $A_{1..3}$, ϕ_{21}, ϕ_{31}

[a] Number of neurons in the hidden layers.

[b] Regularization parameter.

[c] Activation function and solver of optimization, see the `scikit-learn` documentation for their description.

the MLPRegressor was trained with a grid of different hyperparameter values, and cross-validation is used to determine the best combination of hyperparameters. The grid of considered hyperparameter values is shown in Table 2.2. In this search, only neural network architectures with up to two hidden layers of neurons are considered: although adding more hidden layers increases the flexibility of the neural network, doing so does not significantly improve the performance in the tests. Besides the parameters of the network, the different combinations of the 10 independent variables are considered as an additional hyperparameter.

Due to the heavily biased nature of the bulge photometric metallicity distribution (most stars have abundances [Fe/H] ~ -1), the following special sampling method was implemented: the RRL variables are grouped by metallicity in ten 0.2 dex wide bins. The number of stars in each bin is shown in Table 2.3. As can be seen, there are only a few variables with [Fe/H] above 0 and below -2 dex, therefore, these were merged with neighboring bins. In each bin, 10 stars were selected randomly to be part of the validation sample. Of the

Table 2.3: Number of RR Lyrae variables in different metallicity bins

[Fe/H] bin	Number of RRLs
[−0.2, +0.2]	47
[−0.4, −0.2]	102
[−0.6, −0.4]	225
[−0.8, −0.6]	295
[−1.0, −0.8]	487
[−1.2, −1.0]	3888
[−1.4, −1.2]	628
[−1.6, −1.4]	375
[−1.8, −1.6]	84
[−2.2, −1.8]	62

remaining stars, 80 variables were selected randomly with replacement to be part of the training sample, resulting in a training set of 800 data points. As there are less than 90 stars in the most metal-poor and metal-rich bins, some variables at the extremes of the [Fe/H] distributions are selected multiple times due to the selection with replacement. The variables that were not selected for training were added to the cross-validation sample. In case of an uneven distribution of stars between the two extremes in a metallicity bin, the bin was split into sub-bins to guarantee a more even sampling as a function of metallicity.

This separation into training and cross-validation samples guarantees that: (1) the training set is large enough for the training of neural networks; (2) the training set is balanced on the total range of possible abundances; (3) by repeating this separation, variables in bins with few stars get selected for the cross-validation sample at least a few times.

This separation was repeated 40 times, resulting in 40 training and cross-validation samples. The neural networks were trained on all 40 training samples with all combinations of the possible hyperparameters detailed in Table 2.2. In all cases, the predictions of the trained neural networks were calculated for the corresponding cross-validation samples. Finally, for each star and for each hyperparameter combination, the predictions were averaged over the 40 repeats. Every star appeared at least six times in the cross-validation samples.

Due to the imbalanced data set, neural networks with the highest R^2 score (also called the coefficient of determination) calculated on the complete sample of averaged predictions is heavily biased towards solutions always predicting [Fe/H]~ -1, irrespective of the input values of the independent parameters. Therefore, the best hyperparameter combination was selected rather by using the sum of the R^2 scores calculated separately for each of the

abundance bins for the averaged cross-validation predictions. The cross-validation result with the highest score is illustrated by the top panel of Figure 2.14, with features P, A_1, A_2, A_3, ϕ_{21} and ϕ_{31}; hyperparameters $\alpha = 10^0$; two hidden layers with 100 and 20 neurons; activation function relu; and solver lbfgs.

After determining the optimal hyperparameters, 100 new training samples were drawn by randomly selecting 80 stars with replacement from each metallicity bin, but without withholding any stars for cross-validation. The MLPRegressor was trained on each of these samples, and the predicted metallicity is the average of these 100 trained regressors. These trained regressors are used by the algorithm PyMERLIN[7], together with the K_S-band light curve parameters, to determine K_S-band metallicity estimates of RRL stars.

2.4.4 Validation of the metallicity estimates

In order to provide an unbiased evaluation of the performance of a regression model, it is important to test the performance on a test data set not used during the fit for either training or validation. In this case, the model was used to predict the iron abundance from the shape parameters determined directly from the PCA analysis for the variables in Table 2.1. Then, these predictions can be compared to independent determinations of their abundances from the literature.

The cross-validation results in the top panel of Fig 2.14 indicate that the method described in Sect. 2.4.3 gives reasonable photometric abundance estimates with a scatter of ~ 0.22 dex in the $-1.7 <$ [Fe/H] < 0.0 abundance range. The estimated iron abundance is biased for stars below -1.7 towards higher values, probably caused by the combination of overestimation of the I-band photometric abundances, as well as contamination of the parameter space of low-abundance stars with those of higher estimated abundances, due to the higher relative photometric errors in low-amplitude, long-period stars (see Fig. 2.13).

The middle panel of Fig 2.14 compares the photometric abundances of 17 stars from Table 2.1 to the abundances listed in Table 1 of Jurcsik and Kovács (1996). Additionally, the photometric abundance of NSV 660 is compared to the [Fe/H] $= -1.31$ value determined from the SDSS spectra of the SEGUE Stellar Parameter Pipeline (Lee et al. 2011; Szabó et al. 2014). The two values for this sample of stars generally agree. It has to be noted, however, that there is a hint that the abundances of high metallicity RRLs are slightly underestimated.

[7] github.com/gerhajdu/pymerlin

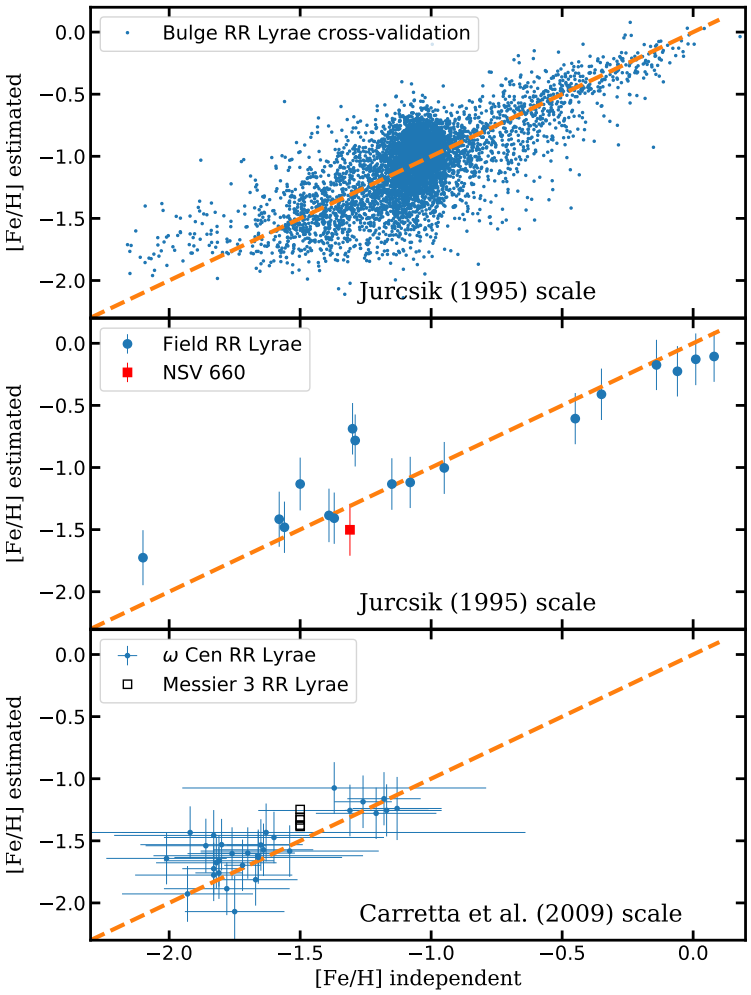

Figure 2.14: *Top:* Cross-validated average K_S-band abundance estimate for 6193 RRLs in the OGLE/VVV sample against the rectified I-band photometric abundance estimates (Section 2.4.2). The dashed orange curve represents equality between the compared abundance estimates. *Middle:* Comparison of estimated photometric abundances of field RRLs (circles) with their spectrophotometric (ΔS-based) abundances from Table 1 of Jurcsik and Kovács (1996) and the star NSV 660 (red square), the field variable with a spectroscopic SEGUE metallicity estimate (Lee et al. 2011; Szabó et al. 2014). Individual errors are estimated as the quadratic sum of the scatter of the 100 abundance estimates (Sect. 2.4.3) and a fixed uncertainty term of 0.2 dex. *Bottom:* Comparison of the estimated photometric abundances of stars in the clusters ω Cen (blue dots) and M3 (black open squares) with the individual spectroscopic measurements (Sollima et al. 2006) and the overall cluster iron abundance (Carretta et al. 2009), respectively.

Furthermore, the [Fe/H] of the most metal-poor star is overestimated, in concordance with the cross-validation results. The abundances of two stars, RR Cet and RR Leo, are overestimated by 0.5 and 0.6 dex, respectively. As there is a single source of photometry for both stars, it can be surmised that the original K_S-band photometry of both stars is affected by systematic trends on their rising branches, causing the estimated light-curve shapes to be deformed, resulting in the overestimation of their abundances.

The bottom panel of Fig 2.14 compares the K_S-band photometric abundances of the RRLs in ω Cen to their spectroscopic measurements by Sollima et al. (2006). The photometric abundances for five variables from the globular cluster M3 are also compared to the average cluster abundance, [Fe/H] = −1.5 dex, given by Carretta et al. (2009). The photometric abundances of ω Cen stars reproduce the bimodal distribution of iron abundances (Sollima et al. 2006). When taking into account the accuracy of individual measurements made by Sollima et al. (2006), there is no significant offset for the metal-rich group at −1.2 dex. However, the mean abundance of the metal-poor group seems to be overestimated by about 0.2 dex. As for the M3 RRLs, their average K_S-band photometric abundance is [Fe/H] = −1.24, 0.26 dex higher than the value given for the cluster by Carretta et al. (2009). Note, however, that the K_S-band photometric estimates are on the Jurcsik (1995) metallicity scale. The conversion between the two scales is estimated by comparing the common clusters with measured abundances between Table 1 of Jurcsik (1995) and Table A1 of Carretta et al. (2009) as shown in Figure 2.15 (the cluster NGC 5927 is neglected, because the difference in metallicity between the two sources is almost 0.8 dex). Linear regression results in the relation:

$$[\mathrm{Fe/H}]_{\mathrm{C09}} = 1.044[\mathrm{Fe/H}]_{\mathrm{J95}} - 0.037, \tag{2.11}$$

with a residual scatter of 0.1 dex. By converting the average estimated abundance to the Carretta et al. (2009) scale, the difference decreases to 0.17 dex. A similar conversion cannot be established between the metallicity scale of Sollima et al. (2006) and that of Jurcsik (1995), but if the former is on a similar scale to that of Carretta et al. (2009), then converting the photometric abundance estimates for ω Cen RRL variables decreases the tension for the lower-metallicity group to 0.1 dex.

During the calibration of the abundance prediction, variables showing the Blazhko effect were discarded, as their modulated light curves lead to systematic biases in their calculated

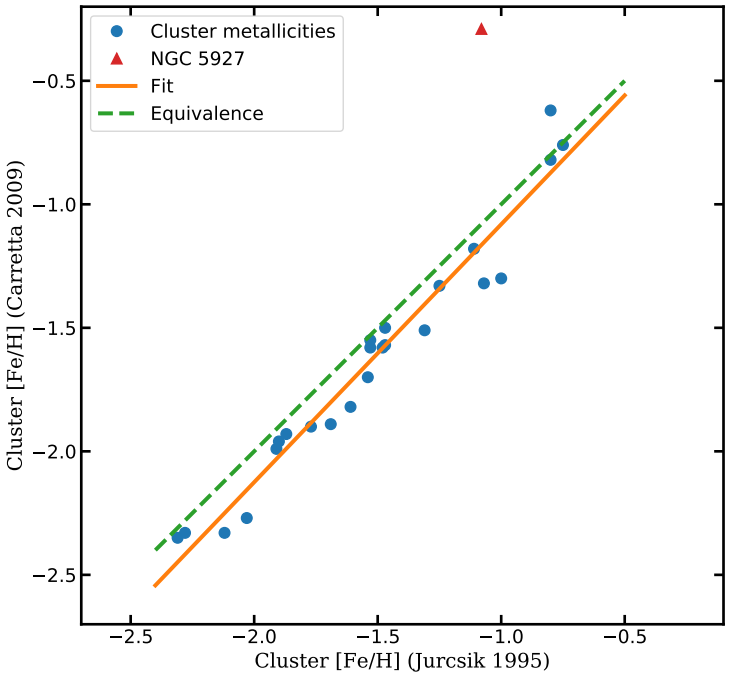

Figure 2.15: The relation established between the Jurcsik (1995) and Carretta et al. (2009) metallicity scales by fitting their commong globular cluster metallicities. Note that the measurement of NGC 5927 is not used, due to the 0.8 dex discrepancy in reported metallicities.

photometric abundances (see, e.g., Fig 4 of Jurcsik and Kovács 1996). Nevertheless, a significant fraction of the total population of RRLs suffers from this light-curve modulation. The nature of the Blazhko effect in the near-IR has not been established until very recently, due to the lack of well-populated light curves covering sufficiently long time spans. Jurcsik et al. (2018b) showed that the modulation is indeed present in the K_S-band light curves, but with diminished amplitudes. There are no (spectroscopic) abundance estimates for the 22 Blazhko RRLs studied in Jurcsik et al. (2018b), but it can be surmised that the majority of them must come from the bulge population of RRLs. Their photometric abundances were calculated based on their individual K_S-band mean light curves. These stars have a mean abundance of [Fe/H]= −1.11 dex with a standard deviation of 0.28 dex. These values are in agreement with those of the original bulge sample, where the mean and standard deviation are [Fe/H]= −1.06 dex and 0.26 dex, respectively. Therefore, it can be surmised that the iron abundances of RRLs should not show systematic biases when estimated with the parameters of the complete, long-term near-IR light curves, even when the Blazhko effect is present.

In summary, the K_S-band light-curve parameters allow the estimation of the iron abun-

dance of RRLs to an accuracy of $0.20-0.25$ dex, with slight hints of systematic trends (at the < 0.05 dex level) in the range of $-1.7 <$ [Fe/H] < 0 (underestimation for high abundances, overestimation for low ones), but abundances lower than this range are systematically overestimated. In order to establish better relations, either more high-quality K_S-band observations of a large sample of RRLs in globular clusters covering a wide metallicity range, and/or K_S-band observations of the hundreds of bright RRLs in the Solar neighborhood, and/or spectroscopic iron abundances for the OGLE+VVV field RRLs, should be obtained, with emphasis on the high and low abundance extremes.

3

RR Lyrae variables in the VVV disk fields

Prviously, as mentioned in Section 1.4, only a quarter of the VVV Galactic disk fields have been searched for RRL stars (Minniti et al. 2017a). The development of a machine-learning algorithm for the classification of RRL stars in the VVV K_S-band light curves Elorrieta et al. (2016), as well as the light-curve fitting and metallicity-estimation techniques developed in Chapter 2, allow the extension of this work, as well as the detailed characterization of the RRL population of the southern Galacic plane.

3.1 Searching for RR Lyrae stars in the VVV disk fields

3.1.1 The VVV disk data set

In contrast to the limited-area study of Minniti et al. (2017a), the complete area of the VVV Galacic disk fields was searched for RRL stars. This area measures approximately $4° \times 57°$ on the southern Galactic sky, covering the mid-plane towards the 4th Galactic quadrant, running parallel on both sides of the Galactic plane. The data were acquired by the VISTA survey telescope at Cerro Paranal between 2010 January 30 and 2015 August 14. Each of the 152 fields comprising the VVV disk area was observed in the *YZJH* photometric bands at least once, while ~ 50 K_S-band time series observations were also obtained for the same fields. As there is a degree of overlapping between neighboring tiles and pawprints, some sky positions can have up to 200 K_S-band measurements.

The analysis of the area was performed on public VVV data products provided by the CASU. Observational data were processed by the VISTA Data Flow System (Emerson et al. 2004; Irwin et al. 2004). The zero-point calibrated magnitudes measured in the five smallest apertures provided by the pipeline were used for the subsequent analysis. Typical limiting

magnitudes are between ∼ 17 to ∼ 18.5 mag in tK_S, and between ∼ 18 to ∼ 20 mag in the J band.

Before the variability search, the light curves of each of the objects on each of the tiles were constructed by cross-correlating individual detections found in the pawprint photometry prepared by CASU using the Starlink Tables Infrastructure Library (STIL, Taylor 2006). The initial source list for each tile was constructed using the 20 highest-quality observing epochs, where each of the multiple detections of the same source were eliminated by using a maximum cross-match radius of 0.36". This process resulted in lists of $6 \cdot 10^5 - 10^6$ sources, depending on the tile. Individual light curves were extracted by positional cross-matching of the source lists with the individual pawprint photometry catalogs of the tiles.

3.1.2 Variability and period search

The procedure described in Section 3.1.1 resulted in approximately $2 \cdot 10^8$ photometric light curves from the VVV disk area. In order to decrease the number of light curves, various photometric variability indices were computed, including Stetson's J index (Equation 1 of Stetson 1996), as well as the ratio of the weighted standard deviation of the time series σ_w and a modified von Neumann index δ (Neumann et al. 1941), defined as:

$$\sigma^2 = \frac{\Sigma_{i=1}^{N-1} (m_{i+1} w_{i+1} - m_i w_i)^2}{\frac{M-1}{M} \Sigma_{i=1}^{N} w_i}, \tag{3.1}$$

where N is the number of measurements, M is the number of non-zero weights, and $w_i = \sigma_i^{-2}$ are the inverse squared total sums of the photometric and zero-point errors.

The 0.1% significance levels were calculated for both indices, and all the sources above either were selected for the next step of analysis, resulting in about ∼ 10^4 variable candidates for each VVV tile.

The light curves were searched for significant periodic signals with the Generalized Lomb-Scargle periodogram (GLS; Zechmeister and Kürster 2009), where significances of the highest signals were estimated analytically, following Zechmeister and Kürster (2009). All stars with periodic signals between 0.35 and 0.9 days with false alarm probabilities (FAP) below 0.1% were selected for further analysis. This selection resulted in approximately 85 000 candidate objects. Each of the light curves in all five apertures was fit with a truncated Fourier series light-curve model using the period found by GLS. The optimal Fourier order and the rejection of outlying light-curve points with a 3σ outlier rejection was determined

following an iterative process (Kovács and Kupi 2007). The optimal aperture is assumed to be where the above process produces the smallest χ^2 of the resulting fits for the five considered apertures.

3.1.3 Classification of RRL candidates

In order to classify the RRL candidates found in Section 3.1.2, the machine-learned RRL classifier developed by Elorrieta et al. (2016) was utilized. This classifier uses the sum of individual decision trees, trained by using an adaptive boosting algorithm, to assign a score to each light curve based on a set of 12 light-curve parameters (see Elorrieta et al. 2016 for the definition of the features and the implementation details). The implementation of this classifier relied upon light curves of RRL in the VVV bulge tiles b293, b294 and b295, originally found by the OGLE survey (Soszyński et al. 2014). Elorrieta et al. (2016) determined that a cut in the score of 0.548 resulted in a precision[1] measure of $\sim$ 0.97, and a recall[2] of $\sim$ 0.90.

Due to the different survey strategies and fewer available light-curve points between the VVV bulge and disk fields, the above score cut should be revised in this application of the RRL classifier. In order to do so, first all objects with a score of 0.5 were selected as an initial sample, resulting in 3379 objects. The light curves of different subsets of these objects were individually and independetly inspected by human experts[3], assigning only integer (0,1) values to the RRL probabilities. By changing the decision boundary in the score, and using the labels assined by the human experts, it has been determined that a score cut at 0.6 results in a a precision of $\sim$ 0.9. This cut results in a second selection of 2147 RRab candidates. As a final step, candidate RRL which have been rejected by at least one human expert were also eliminated from this list, resulting in a final list of 1892 objects.

[1]Defined as the fraction of true positive cases among the total sample of selected objects.

[2]Defined as the number of the true positive cases among the sum of the true positives and false negatives. In other words, it measures the fraction of correctly classified true objects, in this case, RRL stars.

[3]The human experts were: Dr. István Dékány, Dr. Andrés Jordán and Gergely Hajdu.

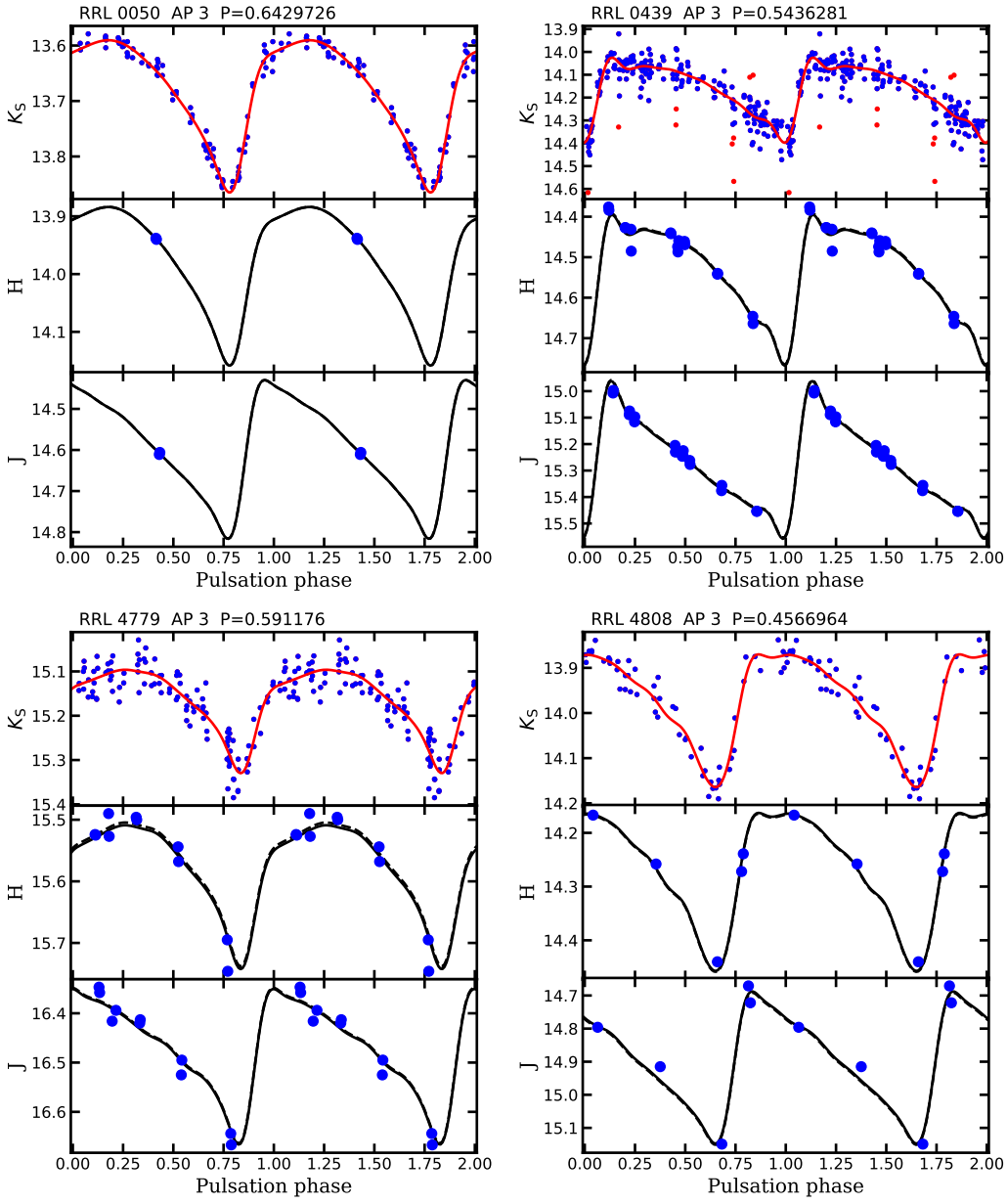

Figure 3.1: Examples of VVV disk RRL light curves fit with the the *PyFiNeR* algorithm. For each star, from top to bottom, the three panels show the estimation K_S, the H and the J-band magnitudes, as outlined in Section 2.3. The name, used aperture and fitted periods (in days) are marked above the top panel for each RRL.

3.1.4 Calculation of individual RRL properties

In order to determine the mean magnitudes and colors of individual RRL stars identified in Section 3.1.3, their light curves in each of the five available apertures were fit utilizing the PyFiNeR algorithm (Section 2.3). For each of the RRL variables, the optimal aperture was determined by selecting the aperture where the fit results in the the smallest value of the Huber cost (Equation 2.5) divided by the number of final (not clipped) light-curve points. Figure 3.1 illustrates the quality of the fitting process on the K_S, H and J-band light curves of four disk RRL stars identified in Section 3.1.3. It has to be noted that most of the RRL sample only have a handful of H and J-band photometry points available, hence the need for the approximation of the J-band light-curve shape (Section 2.2.4), in order to derive accurate J-band mean magnitudes.

Once the K_S-band light-curve properties were determined by the PyFiNeR algorithm, the pulsation period and the principal component amplitudes $U_{1...4}$ were utilized to calculate individual K_S-band photometric [Fe/H] estimates by the PyMERLIN algorithm (Section 2.4). These metallicity measurements allow the detailed characterization of the old RRL population of the Galactic disk.

The distribution of the principal component U_1 (a pseudo-amplitude, see Section 2.2) calculated by the PyFiNeR algorithm as a function of the pulsation period of the VVV disk RRL sample is shown in Figure 3.2. The color scale of the points corresponds to their estimated metal abundances determined by the PyMERLIN algorithm. The panel on the top compares the period distributions of the the bulge RRL variables (Section 2.4.3) with that of the 1892 disk RRL stars. The disk RRL stars have a wider period distribution when compared to the bulge RRL variables (see also Figure 2.13), as expected due to the presumably more inhomogeneous sample of stars in the former.

In order to use period-luminosity-metallicity relations for the determination of distances, the [Fe/H] values were converted to absolute heavy element content Z using the formula:

$$\log Z = [\mathrm{Fe/H}] - 1.538, \tag{3.2}$$

which was obtained by linear regression of a grid of [Fe/H] versus $\log Z$ values, computed for an assumed helium content $Y = 0.245$, alpha-element enhancement $[\alpha/\mathrm{Fe}] = 0.3$, standard solar chemical composition (Grevesse and Sauval 1998) and the chemical element distribution for $[\alpha/\mathrm{Fe}] \neq 0.0$ from Ferguson et al. (2005).

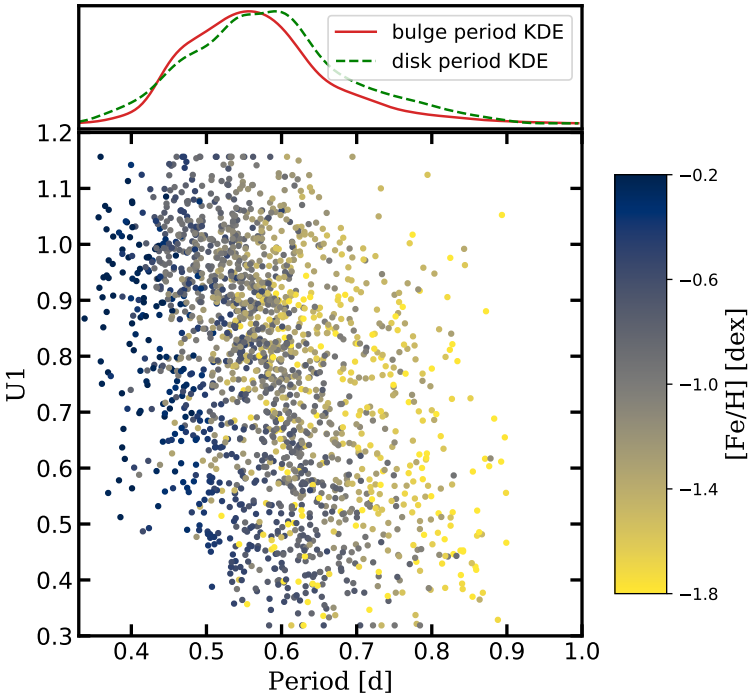

Figure 3.2: Distribution of the principal component U_1 as a function of the pulsation period (pseudo-Bailey diagram) of the VVV disk RRL sample, as calculated by the PyFiNeR algorithm (Section: 2.3). The color scale shows the K_S-band photometric metallicity calculated with the PyMERLIN algorithm (Section 2.4). The top panel compares the period distribution of the disk and bulge (the 6193 RRL variables from Section 2.4.3).

For the calculation of the absolute magnitudes and colors of individual RRL stars, the theoretical PLZ calculations of Catelan et al. (2004) were adopted. The magnitudes of every synthetic star were converted to the VISTA photometric system, utilizing the transformation formulae of Carpenter (2001) and those of CASU (González-Fernández et al. 2018). This process resulted in the following period-luminosity-metallicity relations:

$$M_J = -0.2361 - 1.830 \cdot \log P + 0.1886 \cdot \log Z, \tag{3.3}$$

$$M_H = -0.5539 - 2.302 \cdot \log P + 0.1781 \cdot \log Z, \tag{3.4}$$

$$M_{K_S} = -0.6365 - 2.347 \cdot \log P + 0.1886 \cdot \log Z. \tag{3.5}$$

Equations 3.3 – 3.5 were used to calculate the absolute magnitudes for each RRL in the J, H, K_S bands. The differences between the observed intensity-averaged colors $J - K_S$ and $H - K_S$, and the absolute magnitudes calculated as $M_J - M_{K_S}$ and $M_H - M_{K_S}$, were adopted

as the estimated color excesses $E(J - K_S)$ and $E(H - K_S)$, respecively.

The distances were calculated using the K_S-band distance moduli, corrected for the effect of extinction using the total-to-selective extinction ratios $A(K_S)/E(J - K_S) = 0.49$ or (for stars with no available J-band measurement) $A(K_S)/E(H - K_S) = 1.49$, as measured by Majaess et al. (2016).

It has to be noted that some studies concluded that the extinction law shows significant spatial variations in the optical-near-IR extinction ratios (Nataf et al. 2016). Other studies claim this is not a dominant effect (Schlafly et al. 2016). The former results are partly based upon VVV observations. In Chapter 4, the zero point calibration issues of the VISTA photometry will be explored, putting into question the conclusion of a spatially varying extinction law, as obtained by Nataf et al. (2016).

3.2 Properties of the RRL sample

3.2.1 Spatial distribution

The identified large sample of VVV disk RRL (Section 3.1.3), with their well determined distances and metallicities (Section 3.1.4), allows the characterization of this previously unexplored component of the Galactic disk. The distribution of RRL stars projected onto the Galactic plane in (X, Y) coordinates is shown by Figure 3.3. It is immediately apparent that the spatial distribution of RRL variables is heavily concentrated towards the Galactic center, with the concentration being consistent with recent estimates for the distance of the Galactic center of about 8.3 kpc (Gillessen et al. 2009). As can be seen by the color scale showing the KDE of the RRL positions, as well as contours enhancing its visibility, there is no evidence for any major substructure in their distribution.

In order to better understand the limitations and completeness of the disk RRL sample, the top of Figure 3.4 shows the distribution of the sample on the sky in Galactic l, b coordinates. The color scale and the contours once again correspond to the KDE of the RRL density. In contrast, the color scale of the bottom panel corresponds to the dust extinction map of Schlegel et al. (1998). As can be seen, the density variations are dominated by two effects: (1) the general density of RRL stars steadily decreases as the distance from the Galactic center in Galactic latitude increases; (2) on small scales the density of RRL variables is correlated with areas of lower extinction, meaning that extinction is the most significant

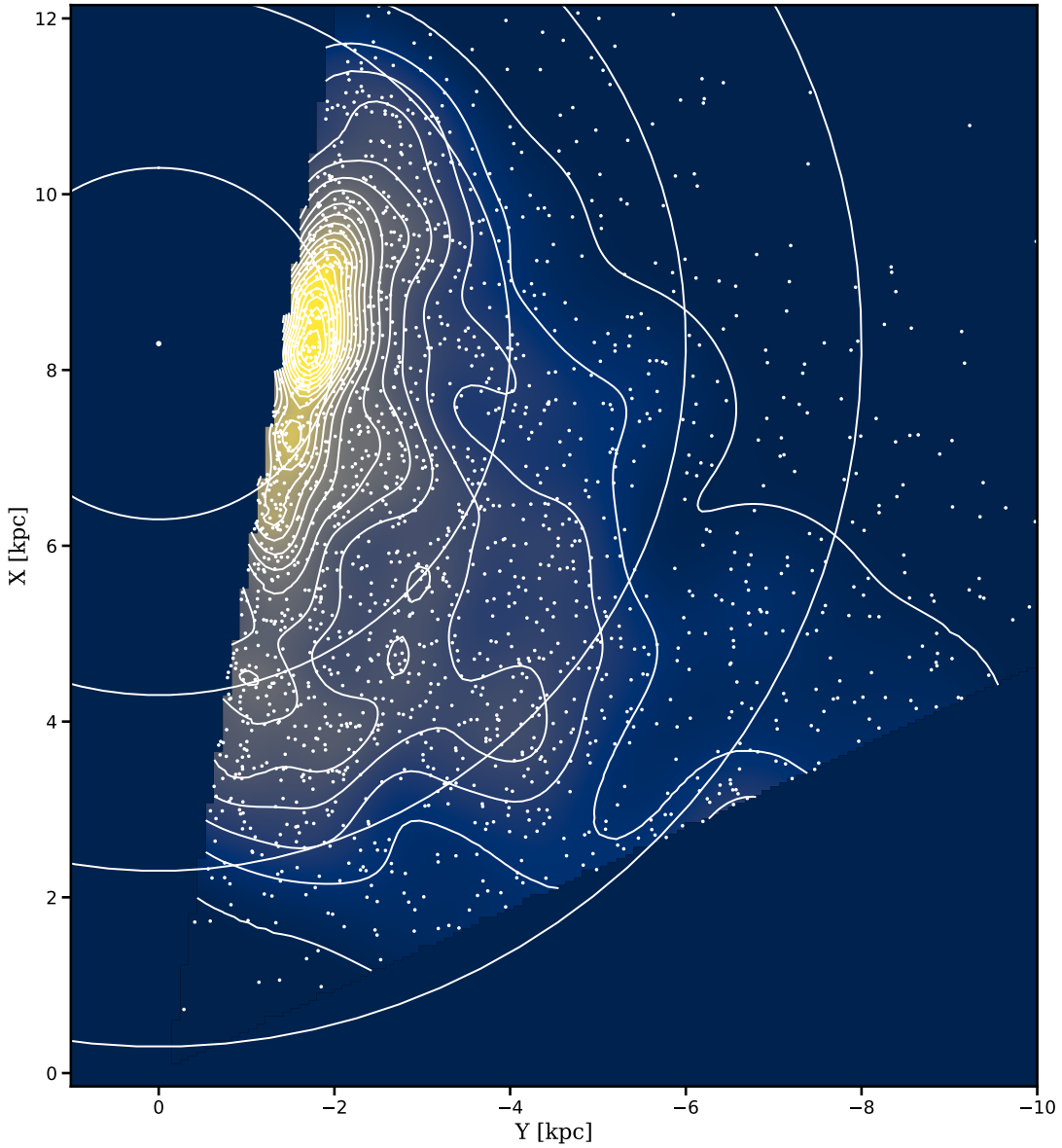

Figure 3.3: Distribution of VVV disk RRL stars projected onto the Galactic plane. Individual RRL variables are marked by white dots. The color scale and the (linearly scaling) contours represent a kernel density estimate of the RRL distribution with a Gaussian kernel size of 400 pc. Concentric circles mark Galactocentric cylindrical coordinates at 2, 4, 6 and 8 kpc radii.

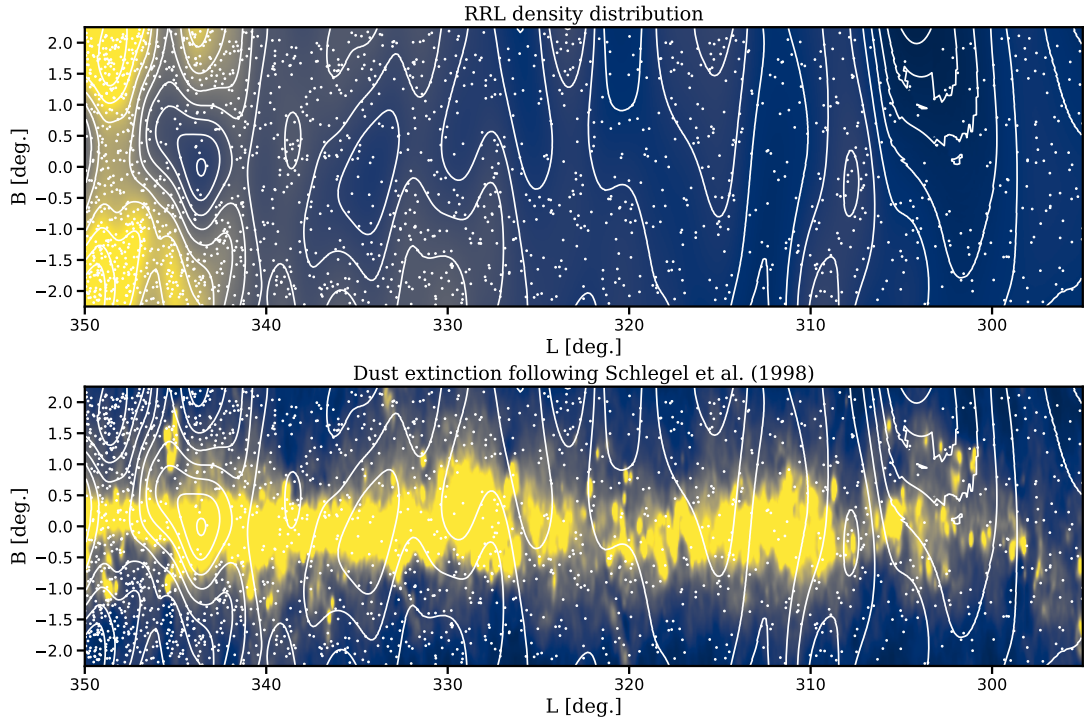

Figure 3.4: Distribution of VVV disk RRL stars in Galactic coordinates. RRL are marked by the white dots. *Top:* The color scale and the (logarithmically scaling) contours represent a kernel density estimate of the distribution with a Gaussian kernel of size 0.°75. *Bottom:* Same as the top panel, but the color scale represents the Schlegel et al. (1998) extinction map in the $1 < A(V) < 15$ mag range.

limiting factor for this sample of RRL stars.

The distribution of the calculated $E(J-K)$ color excesses and the observed K_S magnitudes of the RRL sample is shown in Figure 3.5. RRL stars can have $E(J-K_S)$ color excesses up to ~ 3 magnitudes. Beyond the Galactic bulge, Cepheid variables with $E(J-K_S)$ color excesses of ~ 5 have been found (Dékány et al. 2015b; Matsunaga et al. 2016). However, as Cepheids are intrinsically much brighter stars, they are easier to find than RRL stars at extreme extinction values, where the observation of RRL stars is beyond the capabilities of the VVV survey.

The Galactic center distance of the individual RRL stars were calculated assuming a Sun – Galactic center distance of 8.3 kpc (Gillessen et al. 2009). Using these, Figure 3.6 compares the marginalized heliocentric distances (top panel), as well as the distances (middle panel) to their Galactic plane (Z). The color scale corresponds to the observed $E(J-K_S)$ color excesses in both panels. The top panel reveals the full biasing effect of the extinction: while

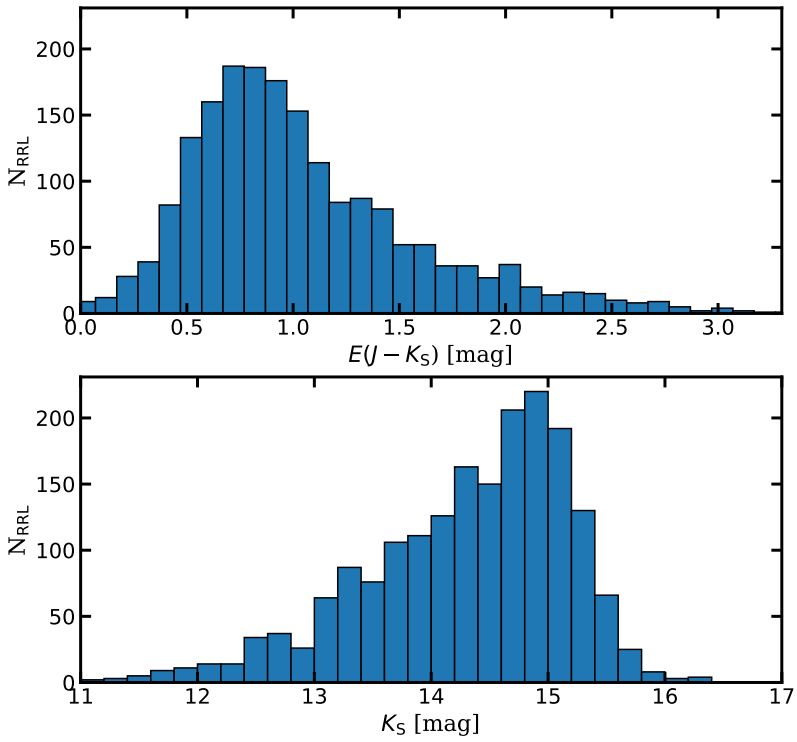

Figure 3.5: *Top:* $E(J-K_S)$ color-excess distribution of the VVV disk RRL sample. *Bottom:* Same as top, but showing the average K_S magnitude distribution.

at high Galactic latitudes RRL can be observed up to a distance of 16 kpc, in the Galactic planes RRL are not detected beyond a distance of $\sim 8-9$ kpc. RRL close to the Sun also go undetected, due to the saturation limit of the VISTA telescope.

The middle panel shows decreased number densities of sources towards the Galactic center at low latitudes as a result of the high extinction experienced in that region (see the bottom panel of Figure 3.4). Furthermore, there seems to be a decrease in the number density of stars which are far from both the Galactic plane and the Galactic center. However, this is caused by another selection effect: stars at high-Z are only present in the sample at large heliocentric distances due to the conical shape of the space covered by the VVV disk fields, where faint magnitude RRL stars can remain undetected.

The bottom panel of Figure 3.6 shows a modified version of the middle panel, where the color scale corresponds to the heliocentric distances of the RRL sample. The strong correlations seen between the distributions of RRL stars in the Galactocentric and heliocentric distances can be credited to the combination of the selection effects described above.

The limited longitude coverage of the VVV survey, combined with the survey's limiting

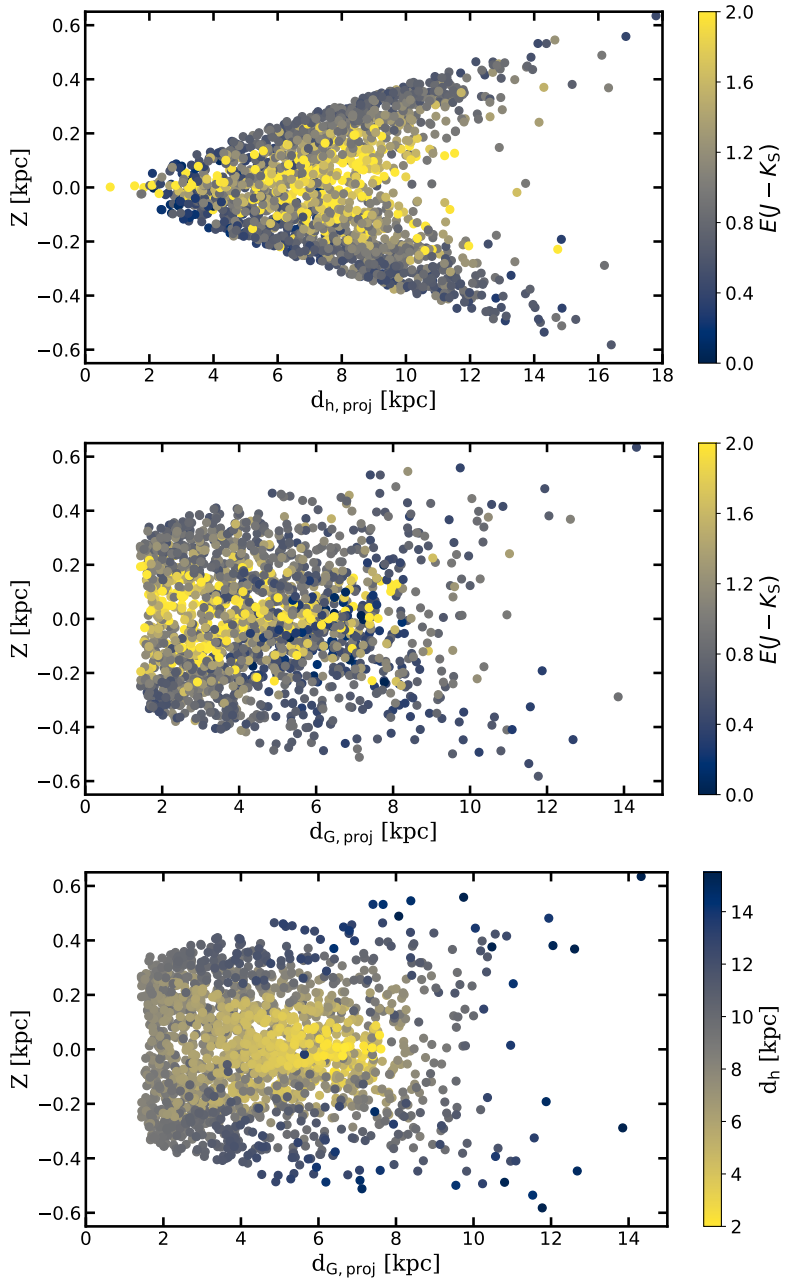

Figure 3.6: *Top:* Galactic plane distance distribution of VVV disk RRL stars as a function of heliocentric distances projected onto the Galactic plane. The color scale represents the $E(J-K_S)$ color excesses. *Middle:* Same as the top panel, but as the function of the projected Galactocentric distance. *Bottom:* Same as the middle panel, but the color scale represents the calculated heliocentric distances.

magnitudes and the selection effect brought about by extinction, results in a sharp drop in the number of RRL stars found beyond ~ 6 kpc from the Galactic center. Such stars only appear in the sample either at high Galactic latitudes, where they are affected by the faint magnitude limit of the survey, or at short heliocentric distances, where the bright magnitude limit and the relatively small sampled volume are the dominant factors in their lower numbers.

Due to the survey biases and the effect of the extinction, a complete analysis of the RRL density across the Galactic bulge is currently not possible. Nevertheless, the changes in their K_S-band photometric metallicities (Section 3.1.4) will be used in Section 3.2.3 to trace the change in the population properties of RRL stars as a function of both the cylindrical Galactic center distance, and the distance from the Galactic plane.

3.2.2 Metallicity distribution

Before assessing the metallicity distribution of RRL stars, only the RRL with the best quality light curves were considered. In order to remove the RRL with the most noisy and uncertain light curves, the RRL classifier score requirement was increased to 0.61. Additionally, the Huber loss function (Equation 2.5) gives information about the total scatter of the light curves during the fitting process. It has been found that most RRL light curves with fractional Huber losses (that is, the value of the Huber loss divided by the number of light-curve points) above a value of 0.001 present uncertain light-curve fits. Therefore, this limit was also adopted to remove uncertain [Fe/H] measurements. Finally, as the original training sample for the calibration of the K_S-band shape – metallicity relation lacks large quantities of very metal-poor or very metalrich RRL stars, and these stars are not expected to be overly abundant in the field, the requirement $-2.2 < \mathrm{[Fe/H]} < 0.2$ was also imposed. These additional quality requirements result in a final sample size of 1628 disk RRL stars.

The disk RRL sample was split in two subsamples based on the projected Galactocentric distances of the objects, with a threshold of $d_G = 4$ kpc, resulting in two subsamples of 785 and 843 objects for the $0\,\mathrm{kpc} < d_G < 4\,\mathrm{kpc}$ (disk-close) and the $4\,\mathrm{kpc} < d_G$ (disk-far) sample, respectively. Figure 3.7 shows the kernel density estimates of the metallicity distribution function of the RRL stars in the complete (1628 stars) disk area, as well as the KDEs of the two samples. As comparison, the MDF of the Galactic bulge is also shown. While both disk subsamples show similarities to each other, there is a relative excess of high metallicity stars in the disk-far sample. There are some apparent similarities between all RRL samples: all

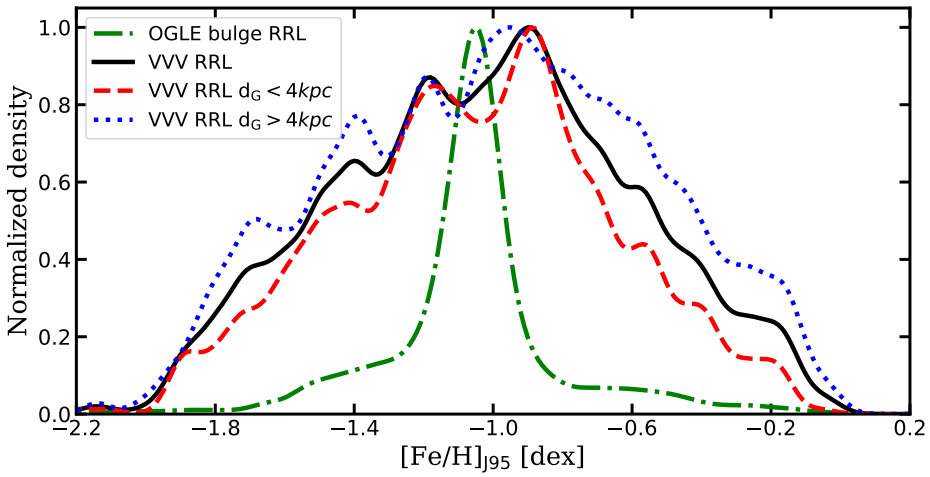

Figure 3.7: Kernel density estimates of the Galactic bulge and Galactic disk RRL samples and subsamples.

of them show multimodality, with the local modes appearing at similar metallicities. All of them show a strong local maximum around −1 dex; however, this component is not overly dominant in the Galactic disk, while in the Galactic bulge the vast majority of RRL stars belong to this group, identified as the dominant bulge RRL population.

In order to quantify the similarities and differences between the populations of RRL stars, Gaussian Mixture Modeling (GMM) was performed on the data subsets with the scikit-learn (Pedregosa et al. 2011) software package, to identify the dominant components of the distributions and their change between the samples. The necessary number of Gaussian components was determined by two approaches. Firstly, the classical information theory approach was applied by calculating the Akaike information criterion (AIC, Akaike 1974). The AIC measures information loss caused by model complexity, and can be used to prevent overfitting the data. The optimal number of Gaussian components k can be found by minimizing the AIC estimate. Furthermore, a K-Fold cross-validation (CV) approach with 20-folds was also applied to determine the number of optimal components.

For the complete disk samples, both the AIC and the CV approaches resulted in the optimal number of Gaussians being 6. For the disk-close and disk-far subsamples, both methods give estimates of 4–6 components. In the case of the CV approach, the found number of optimal components depends heavily on the exact K-fold samples chosen. As all disk sample MDFs are structurally similar, a Gaussian component number $k = 6$ is adopted

Table 3.1: Parameters of the GMM components of the RRL MDFs observed towards the Galactic bulge and the Galactic disk, on the metallicity scale of Jurcsik (1995).

RRL subset	Component	Mean [Fe/H]	σ	Weight
bulge	o1	-0.30 ± 0.05	0.148 ± 0.005	0.03 ± 0.01
	o2	-0.36 ± 0.05	0.120 ± 0.005	0.08 ± 0.01
	o3	-0.92 ± 0.03	0.078 ± 0.005	0.11 ± 0.03
	o4	-1.05 ± 0.03	0.049 ± 0.005	0.49 ± 0.01
	o5	-1.17 ± 0.03	0.074 ± 0.005	0.16 ± 0.03
	o6	-1.42 ± 0.03	0.111 ± 0.005	0.12 ± 0.02
	o7	-1.79 ± 0.05	0.194 ± 0.005	0.02 ± 0.01
disk all d_G	o1	-0.30 ± 0.05	0.130 ± 0.007	0.10 ± 0.01
	o2	-0.62 ± 0.06	0.115 ± 0.002	0.17 ± 0.01
	o3	-0.90 ± 0.07	0.101 ± 0.002	0.25 ± 0.03
	o4	-1.17 ± 0.07	0.100 ± 0.002	0.21 ± 0.01
	o5	-1.44 ± 0.07	0.103 ± 0.003	0.15 ± 0.03
	o6	-1.72 ± 0.06	0.141 ± 0.007	0.11 ± 0.02
disk $d_G < 4$ kpc	o1	-0.36 ± 0.10	0.129 ± 0.010	0.08 ± 0.03
	o2	-0.64 ± 0.09	0.107 ± 0.003	0.15 ± 0.04
	o3	-0.89 ± 0.09	0.089 ± 0.002	0.26 ± 0.03
	o4	-1.16 ± 0.09	0.099 ± 0.002	0.26 ± 0.03
	o5	-1.45 ± 0.10	0.109 ± 0.004	0.16 ± 0.04
	o6	-1.72 ± 0.12	0.147 ± 0.011	0.09 ± 0.04
disk $d_G > 4$ kpc	o1	-0.29 ± 0.08	0.132 ± 0.009	0.12 ± 0.04
	o2	-0.62 ± 0.09	0.120 ± 0.004	0.20 ± 0.04
	o3	-0.92 ± 0.09	0.108 ± 0.003	0.24 ± 0.03
	o4	-1.18 ± 0.10	0.093 ± 0.003	0.16 ± 0.03
	o5	-1.43 ± 0.10	0.092 ± 0.004	0.14 ± 0.03
	o6	-1.71 ± 0.11	0.140 ± 0.010	0.13 ± 0.04

for all of them in the following. For the bulge RRL sample, both approaches determined $k = 7$ to be the optimal number of components.

The results of the GMM analysis is documented in Table 3.1 for all four samples of RRL stars. Likewise, the separation of the RRL metallicity components by the GMM is shown in Figure 3.8. The top left panel shows that the dominant MDF component in the Galactic bulge sample is the one at [Fe/H] ~ -1 dex, corresponding to the dominant RRL component of the bulge, as discussed above. The disk-close (top right panel) and disk-far (bottom right panel) samples share the same metallicity components. However, the high-metallicity components ($-1 <$ [Fe/H]) are more pronounced in the case of the latter. Hence, this observation hints at the existence of a radial metallicity gradient in the oldest populations of the Galactic disk.

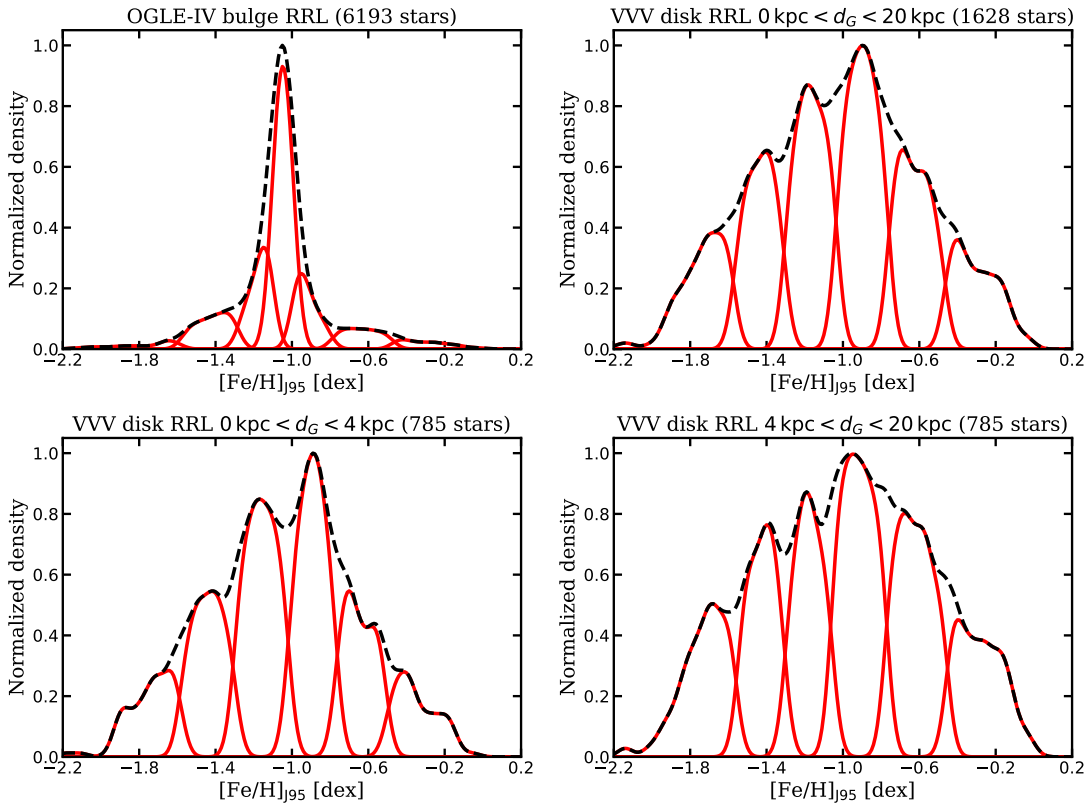

Figure 3.8: *Top left:* The original MDF of the bulge sample of RRL variables (black dashed line) and the KDEs of the seven identified metallicity components (red lines). Note that due to the small number of stars in the most metal-rich and metalpoor bins, they are hard to see. *Bottom left:* Same as the top left panel, but for the complete disk RRL sample. Note the smaller number of GMM components (6), when compared to the bulge sample. *Top right:* Same as in the top left, but for the disk-close sample of RRL stars. *Bottom right:* Same as in the top left, but for the disk-far sample of RRL variables. Note the increased fraction of stars in the two most metal-rich GMM components.

3.2.3 Spatial variations in metallicity

The distribution of the means and weights of the GMM components towards the Galactic bulge are shown in Figure 3.9. Likewise, the components identified in the complete VVV disk RRL sample, as well as the two Galactic center-distance based subsamples, are shown in Figure 3.10. It is remarkable that the means of the two disk subsamples agree with each other to such a degree, given that they have no common objects.

As can be seen by comparing the two disk subsamples shown in Figure 3.10, the higher-metallicity components v1 and v2 become more prominent at larger Galactocentric radii,

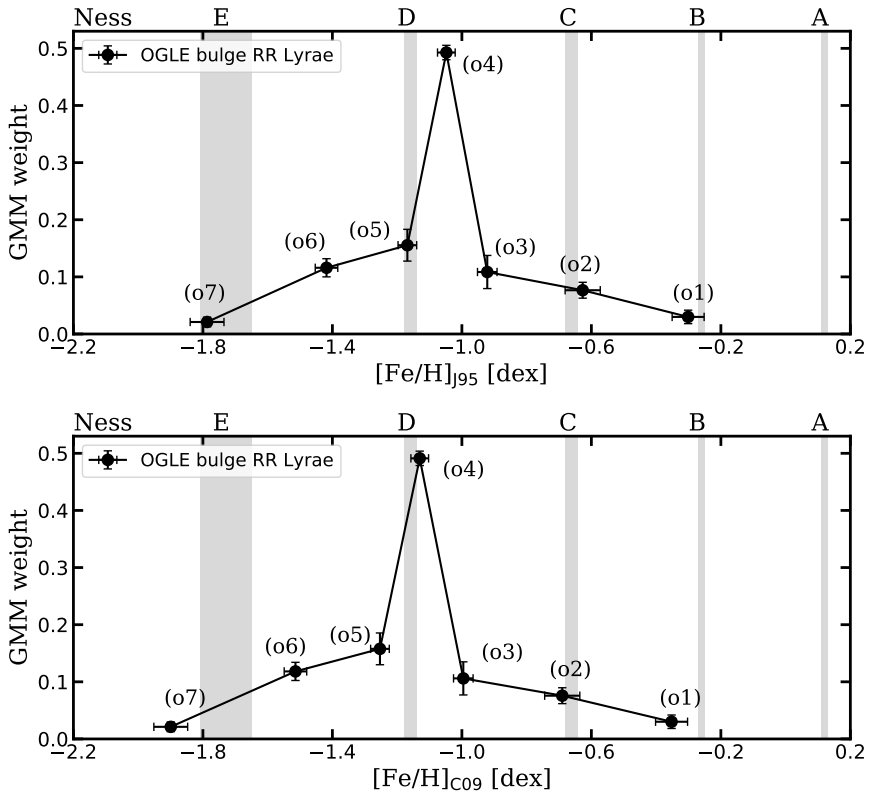

Figure 3.9: The weights and means of the GMM components of RRL MDFs for the 6193 OGLE+VVV RRL stars towards the Galactic bulge from Section 2.4.3. The top panel shows the results on the metallicity scale of Jurcsik (1995), while the bottom panel on that of Carretta et al. (2009). The vertical grey bars show the 1σ intervals of the GMM component maxima found by the ARGOS survey (Ness et al. 2013; their Table 3 for the bulge sample at $b = -5°$, $l = \pm 15°$ and $d_G \leq 3.5$ kpc).

with respect to the lower-metallicity components. This observation points towards a positive metallicity gradient in the RRL distribution of the Galactic disk.

It is interesting that although only a very small fraction of RRL stars contribute to the high- and low-metallicity wings of the MDF of the bulge fields, the means of the corresponding GMM components are very similar to those of the disk field samples. Namely, the positions of the disk components v1, v2, v5 and v6 match those of the bulge components o1, o2, o6 and o7, respectively. The I-band photometric metallicities towards the Galactic bulge sample have a higher precison (Section 2.4) than the K_S-band metallicities of the RRL in the Galactic disk. The AIC and CV analysis determined that one less GMM component is needed for the decomposition of the MDF of the Galactic disk sample than for that of the

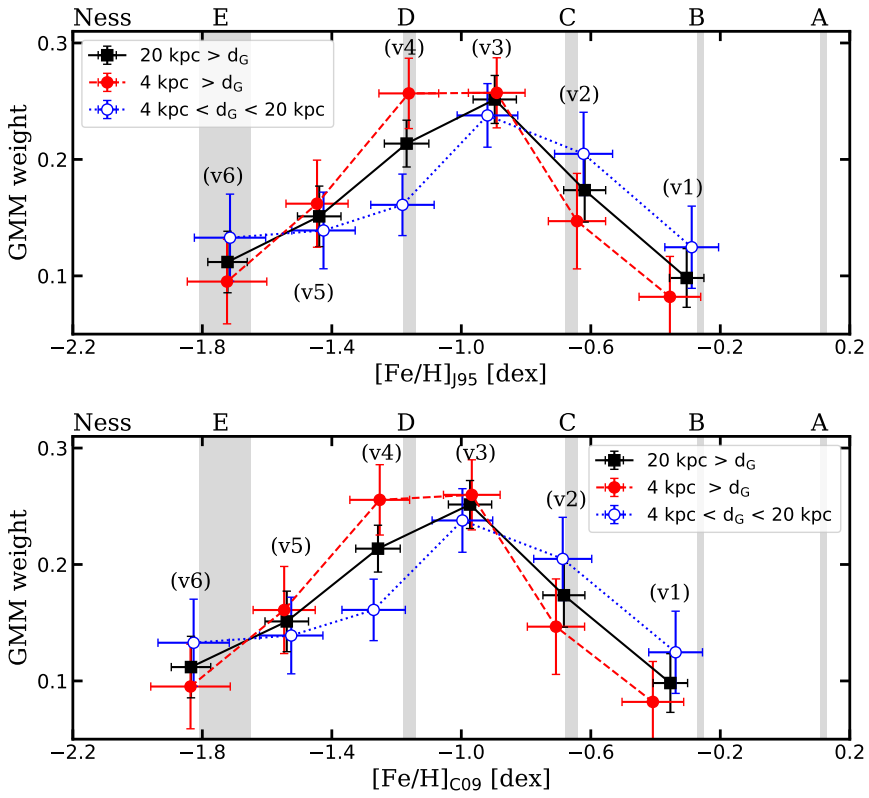

Figure 3.10: The weights and means of the GMM components of RRL MDFs for the 1628 VVV RRL disk stars. Continuous black lines and filled squares represent the total sample, while dashed red lines and filled circles denote the sample of stars closer than 4 kpc to the Galactic center. Dotted blue lines and empty circles show the sample of RRL stars farther than 4 kpc from the Galactic center. As in Figure 3.9, the top panel is on the Jurcsik (1995) scale, while the bottom panel is on the Carretta et al. (2009) metallicity scale, and the vertical grey bars present the GMM results of Ness et al. (2013).

Galactic bulge. Therefore, in the case of the bulge sample, around [Fe/H] ~ -1 the GMM separates the distribution into the three components o3, o4 and o5, while in the Galactic disk they are partially blended into the two GMM components v3 and v4.

Components v3 and v4 are also the most prominent components in all Galactic disk samples. It has to be noted, that in the case of the bulge sample, the question whether the components o3 and o5 are separate from o4 cannot be settled at the current precision of photometric metallicity estimates. The former are much less pronounced than the latter at [Fe/H] ~ -1, while displaying significant overlap with it at the same time. Therefore, more precise metallicity measurements will be needed in the future to study the MDF around its

maximum, both in the case of the Galactic bulge, as well as the Galactic disk RRL variables.

The dependence of the RRL metallicity distribution on the Galactocentric distance and the distance from the Galactic plane can also be investigated by splitting the RRL sample into subsamples by their MDF components. The relative densities of these components as a function of these two quantities are shown by their KDEs in Figure 3.11. While the relative distributions of the metal-poor components v3–v6 follow that of the total sample, the metal-rich components v1 and v2 have higher relative abundances towards larger Galactocentric distances, once again pointing towards a positive metallicity gradient. That is, while lower-metallicity RRL are preferentially found closer to the Galactic center, higher-metallicity disk RRL stars are found farther from it.

Additionally, the bottom panel of Figure 3.11 clearly shows that the high-metallicity RRL components are located preferentially closer to the Galactic disk than the low-metallicity components.

3.2.4 Comparison of the MDF to independent results

The decomposition of the MDF of the RRL variables in the Galactic bulge, as well as in the Galactic disk, provides an opportunity to relate the properties of the populations of these variables to those of other population tracers. The ARGOS survey (Ness et al. 2013) observed and determined the metallicities of red clump (RC) giants towards the inner galaxy with a relatively high spectral resolution of $R \simeq 11\,000$, with a precision of about ~ 0.09 dex.

The highly multimodal MDFs of RC stars were subject to a GMM decomposition, similar to that described in Section 3.2.2, making a direct comparison possible. The ARGOS sample is dominated by objects with [Fe/H] > -0.5, as higher metallicty, younger stellar populations produce higher-mass horizontal branch stars observable as the RC. Meanwhile, older, lower-metallicity populations can produce lower-mass horizontal branch stars, a fraction of which can be observed as RRL variables when they are within the instability strip.

Despite the differences between RRL variables and RC stars, the means of the GMM decompositions display a remarkable similarity. The grey bands in Figures 3.9 and 3.10 show the 1σ estimates of the maxima of RC MDF components, as determined by Ness et al. (2013), while also being labeled in order of decreasing metallicity with letters between A and E. By comparing the distribution of the component means, the following conclusions can be drawn:

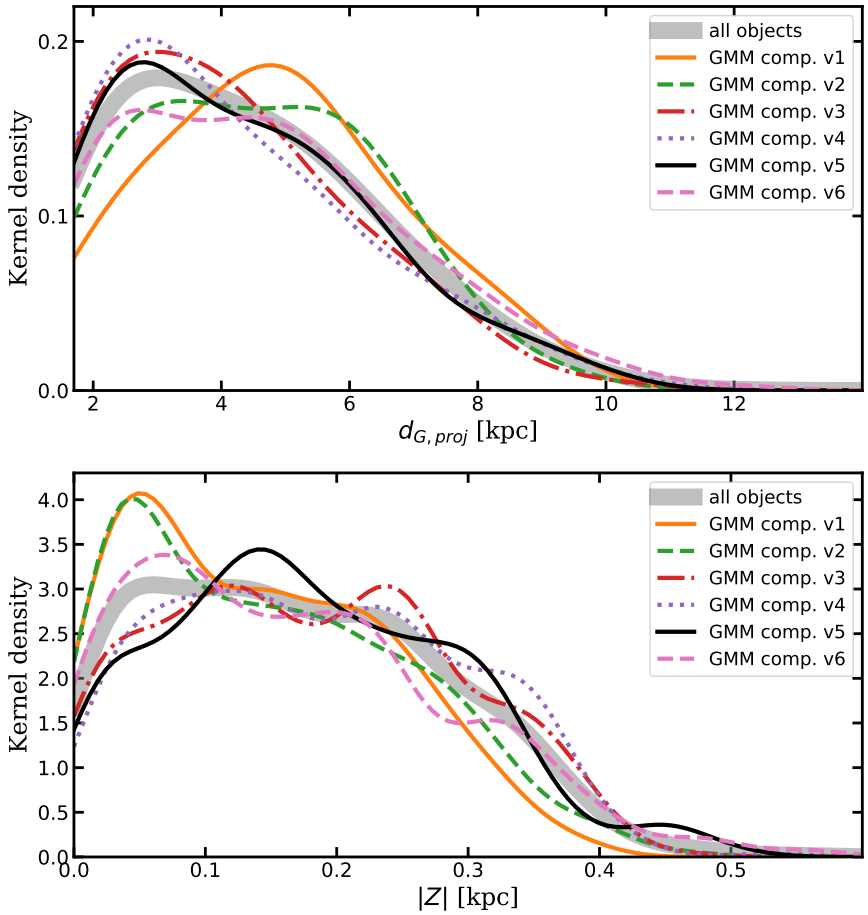

Figure 3.11: *Top:* Kernel density estimates of the disk GMM components as a function of the Galactocentric distance. *Bottom:* Same as the top panel, but as a function of the Galactic plane distance $|Z|$. Note the preference of components v1 and v2 towards larger Galactocentric distances, but smaller Galactic plane distances.

- The ARGOS metallicity component A is missing from both RRL samples.
- The high-metallicity ARGOS MDF components B and C match exactly the position of the RRL components v1 and v2, as well as o1 and o2, for the disk and bulge samples, respecively. The same can be said about the low-metallicity ARGOS MDF components D and E when compared to the v4 and v6, as well as the o5 and o7, components.
- The RRL MDF components v5 and v6 do not have equivalents in the ARGOS MDF decomposition.
- The main difference between the two samples can be found around the maxima of the MDFs of the RRL samples ([Fe/H] ~ -1). The disk component v3 (also corresponding

to the bulge components o3+o4) falls between the ARGOS components C and D.

It has to be emphasized that these results depend upon the comparison of the ARGOS MDF components to those identified in Section 3.2.2 on the metallicity scale of Jurcsik (1995). Unfortunately, the relationship between the ARGOS and other commonly used metallicity scales is not established, but it would not be unexpected if an offset of ~ 0.1 dex existed between the two scales. In order to account for this, the RRL bulge and disk MDF component means are also presented on the metallicity scale of Carretta et al. (2009) on the bottom panels of the Figures 3.9 and 3.10. Despite the progressively greater shift towards lower mean MDF means as the metallicity decreases (following Equation 2.11), the match between the ARGOS and RRL components is mostly preserved, when the statistical uncertainties are also taken into account. However, component o4 of the bulge RRL sample becomes equivalent to the ARGOS component D, instead of o5.

3.3 Discussion and future directions

The large sample of RRL variables identified in the Galactic disk field of the VVV survey, together with the methods developed in Chapter 2, have allowed a detailed, but arguably somewhat biased (see Section 3.2.1), characterization of this component of the Galactic disk. Comparison of the RRL metallicity distribution with the MDF results of the ARGOS survey (Ness et al. 2013) for RC stars (Section 3.2.4) revealed interesting similarities, but also differences between the two.

The ages of RC stars cover a wider range of possible vales, between $1 - 10$ Gyr, besides a wide range of possible metallicities (Girardi and Salaris 2001; Girardi 2016). In contrast, most RRL variables are expected to be older than 10 Gyr (see, e.g., Catelan and Smith 2015, and references therein). When interpreting the MDFs of both types of objects, care must be taken not to confuse these with the real metallicity distributions of the underlying stellar populations. For example, only horizontal branch stars which cross the Instability Strip will produce RRL variables. Layden (1995) estimated the difference between the RR Lyrae Production Function (RPEF) of Galactic environments (i.e., the Galactic disk and the halo) with different characteristic mean metallicities. Compared to the Galactic halo, the RPEF is reduced by a factor of ~ 25 for the metal-poor thick disk ([Fe/H] > -1); by a factor of ~ 40 in the metal-rich thick disk ([Fe/H] > -1); and by a factor of ~ 800 in the thin disk.

Besides the average metallicity of the underlying stellar population, the RPEF is also affected by all the factors that might affect the morphology of the HB, as observed in the globular clusters of the Milky Way and other nearby galaxies. It is unknown whether the disk RRL sample presented here is affected by the Oosterhoff dichotomy (Oosterhoff 1939) observed in Galactic globular clusters. This so-called second parameter problem might be related to the differences between the helium abundances and/or ages of the clusters (Catelan 2009; Gratton et al. 2010).

It can be assumed that similarly to the case of RRL stars, the ARGOS sample of RC stars also represents a biased version of the underlying stellar populations. Due to the incomplete understanding of the combined effect of these factors, it is currently not possible to recover the MDFs of the progenitor stellar populations of either the RRL or the RC stars. Nevertheless, the observed intriguing similarities and differences between the identified GMM components warrant future study. For example, a carefully selected sample of both RC and RRL stars could be observed spectroscopically to determine whether stars in GMM components at similar metallicities have the same alpha-element abundance ratios, which could point towards a common origin of these stars.

By performing cosmological hydrodynamical simulations of a forming disk galaxy, Brook et al. (2012) found that the older, lower-metallicity thick disk has shorter (radial) scale-length than the younger, higher-metallicity thin disk. In the Milky Way, this inside-out formation scenario has been confirmed spectroscopically by Bovy et al. (2012). RRL belonging to the MDF components v1 and v2 are found preferentially at larger Galactocentric distances, as well as lower Galactic plane distances, indicating longer scale lengths and shorter scale heights, in line with this hypothesis. The metallicity of MDF component v4 is consistent with the low-metallicity extreme of the thick-disk stars around the Solar neighborhood (Haywood et al. 2013). The absence of the RRL metallicity component v5 is explained by the very low number of RC stars at this metallicity in the ARGOS survey, while the MDF component v6 is probably mostly composed of halo interloper stars.

The current effort for the characterization of the RRL component of the Galactic disk is restricted by both the possible contamination of the RRL sample with halo interloper stars, as well as the relative inaccuracy of the photometric metallicity estimates. The former can be diminished significantly once proper motion estimates based on PSF photometry become widely available (see, e.g., Contreras Ramos et al. 2017 and Gran et al. 2019 for such

studies on small areas). In order to decrease the latter, ideally, all RRL variables should be observed spectroscopically. However, due to their faintness, this is probably not feasible in the near future. Alternatively, the K_S-band photometric metallicity could be calibrated more precisely by utilizing new light-curve and/or spectroscopic metallicity data, as suggested in Section 2.4.4. Furthermore, in the future, RRL identified in the extended Galactic disk regions of the eXtended VVV survey (VVVX, Minniti 2018) will greatly increase the number of RRL stars identified at larger Galactocentric radii.

Additional kinematical data could also be used to identify RRL variables associated with the "Gaia Sausage" (also called Enceladus), a massive dwarf galaxy accreted onto the Milky Way about 8 – 11 Gyr ago, whose stars were identified by the peculiar, sausage-like distribution of stars in the velocity space provided by the *Gaia* satellite (Belokurov et al. 2018; Helmi et al. 2018). This signature is also present in the RRL stars of the halo (Iorio and Belokurov 2019), with at least a fraction of the VVV disk RRL sample expected to belong to the "Gaia Sausage".

Clementini et al. (2019) classified a large number of variable sources found in the photometric time series of the *Gaia* satellite as RRL variables. As *Gaia* operates in the optical wavelength range, its observations of the Galactic disk RRL variables are constrained by the high amount of reddening found in the Galactic plane. Furthermore, the metallicity estimates provided by Clementini et al. (2019) are based on the calibration of Nemec et al. (2013), which has been shown to exhibit severe biases (see Section 2.4.2), which also necessitates the revision of the photometric metallicity estimates of all *Gaia* RRL variables. Nevertheless, it is expected that in the future, the analysis of the *Gaia* satellite data will provide invaluable contributions to our understanding of the Galactic disk population of RRL variables.

4

Recalibration of VISTA photometry

4.1 VVV photometric inconsistencies

During the study of the RR Lyrae variable population of the VVV disk fields, inspection of the light-curves and fits made by the *PyFiNeR* revealed inconsistencies between the H-band measurements. Figure 4.1 is showing two examples of this. On the left panel, the H-band measurements of the star come from two separate pawprints, while on the right panel the H-band data comes from two separate nights. In both cases, the two H-band measurements for each star are inconsistent with a common H-band average magnitude. J-band measurements seemingly did not exhibit this anomaly, and they were available for almost all stars concerned. Therefore, they were used to calculate color excesses, and hence extinction values towards most individual RR Lyrae stars. However, Cepheid variables in the VVV survey, presented in the next chapter, generally lie in regions of higher extinction, due to their concentration in the Galactic plane. Additionally, Cepheids in the direction of low extinction regions are frequently saturated in VISTA observations, further increasing their preference towards high extinction regions. This results in that many of them have no detections in the shorter J band, making the use of H-band observations for the calculation of extinction unavoidable. Hence, understanding the cause of the H-band inconsistencies and making the necessary corrections is essential for the study of the VVV Cepheid variables.

An additional source of uncertainty has been discovered by Jurcsik et al. (2018b) during their study of the Blazhko effect in the near-infrared. They note that their VVV K_S-band RR Lyrae light curves (selected to contain the highest amount of observations possible) showed zero-point offsets between photometry coming from different pawprints and/or tiles. These offsets have been manually corrected by Jurcsik et al. (2018b), but this additional source of scatter also warrants revision of the VISTA photometry.

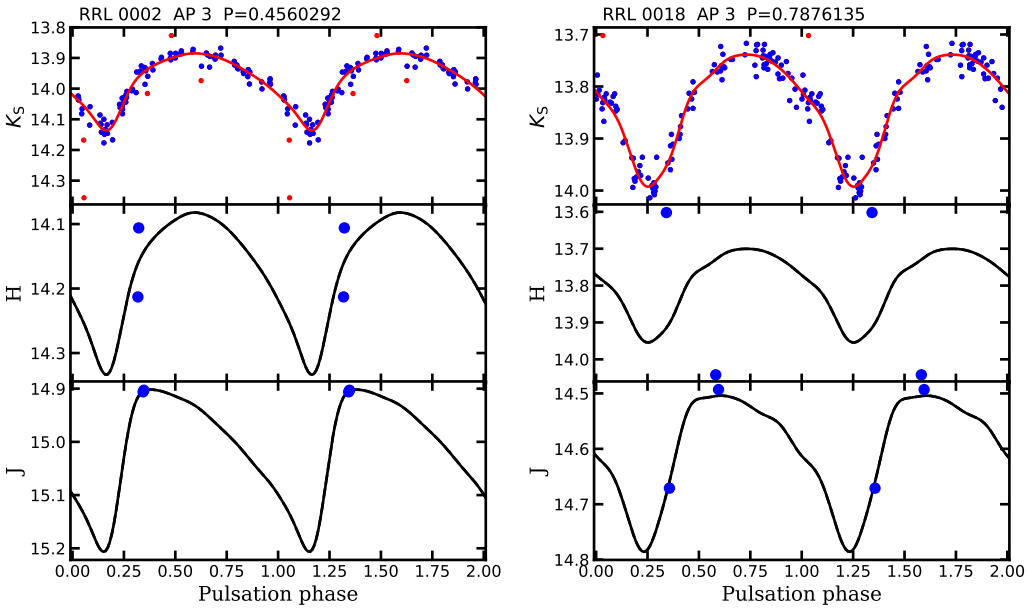

Figure 4.1: The graphics output of the *PyFiNeR* routine for two RR Lyrae stars in VVV tile *d*001 illustrates the inconsistencies of VISTA *H*-band photometric measurements. Note the inconsistent *H*-band measurements (middle panels) in both cases.

4.2 VISTA calibration procedures

All VISTA observations, including those observed during the course of VVV, are processed and calibrated both astrometrically and photometrically by the Cambridge Astronomical Survey Unit (CASU). The general procedures of calibration are based on the pipeline presented in Hodgkin et al. (2009), developed for the Wide Field Camera (WFCAM, Casali et al. 2007), located on the United Kingdom Infrared Telescope (UKIRT) on Mauna Kea, Hawai'i. There have been multiple versions of the photometry provided by CASU, differentiated by version numbers. The details of the calibration of the aperture for version 1.3 and 1.5 are detailed in González-Fernández et al. (2018). These two versions share the same basic photometric procedures, and only differ in the zero points.

The instrumental VISTA photometry, as measured on individual epochs by aperture photometry, are calibrated to the common VISTA photometric system by a zero-point shift, on a pawprint basis with respect to the 2MASS photometry (Skrutskie et al. 2006) of stellar sources converted to the VISTA system (either version 1.3 or 1.5). Changes in the individual detector sensitivities are accounted for by an additional offset, calculated on a monthly basis for every filter.

4.2.1 The 2MASS Point Source catalog in the VISTA system

In order to better investigate the previously discussed photometric anomalies, the 2MASS Point Souce Catalog (Skrutskie et al. 2006) magnitudes in the areas defined as:

$$\text{bulge: } -10.4\,\text{deg} < l < 11\,\text{deg} \text{ and } -10.5\,\text{deg} < b < 5.3\,\text{deg}$$
$$\text{disk: } -65.5\,\text{deg} < l < -9\,\text{deg} \text{ and } -2.4\,\text{deg} < b < 2.4\,\text{deg},$$

have been retrieved from the Infrared Processing & Analysis Center[1]. These two data sets completely encompass the VVV bulge and disk areas, containing approximately 13.7 and 10.6 million sources, respectively.

The transformation equations between the VISTA and 2MASS systems, as given by González-Fernández et al. (2018), are:

$$J_{V,1.3} = J_2 - 0.077 \times (J-H)_2 + 0.010 \times \mathrm{E}(B-V) \tag{4.1}$$
$$H_{V,1.3} = H_2 + 0.032 \times (J-H)_2 + 0.015 \times \mathrm{E}(B-V) \tag{4.2}$$
$$K_{S\,V,1.3} = K_{S\,2} + 0.010 \times (J-K_S)_2 + 0.005 \times \mathrm{E}(B-V), \tag{4.3}$$

for version 1.3 of the photometry. Meanwhile, version 1.5 uses the following equations:

$$J_{V,1.5} = J_2 - 0.031 \times (J-K_S)_2 + 0 \times \mathrm{E}(B-V) \tag{4.4}$$
$$H_{V,1.5} = H_2 + 0.015 \times (J-K_S)_2 + 0 \times \mathrm{E}(B-V) \tag{4.5}$$
$$K_{S\,V,1.5} = K_{S\,2} - 0.006 \times (J-K_S)_2 + 0.005 \times \mathrm{E}(B-V), \tag{4.6}$$

where the subscripts $_2$, $_{V,1.3}$ and $_{V,1.5}$ denote the 2MASS and the VISTA systems versions 1.3 and 1.5, respectively. The color excess $\mathrm{E}(B-V)$ towards each star is estimated by a multi-stage procedure outlined by González-Fernández et al. (2018). First, the Schlegel et al. (1998) maps of $\mathrm{E}(B-V)$, calculated from the infrared emission of dust, are interpolated to the coordinates of each 2MASS source. The $\mathrm{E}(B-V)$ values above 0.1 are modified as: $\mathrm{E}'(B-V) = 0.1 + 0.65 \times (\mathrm{E}(B-V) - 0.1)$, following the correction introduced by Bonifacio et al.

[1] www.ipac.caltech.edu

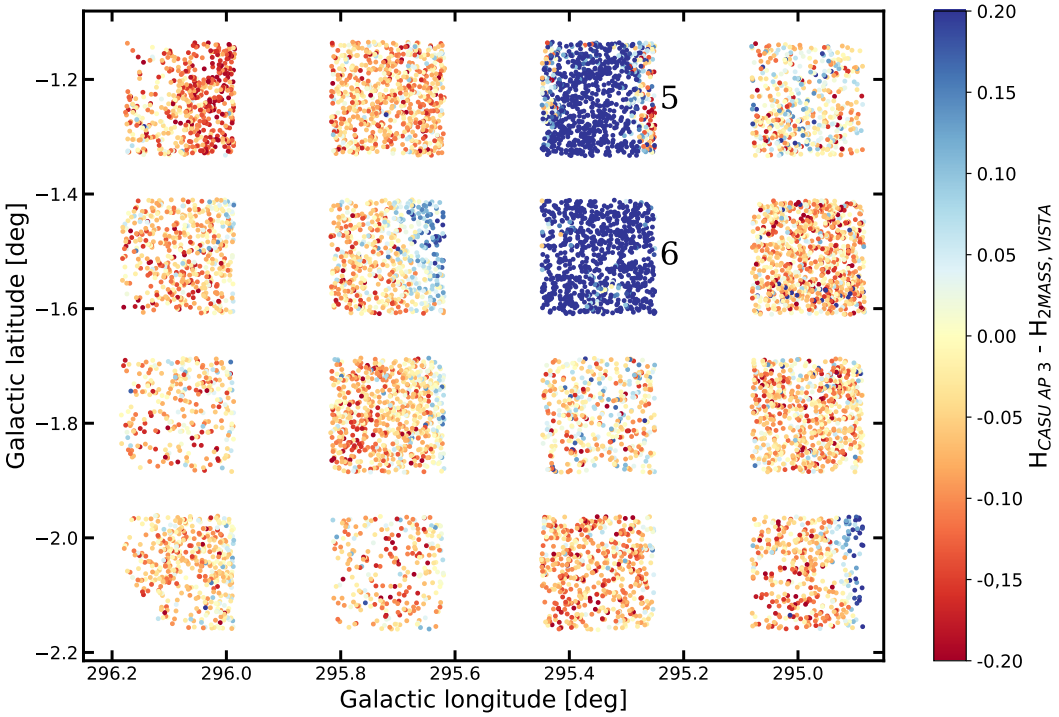

Figure 4.2: Photometric difference between VISTA H-band observations and converted 2MASS magnitudes for one pawprint. Chips 5 and 6 (marked on the figure by numbers) possess systematically fainter magnitudes than all other detectors. Furthermore, some intrachip differences are also evident when comparing the two sources of photometry.

(2000)[2]. Furthermore, as integrated extinction maps can overestimate the extinction towards individual stars, corrected color excess values were calculated for each star by following the procedure outlined in Sect. 4.4 of González-Fernández et al. (2018). These final E($B - V$) values were used to convert the magnitudes of the 2MASS sources using Equations 4.1–4.6 for both the bulge and disk area.

González-Fernández et al. (2018) note that there are small magnitude zero-point offsets between versions 1.3 and 1.5 of VISTA photometry, mainly caused by the change in the transformation Equations 4.1–4.6. The 2MASS photometry, when converted to the VISTA system versions 1.3 and 1.5, show differences consistent with the ones given by González-Fernández et al. (2018) in their Equations C9–C11, validating the conversion performed in

[2]Note that, for version 1.5 of the VISTA photometric system, González-Fernández et al. (2018) use unmodified Schlegel et al. (1998) E($B - V$) values. The only band affected marginally by this correction step is K_S in version 1.5, where the offset in the converted magnitudes is only about 0.001 magnitudes.

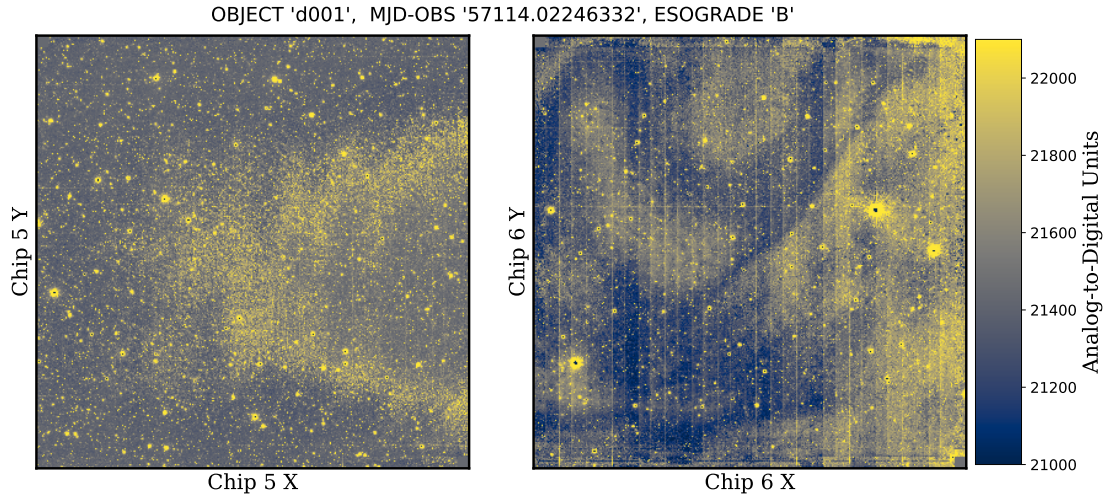

Figure 4.3: The *H*-band (calibrated) images of chips 5 and 6 (marked by numbers), corresponding to the photometric difference map presented in Fig. 4.2. Both chips suffer from dark regions with diminished point-source ADU counts. The high background counts indicate suboptimal observing conditions.

this section[3].

4.3 Revision of *H*-band VISTA observations

The converted 2MASS photometry, described in the previous section, allows an independent check of the photometry delivered by the CASU VISTA pipeline. Figure 4.2 shows the *H*-band magnitude difference between the converted 2MASS magnitudes and VISTA aperture photometry (of aperture 3), as provided by CASU, of individual detected point sources in one particular pawprint of tile *d001*. As can be seen, the magnitudes of stars on chips 5 and 6 are heavily underestimated, while those of other chips end up being slightly overestimated. Some of the other detectors also show small-scale variations in the magnitude difference distribution of stars. The original, reduced observations of chips 5 and 6 are shown in Figure 4.3. Both detectors exhibit very apparent artifacts, probably caused by the non-linearity of these two chips, when the sky background is high. The straight line artifacts on chip 6 are

[3]González-Fernández et al. (2018) calculate these differences on calibrated VISTA photometry over different sky regions, while here this was done for the converted 2MASS magnitudes only in the VVV sky footprint. Hence, these differences are not expected to be exactly the same, owing to the different selection functions of the samples, as well as the extinction terms in Equations 4.1–4.6.

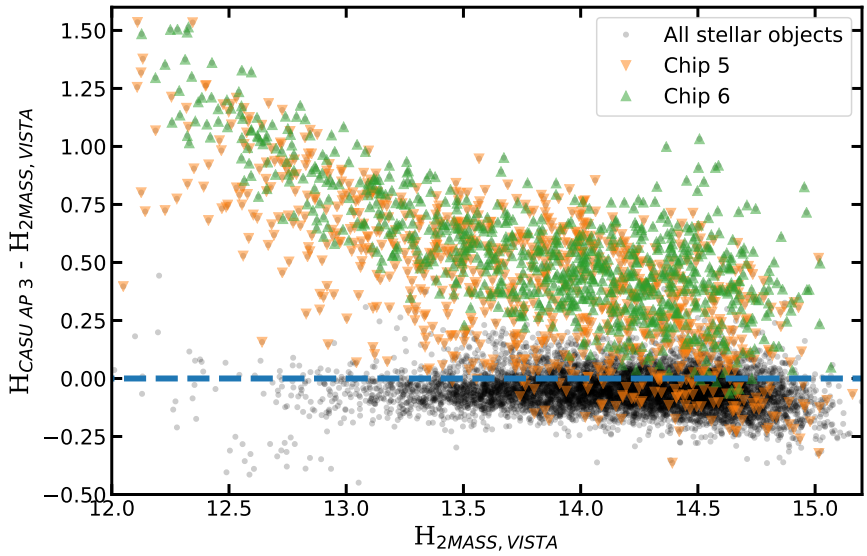

Figure 4.4: *H*-band magnitude differences between VISTA and converted 2MASS observations for the same pawprint as presented in Figs. 4.2 and 4.3. The non-linear behaviour of the VIRCAM chips 5 and 6 results in lower source counts and biased photometry for these detectors. Furthermore, while the photometry from the other detectors is mostly internally consistent, the per pawprint calibration scheme employed by CASU results in a systematic overestimation of the magnitudes of the sources present on them.

most probably caused by the detector manufacturing process. The distribution of magnitude differences for the same *d*001 pawprint are shown on Figure 4.4. In accordance with what has been shown by Fig. 4.2, the magnitudes in chips 5 and 6 are underestimated, while on all other chips, they are generally overestimated. Magnitudes measured in chips 5 and 6 also show a large dispersion and non-linearity, when compared to the converted 2MASS magnitudes. In contrast to this, the observations on the other chips seem to have small dispersions and no apparent non-linearity. Therefore, observations taken by these chips in the *H*-band should be generally reliable after a zero-point correction. Meanwhile, *H*-band observations taken by chips 5 and 6 under suboptimal conditions should be treated as suspect. It has to be noted that similar offsets, apparently caused by unfavorable observing conditions, were not found in the case of the *J* and K_S photometry, as the sky background in those bands is generally lower.

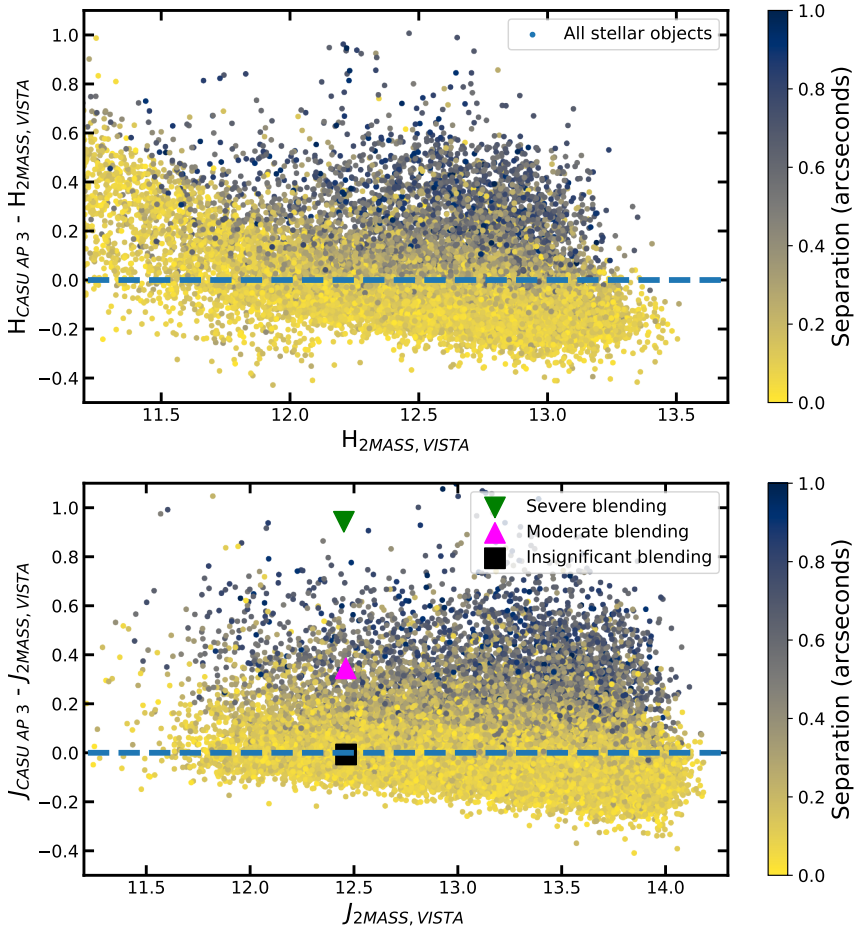

Figure 4.5: *Top: H*-band photometric differences in tile *b307* between the converted 2MASS measurements and the aperture photometry of stellar sources provided by CASU. Note the systematic offset between the majority of the sources and the fiducial line. The color scale denotes the separation between the coordinates of the sources in the 2MASS and VISTA catalogs. *Bottom:* The same as for the top, but for the *J* band. The 2MASS and VISTA images of the three individually marked stars are shown in Figure 4.6.

4.4 VISTA zero point bias in dense stellar fields

The converted 2MASS photometry (Sect. 4.2.1) allows a general revision of the VISTA photometry, not only in the *H*-band. Besides the Galactic center tile *b333*, the densest tile is *b307*, containing most of Baade's Window, resulting in very high stellar densities.

The top and bottom panels of Figure 4.5 compare the magnitudes of stellar sources given by CASU and the converted 2MASS photometry in this tile for one pawprint for the *H* and *J* bands, respectively. The bulk of the magnitudes measured by the VISTA pipeline is

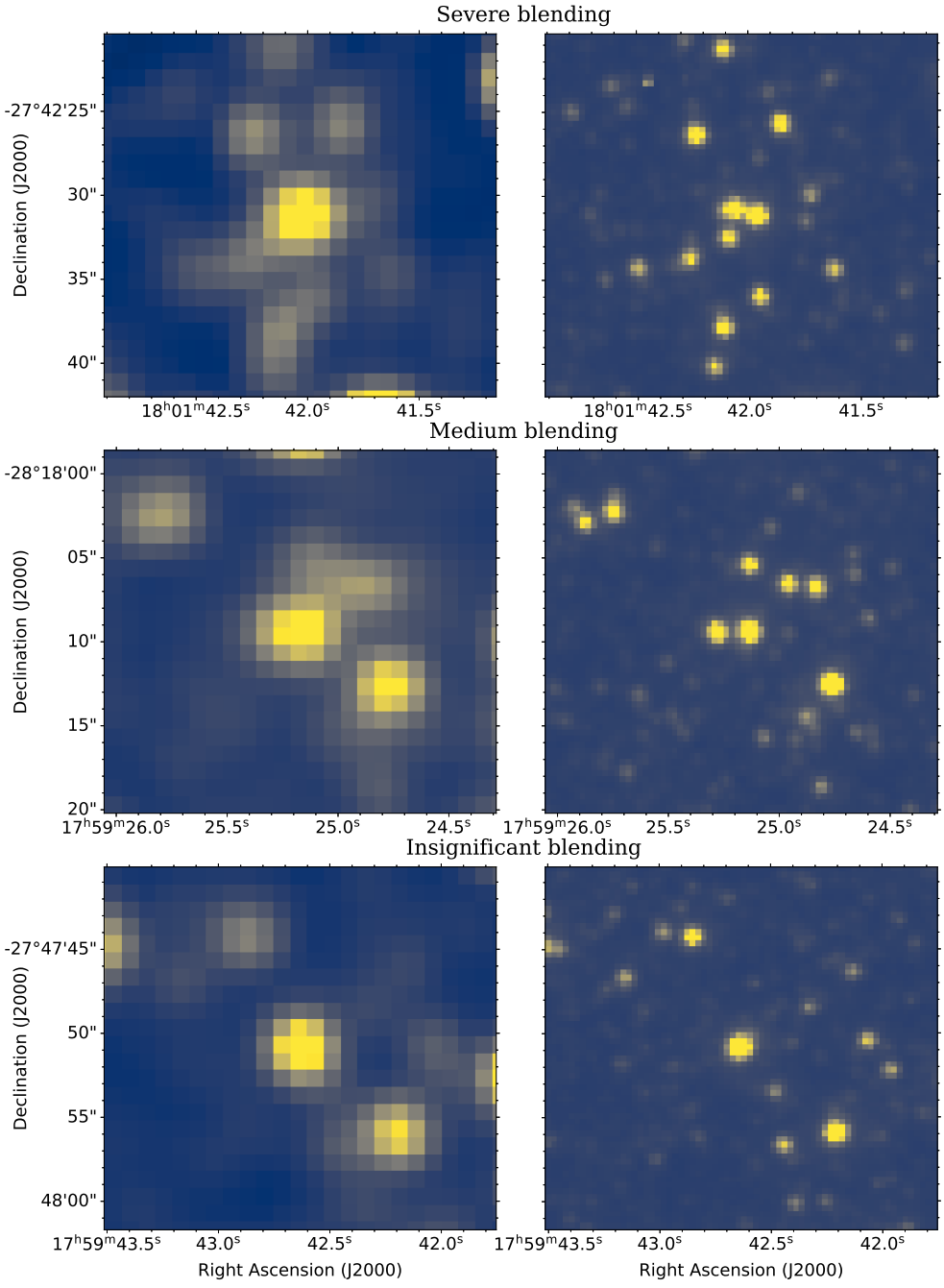

Figure 4.6: Examples for severe, medium, and insignificant cases of blending in 2MASS *J*-band observations, shown by the top, middle, and bottom panels, respectively. The stars correspond to the objects marked on the bottom panel of Figure 4.5. *Left: J*-band images of the 2MASS observations. *Right: J*-band VISTA images.

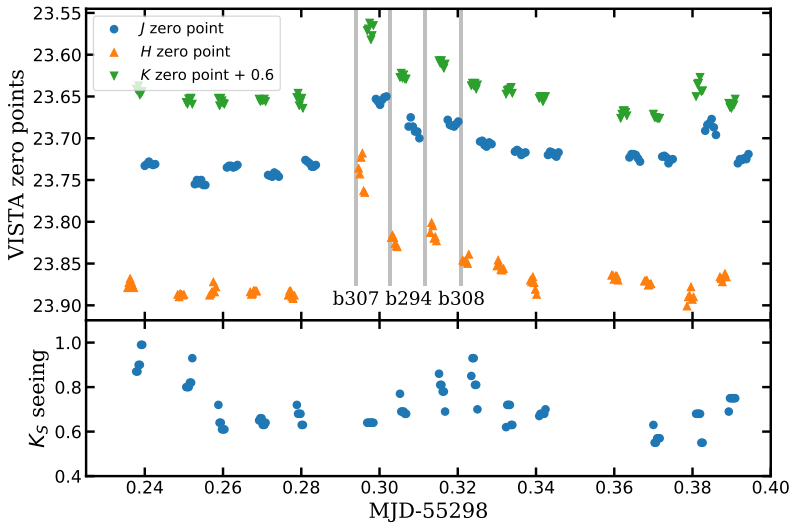

Figure 4.7: *Top:* Evolution of the VIRCAM@VISTA zero points as calculated by the CASU pipeline during a period of 4.2 hours of one night. Zero points in tiles with high stellar density (especially in tiles *b307*, *b294* and *b308*, marked by the vertical grey bars as well as text) are systematically underestimated due to the photometric blending in the 2MASS catalog. *Bottom:* The seeing measured on the K_S-band observations. The changes in seeing and zero points are not correlated.

evidently offset from the fiducial line, when compared to the converted 2MASS magnitudes. Furthermore, as can be seen, many outliers from the bulk of sources are preferentially above the fiducial line, meaning that the same sources are measured significantly dimmer by VISTA than by 2MASS. Visual inspection of images taken by both telescopes revealed that almost all of the latter stars are blended in the 2MASS observations, while typically being resolved into 2 – 4 separate sources by VISTA. Examples of severe, medium and insignificant blending in the 2MASS observations are shown in Figure 4.6, for the individually marked sources in the bottom panel of Figure 4.5.

CASU's processing pipeline calculates the zero points of individual pawprints as the average of the difference between the instrumental magnitudes of stellar sources and their converted 2MASS magnitudes. In fields of high stellar density, several VISTA sources can be blended as one 2MASS source, leading to the photometric zero-point bias revealed here. In order to visualize the effect of this bias, Figure 4.7 illustrates the change in magnitude zero points as measured by the CASU pipeline during ~ 4 hours of VISTA observations of VVV fields. For comparison, the seeing as measured on the K_S-band observations is also shown.

These two quantities are not correlated with each other.

During this particular night, VISTA was obtaining multi-band observations for the VVV survey, hence the availability of all three of the bands J, H and K_S. Near the beginning and near the end of this period VISTA observations were obtained in uncrowded fields far from the Galactic plane, while during the middle the densest VVV tiles were observed. As can be seen, the amplitude of the photometric zero point bias also depends on the filter being used, being the largest in the H band, due to the different blending levels of sources between the bands. Besides low stellar densities, high extinction can also reduce the size of this bias, due to the artificial lowering of the number of detected stellar sources in the 2MASS catalog.

Besides the pawprint catalogs, CASU also provides photometric catalogs for the complete tile observations of VISTA, by performing aperture photometry on the combined images of individual pawprint observations of each of the observed tiles. González-Fernández et al. (2018) describe a grouting process for the correction of the photometric differences between stars found on the same tile, but on different individual pawprints. They attribute these differences mostly to the changing observing conditions, primarily seeing, and hence point-spread function variations between pawprint observations of a VISTA tile. It has to be noted that as each pawprint of a tile covers a slightly different sky region, if there are regions with different amounts of extinction within a tile, the amount of crowding, and hence blending in the 2MASS catalog, will change for different VISTA pawprints. As blending biases the pawprint zero points, different blending levels between VISTA pawprints increases the scatter of zero points measured by the CASU pipeline. This effect can be seen in Figure 4.7 as increased scatter in the pawprint zero points for individual tiles of high stellar density regions. Although the grouting process corrects for the changes in seeing, the photometric catalogs of tiles are expected to be just as biased as those of individual pawprints.

4.5 Recalibration of VVV JHK_S observations

As the photometric biases detailed in Sections 4.3 and 4.4 are mostly manifesting themselves in the zero points of the VIRCAM detectors and the VISTA pawprint observations, respectively, they can mostly be corrected for by a revision of the per detector zero points.

The zero-point calculation procedure described by González-Fernández et al. (2018) allows a relatively large, 1 arcsecond cross-match radius between the 2MASS catalog and the detected VISTA sources. The color scale in Figure 4.5 shows the cross-match distance be-

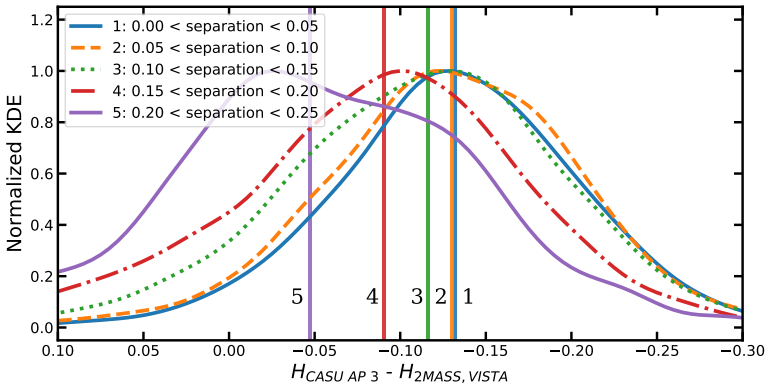

Figure 4.8: Kernel density estimates of magnitude differences in different cross-match bins in the *H*-band in VVV tile *b307* (top panel of Fig. 4.5, for stars fainter than 12.5 magnitudes in the 2MASS catalog). Different lines correspond to different matching radii, as indicated in the inset (in arcsec). The KDEs are normalized to their individual maxima. Vertical bars mark the median values of different bins. Note the progressively fainter magnitudes measured by VISTA with increasing cross-match radius.

tween the VISTA sources and their 2MASS equivalents. Most notably, the 2MASS stellar sources lying above the fiducial lines (hence probably being blends in the 2MASS catalog) generally have higher separations than those belonging to the main locus of unblended stars. Naturally, as many sources are blended in the 2MASS Point Source Catalog towards high stellar density fields, not only their magnitudes, but also their measured positions, are affected by the blending, explaining the higher cross-match distance.

The dependency of the bias on the cross-match radius is examined in Figure 4.8 for the same *H*-band pawprint of VVV tile *b307*, as shown previously on the top panel of Fig. 4.5. It can be seen that for higher cross-match radius bins the VISTA observations are becoming progressively fainter. This is most likely due to the increasing fraction of blending in the 2MASS catalog with increasing cross-match distance. Up to 0.1 arcseconds, the effect of blending is minimal, as shown by the median values of the samples, marked by vertical bars.

A method has been developed in order to correct the magnitude zero-points of individual detectors, while taking into account the specific properties of the VVV survey observations. The calibration process described by González-Fernández et al. (2018) calculates zero points on the pawprint level. The principal reason for this is to guarantee that enough 2MASS stars are available for the robust determination of zero points, even in high Galactic latitude fields.

As VVV fields are generally rich in stars, the zero-point corrections are calculated on a per detector, per pawprint (epoch) basis for each of the five smallest apertures. The main points of this correction method are the following:

- The 2MASS Point Source Catalog converted to the VISTA version 1.5 system (Sect. 4.2.1) is used as the source catalog for the (re)calibration of VISTA observations. The utilized catalogs in the VISTA version 1.5 system, therefore correcting the photometric biases with this 2MASS catalog results results in that the corrected catalogs will be placed in the more recent system. As in some cases non-linearity affects VISTA observations down to 12.5 magnitudes (see upper panel of Fig. 4.5), only observations fainter than this limit are used.

- The converted 2MASS catalog is cross-matched with the unified catalogs of sources detected in each of the VVV tiles with a cross-match radius of 0.15 arcseconds. The small cross-match radius helps in diminishing the effect of the photometric bias due to the different blending characteristics between the 2MASS and VISTA observations (Sect. 4.4). Although this effect becomes non-marginal for cross-match radii larger than 0.1 arcseconds, it is still fairly small until 0.15 arcseconds, and the increased source counts help in deriving a more robust zero-point correction in tiles with lower stellar densities.

- The zero point correction is calculated for each detector and each aperture of each pawprint observation as the median difference between the magnitudes of stellar sources and their corresponding converted 2MASS magnitudes. The median difference is preferred over the mean, as it is a more robust statistic in the presence of outliers (for example, variable stars).

- Individual light curves are corrected for the VISTA zero-point biases by subtracting the appropriate zero-point corrections from their original measured magnitudes.

4.5.1 Effects of the photometric corrections

After calculating the photometric corrections, as described in Section 4.5, the full effect of the zero-point biases described in Sections 4.3 and 4.4 can be evaluated. Figure 4.9 illustrates the effect of correcting the zero points of individual observations upon the light curves of four classical Cepheid variables. In the case of three Cepheids from the moderately

crowded tile *b319*, their average magnitudes have decreased by about 0.1 magnitudes. Furthermore, the scatter in their light curves decreased by about one order of magnitude. The fourth Cepheid illustrates the mostly constant photometric offsets between different chips in the case of a VVV disk tile, which also adds to the light curve scatter. The fact that the characteristics of the light curves have improved so drastically validates the choice of the chip-based recalibration of the photometric zero points.

The spatial distribution of the magnitude offsets is affected by the pawprint pattern, as is the apparent temporal variation of the zero-point bias, most probably due to changes in the observing conditions and variations in telescope pointing. The decreased scatter of the light curves shown by Figure 4.9 is a result of accounting for this fact. Due to this, it is challenging to visualize the effect that the zero-point bias has on the VISTA photometry on small spatial scales. Nevertheless, in order to get an idea of its distribution, both the zero-point corrected, as well as the uncorrected light curves of 19 760 RR Lyrae variables, were fit using the *PyFiNeR* algorithm. The difference between the magnitude averaged mean magnitudes in the J, H, K_S bands, calculated using the bias corrected and the original light curves, is shown by Figure 4.10 towards the Galactic bulge. The derived colors are similarly biased, as shown in Figure 4.11. In the outer regions of the bulge, the magnitude and color differences are minimal, due to the minimal blending of VISTA calibrator stars in the 2MASS catalog. The small systematic differences in each band, seen prominently towards the outer regions of the Galactic bulge, are caused by the systematic difference between the VISTA version 1.3 and v1.5 photometric systems. Towards regions of high stellar density, such as Baade's Window at $l, b \sim 2$ deg, -2 deg, the difference increases drastically. It has to be noted that this difference behaves markedly different from one band to the next, most probably due to the different intrinsic colors and amount of extinction that each star in the converted 2MASS catalog (Section 4.2.1) is subject to.

The diagonal streaks, mostly seen in the J band in the southern side of the bulge, are artefacts of the 2MASS scanning pattern and point towards zero-point offsets on the order of 0.02 mag in those regions.

As can be seen, these photometric biases are an integral part of the reduced VISTA photometric catalogs. Unfortunately, their profound effect necessitates the revision of many of the results obtained by previous use of VVV (VISTA) observations. As both the derived apparent magnitudes and colors vary towards different regions of the Galactic bulge (Figs. 4.10

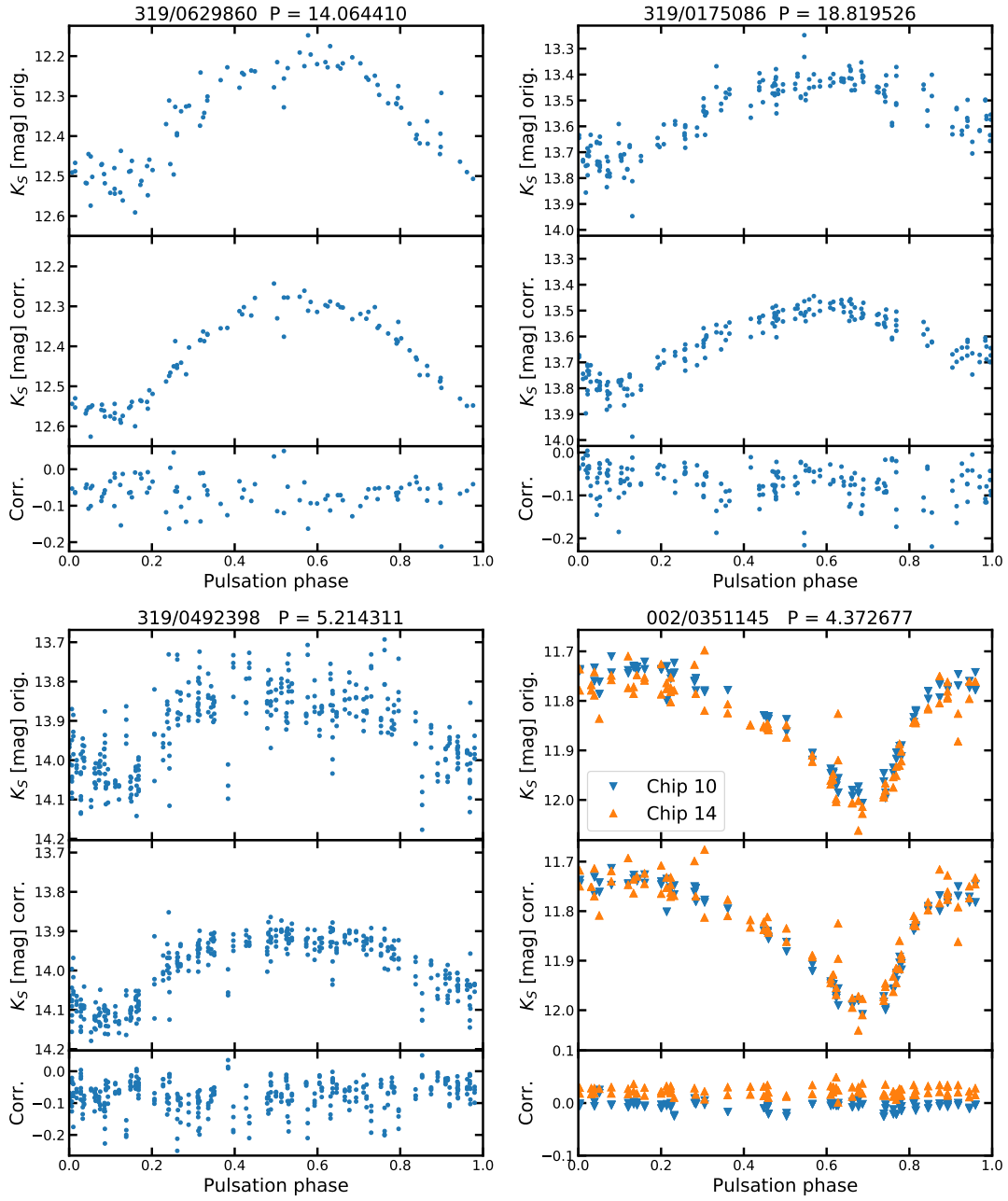

Figure 4.9: Cepheid K_S band light curves, corrected for the VISTA zero-point bias with the method described in Sect. 4.5 in both cases. *Top panels:* The original folded light curves. *Middle panels:* The corrected folded light curves. The magnitude scales are the same as in the corresponding *top* panels. *Bottom panels:* The offsets for individual epochs of observation. Note that the last Cepheid has VVV observations from two separate VISTA detectors.

and 4.11), studies of the extinction law are gravely affected. As an example, Nataf et al. (2016) studied the variation of the extinction law based on combined VVV JHK_S and OGLE IV photometry by tracing the changes and hence color excesses of the red clump stars across the bulge. Their VVV photometry has been derived by the point-source-fitting (PSF) method; however, their zero-point calibration is based on the CASU photometric catalogs, which makes it inherit the spatial zero-point bias structure shown in Figure 4.10. Nataf et al. (2016) claim to have found pairs of sightlines where their respective red clumps occupy the same position on the K_S vs. $V-I$ color-magnitude diagram; however, the red clumps have very different $J-K_S$ colors (their Figs. 5 and 6). Based on this, they claim that the extinction law has a non-standard shape towards the Galactic bulge. As the compared fields are not on the same VVV pawprint or even tile, their J and K_S magnitude zero points and hence every color derived from these quantities are affected by the zero-point bias differently. It can be concluded that a deeper reexamination of the VVV photometry will be required in order to properly establish the nature and shape of the extinction law towards different sight lines.

4.5.2 Recommendations for future usage of VISTA observations

The relatively simple recalibration procedure described in Section 4.5 can only serve as a temporary measure in future studies utilizing VISTA observations. The selection of non-blending 2MASS stars based on the cross-match distance is a very crude method, as it still leaves a considerable fraction of blending stars in the sample used to calculate the zero-point corrections. Ideally, each and every 2MASS star used for the zero-point calibration of VISTA observations should be revised and discarded, if VISTA observations reveal close companions within the 2MASS PSF. This will require a considerable amount of work on the part of CASU.

Unfortunately, PSF photometry is also affected by the blending in the 2MASS catalogs, as so far all studies in the VVV footprint using the PSF technique have relied on the CASU aperture photometry catalogs for zero-point calibration, inheriting their biases as well. As an alternative, usage of the converted 2MASS photometry can be recommended instead for the zero-point calibration of PSF measurements, together with implementing additional steps for rejecting blending stars in the 2MASS catalog. Towards the Galactic center, PSF photometry has the unique power of resolving individual stellar sources in the VVV survey. The centralmost tiles are generally observed right after and/or before more outlying tiles.

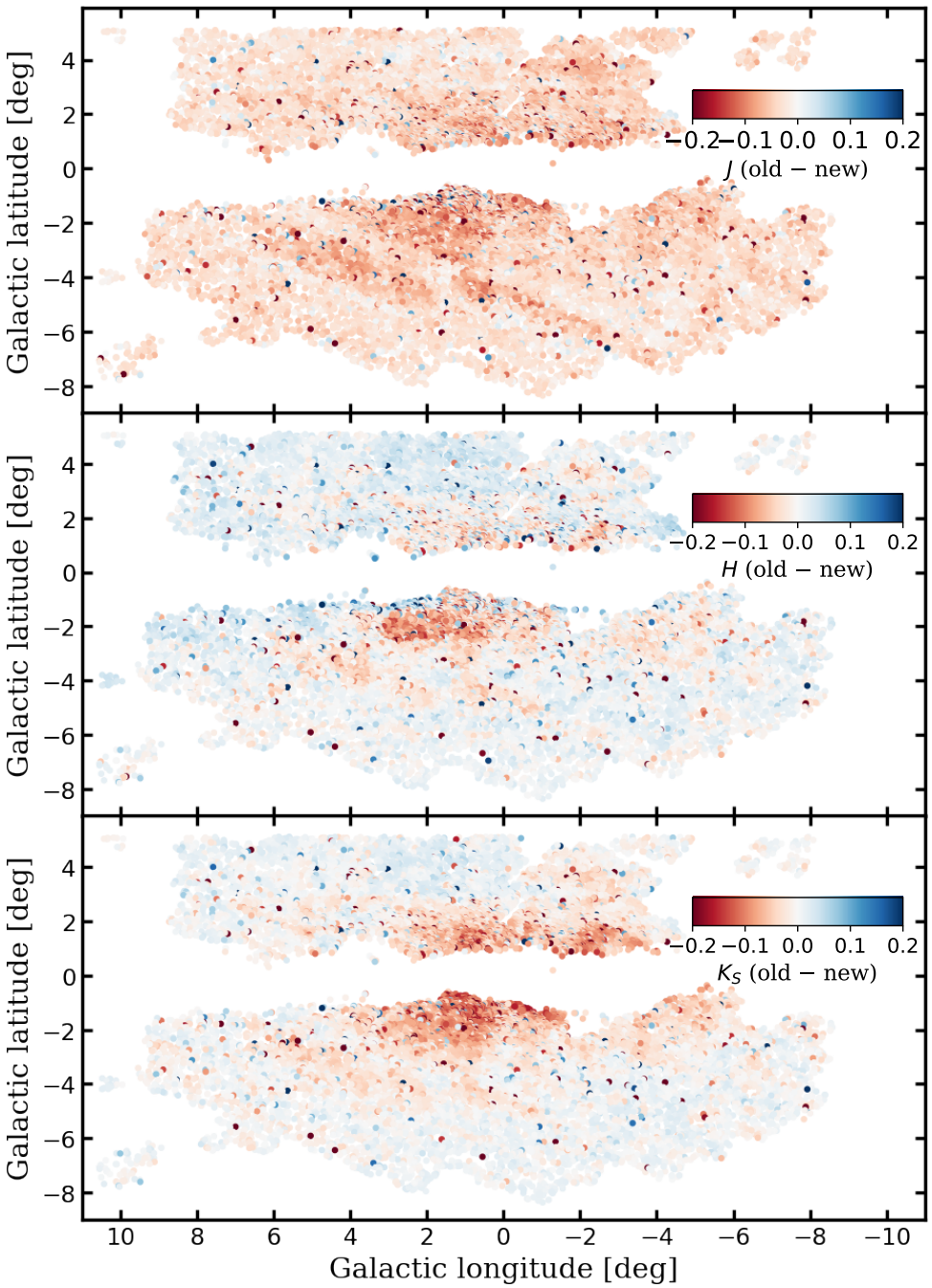

Figure 4.10: Average magnitude differences in the J,H,K_S bands between the original CASU light curves and their corrected equivalents for OGLE RR Lyrae stars.

Therefore, the zero points derived for these flanking tiles could be adopted as the zero points for the photometry of Galactic center tiles.

Alternatively, the photometric catalogs of different tiles and pawprints could be used in an "*Ubercal*"-like calibration scheme (Schlafly et al. 2012; Magnier et al. 2013). However, the photometric zero points will have to be derived with extreme care towards regions of high stellar densities.

The VISTA photometric pipeline described by González-Fernández et al. (2018) is based on that of WFCAM (Hodgkin et al. 2009). As the resolution of the two cameras and telescope systems are quite similar, it is very likely that observations calibrated by CASU for the latter are also affected by the photometric zero-point bias described in Section 4.4 for VIRCAM@VISTA. Therefore, the recommendations described here are also applicable to WFCAM for observations in high stellar density regions. Furthermore, it has to be noted that VISTA observations obtained for other high stellar density fields, such as the VISTA Magellanic Cloud Survey (VMC, Cioni et al. 2011), might be similarly affected by zero-point offsets caused by blending in the 2MASS catalog.

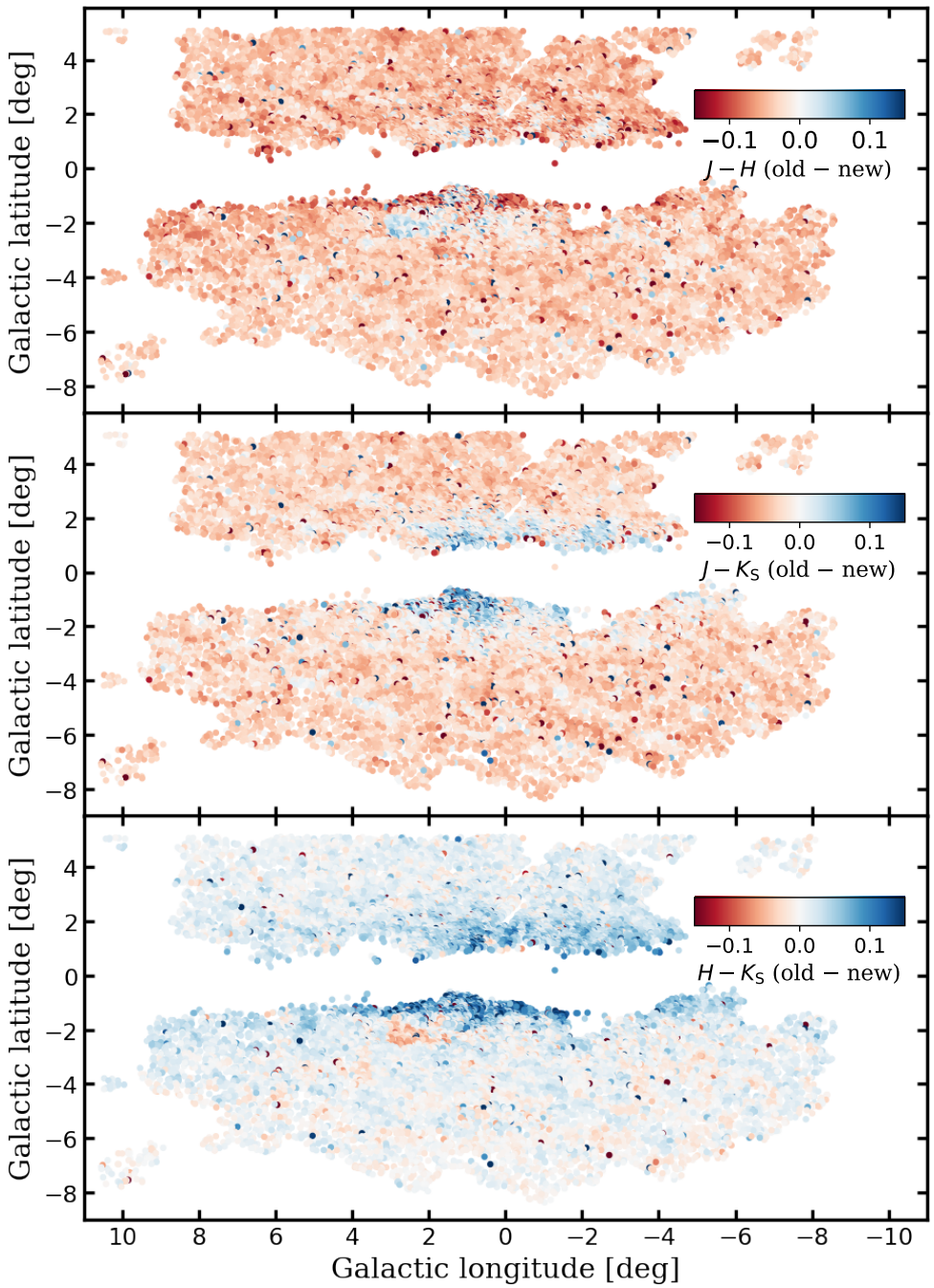

Figure 4.11: As in Figure 4.10, but the three panels from top to bottom show the average $J - H$, $J - K_S$ and $H - K_S$ color differences, respectively.

5

Cepheids in the VVV Survey

5.1 Cepheid search in the corrected VVV photometry

The search for Cepheid variables has been conducted in the complete area covered by the VVV disk fields, covering the same area as the disk RRL search (Section 3.1.1), augmented with a ~ 3 degree wide central strip of the VVV bulge area, including the bulge fields *b313* – *b332* and *b335* – *b354*. The extremely crowded fields *b333* and *b334* (containing the Galactic center region) are excluded from this variability search.

The same procedure as described in Section 3.1.1 was followed in order to construct the light curves of individual objects from the VVV pawprint photometric catalogs, resulting in a catalog of a total of $\sim 3 \cdot 10^8$ point sources. The individual epochs were corrected for the photometric zero-point issues discussed in Chapter 4, resulting in substantially reduced light-curve scatter for many of the crowded VVV tiles (see Figure 4.9).

In order to find the variable sources in the corrected light curves, the same variability index-based search was performed as in the case of the disk RRL stars, as described in Chapter 3.1.2. This preselection resulted on average in a number of $\sim 10^4$ candidate variables sources per VVV tile. These light-curves were searched for periodic signals between 3.8 and 40 days with the GLS method (Zechmeister and Kürster 2009). The upper limit was imposed by the time sampling of VVV disk tiles, as virtually all variables with periods longer than this limit present large light-curve gaps in their photometry. Meanwhile, the lower limit was chosen to minimize the possible contamination of the sample with Cepheids of different pulsation modes, as well as other kinds of variables. Additionally, the criterion $K_{\mathrm{s,min}} - K_{\mathrm{s,max}} - \sigma > 0.15$ mag, where σ is the weighted standard deviation of the measurements, was also adopted to decrease the number of variables with low-amplitude, noisy light curves. This procedure resulted in a final sample of $\sim 4 \cdot 10^4$ objects.

5.1.1 Classification of Cepheid candidates

Although there are many case studies of light-curve classification in the optical wavelengths (see e.g., Kim and Bailer-Jones 2016; as well as references therein), the same cannot be said for near-IR bands, where the diminished amplitudes of pulsating variables pose a particular challenge. In Section 3.1.3, the classifier of Elorrieta et al. (2016) was utilized to search for RRL variables in the VVV disk fields. As no such classifier is available for Cepheid variables, Dr. István Dékány has developed a classifier based on Convolutional Neural Networks (CNN, Lecun and Bengio 1995), in order to classify the Cepheid candidates found in Section 5.1.

As preparation, the light curves of the variables have been fit with a Fourier series of low (2–7) order, using the Huber loss function (Equation 2.5), combined with iterative rejection of outlying points, in order to reach a robust fit of the light curves. The order of the Fourier series is determined by 10-fold cross validation. Finally, the light curve fits have been aligned by the phase of the first Fourier harmonic, and the light curves are sampled in an equidistant grid of 38 phases. These phase points, along with the pulsation period, serve as the input of the CNN classifier, effectively changing the problem to that of computer vision (Lecun and Bengio 1995). This is in contrast to traditional approaches, where the separation between the object classes is searched for as a function of a set of their light-curve parameters (such as the identified periods, the amplitude, color change, standard deviation of the light curve, etc.).

The classifier has been trained to differentiate between Type I and II Cepheids, as well as non-Cepheid variables. The light curves of Classical Cepheids have been assembled from the literature: for Galactic Cepheids, the photometry published by Welch et al. (1984), Laney and Stobie (1992), Barnes et al. (1997), and Monson and Pierce (2011); for the LMC, by Persson et al. (2004); and for the SMC, by Ita et al. (2018). The VVV light curves of Type II Cepheids found by the OGLE survey (Soszyński et al. 2017) have also been extracted to serve as the Type II Cepheid training sample. Each of these Cepheid light curves have been fit with the method described above. The light-curve fits have been visually inspected to reject objects with noisy or sparse photometry. After this step, 140 Classical and 356 Type II Cepheid light curves have been left.

In order to increase the very small number of Classical Cepheids in the training sample, additional variables have been selected from the bulge fields following the method used in

Dékány et al. (2015b). According to it, the color excesses (calculated for the variables under the assumption that they are Cepheids) are compared to the map of Gonzalez et al. (2012), which provides the line-of-sight extinction integrated at the distance of the Galactic bar. Classical Cepheids behind the bulge have larger color excesses than expected from the map of Gonzalez et al. (2012). The most reddened variables have been manually inspected, and a further 48 Classical Cepheids have been selected to supplement the sample of training light curves, for a total of 188 objects.

To assemble a representative sample of variable stars belonging to non-Cepheids, but occupying the same period range as Cepheids (including, for example, semi-contact binaries, anomalous Cepheids, as well as young, active stars), the photometry of the VVV tiles *b292*–*b296* were searched for variables with the same method as described in Section 5.1. These tiles lie towards Baade's Window, and have been covered extensively by the OGLE survey (Udalski et al. 2015), identifying virtually all Classical and Type II Cepheids in it (Soszyński et al. 2017), which have been removed from the variable sample, as well. At the end, a total number of 498 high signal-to-noise ratio, non-Cepheid variables have been selected to serve for the training of the classifier.

The implementation details of the classifier will be discussed in detail in Dékány et al. (in prep.), however, its performance is discussed here. It is traditionally recommended to evaluate the performance of a classifier on a separate test set of objects, which were not used during any of the training or cross-validation steps. Unfortunately, the very small number of Cepheids does not permit this approach: either too much data would have to be withheld during the training, limiting classification accuracy, or the very few withheld objects would not provide a meaningful statistic of the performance of the classifier. This results in having to rely on the cross-validation results obtained during the training of the classifier, which might overestimate its performance. Cross-validation results in an estimated accuracy (fraction of correctly classified objects) of 95 % for all objects, and 97 % for both when only Classical or Type II Cepheids are concerned. The estimated purity (fraction of correctly classified objects) is 93 and 96 % for the Classical and Type II Cepheids, respectively. The recall (completeness) is 93 and 95 % for the Classical and Type II Cepheids, respectively. As discussed above, these values are probably overestimated due to not using a separate test set, therefore the real performance is expected to be lower by about $\sim 5 - 10$ %.

As the training light curves have been preferentially selected to include well defined, high

signal-to-noise ratio light curves, the performance of the classifier is expected to decrease for real data of fainter, noisier light-curves. In order to assess this, synthetic light curves were generated by injecting Cepheid and non-Cepheid light-curve shapes into light curves of a representative sample of of non-varying stars, also selected form the VVV survey. This process resulted in 33 000 light-curves with properties identical to that of real variables the VVV survey, but with known class labels. Applying the classifier to these light curves reveals that the performance metrics stay at around 90 % down to $K_S \sim 14$ mag, but fall rapidly to 80 % at ~ 15 mag. Nevertheless, in case of the cumulative performance indices (calculated until a certain magnitude), neither of them fall below 90 % until $K_S \sim 15$ mag, lending high confidence in the performance of the classifier.

At this step, the classifier is applied to the light-curve sample of Cepheid candidates in the VVV survey discussed in Section 5.1. The light curves of the objects classified as either Classical or Type II Cepheids were visually inspected in order to increase the sample's purity. There are two main categories of light curves rejected at this stage. Firstly, there were cases of light-curve shapes which obviously did not belong to any type of Cepheid variable. These complex light-curve shapes are not present in the training sample of the non-Cepheid variables due to their rarity. In a few cases, the extrapolation of the parameter space done by the classifier results in the (incorrect) classification of such a variable as a Cepheid. However, the number of these variables is very small and they are easily rejected. Secondly, the GLS period search sometimes finds half of the real period of long-period contact or semi-detached binaries, and the iterative rejection of "outlying" points results in a light-curve fit resembling that of a Cepheid.

After the rejection of obviously misclassified Cepheids, a total of 608 variables are classified as Type II, with the majority, 433 objects, being located in the bulge fields of the VVV survey. Out of them, 82 were already found by the OGLE survey (Soszyński et al. 2017), while 4 of them were found and misclassified as Classical Cepheids by Dékány et al. (2015b). The remaining 347 Type II Cepheids are new discoveries. In the disk fields, 6 stars were also found by Chen et al. (2018) with ambiguous classifications, leaving 169 new discoveries.

In the final sample, a total of 686 objects are classified as Classical Cepheids, with 238 being in the bulge, and 448 in the disk regions of the VVV survey. In the bulge, 34 have been found by Dékány et al. (2015b), and 18 by Matsunaga et al. (2016), with 7 being common between the two samples. In the disk fields, Chen et al. (2018) found 15 of the variable objects

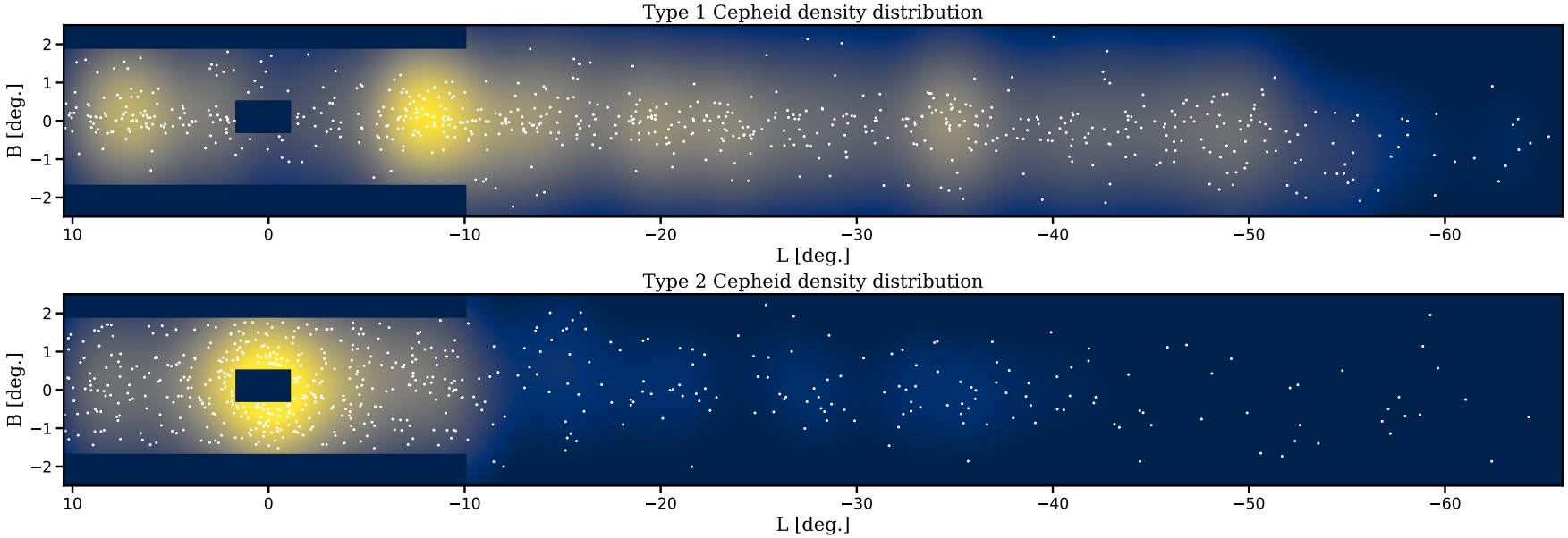

Figure 5.1: *Top:* Distribution of Cepheids classified as Type 1 in Galactic coodinates. The color scale corresponds to kernel density estimate of the distribution with a Gaussian kernel of 1° size. The kernel density is only shown for areas that are covered by the VVV tiles used during the Cepheid search. Note that Type 1 Cepheids are preferentially located below the Galactic plane farther from the Galactic center due to the effect of the Galactic warp. *Bottom:* Same as top, but for the distribution of Type 2 Cepheids. Note the heavy concentration of Type 2 Cepheids towards the Galactic center.

classified here as Classical Cepheids, but only classified 6 of them as such. Therefore, the final number of newly discovered Classical Cepheids is 640.

The distribution of the variables classified as Classical and Type II Cepheids in Galactic coordinates is shown in Figure 5.1. The Classical Cepheids are distributed along the Galactic plane, showing multiple local peaks, with the most significant being the one at $l \sim -8°$. Furthermore, Cepheids in the southern disk, far from the Galactic bulge, are located preferentially below the Galactic plane, following the Galactic warp, in agreement with the results of Skowron et al. (2018) and Chen et al. (2019). In contrast, the Type II Cepheids are heavily concentrated towards the Galactic center.

5.1.2 Cepheid period-luminosity relationships

In order to calculate the most accurate distances possible of both subtypes of Cepheid variables, the choice of the adopted period-luminosity (PL) relations is critical. As far as near-IR PL relations go, due to their importance in the calibration of the cosmic distance scale, Classical Cepheids have received substantial attention over the past few decades (see, e.g., Persson et al. 2004; Ripepi et al. 2012, 2017; Macri et al. 2015; Groenewegen 2018; Riess et al. 2018, as well as references within). In contrast, only a handful of studies have explored near-IR PL relations for Type II Cepheids (Matsunaga et al. 2006, 2009, 2011; Bhardwaj et al. 2017a, 2017c).

The recent *Gaia* Data Release 2 (Gaia Collaboration et al. 2018) promised to provide a highly accurate calibration of the Cepheid PL relations by providing accurate distances to nearby Classical Cepheid variables. However, due to the limited number of stars, as well as the possible offsets in the parallax zero points, this promise has been left unfulfilled so far (Groenewegen 2018; Riess et al. 2018). As the VVV observations have been taken with the VISTA telescope, the same one used by the VMC survey (Cioni et al. 2011), it would be a natural choice to use the PL relationships provided by Ripepi et al. (2012, 2017), for the VVV Cepheids as well. However, the VMC survey did not observe in the H-band, requiring the use of an external source of the PL relation for that band. Furthermore, as discussed in Chapter 4, VISTA photometry suffers from photometric zero-point biases in dense stellar fields. As this bias might also affect the photometry of the VMC survey, it is wise to refrain from utilizing PL relations calibrated on it, until its effect can be properly quantified. Due to this, from hereon, the PL relations calibrated in the 2MASS system using Large Magellanic

Cloud Cepheids by Macri et al. (2015) are adopted[1]:

$$M_{J,\mathrm{2MASS}} = 13.228 - 3.156 \cdot (\log P - 1) - \mu_{\mathrm{LMC}}, \tag{5.1}$$

$$M_{H,\mathrm{2MASS}} = 12.847 - 3.187 \cdot (\log P - 1) - \mu_{\mathrm{LMC}}, \tag{5.2}$$

$$M_{K_{\mathrm{S}},\mathrm{2MASS}} = 12.776 - 3.274 \cdot (\log P - 1) - \mu_{\mathrm{LMC}}, \tag{5.3}$$

where μ_{LMC} is the distance modulus of the Large Magellanic Cloud. Equations 5.1 – 5.3 have been converted into the VISTA system by using the formulas given by González-Fernández et al. (2018):

$$M_{J,\mathrm{VISTA}} = 13.214 - 3.159 \cdot (\log P - 1) - \mu_{\mathrm{LMC}}, \tag{5.4}$$

$$M_{H,\mathrm{VISTA}} = 12.854 - 3.186 \cdot (\log P - 1) - \mu_{\mathrm{LMC}}, \tag{5.5}$$

$$M_{K_{\mathrm{S}},\mathrm{VISTA}} = 12.773 - 3.248 \cdot (\log P - 1) - \mu_{\mathrm{LMC}}. \tag{5.6}$$

Very recently, Pietrzyński et al. (2019) determined the distance of the LMC with a precision of 1%, with a mean distance modulus of $\mu_{\mathrm{LMC}} = 18.477 \pm 0.026$ mag. As this is the most accurate determination of this quantify to date, it has been adopted for the calculation of the absolute magnitude of Cepheids.

Similarly to the situation of the Classical Cepheids, the Type II Cepheid PL relations of Bhardwaj et al. (2017a), as they rely upon VVV observations of Type II Cepheids identified by the OGLE survey (Udalski et al. 2015), might also be biased due to the photometric problems discussed in Chapter 4. Due to this possible factor, usage of the PL relations determined by Bhardwaj et al. (2017c) is preferred, after converting them to the VISTA system (González-Fernández et al. 2018):

$$M_{J,\mathrm{VISTA}} = 15.548 - 2.066 \cdot (\log P - 1) - \mu_{\mathrm{LMC}}, \tag{5.7}$$

$$M_{H,\mathrm{VISTA}} = 15.149 - 2.199 \cdot (\log P - 1) - \mu_{\mathrm{LMC}}, \tag{5.8}$$

$$M_{K_{\mathrm{S}},\mathrm{VISTA}} = 15.067 - 2.233 \cdot (\log P - 1) - \mu_{\mathrm{LMC}}, \tag{5.9}$$

while also adopting the same LMC distance modulus (Pietrzyński et al. 2019), as done for the Classical Cepheids.

[1]The actual values of the coefficients are taken from the erratum published in Macri et al. (2016).

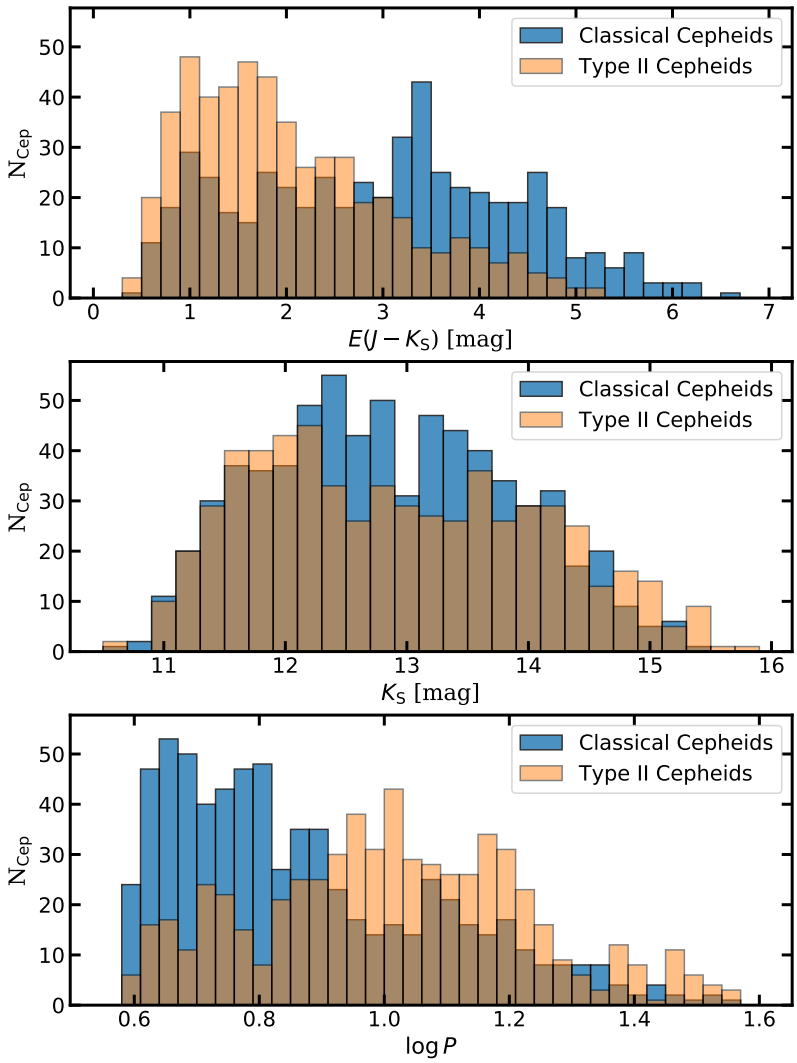

Figure 5.2: *Top:* $E(J-K_S)$ color-excess distribution of the VVV Classical and Type II Cepheid samples. *Middle:* Same as top, but showing the distribution of the average K_S magnitudes. *Bottom:* Same as top, but showing the distribution of the logarithm of the pulsation periods.

The color excesses of the variables classified as Classical and Type II Cepheids were calculated adopting Equations 5.4 – 5.6 and Equations 5.7 – 5.9, respectively. The distribution of the calculated $E(J-K)$ color excesses and the observed K_S magnitudes of both Cepheid samples is shown in the top and middle panels of Figure 5.2.

Classical Cepheids can have $E(J-K_S)$ color excesses reaching ~ 7 magnitudes, beyond which some of them fall below the J-band detection limit of the VVV survey. In contrast, the color excesses of Type II Cepheids only reach about 5 magnitudes. As variables of this type are much fainter than Classical Cepheids, they simply go undetected at larger distances

in the Galactic disk, where they would be expected to have color excesses similar to those of Classical Cepheids. The bulk of the Type II Cepheids come from the Galactic bulge, having $E(J - K_S)$ color excesses in the range of $0.5 - 2.5$ magnitudes, in general agreement with the color excesses of red clump stars (Gonzalez et al. 2012) measured in the same area. The K_S-band magnitude distribution of the two Cepheid samples is similar, ranging from the saturation limit of the VVV survey to about $K_S \sim 15.5$ magnitudes, below which the light curve scatter does not allow good light-curve fits, and hence classification of variable classes.

5.1.3 The near-infrared extinction law

In order to determine the extinction suffered by the individual Cepheid variables (both Classical and Type II), total-to-selective extinction ratios (an extinction law), defined as:

$$R_{K_S JK_S} = A(K_S)/E(J - K_S) \tag{5.10}$$

$$R_{K_S HK_S} = A(K_S)/E(H - K_S), \tag{5.11}$$

must be adopted. The measured ultraviolet, optical and infrared fluxes of nearly all astronomical sources are affected by the reddening, hence our incomplete understanding of its properties is a fundamental limiting factor in astronomy. Due to this, the extinction law has been studied by many authors (see, e.g. Cardelli et al. 1989; Nishiyama et al. 2006, 2009; Majaess et al. 2016; Nataf et al. 2016, Alonso-García et al. 2017; Schlafly et al. 2017, as well as references therein).

The selection of the correct extinction law is critical in studies of objects with high extinction, as is the case here (top panel in Figure 5.2). As an example, Dékány et al. (2015b) claimed the existence of a thin, star-forming stellar disk within the volume of the Galactic bulge, based on distances calculated using the extinction law of Nishiyama et al. (2009). However, Matsunaga et al. (2016) argued that the correct extinction law is that of Nishiyama et al. (2006), resulting in differences of $0.3 - 0.4$ mag in calculated extinction values, and therefore in distance moduli. This results in a difference in calculated distances of $\sim 1 - 1.5$ kpc at ~ 8 kpc.

In order to determine robust distances, the extinction law is determined by utilizing the large number of well-measured Type II Cepheids, concentrated in the Galactic bulge region. Recently, the distance of the Galactic center (Sgr A*) was measured with high precision with

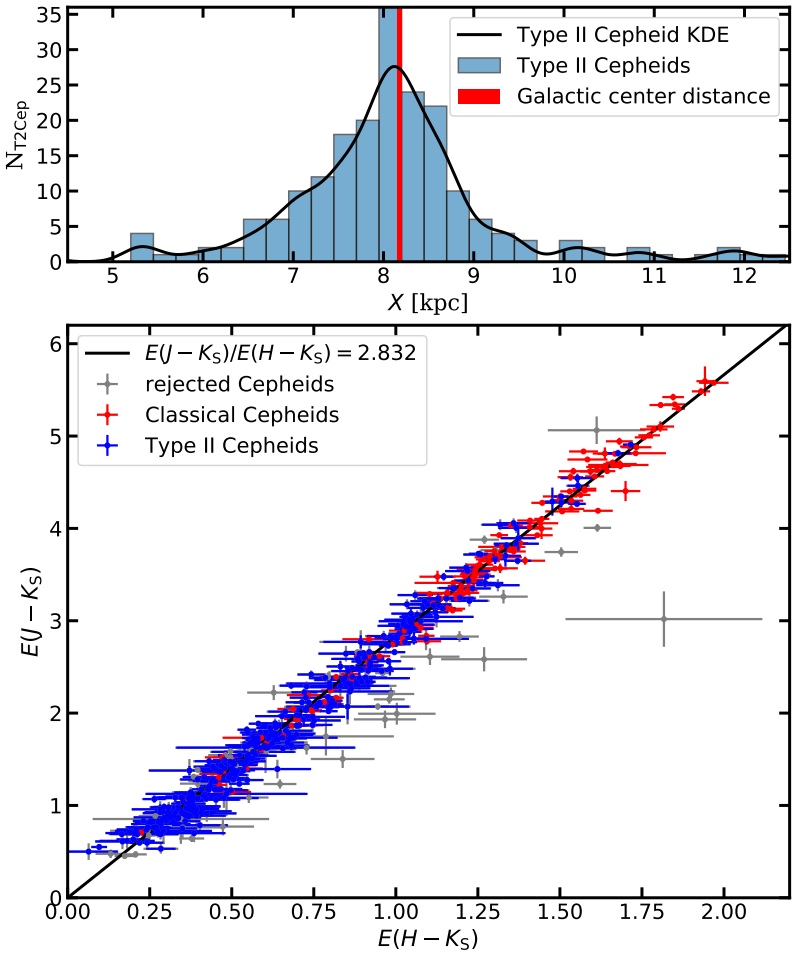

Figure 5.3: *Top:* The optimal total-to-selective extinction ratios are found by optimizing their values, so that the maximum of the kernel density distribution of Type II Cepheid distances in Galactic X coordinates matches that of the Galactic center (vertical red band, Abuter et al. 2019). *Bottom:* The mean selective extinction ratio $E(J-K_S)/E(H-K_S)$ is found by a robust fit of the individual Cepheid extinction ratios towards the Galactic bulge.

the *GRAVITY* instrument of the Very Large Telescope Interferometer to be of 8178 ±25.5 kpc (Abuter et al. 2019). The top panel in Figure 5.3 illustrates the method utilized to determine the extinction ratios. The distances to individual Type II Cepheids (with $\log P < 1.3$) are calculated assuming different values of $R_{K_S JK_S}$. For each of these values, the mean distance of this population of stars is determined by locating the maxima of the kernel density estimates in Galactic X coordinates. The most probable value is found when this distance matches that of the Galactic center (Abuter et al. 2019):

$$R_{K_S JK_S} = 0.528 \pm 0.021\,(\text{stat.}) \pm 0.08\,(\text{sys.}) \tag{5.12}$$

$$R_{K_S HK_S} = 1.488 \pm 0.004\,(\text{stat.}) \pm 0.019\,(\text{sys.}). \tag{5.13}$$

Furthermore, in order to revise these values, the selective extinction ratio $R_{JK_S HK_S} = E(J - K_S)/E(H - K_S)$ can also be determined by fitting their values for individual Cepheids directly. The bottom panel in Figure 5.3 illustrates the calculation of this quantity: in order to determine a more robust value, both the Classical and Type II Cepheids found in the VVV bulge fields are used in the fit. A few of the Classical Cepheids have extreme, $R_h > 25\,\text{kpc}$ calculated heliocentric distances, suggesting either photometric problems or misclassification of non-Cepheid variables, which were also not rejected in the manual revision of the light curves. These variables are omitted from the analysis. In addition, only Type II Cepheids with $\log P < 1.3$ are used during the fit. During the linear regression fit, the most outlying Cepheid variables were removed using iterative rejection with a 3σ threshold, resulting in the estimate: $R_{JK_S HK_S} = 2.832$. Dividing $R_{K_S JK_S}$ with $R_{K_S HK_S}$ (Equations 5.12 and 5.13) results in a slightly different value of 2.818. As $E(H - K_S)$ always has larger relative errors than $E(J - K_S)$, as well as to enforce consistency of distance estimates based on $R_{K_S JK_S}$ and $R_{K_S HK_S}$, from here now on, instead of relying on the value in Equation 5.13, the value:

$$R_{K_S JK_S} = R_{JK_S HK_S} \cdot R_{K_S JK_S} = \frac{E(J - K_S)}{E(H - K_S)} \cdot \frac{A(K_S)}{E(J - K_S)} = \frac{A(K_S)}{E(H - K_S)} = 0.525 \tag{5.14}$$

is adopted as the selective-to-total extinction ratio for the K_S-band calculated from the $E(H - K_S)$ color excess measurements.

For every variable classified as a Type II Cepheid, the absolute magnitudes and the color excesses are calculated using the PL relations in Equations 5.7–5.9. The K_S-band extinction is calculated with Equations 5.12 and 5.14, resulting in two distance estimates for the majority of the variables. These distances are averaged by calculating their weighted arithmetic mean, using the inverse squares of their calculated errors as weights. For the remaining variables, when only the $E(H - K_S)$ or $E(J - K_S)$ color excess is available, the distance calculated using that color excess is adopted as the final value. Figure 5.4 shows the distribution of Type II Cepheids in Galactocentric coordinates (calculated with a Galactic center distance of 8178 kpc; see Abuter et al. 2019). They are centered very strongly on the Galactic center, without any indication that they would be following the Galactic bar, as traced by the bulk

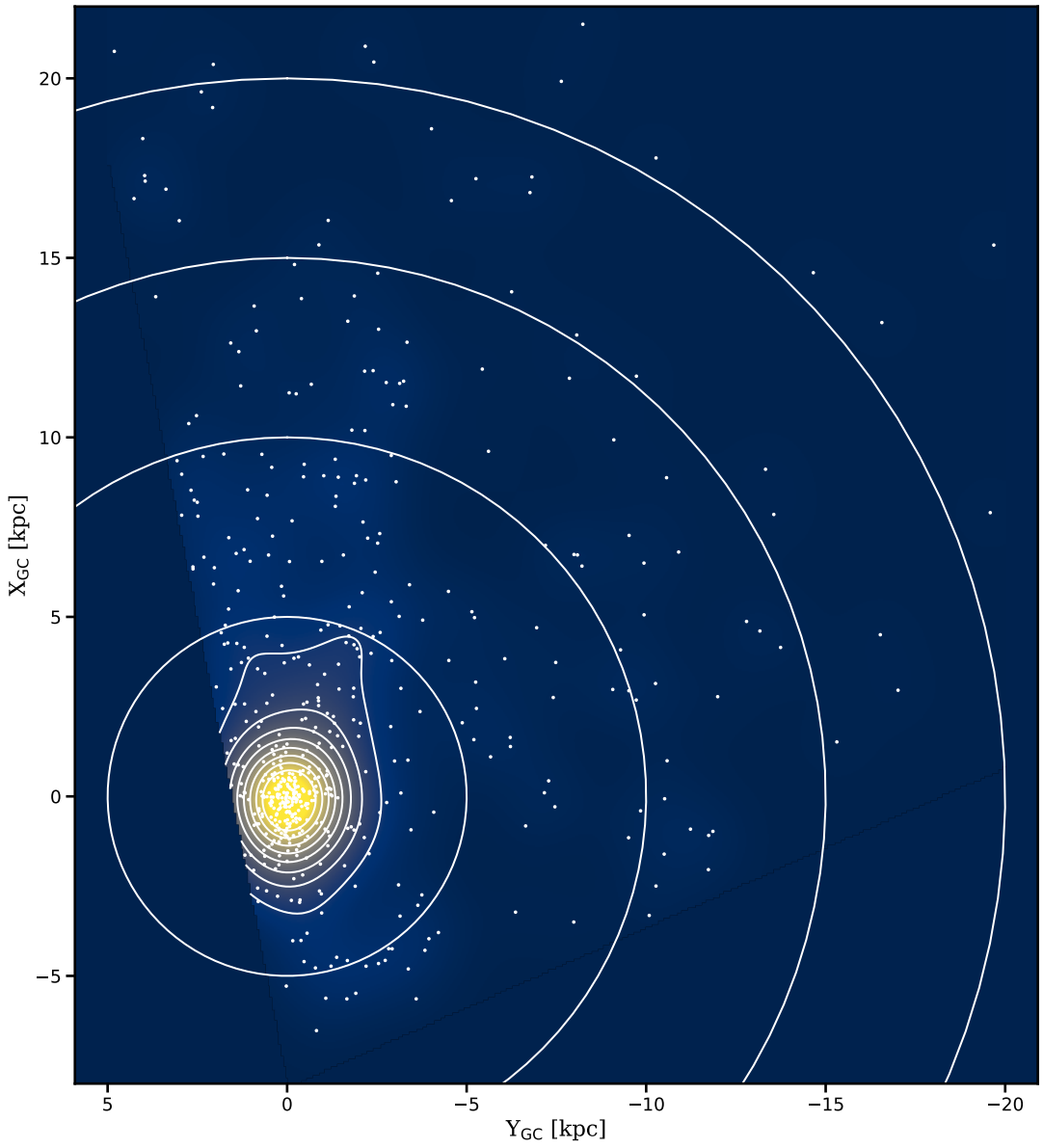

Figure 5.4: Distribution of VVV Type II Cepheid variables projected onto the Galactic plane in Galactocentric coordinates. Individual Cepheid stars are marked by white dots. The color scale and the (linearly scaling) contours represent a kernel density estimate of the Cepheid distribution with a Gaussian kernel size of 0.9 kpc. Concentric circles mark Galactocentric cylindrical distances at 5, 10, 15 and 20 kpc radii.

of the red clump stars within the volume of the bulge (see, e.g., Fig. 10 of Gonzalez et al. 2012). Rather, they follow the spherical distribution of RRL stars (Dékány et al. 2013), confirming that the oldest stellar populations constitute a kinematically hot component of the Galactic bulge volume. Nevertheless, these Type II Cepheids will require further (kinematic and spectroscopic) study in order to definitively establish whether the majority of them is connected to the metal poor ([Fe/H] ~ -0.4) component of the bulge red clump stars, which displays a similar, centrally concentrated distribution (Zoccali et al. 2018). Alternatively, they could be mostly assembled from old halo AGB stars crossing the volume of the bulge.

5.2 Distribution of the VVV classical Cepheid sample

The distances of the Classical Cepheids have been calculated similarly to those of Type II (Section 5.1.3), but by using Equations 5.4–5.6 for the calculation of absolute magnitudes and color excesses. The distribution of the Classical Cepheids in the Galactic plane (assuming the same Galactic center distance as for the Type II Cepheids) is shown in Figure 5.5. It is immediately apparent that the distribution of Classical Cepheids is very different from that of Type II Cepheids (Figure 5.4), as they are not concentrated towards the Galactic center. Furthermore, in contrast to the smooth distribution of RRL stars throughout the Galactic disk (Figure 3.3), Classical Cepheids display significant overdensities. Most of these can also be connected to the overdensities seen in the Galactic longitude distribution (Figure 5.1, top panel) of Classical Cepheids.

The top panel of Figure 5.6 compares the marginalized heliocentric distances of the VVV Classical Cepheids to their Galactic plane distances (Z). The color scale corresponds to the observed $E(H - K_S)$ color excesses. Similarly to the case of the VVV disk RRL variables (Figure 3.6), Classical Cepheids at larger distances from the Galactic plane can only be found at large heliocentric distances. While in the case of RRL variables this was mostly caused by the limited latitude coverage of the VVV disk fields, in the case of Classical Cepheids it is mostly because close by Cepheids at high Galactic latitudes are saturated in the VVV survey, and therefore missed by the variability search. Furthermore, the extinction limits the detection of Classical Cepheids located close to the Galactic plane at projected heliocentric distances $d_{\rm h,proj} \gtrsim 16\,$kpc. There are 62 Classical Cepheids in the VVV sample with neither J nor H-band measurements, with half of them located closer than 0.5° to the Galactic plane. Photometric follow-up is needed to obtain their colors, and hence the distances of

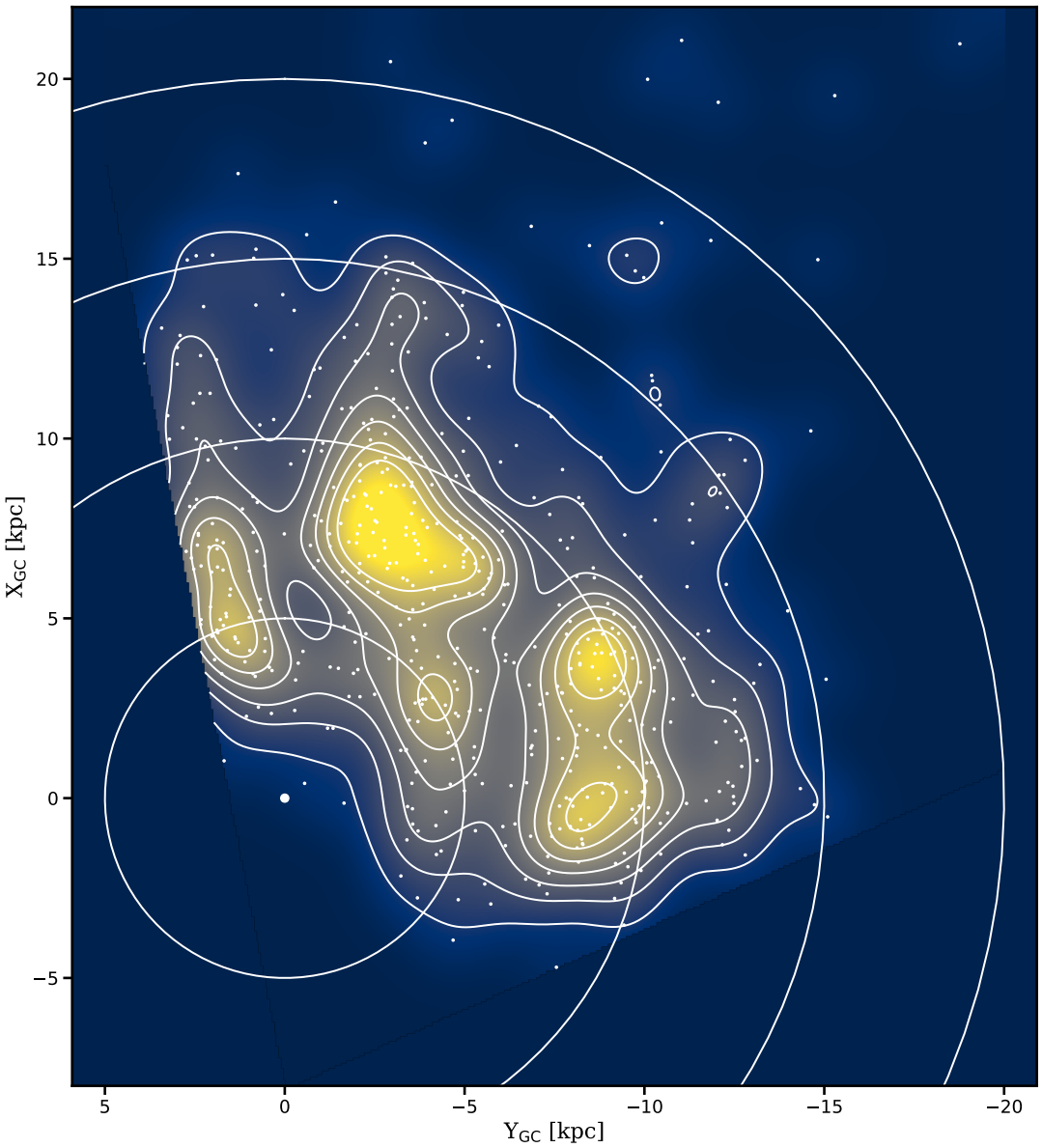

Figure 5.5: As in Figure 5.4, but for the VVV Classical Cepheids.

these variables, hopefully increasing the Galactic plane region spanned by the VVV Classical Cepheid sample.

Using optical observations of much of the southern Galactic disk, the OGLE survey (Udalski et al. 2015) has announced the discovery of hundreds of Classical Cepheids in these regions (Udalski et al. 2018). However, at the time of writing, the catalog of OGLE Galactic Cepheids is still unavailable, preventing the comparison of the two samples. In the future, by combining them, we will have an almost complete view of the Classical Cepheid population of the southern Galactic disk up to a heliocentric distance of $\sim 16\,\mathrm{kpc}$.

The middle panel of Figure 5.6 shows the distances of the VVV Classical Cepheids from the Galactic center. There are very few stars located closer than 2 kpc to the Galactic center, lending support to the suggestion of Matsunaga et al. (2016) that the Galactic bulge did not have recent star formation episodes. Furthermore, at distances $d_{\mathrm{G,proj}} > 6\,\mathrm{kpc}$, the average Galactic plane distance of the VVV Classical Cepheids increases, as they follow the flaring Galactic disk (Li et al. 2019). Feast et al. (2014) had found old, short-period ($P < 4\,\mathrm{d}$) Classical Cepheids at distances of $\sim 2\,\mathrm{kpc}$ from the Galactic plane by following up on the OGLE Cepheid candidates found in the direction of the Galactic bulge. The current search does not cover either the period range or the Galactic latitudes ($|b| > 2.3°$) where these variables can be found, but the flaring of the Classical Cepheids in the disk is unquestionably detected.

The bottom panel of Figure 5.6 shows a modified version of the middle panel, where the color scale corresponds to the heliocentric distances of the Classical Cepheid sample. There is a trend that Cepheids closer to the Galactic center are also found closer to us, due to the restricted Galactic latitude coverage of the VVV survey (also see the covered Galactic region in Figure 5.5), which is combined by the fact that due to saturation, all Cepheids below $d_{\mathrm{h}} \lesssim 6\,\mathrm{kpc}$ will go undetected in the VISTA observations.

Similarly to the RRL variables in the VVV disk region (Section 3.2.1), the distribution of VVV Galactic Cepheids is biased by selection effects. At low Galactic latitudes, high extinction regions decrease the magnitudes of variables below the saturation limit of VISTA. However, Cepheids found at higher Galactic latitude, low extinction regions will still be saturated out to much greater distances. At these distances, in return, Cepheids found close to the Galactic plane will go undetected (at least in the JH bands), due to the combined effect of distance and large amounts of extinction. Although these biases prevent a complete characterization of the population of Classical Cepheids found in the southern part of the far

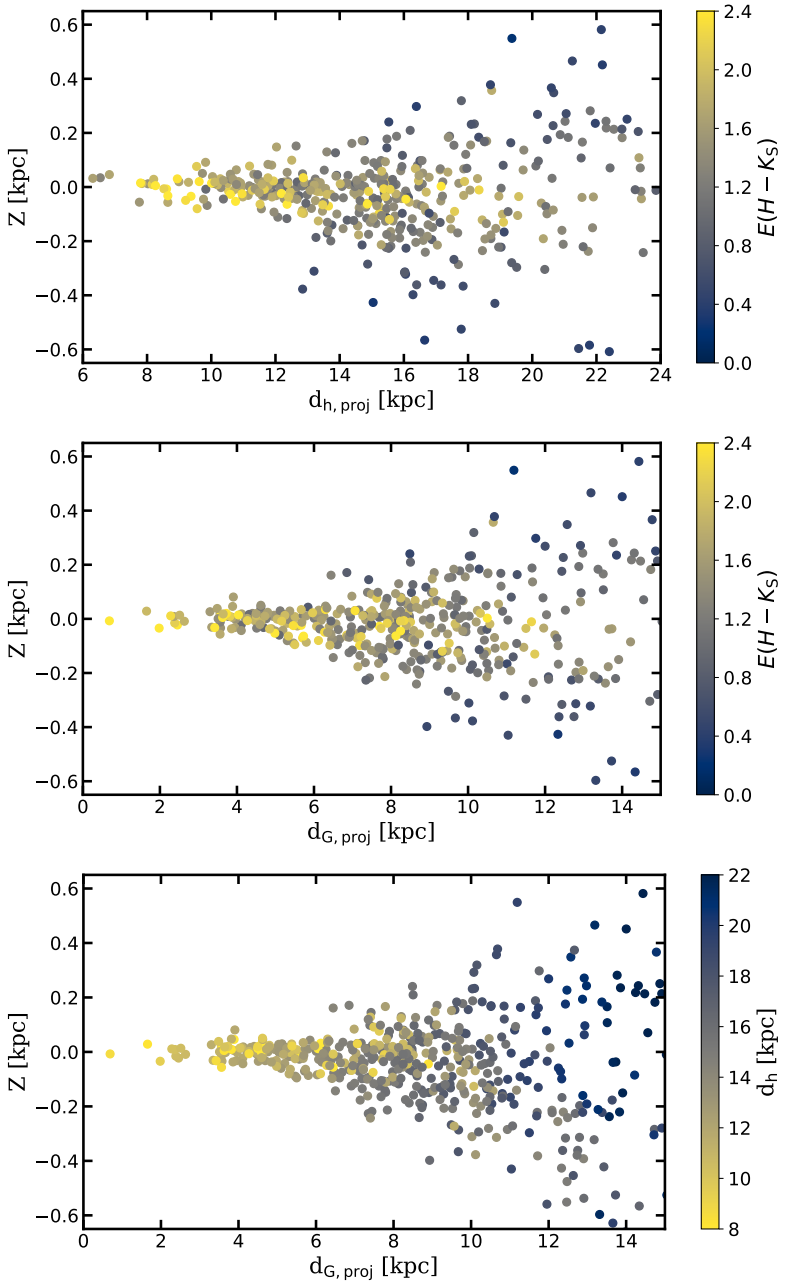

Figure 5.6: *Top:* Galactic plane distance distribution of VVV Classical Cepheid variables as a function of heliocentric distances projected onto the Galactic plane. The color scale represents the $E(H-K_S)$ color excesses. *Middle:* Same as the top panel, but as a function of the projected Galactocentric distance. *Bottom:* Same as the middle panel, but the color scale represents the calculated heliocentric distances.

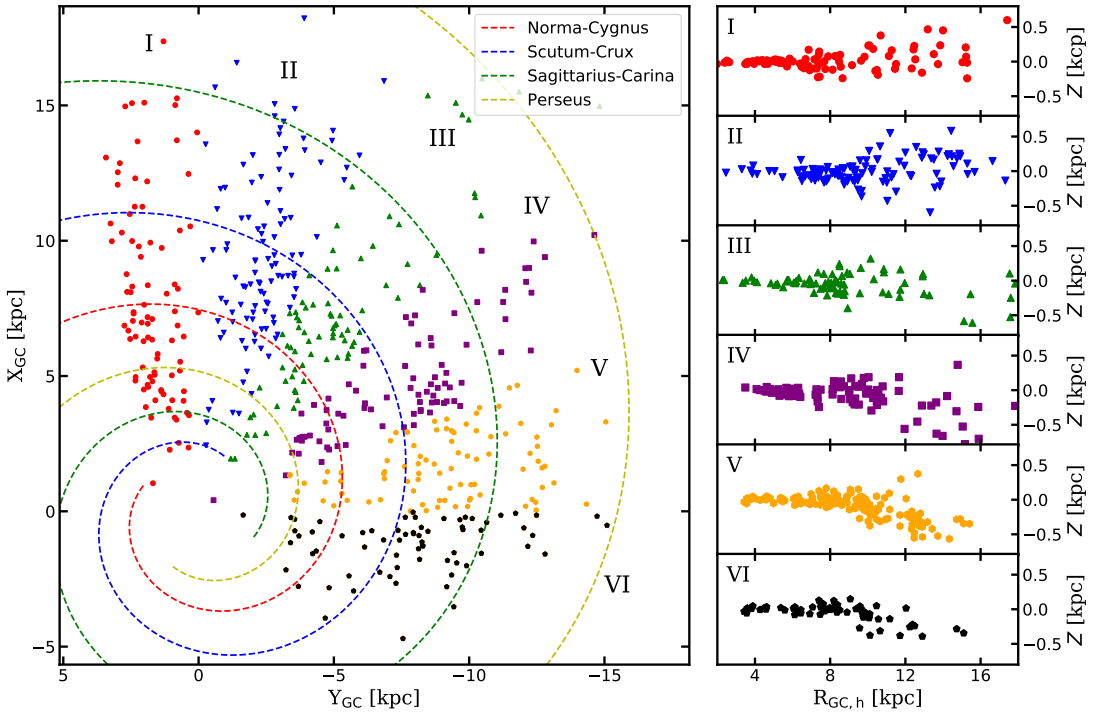

Figure 5.7: *Left:* The distribution of the VVV Classical Cepheid sample in Galactocentric coordinates. Six samples have been defined by Galactocentric longitude, marked by different symbols and colors, as well as by Roman numerals. *Right:* The distances of the Cepheids from the Galactic plane as a function of their Galactocentric distances. The separate panels correspond to the subsamples shown in the left panel. The spiral arms with a pitch angle of −13.1° are reproduced according to their parameters given by Vallée (2017).

Galactic disk, they are still valuable objects for the exploration of the structure of the Milky Way disk. In the future, radial velocity and proper-motion information obtained for them will give us kinematic information about this region, which is otherwise beyond the reach of the *Gaia* satellite, due to the high levels of extinction.

The outer part of the Galactic disk has been well known not to follow the flat shape of the inner part, instead warping upward in the north and downward in the south, as first detected in the 21 cm molecular hydrogen line (Kerr 1957; Oort et al. 1958). Later, this feature was also detected in the stellar component (see, e.g., Reylé et al. 2009; Li et al. 2019, and references therein). Warping is also a common feature of disk galaxies (see, e.g., Reshetnikov et al. 2016). The Galactic warp in the distribution of Classical Cepheids was detected by Skowron et al. (2018), as well as by Chen et al. (2019). Figure 5.7 illustrates the warp in the southern Galactic disk, as traced by the VVV Classical Cepheid sample. The complete sample of variables is

split into 6, roughly equal sized subsamples according to their Galactic polar coordinate ϕ. The right panels show the distances of the Classical Cepheids in each of these subsamples as a function of the projected Galactocentric distances ($R_{GC,h}$). Subsamples IV-VI clearly show the Galactic warp in the southern disk at $R_{GC,h} > 10\,\mathrm{kpc}$, as the majority of Cepheids are found below the Galactic plane, in accordance with the results of Skowron et al. (2018) and Chen et al. (2019). In contrast, subsamples I-III, found in the general direction of the Galactic bulge, show a strongly flared outer disk, in agreement with the detection of highly flared Classical Cepheids by Feast et al. (2014).

5.2.1 Apparent Cepheid pairs

In a pilot study, preceding the current systematic Cepheid search, a twin pair of Cepheids was found in the VVV bulge area by their unusual properties (Dékány et al. 2015a): very close position on the sky (18.″3 difference), almost exactly the same period ($\sim 11.2\,\mathrm{d}$), magnitude ($K_S \sim 12.7\,\mathrm{mag}$) and color excess ($E(H-K_S) \sim 2$). This has led to the conclusion that these two Cepheids are probably still bound to the same young open cluster, where they must have been born. The large amount of extinction makes the unevolved, main sequence stars unobservable, while the evolved stars, such as the detected Classical Cepheids, are bright enough in the near-IR bands to be observed by VISTA, despite the reddening.

Due to the possibility of calibrating the Cepheid PL relation using the distances of their host clusters, Classical Cepheids have been extensively searched for in Galactic open clusters. However, Classical Cepheids seem to be uncommon in Galactic open clusters, with only one cluster known to simultaneously host three Classical Cepheids (NGC 7790; Anderson et al. 2013). In contrast, the massive young clusters of the Magellanic Clouds, due to their higher stellar contents, can have much higher Cepheid numbers (e.g., NGC 1866 has 23 Classical Cepheids; see Welch et al. 1991, as well as Lemasle et al. 2017).

The VVV Classical Cepheid catalog has been searched for apparent Cepheid pairs, which might be indicative of their common origins, namely that they still reside in the open cluster cluster where they were born, or that their parent cluster has dissolved very recently. The candidate pairs are required to be located closer than 5' to each other on the sky. Furthermore, their heliocentric distances are required to agree within their combined $2\,\sigma$ uncertainties. Table 5.1 shows the basic properties of the candidate Classical Cepheid pairs. The candidate *bulge1* is the Cepheid pair found in the invisible open cluster published in

Table 5.1: Basic parameters of candidate Classical Cepheid pairs in the VVV sample. For each Cepheid, the pulsation period, celestial coordinates, as well as the calculated distance is given.

	Cepheid 1			Cepheid 2		
Pair name	Period [d]	Coordinate	distance [kpc]	Period [d]	Coordinate	distance [kpc]
bulge1[a]	11.23424	18:01:24.49 -22:54:44.6	13.3 ± 0.2	11.21709	18:01:25.08 -22:54:28.3	13.4 ± 0.1
bulge2[b]	7.46200	18:04:49.69 -21:19:21.3	16.5 ± 0.1	7.80276	18:04:53.37 -21:16:12.0	15.7 ± 0.2
bulge3[b]	4.40730	17:23:17.55 -35:02:43.2	22.6 ± 0.2	4.36703	17:23:15.93 -35:01:25.4	22.3 ± 0.2
disk1[c]	6.93387	15:46:43.29 -54:28:00.8	13.3 ± 0.1	8.01823	15:46:48.13 -54:28:23.9	12.9 ± 0.1
disk2[c]	14.56058	16:37:28.51 -46:56:58.8	9.3 ± 0.2	13.03596	16:37:33.31 -46:54:29.8	9.6 ± 0.2
disk3[d]	10.47008	16:40:30.82 -46:27:53.9	15.4 ± 0.4	14.58135	16:40:35.24 -46:30:46.3	14.9 ± 0.3
disk4[d]	12.88066	16:48:04.94 -44:48:59.7	10.1 ± 0.2	5.43532	16:48:19.64 -44:48:36.4	10.1 ± 0.1
disk5[c]	10.82043	16:50:47.36 -44:27:15.5	11.5 ± 0.1	8.33871	16:51:10.85 -44:27:00.9	11.5 ± 0.2

[a] The Twin Cepheids published in Dékány et al. (2015a)

[b] Possible pairs, if both Cepheids are in the same IS crossing.

[c] Possible pairs, if the shorter-period Cepheid is in the second IS crossing, while its pairs is in the third.

[d] The periods are too different to be physically associated.

Dékány et al. (2015a). The other two candidate pairs in the VVV bulge fields have Cepheids with similar periods. Calculating their ages with the period-age relations published for second crossing Cepheids by Anderson et al. (2016), and assuming a metallicity $Z = 0.014$, the average ages are $\sim$ 82 Myr and $\sim$ 113 Myr for pairs *bulge2* and *bulge3*, respectively. The Classical Cepheid pairs *disk3* and *disk4* have period differences between them that cannot be explained by stellar evolution, not even when assuming that the shorter-period Cepheid is in the second, while its longer-period pair is in the third, crossing of the IS. However, by assuming this for the pairs *disk1*, *disk2*, *disk4*, average ages of 88 Myr and 94 Myr, 63 Myr and 64 Myr, and 77 Myr and 80 Myr, respectively, are obtained, making this scenario plausible for all three pairs. The VVV Infrared Astrometric Catalogue (VIRAC; Smith et al. 2018) has been searched for the proper motions of the remaining plausible Cepheid pairs. Unfortunately, due to the large distances of the stars, coupled with the fact that VIRAC relies upon the astrometry of pawprint observations given for the aperture photometry (resulting in relatively larger errors), the proper motion differences are currently not sufficient to confirm or to refute the physical association of either of the remaining Classical Cepheid pair candidates.

In the future, as proper motion estimates based on PSF photometry for the whole VVV area becomes available (see, e.g., Contreras Ramos et al. 2017 and Gran et al. 2019), the

proper motions, alongside radial velocity measurements, will be made use of to determine whether these Classical Cepheid pairs are really physically associated.

5.3 The radial age trend of VVV Classical Cepheids

Leveraging the large numbers of newly discovered OGLE Cepheid variables, Skowron et al. (2018) studied the large-scale properties of the Classical Cepheids in the near-side of the Galactic disk. They have found that many Cepheids were formed in three, spatially and temporally well-separated star formation episodes, which have taken place approximately 35, 68 and 91 million years ago (see Figure 3 of Skowron et al. 2018). These episodes are separated radially as well, with the latest one being the one closest to the Galactic center, and the oldest one being the one farthest away from it. They also noted that the positions of the Cepheids do not coincide with the positions of the spiral arms, due to the migration of newly born Cepheid progenitor stars, before they become Classical Cepheids. Furthermore, they modeled the current positions of the identified overdensities with a simple model (see their Figure 4).

In order to compare the results obtained by Skowron et al. (2018) for the near-disk Cepheids to the VVV Classical Cepheid sample, the individual ages of the stars must be estimated. The calculated ages of Cepheid variables depend on the number of the IS crossing they are currently on (Anderson et al. 2016). The period-change rates of Cepheids give information on this: variables are crossing the IS the second time evolving towards the blue, resulting in decreasing periods, after which, in the redward evolution taking place during the third IS crossing, their periods are increasing (see, e.g. Fadeyev 2013, as well as references therein). However, as the Cepheids here are new discoveries, no information is available about their period-change rates, and hence on whether they are second or third crossing Cepheids. Therefore, in order to estimate their individual ages, as well as to be able to compare the results directly to those of Skowron et al. (2018), the crossing-averaged period-age relation for fundamental-mode Classical Cepheids, published by Bono et al. (2005), are adopted in the form:

$$\log t = -0.67 \cdot \log P + 8.31 \qquad (Z = 0.020), \tag{5.15}$$

$$\log t = -0.78 \cdot \log P + 8.41 \qquad (Z = 0.010), \tag{5.16}$$

$$\log t = -0.79 \cdot \log P + 8.49 \qquad (Z = 0.004), \tag{5.17}$$

where t is the age in years and P is the pulsation period in days.

There are currently no spectroscopic iron abundance measurements for any of the newly discovered Classical Cepheids, although in the future, low resolution near-IR spectra might be used to obtain them, following the pioneering work of Inno et al. (2019). Therefore, having no other option, the iron abundances of individual Cepheids are estimated using the young-disk metallicity gradient determined using Cepheid variables by Genovali et al. (2014):

$$[\mathrm{Fe/H}] = 0.47 \ - 0.051\, R_{\mathrm{GC}} \ (\mathrm{in\,kpc}), \tag{5.18}$$

with the slight modification that the lowest metallicity each Cepheid is allowed to have is restricted to [Fe/H] = −0.4, the average metallicity of young stellar populations in the outer parts of the Galactic disk (see, e.g. Genovali et al. 2014, as well as references therein). The absolute heavy element content Z, utilized in Equations 5.15–5.17, is calculated as:

$$\log Z = [\mathrm{Fe/H}] - 1.77, \tag{5.19}$$

a relation which was obtained by linear regression on a grid of [Fe/H] versus $\log Z$ values, obtained assuming a helium content $Y = 0.245$, no alpha-element enhancement ($[\alpha/\mathrm{Fe}] = 0.0$), and standard solar chemical composition (Grevesse and Sauval 1998).

The individual ages of the VVV Classical Cepheids were then estimated by first using Equation 5.18 to estimate the (typical) metallicity at the Galactocentric radius of the Cepheid. This metallicity was then converted to absolute heavy element content Z using Equation 5.19; the age estimates proper were then calculated using Equations 5.15–5.17. For $Z < 0.020$, the ages are interpolated linearly between the two relations closest in Z, and extrapolated in the same way using Equations 5.15 and 5.16 for stars with $Z > 0.020$.

The distribution of estimated Classical Cepheids ages in the Galactic disk is shown in Figure 5.8. Although there is a hint that the younger Cepheids are preferentially located

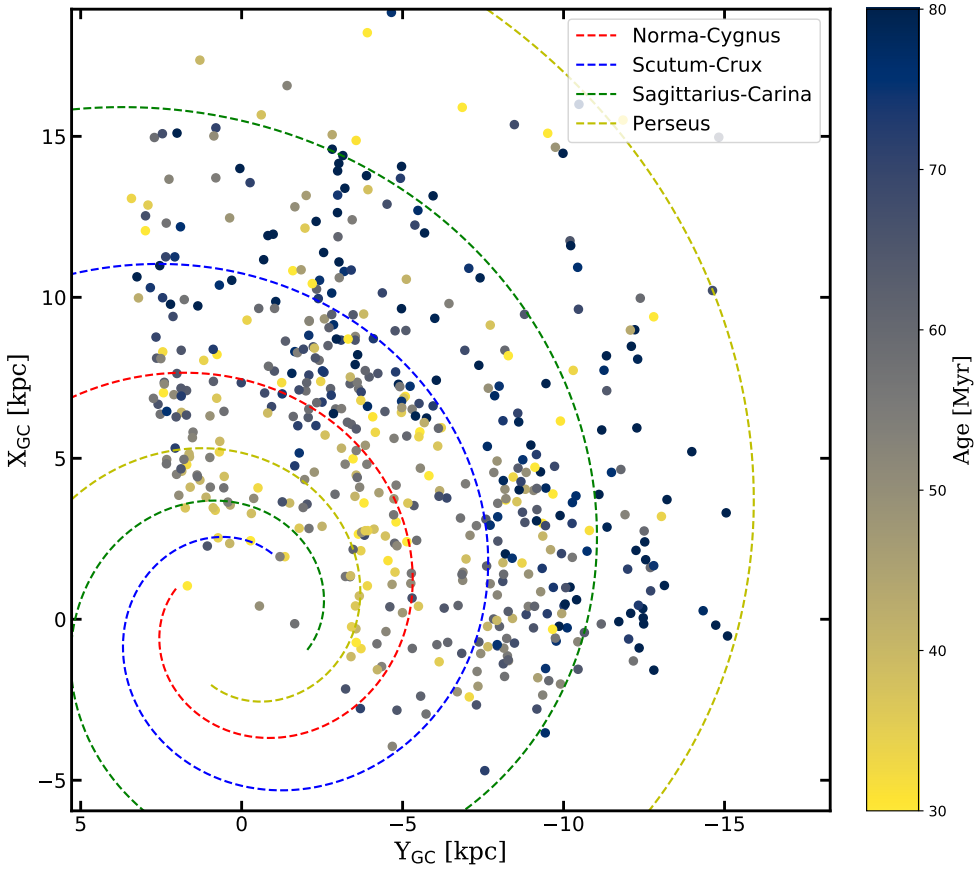

Figure 5.8: The age distribution of VVV Classical Cepheids across the disk in Galactocentric coordinates. The color scale corresponds to the estimated ages of individual Cepheids. The spiral arms with a pitch angle of −13.1° are reproduced according to their parameters given by Vallée (2017).

closer to the Galactic center, when looking at the age distribution as a function of the Galactocentric radius in the left panel of Figure 5.9 the trend becomes clear: similarly to the radial distribution of Cepheid overdensities found by Skowron et al. (2018) in the near-side of the Galactic disk, young Classical Cepheids are indeed located preferentially closer to the Galactic center. The cut in the Cepheids ages at ∼ 90 Myr is caused by the lower limit in the period search of 3.8 days: older Classical Cepheids generally have shorter periods than this limit. The right-hand panels of Figure 5.9 show the kernel density estimates of 6 kpc wide Galactocentric radius bins of the age distribution of the VVV Classical Cepheid sample, further illustrating the small Galactocentric radius preference of younger Classical Cepheids. Finally, it has to be noted that this trend is robust against both the assumptions used to estimate the individual iron abundances, as well as against the choice of the exact Cepheid

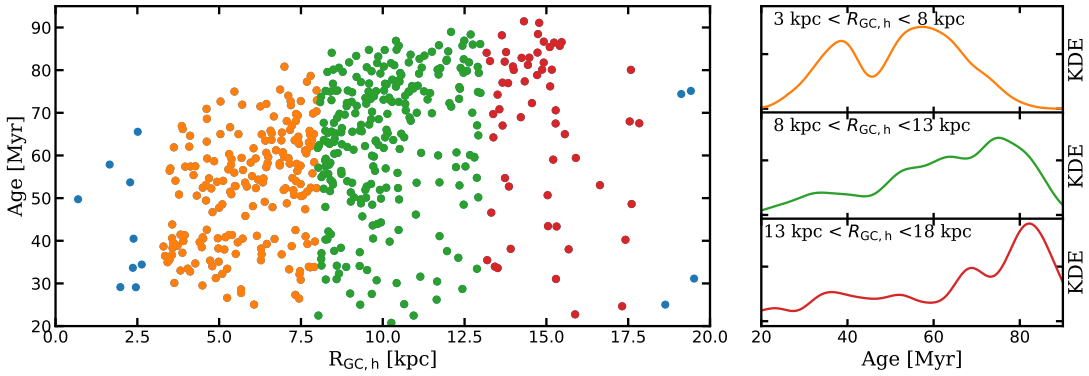

Figure 5.9: *Left:* The calculated ages (Bono et al. 2005) of individual Classical Cepheids as a function of the projected Galactocentric radius. *Right:* Kernel density estimates of the age distributions of Classical Cepheids in three Galactocentric radius subsamples, calculated with a Gaussian kernel of 3 Myr size.

period-age relation. That is, utilizing the relations of Anderson et al. (2016) instead of those of Bono et al. (2005) would result in $\sim$ 50% larger estimated ages, but the age gradient would not disappear.

There is an apparent lack of Cepheids with ages of $\sim$ 45 Myr at all Galactocentric radii, as seen in the left panel of Figure 5.9. According to the period-age relations of Bono et al. (2005), this age corresponds to Cepheids with pulsation periods of $\sim$ 10 d. The distribution of the logarithm of the period of both the Classical as well as the Type II Cepheids in the VVV sample is shown in the bottom panel of Figure 5.2. There is an apparent lack of Classical Cepheids with $\log P \sim 1$, when compared to their numbers at $\log P \sim 0.85$ and $\log P \sim 1.15$, while there is an excess of Type II Cepheids at this period. This might be the effect of a decreased efficiency of the Cepheid classifier described in Section 5.1.1 at this period: the light-curve shapes of Cepheids changes dramatically with the increase in period due to the Hertzsprung progression (Hertzsprung 1926; also see Figure 7 of Soszyński et al. 2010), caused by a the resonance between the fundamental and second overtone modes (Bono et al. 2000; Gastine and Dintrans 2008). This resonance is also metallicity dependent in the sense that at lower metallicities, the Hertzsprung progression happens at shorter periods. Therefore, including lower-metallicity Small Magellanic Cloud Cepheids in the training sample, together with the generally small number of available K_S-band light curves for these objects, might have caused the performance of the classifier to decrease in this period region.

5.4 The VVV Classical Cepheids and the spiral arms

A simple experiment was performed in order to connect individual VVV Classical Cepheids to their original regions of birth, that is, for most of them, the spiral arms, by circularly rotating their current positions, similarly to the work of Dias and Lépine (2005) performed on open clusters in the solar neighborhood. The rotation is performed by using the difference of the Galactic rotation curve and the spiral pattern speed at the radii of individual Cepheids variables, multiplied by their estimated ages (Section 5.3).

In order to perform this rotation , the Galactic rotation curve presented as model B by Mróz et al. (2019) was adopted. The parameters of this curve have been determined using Classical Cepheid variables with *Gaia* proper motions and radial velocities, following a parametric description of the circular velocity curve (Persic et al. 1996). The period of the spiral pattern of the Milky Way has a considerable scatter in the literature (see, e.g., Gerhard 2011, as well as references therein). Therefore, the experiment was performed with different assumed values. Figures 5.10-5.12 show the results of this experiment, compared to the four-arm spiral structure model of Vallée (2017). Some of the maxima in the distribution of the rotated positions of the VVV Cepheids seem to agree well with the (extrapolated) location of the spiral arms, although there are some prominent features that always fall inbetween them. For example, looking at Figure 5.10, there are two overdensities at $Y_{\rm GC}$, $X_{\rm GC} \sim -5\,{\rm kpc}, 7\,{\rm kpc}$, as well as at $Y_{\rm GC}$, $X_{\rm GC} \sim -1\,{\rm kpc}, 6\,{\rm kpc}$, tentatively associated with the Scutum-Crux and Norma-Cygnus arms, respectively. However, these two overdensities are also connected by Cepheids located between the two arms. The differences between the locations of the maxima of the backwards-rotated Classical Cepheids, as compared to the spiral-arm model, may have many contributing factors:

- The real ages of individual Classical Cepheids could be smaller or larger, depending on the number of the IS crossing each Cepheid is currently completing.

- The supposed circular rotation is a necessary simplification in the absence of kinematic data for the individual stars.

- The period-age relation might have larger than anticipated scatter, if there is a spread in the initial rotation speeds of the progenitor main sequence stars of Classical Cepheids (Anderson et al. 2016).

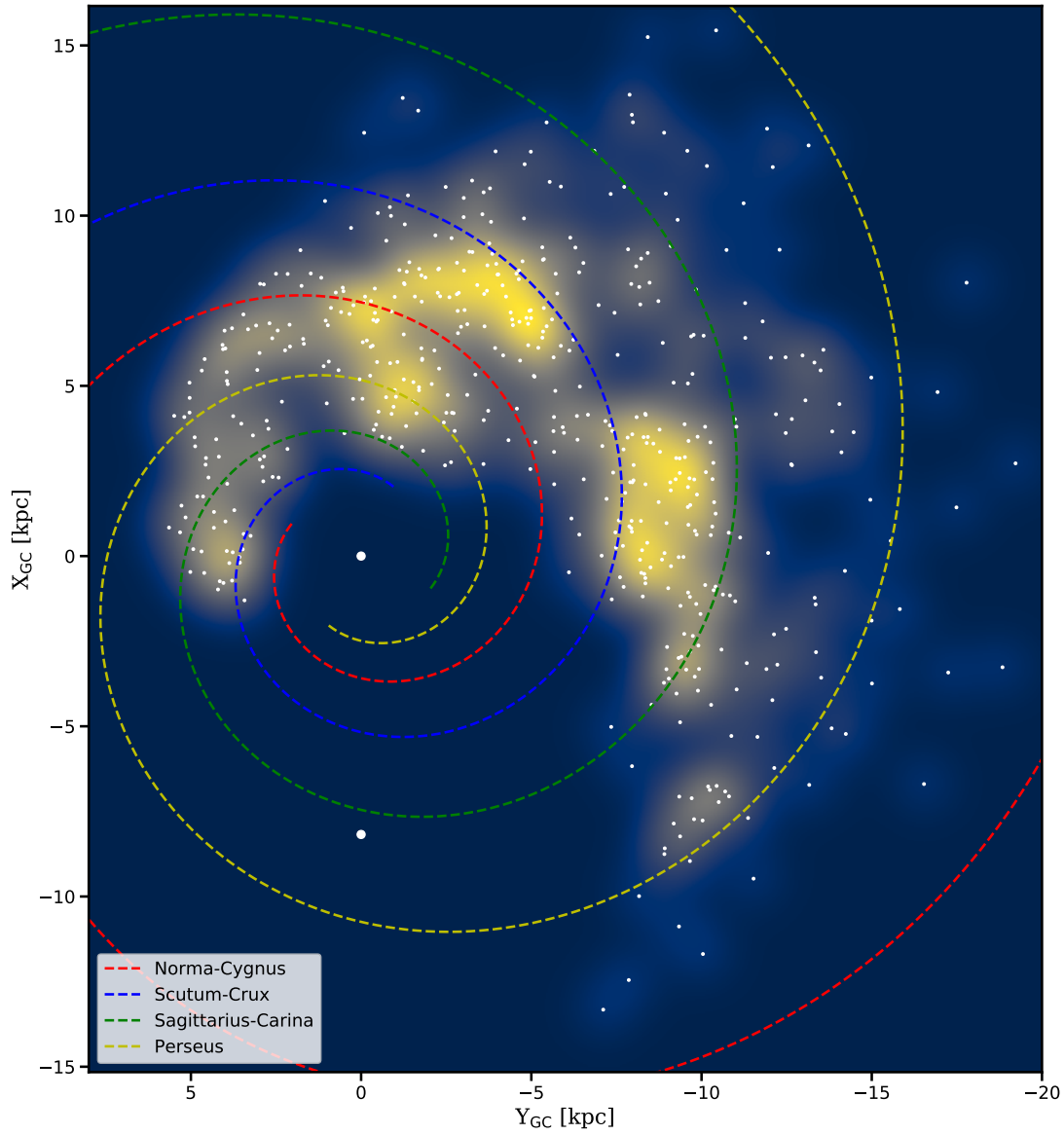

Figure 5.10: Distribution of VVV Classical Cepheid variables in Galactic coordinates, circularly rotated to their supposed regions of birth. The rotation was done with an assumed spiral pattern period of 230 Myr. Individual Cepheid stars are marked by white dots. The color scale represents a kernel density estimate of the Cepheid distribution with a Gaussian kernel size of 0.75 kpc. Large white dots mark the positions of the Galactic center and the Sun. The spiral pattern shown follows the description of the four-arm Milky Way model of Vallée (2017).

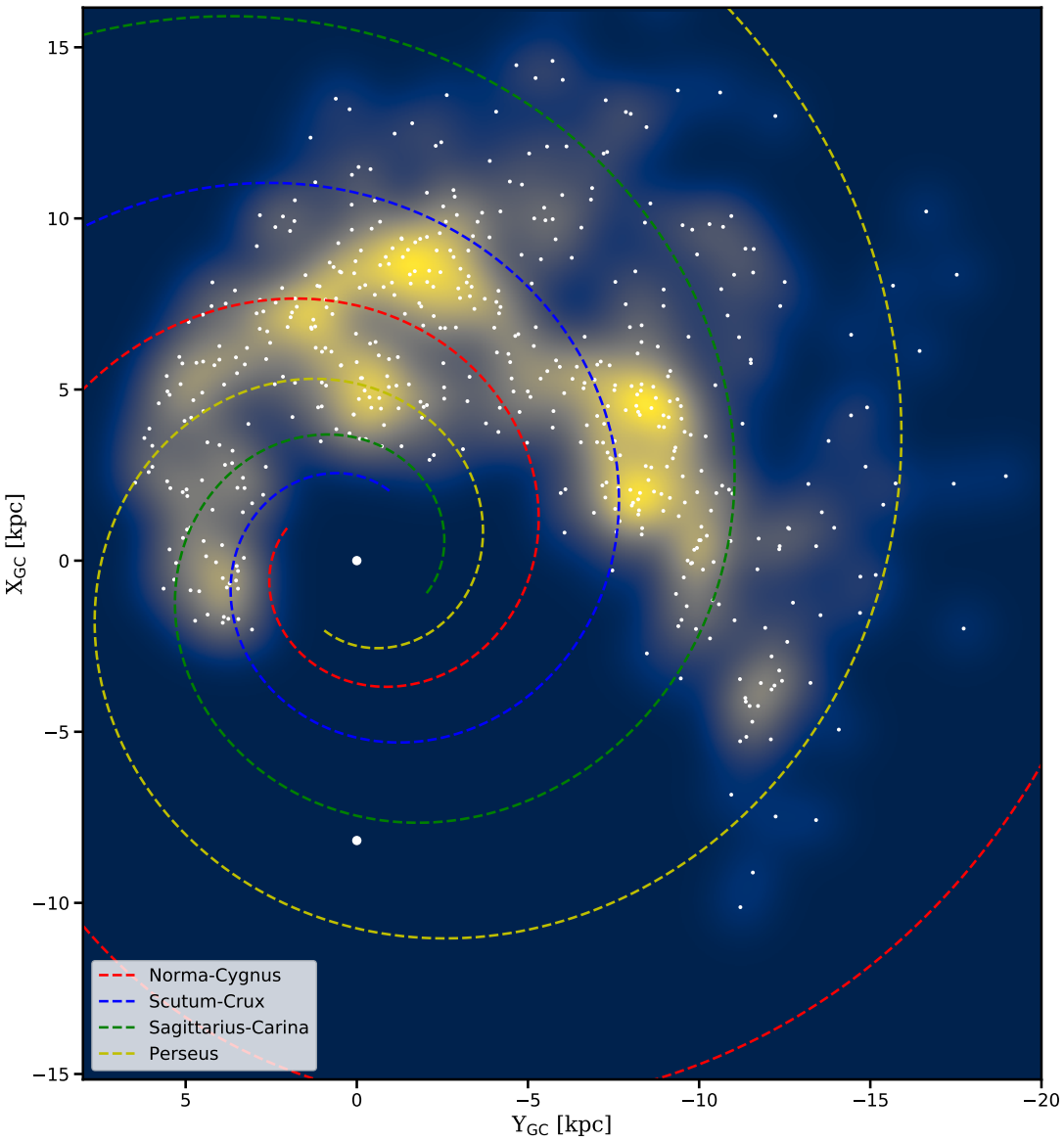

Figure 5.11: Same as Figure 5.10, but assuming a spiral pattern period of 270 Myr.

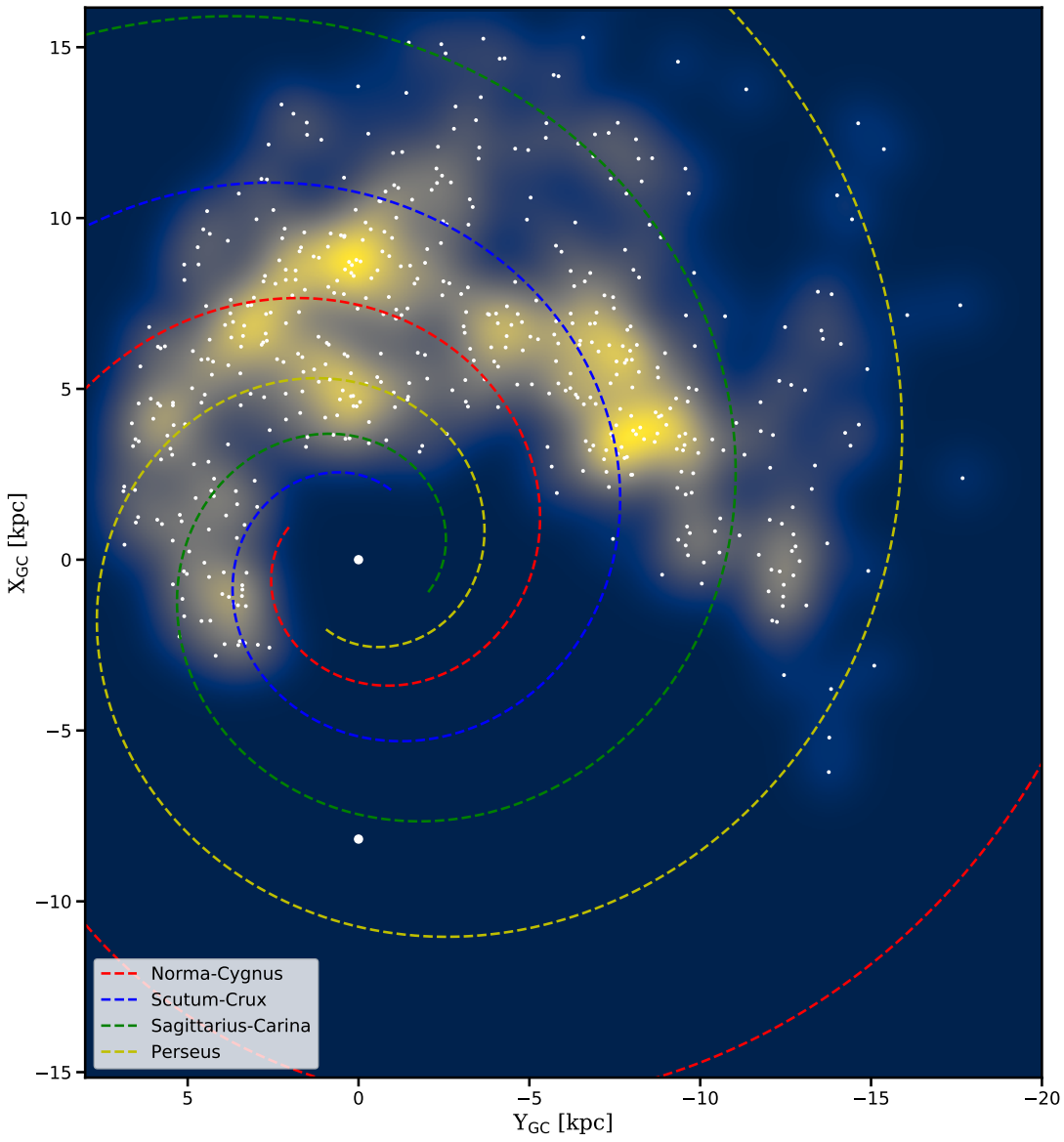

Figure 5.12: Same as Figure 5.10, but assuming a spiral pattern period of 320 Myr.

- The estimated distances might be systematically biased, if there is a small change in the extinction law as a function of Galactic latitude (see, e.g. Figure 17 of Schlafly et al. 2016), propagating into the Galactocentric radii of the Cepheids.
- The pattern speed of spiral arms might be uneven or even resemble ring-like structures in the outer regions of the Milky Way (Mel'nik et al. 2015).
- There could be large inter-arm features or local arms connecting major spiral arms, similarly to the Orion arm (Vázquez et al. 2008) in the Solar neighborhood.

5.5 Summary and future outlook

The recent development of a CNN-based K_S-band Cepheid light-curve classifier (Section 5.1.1; Dékány et al. in prep.) has enabled the discovery of over a thousand Classical and Type II Cepheids in the region of the Galactic plane covered by the VVV survey. By assuming a centrally concentrated distribution around the Galactic center, the large number of bulge Type II Cepheids has allowed the derivation of an extinction law (Section 5.1.3). This was then applied to calculate the distances towards the Classical Cepheids in the far side of the Galactic disk (Section 5.2), revealing a very uneven spatial distribution across the Galactic disc (Figure 5.5), not unlike that of the Cepheids on the near-side of the Milky Way (see, e.g., Figure 7 of Majaess et al. 2009, as well as Figure 1 of Skowron et al. 2018). Furthermore, the Galactic-plane distance distribution of VVV Classical Cepheids at large Galactocentric radii follows the Galactic warp, in agreement with the results of both Skowron et al. (2018) and Chen et al. (2019). The radial distribution of Cepheid ages (Figure 5.9) shows, that on average, Classical Cepheids in the central regions of the Milky Way are younger than in the outer ones, also in agreement with the results of Skowron et al. (2018), but now extended to the far-side of the Milky Way.

A simple experiment was undertaken to try to connect the discovered Classical Cepheids to their original birth regions, by rotating their current Galactic position with the difference between the rotation curve of the Milky Way and the rotation speed of the spiral arms (supposing different values of the spiral arm pattern speed), multiplied with the ages of the Classical Cepheids (Section 5.4). Although some of the Cepheid overdensities seemingly lie upon the spiral arm structure model of Vallée (2017), others fall in inter-arm regions. There are many possible reasons for this, including uncertainties of Cepheid distances and ages,

as well as uncertainties in the spiral arm parameters and pattern speeds. In the future, when the Galactic near-side Cepheid catalogs of Udalski et al. (2018) and Skowron et al. (2018) become available, this experiment will be repeated including all the known Classical Cepheids in the Milky Way, allowing the calibration of some of the uncertainties (i.e., the Cepheid age scale and the spiral pattern speed of the Milky Way, by matching the rotated positions of the Classical Cepheids to the present positions of nearby spiral arms).

The newly discovered Classical Cepheids will provide a unique contribution to our understanding of the Milky Way. They lie in regions inaccessible to the *Gaia* satellite, therefore radial velocity observations from near-IR spectroscopy, combined with spectroscopic metallicity estimates (Inno et al. 2019) and VVV proper motions (Gran et al. 2019), will allow us to gain an otherwise unattainable insight into the structure, kinematics, and recent star-formation history of the far side of the Galactic disk.

6

Conclusions

In this thesis, I presented a summary of the work I developed during my PhD studies for the characterization of the RR Lyrae and Cepheid variables in the VVV survey.

Chapter 2 presented my investigation into the light-curve shapes of RRL variables in the near-infrared: PCA was performed on the folded K_S-band light curves of a set of well-observed RRL variables (Section 2.2). In contrast to previous light-curve shape studies, the light curves were aligned by the light-curve minimum, not the maximum, as that feature is not sharply defined in the K_S-band light curves of RRL stars. The PCA resulted in a compact description of the light-curve shapes, as only four PC amplitudes are sufficient to describe the light-curve shapes with high accuracy. Furthermore, I also showed that the PC amplitudes can be used to predict the light-curve shapes of RRL variables in the J-band with an precision of 0.02 mag (Section 2.2.4), meaning that highly accurate mean magnitudes can be determined for these stars, even if only one J-band light-curve point is available. Using these results, I developed a robust light-curve fitting method to accurately fit the K_S-band light curves of RRL stars in the VVV survey, while simultaneously providing accurate J and H-band mean magnitude estimates (Section 2.3).

As the PCs provide a very compact description of the light-curve shapes of RRL stars in the near-IR, I have investigated whether they can be used to estimate the metallicities of the variables, similarly to how it is done using their Fourier parameters in the optical (Section 2.4). To do so, I used the RRL stars common between the VVV and OGLE surveys towards the Galactic bulge. I fitted the VVV K_S-band light curves with my method to determine their PCs (Section 2.4.1). After calculating the photometric metallicities of the OGLE I-band light curves, I discovered that all of the formulae used so far for this purpose possess biases (Section 2.4.2). I have used the least biased formula, and manually corrected for the remaining bias in the metallicities. Then, I used the K_S-band PC amplitudes and these metallicity estimates to develop a neural network-based algorithm for the estimation of RRL metallicities (Section 2.4.3).

Chapter 3 presents our RRL search in the VVV disk fields (Section 3.1). The machine-learned classifier of Elorrieta et al. (2016) was used to distinguish real RRL variables from other classes of variables (Sectiom 3.1.3). I fit the light-curves of RRL stars with the method developed in Chapter 2, while simultaneously estimating their K_S-band photometric metallicities (Section 3.1.4). Unfortunately, the spatial distribution of this sample is biased by the distribution of extinction of the Galactic disk, as well as the relatively small sky area covered by the VVV disk fields (Section 3.2.1). Nevertheless, the metallicity distribution of the RRL variables in the Galactic disk revealed interesting spatial variations (Section 3.2.3), lending tentative support to the inside-out formation scenario of the thick disk (Brook et al. 2012; Bovy et al. 2012).

In Chapter 4, I investigated the photometric inconsistencies shown by some of the RR Lyrae light curves found in the disk area of the VVV survey in Chapter 3. This revision led to the discovery of a major bias in the photometric zero-point calibration performed by CASU during the calibration of the VISTA photometry, which has a debilitating effect on the quality of the photometry in dense stellar fields. I developed a method to correct for this bias, and applied it to the complete JHK_S observation set of the VVV survey. These corrections have led to a substantial increase in the photometric quality of the light curves.

Chapter 5 includes the search for Cepheid variables (both Classical and Type II) in the Galactic disk area covered by the disk and bulge fields of the VVV survey (Section 5.1), performed on the K_S-band photometry, after correcting it for the photometry bias described in Chapter 4. A newly developed CNN-based classifier was used to distinguish Classical Cepheid, Type II Cepheids, as well as non-Cepheid variables (Section 5.1.1). The spatial distribution of the Type II Cepheids, which are centered heavily around the Galactic center, was used, in conjunction with the distance of Sgr A*, to derive a new extinction law (Section 5.1.3). This extinction law was then used to derive the distances to the Classical Cepheids located in the Galactic disk (Section 5.2), revealing significant spatial clusterings in their distribution, similarly to what has been seen in the near-side of the Galactic disk (Skowron et al. 2018). I presented additional Cepheids pairs, similar to the Twin Cepheids (Dékány et al. 2015a), which might have formed in the same parent open cluster (Section 5.2.1). I investigated the radial age trend of VVV Classical Cepheids (Section 5.3), and found agreement with the results of the OGLE project (Skowron et al. 2018). Finally, I performed an experiment to try to connect the VVV Classical Cepheids to their parent spiral arms by rotating their current

positions, taking into account their ages, the Galactic rotation curve and different values of the pattern speed of the spiral arms. I found that while some of the Cepheid overdensities lie upon the (extrapolated) location of spiral arms, there are also overdensities inbetween, warranting future study.

6.1 Outlook

RRL and Cepheid (both Classical and Type II) provide us with a unique way of looking at the spatial, chemical, as well as temporal evolution of stellar populations. While Classical Cepheids trace young stellar populations, RRL and Type II Cepheids show where the oldest stellar populations are located. The new methods developed un this thesis will be imporved upon in the future, and extended to other areas of study of pulsating variables.

The PCA-based light curve fitting algorithm is a new addition to the various methods developed for the characterization of the light curves of variable stars. In particular, the time-series observations of LSST (LSST Science Collaboration et al. 2009) will provide sparse, multiband light curves of tens of thousands of RRL variables. It will be the subject of a future study, if a light-curve fitting method can be developed that simultaneously fits the light curves of RRL stars in the LSST bands, based on a multi-band PCA of RRL light curves. Furthermore, in the case of RRL variables, the realization that all relations connecting their light-curve shapes to their metallicities are biased, mandates the revision of these relations. As they are widely used to determine the metallicities of old stellar populations that host RRL variables (see, e.g. Skowron et al. 2016), their inaccuracies can have great effect on the conclusions reached by such studies. Therefore, the calibration of accurate relations for RRL stars, even in the optical bands, is highly desirable. In order to achieve this, enlarged samples of spectroscopically observed RR Lyrae will be required. Furthermore, it would be of great interest to explore the extension of the PC-based light-curve fitting technique to other types of variable stars, such as Classical Cepheids.

The unexpected discovery of severe biases in the photometric zero-point calibration of the VISTA observations puts into question many studies that have reached their conclusions using VVV photometry in crowded fields. Studies of the extinction law are especially affected. The majority of these studies will have to be revised, following a careful recalibration of the VVV observations. It is worth repeating here that results based on VMC data, likewise obtained by the VISTA telescope, and processed by CASU, may be similarly

affected.

The large number of newly found Classical Cepheids, on the far side of the Galactic disk, provides unique probes of the Galactic structure and kinematics. Spectroscopic follow-up of the brightest ones has already started, with the aim of investigating the kinematic and chemical properties of the far disk. In this context, the metallicity-determination method using low-resolution spectroscopic near-IR observations presented by Inno et al. (2019) is especially interesting, as it makes the Classical Cepheid sample presented here prime probes of the metallicity distribution of the far side of the Galactic disk. More work is needed in oreder to connect the current distribution of the VVV Cepheids to their original spiral arms, especially with regard to the calibration of the Cepheid period-age relationship, to the Galactic rotation curve, as well as to the extinction law.

In the future, the VVVX survey (Minniti 2018) will play a leading role in the exploration of the Galactic disk by extending substantially the Galactic longitude and latitude coverage of the original VVV survey, allowing the discovery and characterization of RRL and Cepheid variables in ever larger sections of the Milky Way, using the methods presented in this thesis.

ACKNOWLEDGEMENTS

I would like to thank all the support I have received during the wonderful years I have spent at the Pontificia Universidad Católica de Chile, Santiago, as well as during my one year exchange at the Astronomisches Rechen-Institut der Universität Heidelberg.

First and foremost, I would like to thank my supervisor, Prof. Márcio Catelan, for all the support, encouragement, and motivation I have received from him. Secondly, I am also deeply thankful for my co-advisor and exchange host, Prof. Eva Grebel, for allowing me to spend a very productive, wonderful year in Heidelberg.

The work presented in this thesis would not have been possible without that of Dr. István Dékány. We started working together on VVV data at Católica, and continued to do so after he moved to Heidelberg, which has resulted in all the results presented here. I cannot overstate his inspirational effect on my work, as well on my development as an astronomer.

I would like to thank my mother for all the patience she has displayed while I was conducting my studies in Chile and Germany.

I cannot end the acknowledgements without mentioning my comrades among the students, both at the IA, as well as at the ARI. At the former, I could always count upon Diego Calderón, Rodrigo Carvajal, Camila Navarrete, Katerine Joachimi, and many others, to have stimulating conversations about our research, as well as just to remind ourselves to relax, and spend time together, in order to ease our minds of the everyday stress experienced by the majority of students, in this otherwise unforgiving academic environment. During my time at ARI, I have received much support from Michael Hanke, Zdenek Prudil and Josefina Michea in the daily goings of both the university, as well as Heidelberg. I would like to thank my officemates at ARI, Bekdaulet Shukirgaliyev and Clio Bertelli Motta for a wonderful time.

Finally, I would like to acknowledge the “First La Serena School for Data Science"", where I have been introduced for the first time to many of the statistical and data science methods utilized throughout this thesis. In this endevour, the book “Statistics, Data Mining, and Machine Learning in Astronomy” (Ivezić et al. 2014) has also been indispensable.

During my studies, I have been financially supported by CONICYT-PCHA (Doctorado Nacional 2014-63140099), by the Graduate Student Exchange Fellowship Program between

the Institute of Astrophysics of the Pontificia Universidad Católica de Chile and the Zentrum für Astronomie der Universität Heidelberg, funded by the Heidelberg Center in Santiago de Chile and the Deutscher Akademischer Austauschdienst, as well as by the Ministry for Economy, Development, and Tourism's Programa Iniciativa Milenio through grant IC1200009. Additional support is acknowledged by Proyecto Basal PFB-06/2007 and AFB-170002, by FONDECYT grants #1141141 and #1171273, by CONICYT Anillo grant ACT 1101, by CONICYT's PCI program through grant DPI20140066, and the CONICYT's "Beca Asistencia a eventos y cursos cortos para estudiantes de doctorado - convocatoria 2015". Processing and analysis of data were partly performed on the Milky Way supercomputer, which is partly funded by the Sonderforschungsbereich SFB 881 "The Milky Way System" (subproject Z2) of the Deutsche Forschungsgemeinschaft.

AUTHOR'S PUBLICATIONS

This section lists the papers published by Gergely Hajdu as either a first, or as a co-author. In this thesis, each of Chapters 2–5 correspond to one article, either already published or in preparation, with Gergely Hajdu as first or second author. Chapter 2 is based on Hajdu et al. (2018), Chapter 3 on Dékány et al. (2018), Chapter 4 on Hajdu et al. (in prep.), and Chapter 5 on Dékány et al. (in prep.).

Ábrahám, P., Á. Kóspál, M. Kun, O. Fehér, G. Zsidi, J. A. Acosta-Pulido, et al. (2018). "An UXor among FUors: Extinction-related Brightness Variations of the Young Eruptive Star V582 Aur." In: ApJ 853.1, 28, p. 28. DOI: 10.3847/1538-4357/aaa242. arXiv: 1712.04968 [astro-ph.SR].

Angeloni, R., R. Contreras Ramos, M. Catelan, I. Dékány, F. Gran, J. Alonso-García, et al. (2014). "The VVV Templates Project Towards an automated classification of VVV light-curves. I. Building a database of stellar variability in the near-infrared." In: A&A 567, A100, A100. DOI: 10.1051/0004-6361/201423904. arXiv: 1405.4517 [astro-ph.SR].

Contreras Ramos, Rodrigo, Dante Minniti, José G. Fernández-Trincado, Javier Alonso-García, Márcio Catelan, Felipe Gran, et al. (2018a). "The Orbit of the New Milky Way Globular Cluster FSR1716 = VVV-GC05." In: ApJ 863.1, 78, p. 78. DOI: 10.3847/1538-4357/aacd09. arXiv: 1807.02065 [astro-ph.GA].

Dékány, I., D. Minniti, G. Hajdu, J. Alonso-García, M. Hempel, T. Palma, et al. (2015a). "Discovery of a Pair of Classical Cepheids in an Invisible Cluster Beyond the Galactic Bulge." In: ApJ 799.1, L11, p. L11. DOI: 10.1088/2041-8205/799/1/L11. arXiv: 1412.8658 [astro-ph.GA].

Dékány, I., D. Minniti, D. Majaess, M. Zoccali, G. Hajdu, J. Alonso-García, et al. (2015b). "The VVV Survey Reveals Classical Cepheids Tracing a Young and Thin Stellar Disk across the Galaxy's Bulge." In: ApJ 812, L29, p. L29. DOI: 10.1088/2041-8205/812/2/L29. arXiv: 1509.08402.

Dékány, István, Gergely Hajdu, Eva K. Grebel, and Márcio Catelan (in prep.). In:

Dékány, István, Gergely Hajdu, Eva K. Grebel, Márcio Catelan, Felipe Elorrieta, Susana Eyheramendy, et al. (2018). "A Near-infrared RR Lyrae Census along the Southern Galactic Plane: The Milky Way's Stellar Fossil Brought to Light." In: ApJ 857.1, 54, p. 54. DOI: 10.3847/1538-4357/aab4fa. arXiv: 1804.01457 [astro-ph.SR].

Elorrieta, Felipe, Susana Eyheramendy, Andrés Jordán, István Dékány, Márcio Catelan, Rodolfo Angeloni, et al. (2016). "A machine learned classifier for RR Lyrae in the VVV survey." In: A&A 595, A82, A82. DOI: 10.1051/0004-6361/201628700. arXiv: 1610.05707 [astro-ph.IM].

Hajdu, G., M. Catelan, J. Jurcsik, I. Dekany, A. J. Drake, and J. -B. Marquette (2015). "New RR Lyrae variables in binary systems." In: MNRAS 449, pp. L113–L117. DOI: 10.1093/mnrasl/slv024. arXiv: 1502.01318 [astro-ph.SR].

Hajdu, G., J. Jurcsik, Á. Sódor, B. Szeidl, P. Smitola, B. Belucz, et al. (2012). "The Blazhko modulation of TV Boo in 2010." In: *Astronomische Nachrichten* 333.10, p. 1074. DOI: 10.1002/asna.201211826. arXiv: 1211.3640 [astro-ph.SR].

Hajdu, Gergely, István Dékány, Márcio Catelan, Eva K. Grebel, and Johanna Jurcsik (2018). "A Data-driven Study of RR Lyrae Near-IR Light Curves: Principal Component Analysis, Robust Fits, and Metallicity Estimates." In: ApJ 857.1, 55, p. 55. DOI: 10.3847/1538-4357/aab4fd. arXiv: 1804.01456 [astro-ph.SR].

Hajdu, Gergely, István Dékány, Eva K. Grebel, and Márcio Catelan (in prep.). In:

Jurcsik, J. and G. Hajdu (2017). "The first Baade-Wesselink analysis of Blazhko RR Lyrae stars: discrepancies between photometrically and spectroscopically determined radius variations." In: MNRAS 470.1, pp. 617–625. DOI: 10.1093/mnras/stx1302. arXiv: 1705.08537 [astro-ph.SR].

Jurcsik, J., G. Hajdu, and M. Catelan (2018a). "New Galactic Multi-Mode Cepheids from the ASAS-SN Survey." In: *Acta Astron.* 68.4, pp. 341–350. DOI: 10.32023/0001-5237/68.4.2. arXiv: 1812.01575 [astro-ph.SR].

Jurcsik, J., G. Hajdu, I. Dékány, J. Nuspl, M. Catelan, and E. K. Grebel (2018b). "Blazhko modulation in the infrared." In: MNRAS 475.3, pp. 4208–4222. DOI: 10.1093/mnras/sty112. arXiv: 1801.03436 [astro-ph.SR].

Jurcsik, J., G. Hajdu, B. Szeidl, K. Oláh, J. Kelemen, Á. Sódor, et al. (2012a). "Long-term photometric monitoring of RR Lyrae stars in Messier 3." In: MNRAS 419.3, pp. 2173–2194. DOI: 10.1111/j.1365-2966.2011.19868.x. arXiv: 1109.4525 [astro-ph.SR].

Jurcsik, J., Zs. Hurta, Á. Sódor, B. Szeidl, I. Nagy, K. Posztobányi, et al. (2009a). "An extensive photometric study of the Blazhko RR Lyrae star DM Cyg." In: MNRAS 397.1, pp. 350–360. DOI: 10.1111/j.1365-2966.2009.14919.x. arXiv: 0904.4129 [astro-ph.SR].

Jurcsik, J., P. Smitola, G. Hajdu, and J. Nuspl (2014). "On the Modulation of RR Lyrae Stars in the Globular Cluster M3." In: ApJ 797.1, L3, p. L3. DOI: 10.1088/2041-8205/797/1/L3. arXiv: 1411.1987 [astro-ph.SR].

Jurcsik, J., P. Smitola, G. Hajdu, C. Pilachowski, K. Kolenberg, Á. Sódor, et al. (2013). "What is the Difference? Blazhko and Non-Blazhko RRab Stars and the Special Case of V123 in M3." In: ApJ 778.2, L27, p. L27. DOI: 10.1088/2041-8205/778/2/L27. arXiv: 1309.5308 [astro-ph.SR].

Jurcsik, J., P. Smitola, G. Hajdu, Á. Sódor, J. Nuspl, K. Kolenberg, et al. (2015). "Overtone and Multi-mode RR Lyrae Stars in the Globular Cluster M3." In: ApJS 219.2, 25, p. 25. DOI: 10.1088/0067-0049/219/2/25. arXiv: 1504.06215 [astro-ph.SR].

Jurcsik, J., P. Smitola, G. Hajdu, Á. Sódor, J. Nuspl, K. Kolenberg, et al. (2017). "Photometric and radial-velocity time series of RR Lyrae stars in M3: analysis of single-mode variables." In: MNRAS 468, pp. 1317–1337. DOI: 10.1093/mnras/stx382. arXiv: 1702.03264 [astro-ph.SR].

Jurcsik, J., Á. Sódor, G. Hajdu, B. Szeidl, Á. Dózsa, K. Posztobányi, et al. (2012b). "An extensive photometric study of the Blazhko RR Lyrae star RZ Lyr." In: MNRAS 423.2, pp. 993–1005. DOI: 10.1111/j.1365-2966.2012.20772.x. arXiv: 1202.4388 [astro-ph.SR].

Jurcsik, J., Á. Sódor, B. Szeidl, Zs. Hurta, M. Váradi, K. Posztobányi, et al. (2009b). "The Konkoly Blazhko Survey: is light-curve modulation a common property of RRab stars?" In: MNRAS 400.2, pp. 1006–1018. DOI: 10.1111/j.1365-2966.2009.15515.x. arXiv: 0908.1015 [astro-ph.SR].

Kóspál, Á., P. Ábrahám, J. A. Acosta-Pulido, M. M. Dunham, D. García-Álvarez, M. R. Hogerheijde, et al. (2016). "Multiwavelength study of the low-luminosity outburst-

ing young star HBC 722." In: A&A 596, A52, A52. DOI: 10.1051/0004-6361/201528061. arXiv: 1607.05925 [astro-ph.SR].

Lemasle, B., G. Hajdu, V. Kovtyukh, L. Inno, E. K. Grebel, M. Catelan, et al. (2018). "Milky Way metallicity gradient from Gaia DR2 F/1O double-mode Cepheids." In: A&A 618, A160, A160. DOI: 10.1051/0004-6361/201834050. arXiv: 1809.07352 [astro-ph.SR].

Majaess, D., I. Dékány, G. Hajdu, D. Minniti, D. Turner, and W. Gieren (2018). "Establishing the Galactic Centre distance using VVV Bulge RR Lyrae variables." In: Ap&SS 363.6, 127, p. 127. DOI: 10.1007/s10509-018-3346-6. arXiv: 1805.04119 [astro-ph.GA].

Minniti, Dante, Tali Palma, István Dékány, Maren Hempel, Marina Rejkuba, Joyce Pullen, et al. (2017b). "FSR 1716: A New Milky Way Globular Cluster Confirmed Using VVV RR Lyrae Stars." In: ApJ 838, L14, p. L14. DOI: 10.3847/2041-8213/838/1/L14. arXiv: 1703.02033 [astro-ph.GA].

Sódor, Á., G. Hajdu, J. Jurcsik, B. Szeidl, K. Posztobányi, Zs. Hurta, et al. (2012). "The light-curve modulation of XY And and UZ Vir: two Blazhko RR Lyrae stars with additional frequencies." In: MNRAS 427.2, pp. 1517–1529. DOI: 10.1111/j.1365-2966.2012.21837.x. arXiv: 1208.0139 [astro-ph.SR].

Szeidl, B., J. Jurcsik, Á. Sódor, G. Hajdu, and P. Smitola (2012). "Fitting Blazhko light curves." In: MNRAS 424.4, pp. 3094–3100. DOI: 10.1111/j.1365-2966.2012.21461.x. arXiv: 1206.1715 [astro-ph.SR].

REFERENCES

Abuter, R., A. Amorim, M. Bauboeck, J. P. Berger, H. Bonnet, W. Brandner, et al. (2019). "A geometric distance measurement to the Galactic Center black hole with 0.3% uncertainty." In: *arXiv e-prints*, arXiv:1904.05721, arXiv:1904.05721. arXiv: 1904.05721 [astro-ph.GA].

Akaike, H. (1974). "A New Look at the Statistical Model Identification." In: *IEEE Transactions on Automatic Control* 19, pp. 716–723.

Alcock, C., D. R. Alves, A. Becker, D. Bennett, K. H. Cook, A. Drake, et al. (2003). "The MACHO Project Large Magellanic Cloud Variable Star Inventory. XI. Frequency Analysis of the Fundamental-Mode RR Lyrae Stars." In: ApJ 598.1, pp. 597–609. DOI: 10.1086/378689. arXiv: astro-ph/0308019 [astro-ph].

Alexander, Stephen (1852). "On the origin of the forms and the present condition of some of the clusters of stars, and several of the nebulae." In: AJ 2, pp. 95–96. DOI: 10.1086/100227.

Alonso-García, Javier, Dante Minniti, Márcio Catelan, Rodrigo Contreras Ramos, Oscar A. Gonzalez, Maren Hempel, et al. (2017). "Extinction Ratios in the Inner Galaxy as Revealed by the VVV Survey." In: ApJ 849.1, L13, p. L13. DOI: 10.3847/2041-8213/aa92c3. arXiv: 1710.04854 [astro-ph.GA].

Anderson, R. I., H. Saio, S. Ekström, C. Georgy, and G. Meynet (2016). "On the effect of rotation on populations of classical Cepheids. II. Pulsation analysis for metallicities 0.014, 0.006, and 0.002." In: A&A 591, A8, A8. DOI: 10.1051/0004-6361/201528031. arXiv: 1604.05691 [astro-ph.SR].

Anderson, Richard I., Laurent Eyer, and Nami Mowlavi (2013). "Cepheids in open clusters: an 8D all-sky census." In: MNRAS 434.3, pp. 2238–2261. DOI: 10.1093/mnras/stt1160. arXiv: 1212.5119 [astro-ph.GA].

Angeloni, R., R. Contreras Ramos, M. Catelan, I. Dékány, F. Gran, J. Alonso-García, et al. (2014). "The VVV Templates Project Towards an automated classification of VVV light-curves. I. Building a database of stellar variability in the near-infrared." In: A&A

567, A100, A100. DOI: 10.1051/0004-6361/201423904. arXiv: 1405.4517 [astro-ph.SR].

Baade, W. (1944). "The Resolution of Messier 32, NGC 205, and the Central Region of the Andromeda Nebula." In: ApJ 100, p. 137. DOI: 10.1086/144650.

Bailey, Solon Irving (1902). "A discussion of variable stars in the cluster ω Centauri." In: *Annals of Harvard College Observatory* 38, p. 1.

Bailey, Solon Irving and Edward Charles Pickering (1913). "Variable stars in the cluster Messier 3." In: *Annals of Harvard College Observatory* 78, pp. 1–98.6.

Baker, N. and R. Kippenhahn (1962). "The Pulsations of Models of δ Cephei Stars. With 17 Figures in the Text." In: *Zeitschrift fur Astrophysik* 54, p. 114.

Barnes T. G., III, J. A. Fernley, M. L. Frueh, J. G. Navas, T. J. Moffett, and I. Skillen (1997). "BVRIJHK Photometry of Cepheid Variables." In: PASP 109, pp. 645–658. DOI: 10.1086/133927.

Barnes Thomas G., III, Thomas J. Moffett, and Marian L. Frueh (1992). "BVRIJHK Photometry of RR Lyrae Variables." In: *Publications of the Astronomical Society of the Pacific* 104, p. 514. DOI: 10.1086/133026.

Belokurov, V., D. Erkal, N. W. Evans, S. E. Koposov, and A. J. Deason (2018). "Co-formation of the disc and the stellar halo." In: MNRAS 478.1, pp. 611–619. DOI: 10.1093/mnras/sty982. arXiv: 1802.03414 [astro-ph.GA].

Bhardwaj, A., M. Rejkuba, D. Minniti, F. Surot, E. Valenti, M. Zoccali, et al. (2017a). "Galactic bulge population II Cepheids in the VVV survey: period-luminosity relations and a distance to the Galactic centre." In: A&A 605, A100, A100. DOI: 10.1051/0004-6361/201730841. arXiv: 1707.03755 [astro-ph.SR].

Bhardwaj, Anupam, Shashi M. Kanbur, Marcella Marconi, Marina Rejkuba, Harinder P. Singh, and Chow-Choong Ngeow (2017b). "A comparative study of multiwavelength theoretical and observed light curves of Cepheid variables." In: MNRAS 466, pp. 2805–2824. DOI: 10.1093/mnras/stw3256. arXiv: 1612.03806 [astro-ph.GA].

Bhardwaj, Anupam, Lucas M. Macri, Marina Rejkuba, Shashi M. Kanbur, Chow-Choong Ngeow, and Harinder P. Singh (2017c). "Large Magellanic Cloud Near-infrared Synoptic Survey. IV. Leavitt Laws for Type II Cepheid Variables." In: AJ 153.4, 154, p. 154. DOI: 10.3847/1538-3881/aa5e4f. arXiv: 1702.00967 [astro-ph.SR].

Bonifacio, P., S. Monai, and T. C. Beers (2000). "A Search for Stars of Very Low Metal Abundance. V. Photoelectric UBV Photometry of Metal-weak Candidates from the Northern HK Survey." In: AJ 120.4, pp. 2065–2081. DOI: 10.1086/301566. arXiv: astro-ph/0006178 [astro-ph].

Bono, G., F. Caputo, V. Castellani, M. Marconi, J. Storm, and S. Degl'Innocenti (2003). "A pulsational approach to near-infrared and visual magnitudes of RR Lyr stars." In: MNRAS 344, pp. 1097–1106. DOI: 10.1046/j.1365-8711.2003.06878.x. arXiv: astro-ph/0306142 [astro-ph].

Bono, G., M. Marconi, S. Cassisi, F. Caputo, W. Gieren, and G. Pietrzynski (2005). "Classical Cepheid Pulsation Models. X. The Period-Age Relation." In: ApJ 621.2, pp. 966–977. DOI: 10.1086/427744. arXiv: astro-ph/0411756 [astro-ph].

Bono, G., M. Marconi, and R. F. Stellingwerf (2000). "Classical Cepheid pulsation models — VI. The Hertzsprung progression." In: A&A 360, pp. 245–262. arXiv: astro-ph/0006229 [astro-ph].

Bovy, Jo, Hans-Walter Rix, Chao Liu, David W. Hogg, Timothy C. Beers, and Young Sun Lee (2012). "The Spatial Structure of Mono-abundance Sub-populations of the Milky Way Disk." In: ApJ 753.2, 148, p. 148. DOI: 10.1088/0004-637X/753/2/148. arXiv: 1111.1724 [astro-ph.GA].

Braga, V. F., R. Contreras Ramos, D. Minniti, C. E. Ferreira Lopes, M. Catelan, J. H. Minniti, et al. (2019). "New type II Cepheids from VVV data towards the Galactic center." In: *arXiv e-prints*, arXiv:1904.12024, arXiv:1904.12024. arXiv: 1904.12024 [astro-ph.SR].

Brook, C. B., G. S. Stinson, B. K. Gibson, D. Kawata, E. L. House, M. S. Miranda, et al. (2012). "Thin disc, thick disc and halo in a simulated galaxy." In: MNRAS 426.1, pp. 690–700. DOI: 10.1111/j.1365-2966.2012.21738.x. arXiv: 1206.0740 [astro-ph.GA].

Cacciari, C., G. Clementini, and J. A. Fernley (1992). "On the Absolute Magnitude of RR Lyrae Stars: UU Ceti, RV Phoenicis, and W Tucanae." In: ApJ 396, p. 219. DOI: 10.1086/171711.

Cardelli, Jason A., Geoffrey C. Clayton, and John S. Mathis (1989). "The Relationship between Infrared, Optical, and Ultraviolet Extinction." In: ApJ 345, p. 245. DOI: 10.1086/167900.

Carpenter, John M. (2001). "Color Transformations for the 2MASS Second Incremental Data Release." In: AJ 121, pp. 2851–2871. DOI: 10.1086/320383. arXiv: astro-ph/0101463 [astro-ph].

Carretta, E., A. Bragaglia, R. Gratton, V. D'Orazi, and S. Lucatello (2009). "Intrinsic iron spread and a new metallicity scale for globular clusters." In: A&A 508.2, pp. 695–706. DOI: 10.1051/0004-6361/200913003. arXiv: 0910.0675 [astro-ph.GA].

Casali, M., A. Adamson, C. Alves de Oliveira, O. Almaini, K. Burch, T. Chuter, et al. (2007). "The UKIRT wide-field camera." In: A&A 467.2, pp. 777–784. DOI: 10.1051/0004-6361:20066514.

Catelan, M. (2009). "Horizontal branch stars: the interplay between observations and theory, and insights into the formation of the Galaxy." In: Ap&SS 320, pp. 261–309. DOI: 10.1007/s10509-009-9987-8. arXiv: astro-ph/0507464 [astro-ph].

Catelan, M., D. Minniti, P. W. Lucas, I. Dékány, R. K. Saito, R. Angeloni, et al. (2013). "Stellar Variability in the VVV survey." In: *arXiv e-prints*, arXiv:1310.1996, arXiv:1310.1996. arXiv: 1310.1996 [astro-ph.SR].

Catelan, M., Barton J. Pritzl, and Horace A. Smith (2004). "The RR Lyrae Period-Luminosity Relation. I. Theoretical Calibration." In: *The Astrophysical Journal Supplement Series* 154, pp. 633–649. DOI: 10.1086/422916. arXiv: astro-ph/0406067 [astro-ph].

Catelan, M. and H. A. Smith (2015). *Pulsating Stars*.

Chen, Xiaodian, Shu Wang, Licai Deng, Richard de Grijs, Chao Liu, and Hao Tian (2019). "An intuitive 3D map of the Galactic warp's precession traced by classical Cepheids." In: *Nature Astronomy* 3, pp. 320–325. DOI: 10.1038/s41550-018-0686-7. arXiv: 1902.00998 [astro-ph.GA].

Chen, Xiaodian, Shu Wang, Licai Deng, Richard de Grijs, and Ming Yang (2018). "Wide-field Infrared Survey Explorer (WISE) Catalog of Periodic Variable Stars." In: ApJS 237.2, 28, p. 28. DOI: 10.3847/1538-4365/aad32b. arXiv: 1807.03448 [astro-ph.SR].

Cioni, M. -R. L., G. Clementini, L. Girardi, R. Guand alini, M. Gullieuszik, B. Miszalski, et al. (2011). "The VMC survey. I. Strategy and first data." In: A&A 527, A116, A116. DOI: 10.1051/0004-6361/201016137. arXiv: 1012.5193 [astro-ph.CO].

Clement, Christine M., Adam Muzzin, Quentin Dufton, Thivya Ponnampalam, John Wang, Jay Burford, et al. (2001). "Variable Stars in Galactic Globular Clusters." In: AJ 122, pp. 2587–2599. DOI: 10.1086/323719. arXiv: astro-ph/0108024 [astro-ph].

Clementini, G., V. Ripepi, R. Molinaro, A. Garofalo, T. Muraveva, L. Rimoldini, et al. (2019). "Gaia Data Release 2. Specific characterisation and validation of all-sky Cepheids and RR Lyrae stars." In: A&A 622, A60, A60. DOI: 10.1051/0004-6361/201833374.

Cohen, Judith G. and Jorge Meléndez (2005). "Abundances in a Large Sample of Stars in M3 and M13." In: AJ 129, pp. 303–329. DOI: 10.1086/426369. arXiv: astro-ph/0409725 [astro-ph].

Contreras Ramos, R., M. Zoccali, F. Rojas, A. Rojas-Arriagada, M. Gárate, P. Huijse, et al. (2017). "Proper motions in the VVV Survey: Results for more than 15 million stars across NGC 6544." In: A&A 608, A140, A140. DOI: 10.1051/0004-6361/201731462. arXiv: 1709.07919 [astro-ph.GA].

Contreras Ramos, Rodrigo, Dante Minniti, Felipe Gran, Manuela Zoccali, Javier Alonso-García, Pablo Huijse, et al. (2018b). "The VVV Survey RR Lyrae Population in the Galactic Center Region." In: ApJ 863.1, 79, p. 79. DOI: 10.3847/1538-4357/aacf90. arXiv: 1807.04303 [astro-ph.GA].

Cox, J. P., A. N. Cox, K. H. Olsen, D. S. King, and D. D. Eilers (1966). "Self-Excited Radial Oscillations in Thin Stellar Envelopes. I." In: ApJ 144, p. 1038. DOI: 10.1086/148702.

Das, S., A. Bhardwaj, S. M. Kanbur, H. P. Singh, and M. Marconi (2018). "On the variation of light-curve parameters of RR Lyrae variables at multiple wavelengths." In: MNRAS 481, pp. 2000–2017. DOI: 10.1093/mnras/sty2358. arXiv: 1808.08165 [astro-ph.SR].

Deb, S. and H. P. Singh (2009). "Light curve analysis of variable stars using Fourier decomposition and principal component analysis." In: A&A 507, pp. 1729–1737. DOI: 10.1051/0004-6361/200912851. arXiv: 0903.3500 [astro-ph.SR].

Dékány, I., D. Minniti, M. Catelan, M. Zoccali, R. K. Saito, M. Hempel, et al. (2013). "VVV Survey Near-infrared Photometry of Known Bulge RR Lyrae Stars: The Distance to the Galactic Center and Absence of a Barred Distribution of the Metal-poor Population." In: ApJ 776, L19, p. L19. DOI: 10.1088/2041-8205/776/2/L19. arXiv: 1309.5933 [astro-ph.GA].

Dékány, I., D. Minniti, G. Hajdu, J. Alonso-García, M. Hempel, T. Palma, et al. (2015a). "Discovery of a Pair of Classical Cepheids in an Invisible Cluster Beyond the Galactic Bulge." In: ApJ 799.1, L11, p. L11. DOI: 10.1088/2041-8205/799/1/L11. arXiv: 1412.8658 [astro-ph.GA].

Dékány, I., D. Minniti, D. Majaess, M. Zoccali, G. Hajdu, J. Alonso-García, et al. (2015b). "The VVV Survey Reveals Classical Cepheids Tracing a Young and Thin Stellar Disk across the Galaxy's Bulge." In: ApJ 812, L29, p. L29. DOI: 10.1088/2041-8205/812/2/L29. arXiv: 1509.08402.

Dékány, István, Gergely Hajdu, Eva K. Grebel, and Márcio Catelan (in prep.). In:

Dékány, István, Gergely Hajdu, Eva K. Grebel, Márcio Catelan, Felipe Elorrieta, Susana Eyheramendy, et al. (2018). "A Near-infrared RR Lyrae Census along the Southern Galactic Plane: The Milky Way's Stellar Fossil Brought to Light." In: ApJ 857.1, 54, p. 54. DOI: 10.3847/1538-4357/aab4fa. arXiv: 1804.01457 [astro-ph.SR].

Dias, Wilton S. and J. R. D. Lépine (2005). "Direct Determination of the Spiral Pattern Rotation Speed of the Galaxy." In: ApJ 629.2, pp. 825–831. DOI: 10.1086/431456. arXiv: astro-ph/0503083 [astro-ph].

Drake, A. J., M. Catelan, S. G. Djorgovski, G. Torrealba, M. J. Graham, V. Belokurov, et al. (2013). "Probing the Outer Galactic Halo with RR Lyrae from the Catalina Surveys." In: ApJ 763.1, 32, p. 32. DOI: 10.1088/0004-637X/763/1/32. arXiv: 1211.2866 [astro-ph.GA].

Elitzur, M. (1992). *Astronomical masers*. Vol. 170. DOI: 10.1007/978-94-011-2394-5.

Elorrieta, Felipe, Susana Eyheramendy, Andrés Jordán, István Dékány, Márcio Catelan, Rodolfo Angeloni, et al. (2016). "A machine learned classifier for RR Lyrae in the VVV survey." In: A&A 595, A82, A82. DOI: 10.1051/0004-6361/201628700. arXiv: 1610.05707 [astro-ph.IM].

Emerson, James P., Mike J. Irwin, Jim Lewis, Simon Hodgkin, Dafydd Evans, Peter Bunclark, et al. (2004). "VISTA data flow system: overview." In: *Optimizing Scientific Return for Astronomy through Information Technologies*. Ed. by Peter J. Quinn and Alan Bridger. Vol. 5493. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pp. 401–410. DOI: 10.1117/12.551582.

Ensor, T., J. Cami, N. H. Bhatt, and A. Soddu (2017). "A Principal Component Analysis of the Diffuse Interstellar Bands." In: ApJ 836, 162, p. 162. DOI: 10.3847/1538-4357/aa5b84. arXiv: 1701.06180 [astro-ph.GA].

Erroz-Ferrer, Santiago, Johan H. Knapen, Ryan Leaman, Mauricio Cisternas, Joan Font, John E. Beckman, et al. (2015). "Hα kinematics of S^{4}G spiral galaxies - II. Data description and non-circular motions." In: MNRAS 451.1, pp. 1004–1024. DOI: 10.1093/mnras/stv924. arXiv: 1504.06282 [astro-ph.GA].

Fadeyev, Yu. A. (2013). "Evolution and pulsation period change in the Large Magellanic Cloud Cepheids." In: *Astronomy Letters* 39.11, pp. 746–752. DOI: 10.1134/S1063773713110030. arXiv: 1307.5966 [astro-ph.SR].

Feast, Michael W., John W. Menzies, Noriyuki Matsunaga, and Patricia A. Whitelock (2014). "Cepheid variables in the flared outer disk of our galaxy." In: Nature 509.7500, pp. 342–344. DOI: 10.1038/nature13246. arXiv: 1406.7660 [astro-ph.GA].

Ferguson, Jason W., David R. Alexander, France Allard, Travis Barman, Julia G. Bodnarik, Peter H. Hauschildt, et al. (2005). "Low-Temperature Opacities." In: ApJ 623.1, pp. 585–596. DOI: 10.1086/428642. arXiv: astro-ph/0502045 [astro-ph].

Fernley, J. A., A. E. Lynas-Gray, I. Skillen, R. F. Jameson, F. Marang, D. Kilkenny, et al. (1989). "The absolute magnitudes of RR Lyrae stars - I. X Arietis." In: MNRAS 236, pp. 447–473. DOI: 10.1093/mnras/236.3.447.

Fernley, J. A., I. Skillen, R. F. Jameson, T. G. Barnes, D. Kilkenny, and G. Hill (1990). "The absolute magnitudes of RR Lyrae stars - IV. V 445 , SS Leo and VY Serpentis." In: MNRAS 247, p. 287.

Ferreira Lopes, C. E., I. Dékány, M. Catelan, N. J. G. Cross, R. Angeloni, I. C. Leão, et al. (2015). "The WFCAM multiwavelength Variable Star Catalog." In: A&A 573, A100, A100. DOI: 10.1051/0004-6361/201423793. arXiv: 1408.4137 [astro-ph.SR].

Fresneau, A., A. E. Vaughan, and R. W. Argyle (2005). "Density Wave Streaming Motions in Stars along the Sagittarius Spiral Arm." In: AJ 130.6, pp. 2701–2716. DOI: 10.1086/497288.

Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, et al. (2018). "Gaia Data Release 2. Summary of the contents and survey properties."

In: A&A 616, A1, A1. DOI: 10.1051/0004-6361/201833051. arXiv: 1804.09365 [astro-ph.GA].

Galaz, Gaspar and Valerie de Lapparent (1998). "The ESO-Sculptor Survey: spectral classification of galaxies with Z < 0.5." In: A&A 332, pp. 459–478. arXiv: astro-ph/9711093 [astro-ph].

Gastine, T. and B. Dintrans (2008). "Direct numerical simulations of the κ-mechanism. II. Nonlinear saturation and the Hertzsprung progression." In: A&A 490.2, pp. 743–752. DOI: 10.1051/0004-6361:200809891. arXiv: 0804.0345 [astro-ph].

Genovali, K., B. Lemasle, G. Bono, M. Romaniello, M. Fabrizio, I. Ferraro, et al. (2014). "On the fine structure of the Cepheid metallicity gradient in the Galactic thin disk." In: A&A 566, A37, A37. DOI: 10.1051/0004-6361/201323198. arXiv: 1403.6128 [astro-ph.GA].

Gerhard, O. (2011). "Pattern speeds in the Milky Way." In: *Memorie della Societa Astronomica Italiana Supplementi* 18, p. 185. arXiv: 1003.2489 [astro-ph.GA].

Gillessen, S., F. Eisenhauer, S. Trippe, T. Alexand er, R. Genzel, F. Martins, et al. (2009). "Monitoring Stellar Orbits Around the Massive Black Hole in the Galactic Center." In: ApJ 692.2, pp. 1075–1109. DOI: 10.1088/0004-637X/692/2/1075. arXiv: 0810.4674 [astro-ph].

Gilmore, G. and N. Reid (1983). "New light on faint stars. III - Galactic structure towards the South Pole and the Galactic thick disc." In: MNRAS 202, pp. 1025–1047. DOI: 10.1093/mnras/202.4.1025.

Girardi, Léo (2016). "Red Clump Stars." In: ARA&A 54, pp. 95–133. DOI: 10.1146/annurev-astro-081915-023354.

Girardi, Léo and Maurizio Salaris (2001). "Population effects on the red giant clump absolute magnitude, and distance determinations to nearby galaxies." In: MNRAS 323.1, pp. 109–129. DOI: 10.1046/j.1365-8711.2001.04084.x. arXiv: astro-ph/0007343 [astro-ph].

González-Fernández, C., S. T. Hodgkin, M. J. Irwin, E. González-Solares, S. E. Koposov, J. R. Lewis, et al. (2018). "The VISTA ZYJHKs photometric system: calibration from 2MASS." In: MNRAS 474, pp. 5459–5478. DOI: 10.1093/mnras/stx3073. arXiv: 1711.08805 [astro-ph.IM].

Gonzalez, O. A., M. Rejkuba, M. Zoccali, E. Valenti, D. Minniti, M. Schultheis, et al. (2012). "Reddening and metallicity maps of the Milky Way bulge from VVV and 2MASS. II. The complete high resolution extinction map and implications for Galactic bulge studies." In: A&A 543, A13, A13. DOI: 10.1051/0004-6361/201219222. arXiv: 1204.4004 [astro-ph.GA].

Gran, F., D. Minniti, R. K. Saito, M. Zoccali, O. A. Gonzalez, C. Navarrete, et al. (2016). "Mapping the outer bulge with RRab stars from the VVV Survey." In: A&A 591, A145, A145. DOI: 10.1051/0004-6361/201527511. arXiv: 1604.01336 [astro-ph.GA].

Gran, F., M. Zoccali, R. Contreras Ramos, E. Valenti, A. Rojas-Arriagada, J. A. Carballo-Bello, et al. (2019). "Globular cluster candidates in the Galactic bulge: Gaia and VVV view of the latest discoveries." In: *arXiv e-prints*, arXiv:1904.10872, arXiv:1904.10872. arXiv: 1904.10872 [astro-ph.GA].

Gratton, R. G., E. Carretta, A. Bragaglia, S. Lucatello, and V. D'Orazi (2010). "The second and third parameters of the horizontal branch in globular clusters." In: A&A 517, A81, A81. DOI: 10.1051/0004-6361/200912572. arXiv: 1004.3862 [astro-ph.SR].

Gratton, Raffaele G., Eugenio Carretta, and Angela Bragaglia (2012). "Multiple populations in globular clusters. Lessons learned from the Milky Way globular clusters." In: *Astronomy and Astrophysics Review* 20, 50, p. 50. DOI: 10.1007/s00159-012-0050-3. arXiv: 1201.6526 [astro-ph.SR].

Grevesse, N. and A. J. Sauval (1998). "Standard Solar Composition." In: Space Sci. Rev. 85, pp. 161–174. DOI: 10.1023/A:1005161325181.

Groenewegen, M. A. T. (2018). "The Cepheid period-luminosity-metallicity relation based on Gaia DR2 data." In: A&A 619, A8, A8. DOI: 10.1051/0004-6361/201833478. arXiv: 1808.05796 [astro-ph.SR].

Hajdu, Gergely, István Dékány, Márcio Catelan, Eva K. Grebel, and Johanna Jurcsik (2018). "A Data-driven Study of RR Lyrae Near-IR Light Curves: Principal Component Analysis, Robust Fits, and Metallicity Estimates." In: ApJ 857.1, 55, p. 55. DOI: 10.3847/1538-4357/aab4fd. arXiv: 1804.01456 [astro-ph.SR].

Hajdu, Gergely, István Dékány, Eva K. Grebel, and Márcio Catelan (in prep.). In:

Haykin, Simon S. (2009). *Neural networks and learning machines*. Third. Upper Saddle River, NJ: Pearson Education.

Haywood, Misha, Paola Di Matteo, Matthew D. Lehnert, David Katz, and Ana Gómez (2013). "The age structure of stellar populations in the solar vicinity. Clues of a two-phase formation history of the Milky Way disk." In: A&A 560, A109, A109. DOI: 10.1051/0004-6361/201321397. arXiv: 1305.4663 [astro-ph.GA].

Helmi, Amina, Carine Babusiaux, Helmer H. Koppelman, Davide Massari, Jovan Veljanoski, and Anthony G. A. Brown (2018). "The merger that led to the formation of the Milky Way's inner stellar halo and thick disk." In: Nature 563.7729, pp. 85–88. DOI: 10.1038/s41586-018-0625-x. arXiv: 1806.06038 [astro-ph.GA].

Hertzsprung, E. (1926). "On the relation between period and form of the light-curve of variable stars of the δ Cephei type." In: *Bulletin of the Astronomical Institutes of the Netherlands* 3, p. 115.

Hodgkin, S. T., M. J. Irwin, P. C. Hewett, and S. J. Warren (2009). "The UKIRT wide field camera ZYJHK photometric system: calibration from 2MASS." In: MNRAS 394, pp. 675–692. DOI: 10.1111/j.1365-2966.2008.14387.x. arXiv: 0812.3081 [astro-ph].

Hotelling, H. (1933). "Analysis of a Complex of Statistical Variables with Principal Components." In: *Journal of Educational Psychology* 24, pp. 417–441.

Hubble, E. P. (1926). "Extragalactic nebulae." In: ApJ 64, pp. 321–369. DOI: 10.1086/143018.

Inno, L., M. A. Urbaneja, N. Matsunaga, G. Bono, M. Nonino, V. P. Debattista, et al. (2019). "First metallicity determination from near-infrared spectra for five obscured Cepheids discovered in the inner disc." In: MNRAS 482.1, pp. 83–97. DOI: 10.1093/mnras/sty2661. arXiv: 1805.03212 [astro-ph.GA].

Iorio, Giuliano and Vasily Belokurov (2019). "The shape of the Galactic halo with Gaia DR2 RR Lyrae. Anatomy of an ancient major merger." In: MNRAS 482.3, pp. 3868–3879. DOI: 10.1093/mnras/sty2806. arXiv: 1808.04370 [astro-ph.GA].

Irwin, Mike J., Jim Lewis, Simon Hodgkin, Peter Bunclark, Dafydd Evans, Richard McMahon, et al. (2004). "VISTA data flow system: pipeline processing for WFCAM and VISTA." In: *Optimizing Scientific Return for Astronomy through Information Technologies*. Ed. by Peter J. Quinn and Alan Bridger. Vol. 5493. Society of Photo-Optical

Instrumentation Engineers (SPIE) Conference Series, pp. 411–422. DOI: 10.1117/12.551449.

Ita, Yoshifusa, Noriyuki Matsunaga, Toshihiko Tanabé, Yoshikazu Nakada, Daisuke Kato, Takahiro Nagayama, et al. (2018). "A near-infrared variable star survey in the Magellanic Clouds: the Small Magellanic Cloud data." In: MNRAS 481.3, pp. 4206–4220. DOI: 10.1093/mnras/sty2539. arXiv: 1810.01617 [astro-ph.SR].

Ivezić, Željko, Andrew J. Connelly, Jacob T. Vanderplas, and Alexander Gray (2014). *Statistics, Data Mining, and Machine Learning in Astronomy*.

Jeffery, C. S. and H. Saio (2016). "Radial pulsation as a function of hydrogen abundance." In: MNRAS 458.2, pp. 1352–1373. DOI: 10.1093/mnras/stw388. arXiv: 1602.04964 [astro-ph.SR].

Jeon, Y.-B., C.-C. Ngeow, and J. M. Nemec (2014). "Ground-based photometry for 42 Kepler-field RR Lyrae stars." In: *Precision Asteroseismology*. Ed. by J. A. Guzik, W. J. Chaplin, G. Handler, and A. Pigulski. Vol. 301. IAU Symposium, pp. 427–428. DOI: 10.1017/S1743921313014889.

Jones, Rodney V., Bruce W. Carney, and Jon P. Fulbright (1996). "Template K Light Curves for RR Lyrae Stars." In: *Publications of the Astronomical Society of the Pacific* 108, p. 877. DOI: 10.1086/133809.

Jones, Rodney V., Bruce W. Carney, and David W. Latham (1988). "The Baade-Wesselink Method and the Distances to RR Lyrae Stars. VI. The Field Stars RS Bootis, TW Herculis, VY Serpentis, and UU Virginis, and the Absolute Magnitudes of RR Lyrae Stars." In: ApJ 332, p. 206. DOI: 10.1086/166647.

Jones, Rodney V., Bruce W. Carney, David W. Latham, and Robert L. Kurucz (1987a). "The Baade-Wesselink Method and the Distances to RR Lyrae Stars. II. The Field Star X ARIETIS." In: ApJ 312, p. 254. DOI: 10.1086/164867.

— (1987b). "The Baade-Wesselink Method and the Distances to RR Lyrae Stars. III. The Field Star SW Draconis." In: ApJ 314, p. 605. DOI: 10.1086/165090.

Jones, Rodney V., Bruce W. Carney, Jesper Storm, and David W. Latham (1992). "The Baade-Wesselink Method and the Distances to RR Lyrae Stars. VII. The Field Stars SW Andromedae and DX Delphini and a Comparison of Recent Baade-Wesselink Analyses." In: ApJ 386, p. 646. DOI: 10.1086/171046.

Jordán, Andrés, Néstor Espinoza, Markus Rabus, Susana Eyheramendy, David K. Sing, Jean-Michel Désert, et al. (2013). "A Ground-based Optical Transmission Spectrum of WASP-6b." In: ApJ 778, 184, p. 184. DOI: 10.1088/0004-637X/778/2/184. arXiv: 1310.6048 [astro-ph.EP].

Joy, Alfred H. (1937). "Radial Velocities of Cepheid Variable Stars." In: ApJ 86, p. 363. DOI: 10.1086/143874.

— (1940). "Spectroscopic Observations of Barnard's Variable in Messier 3." In: ApJ 92, p. 396. DOI: 10.1086/144228.

Jurcsik, J. (1995). "Revision of the [Fe/H] Scales Used for Globular Clusters and RR Lyrae Variables." In: *Acta Astron.* 45, pp. 653–660.

Jurcsik, J., J. M. Benko, and B. Szeidl (2002). "Blazhko stars: Always distorted?" In: A&A 390, pp. 133–137. DOI: 10.1051/0004-6361:20020654. arXiv: astro-ph/0204453 [astro-ph].

Jurcsik, J., G. Hajdu, I. Dékány, J. Nuspl, M. Catelan, and E. K. Grebel (2018b). "Blazhko modulation in the infrared." In: MNRAS 475.3, pp. 4208–4222. DOI: 10.1093/mnras/sty112. arXiv: 1801.03436 [astro-ph.SR].

Jurcsik, J. and G. Kovács (1996). "Determination of [Fe/H] from the light curves of RR Lyrae stars." In: A&A 312, pp. 111–120.

Jurcsik, J., P. Smitola, G. Hajdu, Á. Sódor, J. Nuspl, K. Kolenberg, et al. (2017). "Photometric and radial-velocity time series of RR Lyrae stars in M3: analysis of single-mode variables." In: MNRAS 468, pp. 1317–1337. DOI: 10.1093/mnras/stx382. arXiv: 1702.03264 [astro-ph.SR].

Kalinova, V., D. Colombo, E. Rosolowsky, R. Kannan, L. Galbany, R. García-Benito, et al. (2017). "Towards a new classification of galaxies: principal component analysis of CALIFA circular velocity curves." In: MNRAS 469, pp. 2539–2594. DOI: 10.1093/mnras/stx901. arXiv: 1509.03352 [astro-ph.GA].

Kanbur, S. M., D. Iono, N. R. Tanvir, and M. A. Hendry (2002). "On the use of Principal Component Analysis in analysing Cepheid light curves." In: MNRAS 329, pp. 126–134. DOI: 10.1046/j.1365-8711.2002.04948.x. arXiv: astro-ph/0108423 [astro-ph].

Kanbur, S. M. and H. Mariani (2004). "Principal-component analysis of RR Lyrae light curves." In: MNRAS 355, pp. 1361–1368. DOI: 10.1111/j.1365-2966.2004.08411.x. arXiv: astro-ph/0407020 [astro-ph].

Kapteyn, J. C. (1890). "Südliche muthmasslich veränderliche Sterne." In: *Astronomische Nachrichten* 125.11, p. 165. DOI: 10.1002/asna.18901251103.

Kerr, F. J. (1957). "A Magellanic effect on the galaxy." In: AJ 62, pp. 93–93. DOI: 10.1086/107466.

— (1962). "Galactic velocity models and the interpretation of 21-cm surveys." In: MNRAS 123, p. 327. DOI: 10.1093/mnras/123.4.327.

Kim, Dae-Won and Coryn A. L. Bailer-Jones (2016). "A package for the automated classification of periodic variable stars." In: A&A 587, A18, A18. DOI: 10.1051/0004-6361/201527188. arXiv: 1512.01611 [astro-ph.IM].

Kim, Dae-Won, Pavlos Protopapas, Coryn A. L. Bailer-Jones, Yong-Ik Byun, Seo-Won Chang, Jean-Baptiste Marquette, et al. (2014). "The EPOCH Project. I. Periodic variable stars in the EROS-2 LMC database." In: A&A 566, A43, A43. DOI: 10.1051/0004-6361/201323252. arXiv: 1403.6131 [astro-ph.IM].

Kim, Yonghwi and Woong-Tae Kim (2014). "Gaseous spiral structure and mass drift in spiral galaxies." In: MNRAS 440.1, pp. 208–224. DOI: 10.1093/mnras/stu276. arXiv: 1402.2291 [astro-ph.GA].

Koo, Bon-Chul, Geumsook Park, Woong-Tae Kim, Myung Gyoon Lee, Dana S. Balser, and Trey V. Wenger (2017). "Tracing the Spiral Structure of the Outer Milky Way with Dense Atomic Hydrogen Gas." In: PASP 129.979, p. 094102. DOI: 10.1088/1538-3873/aa7c08. arXiv: 1706.10084 [astro-ph.GA].

Kovács, G. and G. Kupi (2007). "Computation of the Fourier parameters of RR Lyrae stars by template fitting." In: A&A 462.3, pp. 1007–1016. DOI: 10.1051/0004-6361:20066050. arXiv: astro-ph/0610823 [astro-ph].

Kovács, G. and E. Zsoldos (1995). "A new method for the determination of [Fe/H] in RR Lyrae stars." In: A&A 293, pp. L57–L60.

Kraft, Robert P., Christopher Sneden, G. E. Langer, and Charles F. Prosser (1992). "Oxygen Abundances in Halo Giants. II. Giants in the Globular Clusters M13 and M3 and The Intermediately Metal-Poor Halo Field." In: AJ 104, p. 645. DOI: 10.1086/116261.

Laney, C. D. and R. S. Stobie (1992). "JHKL observations of galactic cepheids." In: A&AS 93, pp. 93–120.

Layden, A. C. (1994). "The metallicities and kinematics of RR Lyrae variables, 1: New observations of local stars." In: AJ 108, pp. 1016–1041. DOI: 10.1086/117132.

Layden, Andrew C. (1995). "The Metallicities and Kinematics of RR Lyrae Variables III. On the Production of Metal-Rich RR Lyrae Stars." In: AJ 110, p. 2312. DOI: 10.1086/117691.

Leavitt, Henrietta S. and Edward C. Pickering (1912). "Periods of 25 Variable Stars in the Small Magellanic Cloud." In: *Harvard College Observatory Circular* 173, pp. 1–3.

Lecun, Yann and Yoshua Bengio (1995). "Convolutional Networks for Images, Speech and Time Series." In: *The Handbook of Brain Theory and Neural Networks*. Ed. by Michael A. Arbib. The MIT Press, pp. 255–258.

Lee, Young Sun, Timothy C. Beers, Carlos Allende Prieto, David K. Lai, Constance M. Rockosi, Heather L. Morrison, et al. (2011). "The SEGUE Stellar Parameter Pipeline. V. Estimation of Alpha-element Abundance Ratios from Low-resolution SDSS/SEGUE Stellar Spectra." In: AJ 141, 90, p. 90. DOI: 10.1088/0004-6256/141/3/90. arXiv: 1010.2934 [astro-ph.IM].

Lemasle, B., M. A. T. Groenewegen, E. K. Grebel, G. Bono, G. Fiorentino, P. François, et al. (2017). "Detailed chemical composition of classical Cepheids in the LMC cluster NGC 1866 and in the field of the SMC." In: A&A 608, A85, A85. DOI: 10.1051/0004-6361/201731370. arXiv: 1709.03083 [astro-ph.GA].

Li, Chengdong, Gang Zhao, Yunpeng Jia, Shilong Liao, Chengqun Yang, and Qixun Wang (2019). "Flare and Warp of Galactic Disk with OB Stars from Gaia DR2." In: ApJ 871.2, 208, p. 208. DOI: 10.3847/1538-4357/aafa17.

Liu, Tianxing and Kenneth A. Janes (1989). "The Luminosity Scale of RR Lyrae Stars with the Baade-Wesselink Method. I. Photometry and Radial Velocities." In: *The Astrophysical Journal Supplement Series* 69, p. 593. DOI: 10.1086/191322.

López-Corredoira, M. and J. Molgó (2014). "Flare in the Galactic stellar outer disc detected in SDSS-SEGUE data." In: A&A 567, A106, A106. DOI: 10.1051/0004-6361/201423706. arXiv: 1405.7649 [astro-ph.GA].

LSST Science Collaboration, Paul A. Abell, Julius Allison, Scott F. Anderson, John R. Andrew, J. Roger P. Angel, et al. (2009). "LSST Science Book, Version 2.0." In: *arXiv e-prints*, arXiv:0912.0201, arXiv:0912.0201. arXiv: 0912.0201 [astro-ph.IM].

Macri, Lucas M., Chow-Choong Ngeow, Shashi M. Kanbur, Salma Mahzooni, and Michael T. Smitka (2015). "Large Magellanic Cloud Near-Infrared Synoptic Survey. I. Cepheid Variables and the Calibration of the Leavitt Law." In: AJ 149.4, 117, p. 117. DOI: 10.1088/0004-6256/149/4/117. arXiv: 1412.1511 [astro-ph.SR].

— (2016). "Erratum: "Large Magellanic Cloud Near-infrared Synoptic Survey. I. Cepheid Variables and the Calibration of the Leavitt Law" <A href="/abs/2015149...117.....M">(2015 AJ, 149, 117)</A>." In: AJ 151.2, 48, p. 48. DOI: 10.3847/0004-6256/151/2/48.

Madore, Barry F. and Wendy L. Freedman (1991). "The Cepheid Distance Scale." In: PASP 103, p. 933. DOI: 10.1086/132911.

Magnier, E. A., E. Schlafly, D. Finkbeiner, M. Juric, J. L. Tonry, W. S. Burgett, et al. (2013). "The Pan-STARRS 1 Photometric Reference Ladder, Release 12.01." In: *The Astrophysical Journal Supplement Series* 205.2, 20, p. 20. DOI: 10.1088/0067-0049/205/2/20. arXiv: 1303.3634 [astro-ph.IM].

Majaess, D. J., D. G. Turner, and D. J. Lane (2009). "Characteristics of the Galaxy according to Cepheids." In: MNRAS 398.1, pp. 263–270. DOI: 10.1111/j.1365-2966.2009.15096.x. arXiv: 0903.4206 [astro-ph.GA].

Majaess, D., I. Dékány, G. Hajdu, D. Minniti, D. Turner, and W. Gieren (2018). "Establishing the Galactic Centre distance using VVV Bulge RR Lyrae variables." In: Ap&SS 363.6, 127, p. 127. DOI: 10.1007/s10509-018-3346-6. arXiv: 1805.04119 [astro-ph.GA].

Majaess, D., D. Turner, I. Dékány, D. Minniti, and W. Gieren (2016). "Constraining dust extinction properties via the VVV survey." In: A&A 593, A124, A124. DOI: 10.1051/0004-6361/201628763. arXiv: 1607.08623 [astro-ph.GA].

Marconi, M., G. Coppola, G. Bono, V. Braga, A. Pietrinferni, R. Buonanno, et al. (2015). "On a New Theoretical Framework for RR Lyrae Stars. I. The Metallicity Dependence." In: ApJ 808, 50, p. 50. DOI: 10.1088/0004-637X/808/1/50. arXiv: 1505.02531 [astro-ph.SR].

Matsunaga, Noriyuki, Michael W. Feast, Giuseppe Bono, Naoto Kobayashi, Laura Inno, Takahiro Nagayama, et al. (2016). "A lack of classical Cepheids in the inner part of the Galactic disc." In: MNRAS 462.1, pp. 414–420. DOI: 10.1093/mnras/stw1548. arXiv: 1606.07943 [astro-ph.GA].

Matsunaga, Noriyuki, Michael W. Feast, and John W. Menzies (2009). "Period-luminosity relations for type II Cepheids and their application." In: MNRAS 397.2, pp. 933–942. DOI: 10.1111/j.1365-2966.2009.14992.x. arXiv: 0904.4701 [astro-ph.SR].

Matsunaga, Noriyuki, Michael W. Feast, and Igor Soszyński (2011). "Period-luminosity relations of type II Cepheids in the Magellanic Clouds." In: MNRAS 413.1, pp. 223–234. DOI: 10.1111/j.1365-2966.2010.18126.x. arXiv: 1012.0098 [astro-ph.SR].

Matsunaga, Noriyuki, Hinako Fukushi, Yoshikazu Nakada, Toshihiko Tanabé, Michael W. Feast, John W. Menzies, et al. (2006). "The period-luminosity relation for type II Cepheids in globular clusters." In: MNRAS 370.4, pp. 1979–1990. DOI: 10.1111/j.1365-2966.2006.10620.x. arXiv: astro-ph/0606609 [astro-ph].

McPherson, Alistair M., Andrew Born, Will Sutherland, Jim Emerson, Bryan Little, Paul Jeffers, et al. (2006). "VISTA: project status." In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Vol. 6267, p. 626707. DOI: 10.1117/12.671352.

Mel'nik, A. M., P. Rautiainen, L. N. Berdnikov, A. K. Dambis, and A. S. Rastorguev (2015). "Classical Cepheids in the Galactic outer ring R_1R_2'." In: *Astronomische Nachrichten* 336.1, p. 70. DOI: 10.1002/asna.201412137. arXiv: 1411.3558 [astro-ph.GA].

Minniti, D., P. W. Lucas, J. P. Emerson, R. K. Saito, M. Hempel, P. Pietrukowicz, et al. (2010). "VISTA Variables in the Via Lactea (VVV): The public ESO near-IR variability survey of the Milky Way." In: *New A* 15, pp. 433–443. DOI: 10.1016/j.newast.2009.12.002. arXiv: 0912.1056 [astro-ph.GA].

Minniti, Dante (2018). "Mapping the Milky Way in the Near-IR: The Future of the VVV Survey." In: *The Vatican Observatory, Castel Gandolfo: 80th Anniversary Celebration*. Vol. 51, p. 63. DOI: 10.1007/978-3-319-67205-2_4.

Minniti, Dante, István Dékány, Daniel Majaess, Tali Palma, Joyce Pullen, Marina Rejkuba, et al. (2017a). "Characterization of the VVV Survey RR Lyrae Population across the

Southern Galactic Plane." In: AJ 153, 179, p. 179. DOI: 10.3847/1538-3881/aa5be4. arXiv: 1703.01711 [astro-ph.GA].

Minniti, Dante, Tali Palma, István Dékány, Maren Hempel, Marina Rejkuba, Joyce Pullen, et al. (2017b). "FSR 1716: A New Milky Way Globular Cluster Confirmed Using VVV RR Lyrae Stars." In: ApJ 838, L14, p. L14. DOI: 10.3847/2041-8213/838/1/L14. arXiv: 1703.02033 [astro-ph.GA].

Monson, Andrew J. and Michael J. Pierce (2011). "Near-infrared (JHK) Photometry of 131 Northern Galactic Classical Cepheids." In: ApJS 193.1, 12, p. 12. DOI: 10.1088/0067-0049/193/1/12.

Morgan, Siobahn M., Jennifer N. Wahl, and Rachel M. Wieckhorst (2007). "[Fe/H] relations for c-type RR Lyrae variables based upon Fourier coefficients." In: MNRAS 374.4, pp. 1421–1426. DOI: 10.1111/j.1365-2966.2006.11247.x. arXiv: astro-ph/0506348 [astro-ph].

Morgan, W. W., S. Sharpless, and D. Osterbrock (1952). "Some features of galactic structure in the neighborhood of the Sun." In: AJ 57, pp. 3–3. DOI: 10.1086/106673.

Mróz, Przemek, Andrzej Udalski, Dorota M. Skowron, Jan Skowron, Igor Soszyński, Paweł Pietrukowicz, et al. (2019). "Rotation Curve of the Milky Way from Classical Cepheids." In: ApJ 870.1, L10, p. L10. DOI: 10.3847/2041-8213/aaf73f. arXiv: 1810.02131 [astro-ph.GA].

Nataf, David M., Oscar A. Gonzalez, Luca Casagrande, Gail Zasowski, Christopher Wegg, Christian Wolf, et al. (2016). "Interstellar extinction curve variations towards the inner Milky Way: a challenge to observational cosmology." In: MNRAS 456.3, pp. 2692–2706. DOI: 10.1093/mnras/stv2843. arXiv: 1510.01321 [astro-ph.SR].

Navarrete, C., R. Contreras Ramos, M. Catelan, C. M. Clement, F. Gran, J. Alonso-García, et al. (2015). "Updated census of RR Lyrae stars in the globular cluster ω Centauri (NGC 5139)." In: A&A 577, A99, A99. DOI: 10.1051/0004-6361/201424838. arXiv: 1501.02286 [astro-ph.SR].

Nemec, J. M., J. G. Cohen, V. Ripepi, A. Derekas, P. Moskalik, B. Sesar, et al. (2013). "Metal Abundances, Radial Velocities, and Other Physical Characteristics for the RR Lyrae Stars in The Kepler Field." In: ApJ 773, 181, p. 181. DOI: 10.1088/0004-637X/773/2/181. arXiv: 1307.5820 [astro-ph.SR].

Ness, M., K. Freeman, E. Athanassoula, E. Wylie-de-Boer, J. Bland-Hawthorn, M. Asplund, et al. (2013). "ARGOS - III. Stellar populations in the Galactic bulge of the Milky Way." In: MNRAS 430.2, pp. 836–857. DOI: 10.1093/mnras/sts629. arXiv: 1212.1540 [astro-ph.GA].

Neumann, J. von, R. H. Kent, H. R. Bellinson, and B. I. Hart (1941). "The Mean Square Successive Difference." In: *Ann. Math. Statist.* 12.2, pp. 153–162. DOI: 10.1214/aoms/1177731746. URL: https://doi.org/10.1214/aoms/1177731746.

Nishiyama, Shogo, Tetsuya Nagata, Nobuhiko Kusakabe, Noriyuki Matsunaga, Takahiro Naoi, Daisuke Kato, et al. (2006). "Interstellar Extinction Law in the J, H, and K_s Bands toward the Galactic Center." In: ApJ 638.2, pp. 839–846. DOI: 10.1086/499038. arXiv: astro-ph/0601174 [astro-ph].

Nishiyama, Shogo, Motohide Tamura, Hirofumi Hatano, Daisuke Kato, Toshihiko Tanabé, Koji Sugitani, et al. (2009). "Interstellar Extinction Law Toward the Galactic Center III: J, H, K_S Bands in the 2MASS and the MKO Systems, and 3.6, 4.5, 5.8, 8.0 μm in the Spitzer/IRAC System." In: ApJ 696.2, pp. 1407–1417. DOI: 10.1088/0004-637X/696/2/1407. arXiv: 0902.3095 [astro-ph.GA].

Oort, J. H., F. J. Kerr, and G. Westerhout (1958). "The galactic system as a spiral nebula (Council Note)." In: MNRAS 118, p. 379. DOI: 10.1093/mnras/118.4.379.

Oosterhoff, P. T. (1939). "Some remarks on the variable stars in globular clusters." In: *The Observatory* 62, pp. 104–109.

Paraficz, D., F. Courbin, A. Tramacere, R. Joseph, R. B. Metcalf, J. -P. Kneib, et al. (2016). "The PCA Lens-Finder: application to CFHTLS." In: A&A 592, A75, A75. DOI: 10.1051/0004-6361/201527971. arXiv: 1605.04309 [astro-ph.GA].

Pearson, K. (1901). "On lines and planes of closest fit to points in space." In: *Philosophical Magazine* 2, pp. 559–572.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al. (2011). "Scikit-learn: Machine Learning in Python." In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Persic, Massimo, Paolo Salucci, and Fulvio Stel (1996). "The universal rotation curve of spiral galaxies - I. The dark matter connection." In: MNRAS 281.1, pp. 27–47. DOI: 10.1093/mnras/281.1.27. arXiv: astro-ph/9506004 [astro-ph].

Persson, S. E., Barry F. Madore, W. Krzemiński, Wendy L. Freedman, M. Roth, and D. C. Murphy (2004). "New Cepheid Period-Luminosity Relations for the Large Magellanic Cloud: 92 Near-Infrared Light Curves." In: AJ 128.5, pp. 2239–2264. DOI: 10.1086/424934.

Pickering, E. C., H. R. Colson, W. P. Fleming, and L. D. Wells (1901). "Sixty-four new variable stars." In: ApJ 13. DOI: 10.1086/140808.

Pietrzyński, G., D. Graczyk, A. Gallenne, W. Gieren, I. B. Thompson, B. Pilecki, et al. (2019). "A distance to the Large Magellanic Cloud that is precise to one per cent." In: Nature 567.7747, pp. 200–203. DOI: 10.1038/s41586-019-0999-4. arXiv: 1903.08096 [astro-ph.GA].

Preston, George W. (1959). "A Spectroscopic Study of the RR Lyrae Stars." In: ApJ 130, p. 507. DOI: 10.1086/146743.

Prudil, Z. and M. Skarka (2017). "Blazhko effect in the Galactic bulge fundamental mode RR Lyrae stars - I. Incidence rate and differences between modulated and non-modulated stars." In: MNRAS 466, pp. 2602–2613. DOI: 10.1093/mnras/stw3231. arXiv: 1701.00782 [astro-ph.SR].

Reshetnikov, Vladimir P., Aleksandr V. Mosenkov, Alexei V. Moiseev, Sergey S. Kotov, and Sergey S. Savchenko (2016). "Galaxies with conspicuous optical warps." In: MNRAS 461.4, pp. 4233–4245. DOI: 10.1093/mnras/stw1623. arXiv: 1607.02274 [astro-ph.GA].

Reylé, C., D. J. Marshall, A. C. Robin, and M. Schultheis (2009). "The Milky Way's external disc constrained by 2MASS star counts." In: A&A 495.3, pp. 819–826. DOI: 10.1051/0004-6361/200811341. arXiv: 0812.3739 [astro-ph].

Reyner, S., S. M. Kanbur, C. Ngeow, and C. Morgan (2010). "Approximating RR Lyrae light curves using cubic polynomials." In: MNRAS 407, pp. 1801–1806. DOI: 10.1111/j.1365-2966.2010.17017.x. arXiv: 1005.2345 [astro-ph.SR].

Riess, Adam G., Stefano Casertano, Wenlong Yuan, Lucas Macri, Beatrice Bucciarelli, Mario G. Lattanzi, et al. (2018). "Milky Way Cepheid Standards for Measuring Cosmic Distances and Application to Gaia DR2: Implications for the Hubble Constant." In: ApJ 861.2, 126, p. 126. DOI: 10.3847/1538-4357/aac82e. arXiv: 1804.10655 [astro-ph.CO].

Ripepi, V., M. I. Moretti, M. Marconi, G. Clementini, M. -R. L. Cioni, J. B. Marquette, et al. (2012). "The VMC survey - V. First results for classical Cepheids." In: MNRAS 424.3, pp. 1807–1816. DOI: 10.1111/j.1365-2966.2012.21274.x. arXiv: 1204.2273 [astro-ph.CO].

Ripepi, Vincenzo, Maria-Rosa L. Cioni, Maria Ida Moretti, Marcella Marconi, Kenji Bekki, Gisella Clementini, et al. (2017). "The VMC survey - XXV. The 3D structure of the Small Magellanic Cloud from Classical Cepheids." In: MNRAS 472.1, pp. 808–827. DOI: 10.1093/mnras/stx2096. arXiv: 1707.04500 [astro-ph.GA].

Sanna, Alberto, Mark J. Reid, Thomas M. Dame, Karl M. Menten, and Andreas Brunthaler (2017). "Mapping spiral structure on the far side of the Milky Way." In: *Science* 358.6360, pp. 227–230. DOI: 10.1126/science.aan5452. arXiv: 1710.06489 [astro-ph.GA].

Schlafly, E. F., D. P. Finkbeiner, M. Jurić, E. A. Magnier, W. S. Burgett, K. C. Chambers, et al. (2012). "Photometric Calibration of the First 1.5 Years of the Pan-STARRS1 Survey." In: ApJ 756.2, 158, p. 158. DOI: 10.1088/0004-637X/756/2/158. arXiv: 1201.2208 [astro-ph.IM].

Schlafly, E. F., A. M. Meisner, A. M. Stutz, J. Kainulainen, J. E. G. Peek, K. Tchernyshyov, et al. (2016). "The Optical-infrared Extinction Curve and Its Variation in the Milky Way." In: ApJ 821.2, 78, p. 78. DOI: 10.3847/0004-637X/821/2/78. arXiv: 1602.03928 [astro-ph.GA].

Schlafly, E. F., J. E. G. Peek, D. P. Finkbeiner, and G. M. Green (2017). "Mapping the Extinction Curve in 3D: Structure on Kiloparsec Scales." In: ApJ 838.1, 36, p. 36. DOI: 10.3847/1538-4357/aa619d. arXiv: 1612.02818 [astro-ph.GA].

Schlegel, David J., Douglas P. Finkbeiner, and Marc Davis (1998). "Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds." In: ApJ 500.2, pp. 525–553. DOI: 10.1086/305772. arXiv: astro-ph/9710327 [astro-ph].

Schwarzschild, Martin (1940). "On the variables in Messier 3." In: *Harvard College Observatory Circular* 437, pp. 1–12.

Sesar, Branimir, Nina Hernitschek, Sandra Mitrović, Željko Ivezić, Hans-Walter Rix, Judith G. Cohen, et al. (2017). "Machine-learned Identification of RR Lyrae Stars from

Sparse, Multi-band Data: The PS1 Sample." In: AJ 153.5, 204, p. 204. DOI: 10.3847/1538-3881/aa661b. arXiv: 1611.08596 [astro-ph.GA].

Sesar, Branimir, Željko Ivezić, J. Scott Stuart, Dylan M. Morgan, Andrew C. Becker, Sanjib Sharma, et al. (2013). "Exploring the Variable Sky with LINEAR. II. Halo Structure and Substructure Traced by RR Lyrae Stars to 30 kpc." In: AJ 146.2, 21, p. 21. DOI: 10.1088/0004-6256/146/2/21. arXiv: 1305.2160 [astro-ph.GA].

Simon, N. R. and A. S. Lee (1981). "The structural properties of cepheid light curves." In: ApJ 248, pp. 291–297. DOI: 10.1086/159153.

Singh, Harinder P., Ravi K. Gulati, and Ranjan Gupta (1998). "Stellar Spectral Classification using Principal Component Analysis and Artificial Neural Networks." In: MNRAS 295, pp. 312–318. DOI: 10.1046/j.1365-8711.1998.01255.x.

Skillen, I., J. A. Fernley, R. S. Stobie, and R. F. Jameson (1993). "The absolute magnitudes of RR Lyrae stars - V. WY Antliae, W Crateris, RV Octantis and BB Puppis." In: MNRAS 265, pp. 301–315. DOI: 10.1093/mnras/265.2.301.

Skowron, D. M., J. Skowron, P. Mróz, A. Udalski, P. Pietrukowicz, I. Soszyński, et al. (2018). "A New 3-D Picture of the Milky Way with Classical Cepheids." In: *arXiv e-prints*, arXiv:1806.10653, arXiv:1806.10653. arXiv: 1806.10653 [astro-ph.GA].

Skowron, D. M., I. Soszyński, A. Udalski, M. K. Szymański, P. Pietrukowicz, J. Skowron, et al. (2016). "OGLE-ing the Magellanic System: Photometric Metallicity from Fundamental Mode RR Lyrae Stars." In: *Acta Astron.* 66, pp. 269–292. arXiv: 1608.00013 [astro-ph.SR].

Skrutskie, M. F., R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, et al. (2006). "The Two Micron All Sky Survey (2MASS)." In: AJ 131.2, pp. 1163–1183. DOI: 10.1086/498708.

Smith, Horace A. (2004). *RR Lyrae Stars*.

Smith, L. C., P. W. Lucas, R. Kurtev, R. Smart, D. Minniti, J. Borissova, et al. (2018). "VIRAC: the VVV Infrared Astrometric Catalogue." In: MNRAS 474.2, pp. 1826–1849. DOI: 10.1093/mnras/stx2789. arXiv: 1710.08919 [astro-ph.SR].

Smolec, R. (2005). "Metallicity Dependence of the Blazhko Effect." In: *Acta Astron.* 55, pp. 59–84. arXiv: astro-ph/0503614 [astro-ph].

Sollima, A., J. Borissova, M. Catelan, H. A. Smith, D. Minniti, C. Cacciari, et al. (2006). "New Metallicities of RR Lyrae Stars in ω Centauri: Evidence for a Non-He-enhanced Metal-intermediate Population." In: ApJ 640, pp. L43–L46. DOI: 10.1086/503099. arXiv: astro-ph/0602055 [astro-ph].

Soszyński, I., R. Poleski, A. Udalski, M. K. Szymański, M. Kubiak, G. Pietrzyński, et al. (2010). "The Optical Gravitational Lensing Experiment. The OGLE-III Catalog of Variable Stars. VII. Classical Cepheids in the Small Magellanic Cloud." In: *Acta Astron.* 60.1, pp. 17–39. arXiv: 1003.4518 [astro-ph.SR].

Soszyński, I., A. Udalski, M. K. Szymański, P. Pietrukowicz, P. Mróz, J. Skowron, et al. (2014). "Over 38000 RR Lyrae Stars in the OGLE Galactic Bulge Fields." In: *Acta Astron.* 64, pp. 177–196. arXiv: 1410.1542 [astro-ph.SR].

Soszyński, I., A. Udalski, M. K. Szymański, Ł. Wyrzykowski, K. Ulaczyk, R. Poleski, et al. (2017). "The OGLE Collection of Variable Stars. Classical, Type II, and Anomalous Cepheids toward the Galactic Center." In: *Acta Astron.* 67.4, pp. 297–316. DOI: 10.32023/0001-5237/67.4.1. arXiv: 1712.01307 [astro-ph.SR].

Stetson, Peter B. (1996). "On the Automatic Determination of Light-Curve Parameters for Cepheid Variables." In: PASP 108, p. 851. DOI: 10.1086/133808.

Suntzeff, N. B., R. P. Kraft, and T. D. Kinman (1994). "Summary of Delta S metallicity measurements for bright RR Lyrae variables observed at Lick Observatory and KPNO between 1972 and 1987." In: ApJS 93, pp. 271–285. DOI: 10.1086/192055.

Szabó, Róbert, Željko Ivezić, László L. Kiss, Zoltán Kolláth, Lynne Jones, Branimir Sesar, et al. (2014). "High-precision 2MASS JHK$_s$ Light Curves and Other Data for RR Lyrae Star SDSS J015450 + 001501: Strong Constraints for Nonlinear Pulsation Models." In: ApJ 780, 92, p. 92. DOI: 10.1088/0004-637X/780/1/92. arXiv: 1311.1385 [astro-ph.SR].

Tanvir, N. R., M. A. Hendry, A. Watkins, S. M. Kanbur, L. N. Berdnikov, and C. C. Ngeow (2005). "Determination of Cepheid parameters by light-curve template fitting." In: MNRAS 363, pp. 749–762. DOI: 10.1111/j.1365-2966.2005.09466.x. arXiv: astro-ph/0504408 [astro-ph].

Taylor, M. B. (2006). "STILTS - A Package for Command-Line Processing of Tabular Data." In: *Astronomical Data Analysis Software and Systems XV*. Ed. by C. Gabriel,

C. Arviset, D. Ponz, and S. Enrique. Vol. 351. Astronomical Society of the Pacific Conference Series, p. 666.

Tibshirani, R. (1996). "Regression shrinkage and selection via the Lasso." In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288. ISSN: 0035-9246.

Torrealba, G., M. Catelan, A. J. Drake, S. G. Djorgovski, R. H. McNaught, V. Belokurov, et al. (2015). "Discovery of 9000 new RR Lyrae in the southern Catalina surveys." In: MNRAS 446.3, pp. 2251–2266. DOI: 10.1093/mnras/stu2274. arXiv: 1410.7653 [astro-ph.GA].

Turner, David G. (1996). "The Progenitors of Classical Cepheid Variables." In: JRASC 90, p. 82.

Udalski, A., I. Soszyński, P. Pietrukowicz, M. K. Szymański, D. M. Skowron, J. Skowron, et al. (2018). "OGLE Collection of Galactic Cepheids." In: *Acta Astron.* 68.4, pp. 315–339. DOI: 10.32023/0001-5237/68.4.1. arXiv: 1810.09489 [astro-ph.SR].

Udalski, A., M. K. Szymański, and G. Szymański (2015). "OGLE-IV: Fourth Phase of the Optical Gravitational Lensing Experiment." In: *Acta Astron.* 65, pp. 1–38. arXiv: 1504.05966 [astro-ph.SR].

Valcarce, A. A. R. and M. Catelan (2011). "Formation of multiple populations in globular clusters: another possible scenario." In: A&A 533, A120, A120. DOI: 10.1051/0004-6361/201116955. arXiv: 1106.6082 [astro-ph.SR].

Vallée, Jacques P. (2017). "A guided map to the spiral arms in the galactic disk of the Milky Way." In: *The Astronomical Review* 13.3-4, pp. 113–146. DOI: 10.1080/21672857.2017.1379459. arXiv: 1711.05228 [astro-ph.GA].

Vázquez, R. A., J. May, G. Carraro, L. Bronfman, A. Moitinho, and G. Baume (2008). "Spiral Structure in the Outer Galactic Disk. I. The Third Galactic Quadrant." In: ApJ 672, pp. 930–939. DOI: 10.1086/524003. arXiv: 0709.3973.

Villalobos, Álvaro and Amina Helmi (2008). "Simulations of minor mergers - I. General properties of thick discs." In: MNRAS 391.4, pp. 1806–1827. DOI: 10.1111/j.1365-2966.2008.13979.x.

Walker, Alistair R. and Donald M. Terndrup (1991). "The Metallicity of RR Lyrae Stars in Baade's Window." In: ApJ 378, p. 119. DOI: 10.1086/170411.

Welch, D. L., F. Wieland, C. W. McAlary, R. McGonegal, B. F. Madore, R. A. McLaren, et al. (1984). "JHK observations of classical Cepheids." In: ApJS 54, pp. 547–579. DOI: 10.1086/190943.

Welch, Douglas L., Mario Mateo, Patrick Cote, Philippe Fischer, and Barry F. Madore (1991). "The Variable Stars of NGC 1866." In: AJ 101, p. 490. DOI: 10.1086/115700.

Yip, C. W., A. J. Connolly, D. E. Vanden Berk, Z. Ma, J. A. Frieman, M. SubbaRao, et al. (2004). "Spectral Classification of Quasars in the Sloan Digital Sky Survey: Eigenspectra, Redshift, and Luminosity Effects." In: AJ 128, pp. 2603–2630. DOI: 10.1086/425626. arXiv: astro-ph/0408578 [astro-ph].

Zechmeister, M. and M. Kürster (2009). "The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms." In: A&A 496.2, pp. 577–584. DOI: 10.1051/0004-6361:200811296. arXiv: 0901.2573 [astro-ph.IM].

Zoccali, M., E. Valenti, and O. A. Gonzalez (2018). "Weighing the two stellar components of the Galactic bulge." In: A&A 618, A147, A147. DOI: 10.1051/0004-6361/201833147. arXiv: 1807.06377 [astro-ph.GA].