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

Put forward by
Steffi Xiang-Ting Yen
born in: Washington D.C., USA
Oral examination: 9 December 2019

Peering into the Milky Way disk:
Gaia's Perspective of our Galaxy's Open Clusters

Referees: Priv. Doz. Dr. Sabine Reffert
Prof. Dr. Eva Grebel

Abstract

In the pre-Gaia era, one of the largest collections of open clusters and their parameters was the Milky Way Star Cluster (MWSC) catalog, which consisted of 2808 open clusters. This sample was nearly complete up to a distance of about 1.8 kpc from the Sun, with the exception of a subset of old nearby clusters. However, with the unprecedented precision of astrometric and photometric data from Gaia, a more accurate census on the number of true open clusters can be achieved, also providing improved cluster parameters and detection of new clusters. With this aim, I developed an automated cluster characterization pipeline to consistently determine cluster membership from astrometry and cluster parameters via isochrone fitting to cluster multi-band photometry. Using Gaia DR1/TGAS data, I analyzed 24 nearby open clusters and found evidence for the non-existence of some clusters. With Gaia DR2 data, I reanalyzed the full MWSC cluster sample, successfully obtaining cluster memberships and parameters for 1873 clusters and denying the existence of 912 clusters. This is the first study, to date, to homogeneously analyze the largest catalog of open clusters. My results show that unlike previously thought, the open cluster census is very incomplete, even at the smallest distances, and thus a dedicated search for new clusters is required in order to gain a full understanding of the open cluster population.

Zusammenfassung

In der Prä-Gaia Ära war der Katalog der offenen Sternhaufen in der Milchstraße (MWSC, engl. Milky Way Star Clusters) eine der größten Sammlungen von offenen Sternhaufen und deren Kennwerten. Er beinhaltete 2808 offene Sternhaufen und war, mit Ausnahme einiger alter, naher Sternhaufen, fast vollständig bis zu Entfernungen von etwa 1,8 kpc von der Sonne. Mit der bis dahin unerreicht hohen astrometrischen und photometrischen Präzision der Gaiadaten kann ein genauerer Zensus der Zahl der wahren offenen Sternhaufen erreicht werden, welcher auch verbesserte Parameter sowie die Entdeckung neuer Sternhaufen bereit hält. Mit diesem Ziel habe ich eine automatische Routine geschrieben, welche einheitlich die stellaren Mitglieder mittels Astrometrie sowie die Parameter mittels Anpassen der Isochronen an Multibandphotometrie bestimmt. Mit den Gaia DR1/TGAS Daten habe ich 24 nahe offene Sternhaufen analysiert und Belege für deren Inexistenz gefunden. Mit den Gaia DR2 Daten habe ich den kompletten MWSC Katalog neu analysiert und erfolgreich stellare Mitglieder und Parameter für 1873 Sternhaufen bestimmt sowie die Existenz von 912 Sternhaufen in Abrede gestellt. Dies ist bis heute die erste Studie, welche den größten Katalog offener Sternhaufen einheitlich analysiert. Meine Ergebnisse zeigen, dass entgegen vorheriger Meinungen der Zensus der offenen Sternhaufen sehr unvollständig ist – selbst bei kleinsten Entfernungen. Daher ist eine dedizierte Suche nach offenen Sternhaufen nötig um ein vollständiges Verständnis des Bestands der offenen Sternhaufen zu erhalten.

To my parents

"Be glad of life because it gives you the chance to love, to work, to play, and to look up at the stars."

– Henry van Dyke

Contents

Listing of Figures

Listing of tables

1

Open Clusters in the Pre-Gaia Era

"Look up." Gazing up at the night sky, your eyes fall upon one of the most spectacular wonders of our world: a glistening sea of stars, spread across the largest and darkest black canvas. As you gaze in amazement, your eyes start to connect some of the brightest stars, making shapes you learned in geometry class and simple representations of things in our every day life, like animals, people, and objects. *"Can you trace out a bear or bull?"* Among one of these constellations, you notice a tight bundle of bright stars. *"Look there, that cluster of stars is the Pleiades."* These collections of stars are known as open star clusters.

Most stars form in embedded aggregates within dense molecular clouds (Lada & Lada 2003). If the star formation efficiency and time scale of gas dispersal are favorable and if the aggregate is massive enough, then it will emerge from its parent molecular cloud as a bound open cluster. Open clusters lie within the disk of its host galaxy, exposing them to a hostile surrounding environment. Experiencing internal dynamical evolution, as well as tidal interactions within the galactic disk, the open cluster will inevitably dissolve over a few hundred million years. As a cluster becomes unbound, it releases its stars into the field, contributing to the stellar population of the galactic disk.

1.1 Characteristics of open clusters

Open clusters are groups of up to 10^5 population I stars, loosely bound by their mutual gravitational attraction. They are irregularly-shaped, roughly up to 20 parsecs in diameter, with an average stellar density of $\sim 0.1 - 10$ stars/pc^3. Located in the galactic disk, the stars of an open cluster will stay together for up to roughly 10^9 years.

Classical identifications of open clusters relied on star counts, which involve counting stars and comparing the density of cluster stars to the density of the surround field stars. However, developments in identification methods have since emerged, focusing on calculating membership probabilities based on proper motions (e.g., Vasilevskis et al. 1958, Sanders 1971, Cabrera-Caño & Alfaro 1985). Photometry has also become a powerful resource in distinguishing between members and non-members of a cluster (e.g., Vogt 1971, Moffat 1972, Baade 1983).

1.2 Open cluster parameters: Techniques

Fundamental parameters such as age, distance, reddening, and metallicity can be determined more accurately for an open cluster than for a single star. The most common method for deriving the fundamental cluster parameters is main-sequence fitting, in which theoretical isochrones are compared to the cluster main sequence in color-magnitude diagrams (e.g. Chaboyer et al. 1996).

An isochrone is a stellar model representing stars of the same age across a range of masses; it is the opposite of an evolutionary track, which represents stars of the same mass, but varying ages. Various isochrone sets are available in the literature, as each include different stellar physics. Netopil et al. (2015) compared three commonly used isochrone sets in the optical photometric bands, shown in Fig. 1.1, and concluded that while slight differences exist, for example, the discrepancies in absolute V magnitudes for the blue hook and subgiant branch, overall, the isochrones are compatible and should not lead to significant differences in age determination. This thesis uses the most recent Padova isochrone set, PARSEC version 1.2S (Bressan et al. 2012), in multiple photometric bands is used.

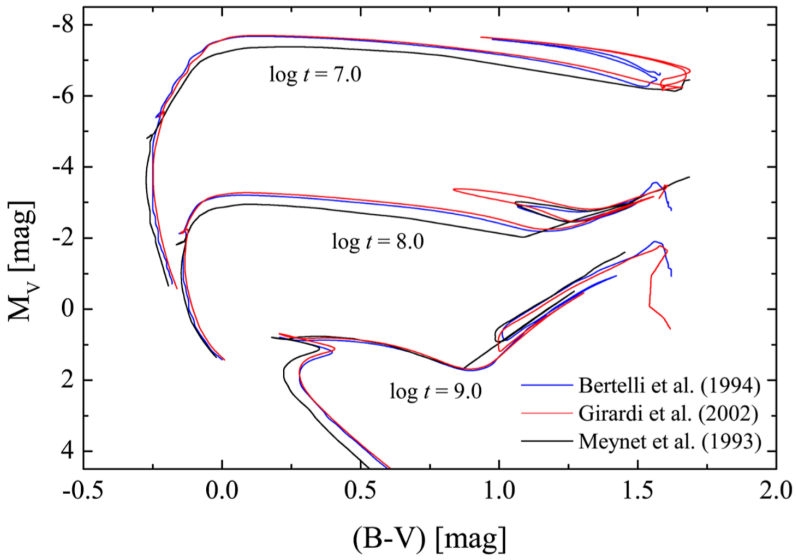

Figure 1.1: Comparison of three isochrone sets in optical photometry from Netopil et al. (2015).

Age

In a color-magnitude diagram (CMD), hot blue stars lie to the left and cool red stars to the right. The main sequence, where stars spend about 90% of their lifetime, stretches diagonally from the lower right to the upper left of the diagram. In this phase, stars are converting hydrogen to helium in their cores. Once the hydrogen in the core has been used up, the core begins to contract, fusing hydrogen in a shell around an inert core. This process causes the envelope to expand and cool, causing the star to move to the right, off of the main sequence; this is known as the main sequence turn off point. The more massive stars, initially located in the upper left portion of the main sequence burn their fuel faster, leaving the main sequence earlier, and so, this point reveals the age of the star. Thus, the age of a cluster can be determined by matching an isochrone to the cluster sequence in the CMD, especially to the brightest and most massive star(s) on the main sequence, in this turn-off regime.

Distance

While cluster distances can be computed directly from trigonometric parallax measurements, the classical approach to deriving cluster distances also rely on the magnitude of the stars. The distance modulus, $(m - M)$, is the difference between the

apparent magnitude m and the absolute magnitude M of an astronomical object, and also gives the distance d in parsecs to the object by the following equation:

$$m - M = 5 \log_{10}(d/10). \tag{1.1}$$

The expression on the right for the distance modulus is derived by using the flux ratio between to stars and their apparent magnitudes.

Reddening

The apparent magnitude of stars also need to be corrected from effects of interstellar absorption. The light emitted by stars will encounter many dust grains in the interstellar medium as they travel through the universe. These dust grains have diameters comparable to the wavelength of blue light, which means they absorb and scatter the blue light of stars, making the starlight we observe appear redder. The amount of reddening of an object is denoted by $E(B-V)$ and is quantified by comparing the color index $(B-V)$ of an object to its true color index $(B-V)_0$:

$$E(B-V) = (B-V) - (B-V)_0. \tag{1.2}$$

The scattering of light by dust grains in the interstellar medium also causes distant stars to appear dimmer, an effect known as extinction A. Linked with the effects of reddening, the more reddened a star, the larger the extinction and the dimmer its appearance. Eq. 1.2 can be rewritten with respect to the extinction in each band: $E(B-V) = A_B - A_V$, where A_B and A_V are the total extinction in the B and V bands. Together, these effects are related via the total-to-selective extinction ratio R_V, expressed as:

$$R_V = \frac{A_V}{E(B-V)}, \tag{1.3}$$

where the subscript V denotes measurements in the visual band. In the Milky Way, the average value for R_V is 3.1 (Schultz & Wiemer 1975).

To account for the loss of light caused by extinction, the magnitude must be brightened, so Eq. 1.1 becomes:

$$d = 10^{0.2(m_V - M_V + 5 - A_V)}. \tag{1.4}$$

Traditionally, main-sequence fitting is manually accomplished with a series of

steps requiring visual checks of CMDs. First, reddening is determined by adjusting the zero age main sequence to the observed color-color diagram (e.g. $(B-V)$ vs. $(U-B)$) of a cluster (Monteiro et al. 2010). Then, with the reddening fixed, the distance and age are adjusted to match the cluster sequence in a CMD and to an isochrone. This fit-by-eye approach has been used because isochrones do not have a simple parametric form, allowing a typical least square technique to be applied (Monteiro et al. 2010).

1.3 Catalogs of cluster parameters

In order to use open clusters as probes of the Galactic disk, a large homogenous catalog of cluster parameters is required. The Lund catalogue was the first compilation of open clusters and their parameters from the literature (Lynga 1982). It consisted of roughly 1200 clusters, of which 400 had heterogeneous estimates of cluster parameters. The WEBDA open cluster database (Mermilliod 1993) emerged next and included most of the information from the Lund catalogue, as well as data for the cluster stars. The Dias et al. (2002) catalog extended the Lund catalogue even further, including parameters for hundreds of optically visible clusters. While these catalogs were great steps forward in working towards a comprehensive assessment of the Galactic cluster population, the cluster parameters came from individual studies, which used observations from various telescopes and a variety of methods to derive cluster parameters, resulting in a heterogeneous parameter set, which caused uncertainty in cluster statistics.

The wealth of information from large, all-sky photometric surveys led to the creation of homogeneous cluster catalogs, where the methods used to derive membership and cluster parameters combine the techniques aforementioned with some optimization of the results by eye. The first catalogs deriving homogenous cluster parameters relied on optical photometry, including the works by Platais et al. (1998), Chereul et al. (1999), Dias et al. (2002), and Kharchenko et al. (2005a,b). Infrared photometry from the 2MASS (Skrutskie et al. 2006) survey then allowed for the discovery of many new, previously undetected open clusters (e.g., Dutra & Bica 2001, Bica et al. 2003, Dutra et al. 2003, Koposov et al. 2008, Glushkova et al. 2010, Bukowiecki et al. 2011). The Milky Way Star Clusters (MWSC) catalog (Kharchenko et al. 2013) is the largest-to-date census on the star cluster populations within the Milky Way. Starting with an input list of 3784 objects compiled from the literature, it used PPMXL (Röser et al. 2010) and 2MASS data to derive

cluster parameters for a total of 3006 objects. For nearly half of the cluster sample, it provides fundamental cluster parameters for the very first time. Comprised of 2808 open clusters, it was determined to be nearly complete to ∼1.8 kpc[1] from the Sun.

1.4 Open cluster parameters: Challenges

While cluster catalogs contain a wealth of information that can provide insights into Milky Way disk, the reliability of their results have been questioned (Netopil et al. 2015; Carraro et al. 2017). Netopil et al. (2015) have compared the parameters of seven open clusters derived from seven different surveys and found large discrepancies amongst their results, as seen in Fig. 1.2.

First of all, the identification of open clusters in itself presents a challenge. A spatial overdensity is only an initial indication of the possible existence of a cluster (Carraro 2006). Within the Galactic disk, many positional overdensities exist due to chance alignments and random fluctuations in extinction (Carraro et al. 2017). Therefore, the definition of a cluster must also require an overdensity in astrometric spaces and a well-defined photometric sequence in the CMD (Carraro 2006). While this thesis is not concerned with the identification of new clusters, it does discuss the identification and confirmation of previously known open clusters in Ch. 4.

One of the biggest challenges in determining reliable cluster parameters is accurately selecting cluster members, since the large number of stars in the Galactic disk causes significant confusion. Without good quality measurements of stellar astrometry, it is quite impossible to separate cluster members from the background stellar population, as evidenced by the different isochrone placements in Fig. 1.2; where each of the studies classified different stars to the cluster NGC 2158. Inclusion of field stars and unresolved binaries gives width to cluster sequence, which makes visual fitting of isochrones more challenging. The varying amount of interstellar extinction in the Galactic disk also affects the detectability of stars, which can lead to incomplete cluster memberships.

Furthermore, the dissolution of clusters and the uncertainties in cluster membership create selection biases. These biases prevent the detection of low-mass open clusters, which are faint and poorly populated, and also the detection of old open

[1]Spoiler: This cluster completeness limit is no longer true, due to the discovery of many new nearby open clusters (see Sec. 2.2) and our new open cluster census in Sec. 4.5.

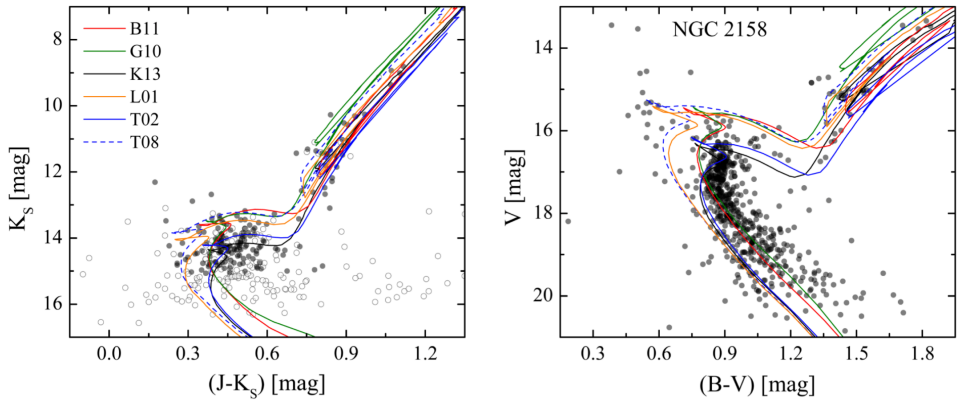

Figure 1.2: Figure 7 from Netopil et al. (2015). Color-magnitude diagrams for NGC 2158. Filled circles in the NIR CMD represents stars with a 2MASS photometric quality flag of A ($S/N > 10$) and open circles are stars with a lower photometry quality. The colored isochrones represent the cluster parameters derived from the surveys investigated by Netopil et al. (2015).

clusters, which are less populated than younger clusters due to dynamical evolution (Moraux 2016).

It is clear that a comprehensive assessment of open clusters and their parameters requires a large, accurate, and precise kinematic, photometric, and spectroscopic dataset, as well as a fully automated analysis procedure to derive parameters homogeneously. Fortunately, for the former, such a dataset now exists.

2

Open Clusters in the Gaia Era

2.1 Gaia Mission

Gaia is a satellite built by the European Space Agency (ESA), who began its construction in 2001. Launched on December 19, 2013, the main objective of the Gaia mission is to create the largest and most precise three-dimensional map of the Milky Way, shedding light on its origin and evolution. Gaia consists of two optical telescopes, which are connected to three scientific instruments to precisely measure the position, velocities, and spectrum of stars. Constantly rotating and collecting data, Gaia will achieve unprecedented measurements of astrometry and photometry by repeatedly measuring each star more than 70 times over the span of five years. Gaia aims to provide astrometry and three-band photometry for more than one billion stars in our Galaxy.

2.2 Gaia's Impact on Open Clusters

The first two Gaia data releases provide remarkable data sets with which questions regarding the origin, structure, and evolutionary history of our Galaxy can be answered, and also insights into stellar formation and evolution have been obtained. Located in the disk of our Milky Way, open clusters are among the perfect objects

to directly benefit from Gaia data.

Gaia data has already discovered around 150 new open clusters (Castro-Ginard et al. 2018; Cantat-Gaudin et al. 2018a; Castro-Ginard et al. 2019), some of which are nearby, discarded previously known clusters (Kos et al. 2018; Cantat-Gaudin et al. 2018a), charted the kinematics of clusters (Soubiran et al. 2018), and revealed the dissolution of open clusters via observations of tidal tails (Röser et al. 2019; Meingast & Alves 2019; Röser & Schilbach 2019; Fürnkranz et al. 2019; Tang et al. 2019). Data from Gaia will mitigate many of the problems mentioned in Sec. 1.4. Briefly, the following improvements are expected:

Cleaner Membership Lists

The precise astrometry and kinematics will allow us to more easily distinguish between true cluster members and non-member field stars. Higher quality photometry also enables discrimination of non-cluster members in the CMD. Fainter photometry will also allow the addition of many late-type stars, which have previously been undetected.

Retraction of Previous Open Clusters

Consequently, the highly precise astrometric data is also expected to lead to retractions of previously known open clusters. Clusters that were previously identified based on their overdensities in position will be confirmed or unconfirmed by Gaia's precise astrometry.

Improved Cluster Parameters

More accurate cluster membership and precise astrometry will yield more accurate cluster positions, distances, sizes, and proper motions, as well as less contaminated color-magnitude diagrams. Together with exceptional photometry, uncontaminated and thinner cluster main sequences will be seen in the CMDs, as illustrated by Gaia Collaboration et al. (2018a) in Fig. 2.1, subsequently allowing for more accurate estimates of cluster ages and reddenings.

New Open Clusters

Gaia data will also have a profound impact on the detection of many new clusters, which were previously undetectable. In particular, the high accuracies of the proper

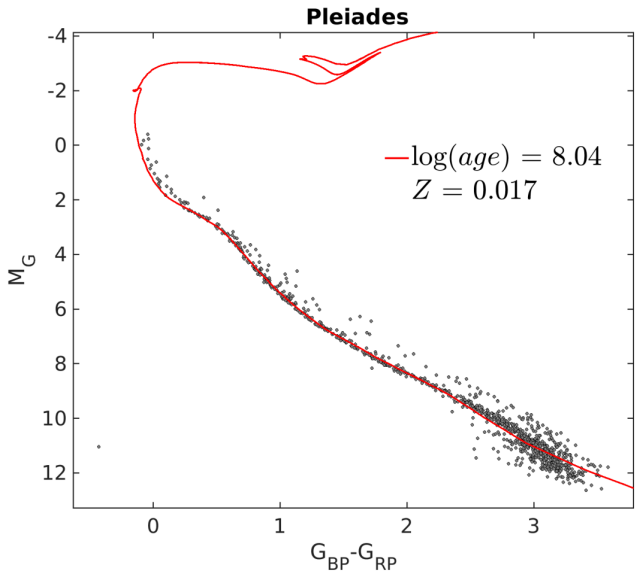

Figure 2.1: Color-magnitude diagram of the Pleiades with a PARSEC isochrone from Gaia Collaboration et al. (2018a).

motions will allow clusters to be easily identified by overdensities in proper motion space. This may bring forth the previously "missing" nearby (distance < 1 kpc from the Sun) old (age > 1 Gyr) clusters (Kharchenko et al. 2013; Schmeja et al. 2014). Gaia parallaxes could also extend the census of open clusters to distances greater than 8 kpc from the Sun.

Understanding of the Milky Way and Stars

A homogeneous census of open clusters will further our understanding of the structure, chemical composition, formation, and evolution of the Milky Way disk. This includes a deeper look into the spiral structure, galactic spatial scales (e.g. scale height of the disk), kinematics, and dynamics of the disk.

2.3 Open Cluster Parameters: New Techniques and Results

2.3.1 New Techniques

Cluster membership relies on the analysis of stellar positions and proper motions. In contrast to the methods aforementioned in Sec. 1.1, statistical approaches nowadays, are developed to derive memberships by combining all available kinematic and

photometric data (e.g., Malo et al. 2013, Sarro et al. 2014, Seleznev 2016). These new techniques allow the inclusion of uncertainties and can also make use of other observables, such as radial velocities and parallaxes.

While studies of individual clusters still make use of manual selection and verification of cluster members and isochrone fitting, studies focused on analyzing many clusters, will require more sophisticated, unsupervised statistical approaches. Several new methods have been developed for open cluster analysis in this era of large, multi-dimensional data sets (von Hippel et al. 2006; Monteiro et al. 2010; Krone-Martins & Moitinho 2014; Perren et al. 2015).

To determine which stars belong to a cluster, Krone-Martins & Moitinho (2014) developed UPMASK, an unsupervised photometric membership assignment technique for stellar clusters. UPMASK is a data-driven approach, relying solely on stellar positions and photometry to determine membership. It also assumes that members of a cluster share common properties and are more spatially concentrated on the sky, in contrast to the random spatial distribution of field stars. UPMASK first runs a principle component analysis on the photometric data to select the most significant principal components. Then a k-means clustering analysis is performed on the selected principal components to identify small groups of stars. The stars in each grouping is then tested for clustering in positional space using kernel density estimations. If they are clustered, they are kept for the next iteration. If not, they are marked as field stars and no longer considered for analysis. Tests on real data demonstrated the ability of UPMASK to properly locate and determine cluster membership of two overlapping clusters in the same field (Krone-Martins & Moitinho 2014). UPMASK can be easily adjusted to work with other types of data, such as proper motions and parallaxes (Cantat-Gaudin et al. 2018a,b).

Recognizing the need for more sophisticated isochrone fitting algorithms in this era of big data, von Hippel et al. (2006) created an open-source software program to obtain cluster parameters from photometry using a bayesian approach. The most recent version of their software is called Bayesian Analysis for Stellar Evolution with Nine Parameters (BASE-9), which uses a Markov Chain Monte Carlo (MCMC) approach to estimate the posterior probability distribution for six cluster parameters: age, metallicitiy, helium abundance, distance modulus, line-of-sight absorption, and parameters of the initial final mass relation (IFMR). BASE-9 also uses numerical integration to compute three stellar properties: primary mass, secondary mass (for binaries), and the cluster membership probability.

2.3.2 New Open Cluster Parameters

These new methods have been applied on Gaia data and have returned new results on our galaxy's open clusters. Applying UPMASK to TGAS astrometry and UCAC4 proper motions, Cantat-Gaudin et al. (2018b) derived new cluster proper motions and parallaxes for 128 open clusters. For 26 of these clusters, BASE-9 was used to derive cluster parameters.

With DR2, Cantat-Gaudin et al. (2018a) started with a compilation of 3328 known open clusters from the Dias et al. (2002) and Kharchenko et al. (2013) catalogs and applied UPMASK to DR2 data within these cluster fields to obtain a new list of members and cluster proper motions and parallaxes for 1229 clusters, 60 of which are newly discovered clusters found in the same field as known clusters. This is an eye-opening discovery as it drastically reduces the Galactic open cluster census to 37% of its original size. Taking advantage of these cluster memberships, Bossini et al. (2019) have selected 269 open clusters, with low extinction ($A_V < 2.5$ mag) and old age ($t >$10 Myr) according to MWSC and DAML catalogs, and derived new cluster parameters using BASE-9 on DR2 photometry of the clusters. While these two studies demonstrate the potential of open cluster science with DR2, neither compute new cluster parameters for the full known open cluster population.

2.4 Our Study with Gaia

Inspired by the work put in by the authors of the MWSC catalog and the plethora of extremely precise stellar data from Gaia, this thesis aims to develop a new tool for characterizing open clusters and to create an updated catalog of open clusters and their parameters. Work for this thesis began prior to the start of the Gaia era. In preparation for DR1, I began by testing a simple routine to determine the best-fitting isochrone to the V vs. $B-V$ CMD of a few well-established MWSC clusters. This exercise illustrated the potential in developing a procedure to fit isochrones in a statistical and qualitative way, superseding the traditional fit-by-eye approach. Over the past four years, the cluster characterization pipeline has gone through many iterations, from using a χ_2 minimization approach on five bands of photometric data and stellar data manually crossmatched and downloaded per cluster, to a maximum likelihood approach which uses up to eight bands of photometry and stellar data retrieved automatically within seconds. I have implemented countless changes to the pipeline, as each data release brought more data and new, unexpected challenges. The next chapters will describe two main versions of the characterization pipeline

and demonstrate their performance with analysis on the MWSC clusters.

In Chapter 3, I describe the first version of the automated cluster characterization pipeline and, using data primarily from Gaia DR1/TGAS, I show results for 24 clusters within 333 pc. In Chapter 4, I describe the current version of the automated cluster characterization pipeline, which features many significant improvements over the first. Using primarily DR2 data as inputs into the pipeline, I reanalyzed all MWSC clusters and present a new catalog of these open clusters and their parameters. Finally, in Chapter 5, I discuss additional improvements to implement in the pipeline, further analysis that can be done with our newly assembled cluster catalog, and the impact of Gaia DR3 on open clusters.

3

Characterization of nearby open clusters with Gaia DR1/TGAS and HSOY

Based on Yen et al. (2018) published in Astronomy & Astrophysics

3.1 Motivation

Open clusters are the keys to unlocking the mysteries of stellar evolution, and the structure and chemical evolution of our Galactic disk. In order to study the Milky Way disk with open clusters, a large amount of stellar data and a homogeneous set of cluster parameters (age, distance, and reddening) are required.

Compilations of open clusters and their parameters from the literature have been assembled, for example Ruprecht et al. (1981) and Lynga (1982). Information for these clusters come from individual studies, which use a variety of methods to derive parameters, resulting in a heterogeneous parameter set. In general, cluster membership relies on the analysis of positions and kinematics: proper motions and radial velocities. Classical methods vary from using visible groupings of stars to calculating probabilities based on proper motions (e.g., Vasilevskis et al. 1958, Sanders 1971, Cabrera-Caño & Alfaro 1985). Photometry can also be used to distinguish member

and field stars of a cluster (e.g., Vogt 1971, Moffat 1972, Baade 1983). Nowadays, rigorous mathematical and statistical approaches are developed to derive memberships using kinematics and photometry (e.g., Malo et al. 2013, Sarro et al. 2014). Cluster distances can be determined directly from trigonometric parallax measurements. However, the most common approach to deriving cluster distance, which subsequently also provides estimates of age and reddening, uses photometric data by fitting theoretical isochrones to the cluster color-magnitude diagram (CMD).

The wealth of information from large photometric surveys has led to the creation of homogeneous cluster catalogs, where the methods used to derive membership and cluster parameters combine the techniques aforementioned with some optimization of the results by eye. The first catalogs deriving homogenous cluster parameters relied on optical photometry, including the works by Becker & Fenkart (1971), Janes & Adler (1982), Platais et al. (1998), Chereul et al. (1999), Dias et al. (2002), and Kharchenko et al. (2005a,b). Infrared photometry from the 2MASS (Skrutskie et al. 2006) survey then allowed for the discovery of many new open clusters (e.g., Dutra & Bica 2001, Bica et al. 2003, Dutra et al. 2003, Koposov et al. 2008, Glushkova et al. 2010, Bukowiecki et al. 2011). The Milky Way Star Cluster (MWSC) catalog (Kharchenko et al. 2013) is the largest-to-date census on the star cluster populations within the Milky Way. It is nearly complete to ∼1.8 kpc from the Sun and contains 2808 open clusters.

While these catalogs have analyzed hundreds or thousands of open clusters in a homogeneous way, large discrepancies exist between the final cluster parameters reported by these catalogs, as illustrated by Netopil et al. (2015). Some of the limitations of these catalogs include: small number of cluster members, low accuracy in proper motion, precision and/or accuracy of photometry, or use of different isochrones. It is clear that a comprehensive assessment requires accurate photometric and kinematic data at least; spectroscopic data would also be a great benefit. Taking advantage of the large amount of stellar data available (photometry from large all-sky surveys and refined astrometry from space missions), we have developed an automated pipeline to consistently determine cluster membership and fit the fundamental cluster parameters: distance d, reddening $E(B-V)$, and age $\log t$, where t is in years. Due to the limited size and precision of the TGAS catalog (Gaia Collaboration, Brown et al. 2016), we performed our analysis on 24 nearby open clusters. These clusters are generally well-studied and are located within 333 pc as given in both MWSC and the Catalogue of Open Cluster Data (COCD, Kharchenko et al. 2005a,b). As the typical error of TGAS parallaxes is 0.3 mas, studying clusters

with parallaxes greater than 3 mas allows us to use accurate stellar parallaxes, with errors less than 10%. The names and MWSC identifiers of the clusters are provided in Table 3.1.

This pipeline was developed to ascertain the possibility of an automated isochrone fitting routine that reliably determines cluster membership and parameters. We note that the techniques described in this paper are specifically designed for working with the current data available and its limitations.

In Section 3.2, we describe our input data. The details of our cluster characterization pipeline is described in Section 3.3. In Section 3.4, we discuss our results and compare them with the literature. Finally, a summary of our technique and first results are given in Section 3.5.

3.2 Data

The basis of our data set is largely rooted in the cluster field star lists of the MWSC catalog. The selection area around each cluster has a radius of $r_{\rm a} = r_{\rm cl} + 0.3°$, where $r_{\rm cl}$ is taken from the literature (Kharchenko et al. 2012). The primary stellar data for MWSC was compiled from the PPMXL (Röser et al. 2010) and 2MASS (Skrutskie et al. 2006) all-sky catalogs. The positions from the PPMXL catalog are used and supplemented with Hipparcos (van Leeuwen 2007), in order to recover any missing bright stars.

Building upon MWSC, we used B and V photometry from the All-Sky Compiled Catalogue of 2.5 million stars (ASCC-2.5, Kharchenko 2001), which are based mainly on Hipparcos and Tycho-2 (Høg et al. 2000). We also incorporated 2MASS JHK_s and Gaia DR1 (Gaia Collaboration, Brown et al. 2016) G band photometry, for a total of up to six bands for each star. The median uncertainties in G magnitudes range from the mmag level to 0.03 mag (Gaia Collaboration, van Leeuwen et al. 2017a; Gaia Collaboration, Evans et al. 2017). In order to also account for systematics in the G magnitudes of the brighter stars, we have adopted a conservative error of 0.03 mag for all stars.

Furthermore, we included precise stellar astrometric data from TGAS (Gaia Collaboration, Brown et al. 2016) and HSOY (Altmann et al. 2017). TGAS supplies proper motions and parallaxes for roughly 2 million Tycho-2 stars. HSOY, which combines positions from Gaia DR1 and data from PPMXL, gives proper motions for 583 million stars. Hipparcos proper motions were also used to recover any missing bright stars. By combining six-band photometric measurements, proper motions,

and parallaxes for the stars, we are able to better constrain and determine cluster membership and parameters.

3.2.1 Isochrone models

For this work, we used the Padova isochrone set: PARSEC version 1.2S (Bressan et al. 2012) in the Johnson BV, 2MASS JHK_s, and Gaia G photometric systems with $Z = Z_{\odot} = 0.0152$ (Caffau et al. 2009, 2011). Most open cluster studies based on isochrone fitting assume solar metallicity for simplicity because cluster metallicities are known for very few open clusters. In the updated catalog by Dias et al. (2002), this parameter is available for roughly 13% of the ∼2000 clusters. Of our 24 clusters, only 14 have metallicities listed in the MWSC. The mean of these metallicities is −0.09 dex, which is close to solar metallicity. Furthermore, considering these values also have some error, it is reasonable to assume solar metallicity for all clusters. Nevertheless, for clusters with highly nonsolar metallicities, this will introduce a small bias to the derived parameters. Our isochrone set spans the age range $6.6 \leq \log t \leq 10.1$, at step sizes $\Delta \log t = 0.01$.

We have also constructed the corresponding zero-age main sequence (ZAMS) for this isochrone set using the evolutionary tracks and "ptcri" file [1]. The age at which the ZAMS occurs for each mass in the isochrone set can be determined by matching the MS_BEG point for a given mass in the ptcri file to the age at the MS_BEG point in the evolutionary track for that mass. With a list of ages for each mass, the ZAMS B, V, J, H, K_s, and G magnitudes can then be determined by finding the matching age and mass combination in the isochrone set.

3.3 Pipeline methodology

Our automated cluster characterization pipeline consistently determines cluster membership and fits the fundamental cluster parameters: distance, reddening, and age. The pipeline follows a sequence of procedures with two main segments: (1) membership determination and (2) isochrone fitting and membership refinement. Cluster membership is first determined through iterative proper motion, parallax, and photometric selections. After the initial membership selection, isochrones are fitted to the photometric observations of cluster members to determine cluster parameters and membership is further refined by removing highly discordant stars. This segment is iterated until membership and cluster parameters are consistent.

[1] http://people.sissa.it/∼sbressan/parsec.html

The pipeline relies on a χ^2 minimization to fit the ZAMS and isochrones to the cluster photometry. The minimization method used is the Levenberg-Marquardt as provided by LMFIT (Newville et al. 2016). We chose to use a least squares method because it is a reliable and commonly used fitting approach. In future versions of our pipeline we may implement a maximum likelihood method as Monteiro et al. (2010), Dias et al. (2012), and Palmer et al. (2014) have used, should it turn out to be more robust. For now, the least squares fitting has produced reliable results for all nearby clusters, as quantified in Sect. 3.4.

3.3.1 Proper motion and parallax selections

The proper motion selection routine first computes the weighted mean cluster TGAS proper motion using TGAS proper motions of the most probable cluster members from the MWSC. In the MWSC, the most probable members are defined as stars with a combined probability of kinematic or proper motion, photometric, and spatial components, greater than 0.61 (Kharchenko et al. 2013). The weighted mean cluster TGAS proper motion in RA and DEC are denoted by $\bar{\mu}_{\alpha^*,\mathrm{T}}$ and $\bar{\mu}_{\delta,\mathrm{T}}$ respectively, where $\alpha^* = \alpha \cdot \cos\delta$.

Considering that the median standard proper motion uncertainty of TGAS is 1.2 mas yr^{-1} (Lindegren et al. 2016), all stars within 2 mas yr^{-1} of the weighted mean cluster TGAS proper motion are selected as candidate members. For 10% of TGAS stars, the standard proper motion error is more than 2.7 mas yr^{-1} (Lindegren et al. 2016), which means there are potential cluster members with larger proper motion errors that fall outside the 2 mas yr^{-1} radius. To recover these stars, a factor of 2.5 is applied to the first radius, and so stars within 5 mas yr^{-1} are also considered cluster candidates if their 3σ proper motion error ellipse is consistent with the weighted mean cluster proper motion.

While the proper motion precision varies across the sky (Lindegren et al. 2016), the limits employed here are general and optimal for our cluster sample. Using a smaller value for the inner selection radius would lead to missing members because although some clusters in our sample have very small proper motion errors, roughly 0.2 mas yr^{-1}, there are systematics to consider on top of the formal error. Furthermore, there likely exists some internal dispersion in the proper motions for cluster members, as they are not expected to have exactly the same proper motions. On the other hand, the inner selection circle cannot be made too large, otherwise too many nonmembers would be included, especially for clusters with larger proper motion

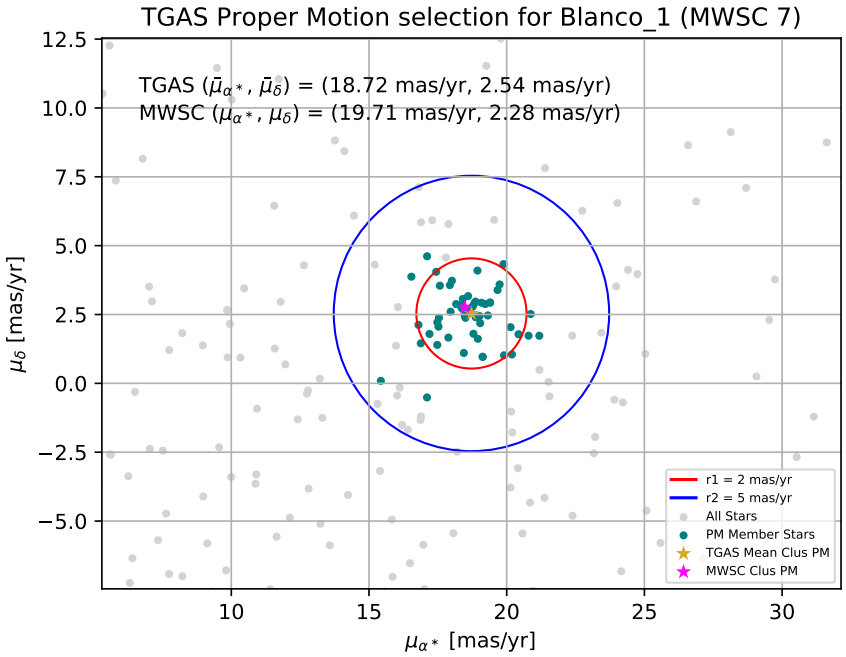

Figure 3.1: Result of the TGAS proper motion selection for Blanco 1. The final Blanco 1 TGAS proper motion is (18.72 mas yr^{-1}, 2.54 mas yr^{-1}). The teal points represent the proper motion members, where all stars within the 2 mas yr^{-1} radius (red circle) of the mean cluster proper motion are selected and the stars within 5 mas yr^{-1} (blue circle) are only selected if their 3σ errors are consistent with the mean cluster proper motion.

errors. The inner selection radius corresponds to the largest median proper motion error present. Taking these considerations into account, an inner radius of 2 mas yr^{-1} is a good compromise for all clusters, regardless of their sky position.

This first selection allows the MWSC cluster members to be identified. But in order to account for systematic offsets between the data sets, the procedure is iterated. With new cluster membership, the weighted mean TGAS proper motion is recomputed and the selection process is repeated. This routine is iterated, on average, once or twice until the cluster membership is unchanged. Illustrations of the result of this procedure are shown in Fig. 3.1 (for Blanco 1) and in Appendix B (for all clusters).

Next, a parallax selection is performed on the TGAS proper-motion candidates to further refine cluster membership. As an initial estimate, the cluster parallax is first computed from the cluster's MWSC distance. Stars are considered cluster candidates if the 3σ parallax error of the star lies within the cluster parallax. Since

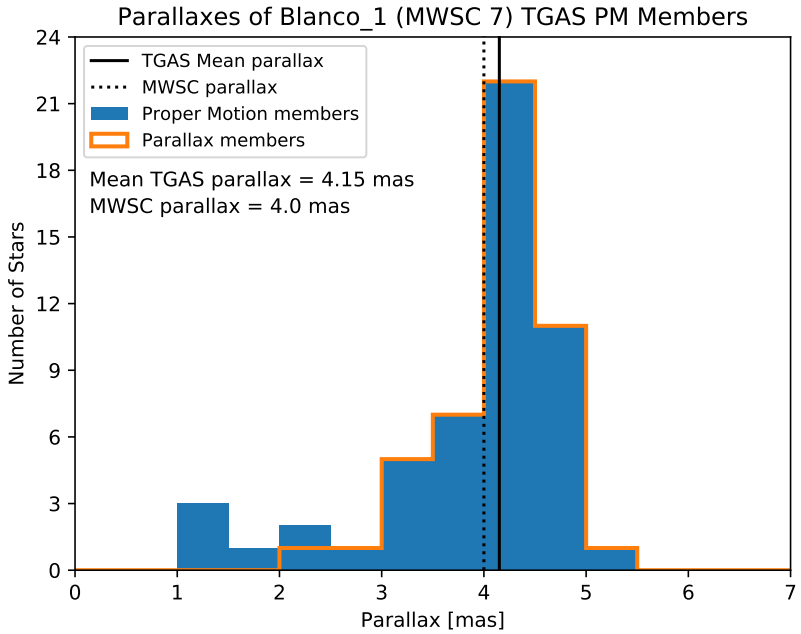

Figure 3.2: Result of the TGAS parallax selection for Blanco 1 proper motion-selected stars. The final weighted mean TGAS parallax for Blanco 1 is 4.15 mas. The orange outline illustrates the stars with 3σ errors consistent with the mean parallax; these stars are the TGAS astrometrically-selected candidates of Blanco 1.

a systematic difference in parallaxes may exist between MWSC and TGAS, we computed the weighted mean cluster parallax, $\bar{\varpi}_{\rm T}$, for the initial parallax-selected stars. Again, the 3σ parallax error of the stars are compared to the $\bar{\varpi}_{\rm T}$ to determine membership. The parallax selection is also iterated until cluster membership no longer changes.

In the cases where there is no clear over density in the cluster parallaxes, or if there is a notable peak at the parallax of field stars, so the resulting $\bar{\varpi}_{\rm T}$ differs by more than 3σ from the MWSC parallax, the parallax iteration is canceled. Instead, stars are considered cluster candidates if their 3σ parallax error is consistent with the MWSC parallax. Kovaleva et al. (2017) have shown that MWSC and TGAS parallaxes are compatible within 2 kpc from the Sun. The final selected cluster parallaxes are shown in Fig. 3.2 for Blanco 1 and illustrate that some accepted TGAS proper motion candidates, from the previous selection, have been eliminated. TGAS proper motion and parallax diagrams for all clusters can be found in Appendix B. We note that the parallaxes obtained at this stage are solely used as initial estimates. Final cluster distances will be computed from isochrone fitting, independent of these trigonometric parallaxes.

After defining the TGAS astrometric candidates, stars with HSOY proper motions are considered for membership. If a star in our data set has both TGAS and HSOY proper motions, their membership relies only on the more precise TGAS data. Moreover, stars that did not meet the TGAS selection criteria are no longer be considered for membership even if they may have consistent HSOY proper motions. For Gaia $G \leq \sim$16 mag, the mean standard error in HSOY proper motions is about 2 mas yr^{-1} for declinations greater than $-30°$ and 3 mas yr^{-1} for declinations less than $-30°$ (Altmann et al. 2017). Considering these errors, the same proper motion selection cuts as used for TGAS are applied. As the HSOY catalog is much larger than TGAS and because there are no clear cluster overdensities in HSOY proper motion space, the HSOY selection is not iterated.

Our selection considers the majority of the HSOY cluster candidates; the missing HSOY candidates are preferentially the fainter stars with $G > 16$ mag, which have standard HSOY proper motion errors larger than 3 mas yr^{-1}. At this stage we are not concerned with completeness, that is, obtaining all possible cluster members, because we do not derive the cluster mass function. We have focused on obtaining true cluster members, so the cluster parameters are accurately determined.

Lastly, we recover the bright stars missing from the cluster, that is, with no TGAS or HSOY proper motions, by using Hipparcos proper motions. These bright stars are crucial for accurate determination of the cluster age. The median proper motion error for stars with broad-band Hipparcos magnitudes brighter than 9 mag is about 0.88 mas yr^{-1} (van Leeuwen 2007). Taking these errors into account, the same proper motion cuts as used for TGAS are applied. The Hipparcos proper motion selection is not iterated because there are very few stars with Hipparcos proper motions.

3.3.2 Photometric selection

The second procedure in the pipeline is a photometric membership selection. Before cluster membership is further refined, any highly evolved stars, such as red giant branch (RGB) stars or supergiants, are identified and temporarily removed. The highly evolved members are temporarily removed because we perform initial fitting, to determine rough estimates for distance and reddening, with the ZAMS (see Sect. 3.3), and since these stars have evolved off the main sequence, they should not be considered in this initial fit. To select the highly evolved stars, the ZAMS is placed in the V vs. $B-V$ CMD at the cluster's MWSC $E(B-V)$ and d values with an

offset of $\Delta V = -5$ mag, which nicely divides the cluster main sequence and the upper right corner of the cluster CMD, where giant stars are located. Stars are identified as evolved if they have V magnitudes brighter than the offset ZAMS. This usually removes all potentially highly evolved stars; repeating a similar procedure in the infrared CMDs does not yield additional stars to remove.

It is important to note that TGAS parallaxes are currently only used for membership selection and not in the pipeline for distance determination because the measurements are available for very few stars. This will certainly change in the next version of the pipeline, after the release of Gaia DR2. Nevertheless, this general isochrone fitting method, which determines cluster distance independent of using TGAS parallaxes (except for membership determination), will still be relevant to use after Gaia DR2; in particular, for the analysis of very distant clusters, which will have poor parallax measurements in Gaia DR2.

Since the HSOY catalog does not provide parallaxes for the stars, a given cluster might end up with hundreds to thousands of candidate members after the HSOY proper motion selection, many of which are field stars and obvious nonmembers. To eliminate the majority of these nonmembers, the ZAMS is placed in the J vs. $G - J$ CMD using the cluster's MWSC $E(B - V)$ and d values with an offset of $\Delta J = +1.0$ mag, removing candidates with J magnitudes greater than this offset and $G - J > 0.5$ mag. Candidates with J magnitudes greater than this offset and $G - J < 0.5$ mag are kept because these are bright stars near the main sequence turn-off and are essential to proper age determination. The ZAMS is also offset by $\Delta(G - J) = +0.5$ mag and candidates with $(G - J)$ magnitudes greater than this offset are removed. This process is repeated in the K_s vs. $J - K_s$ CMD with a $\Delta K_s = +1.0$ mag and $\Delta(J - K_s) = +0.2$ mag offset of the ZAMS to remove additional obvious nonmembers, but taking care to keep the stars above the ΔK_s offset if $J - K_s < 0.5$ mag. These thresholds were determined after varying the values and looking at the CMDs of our 24-cluster sample. The cuts needed to be such that clear cluster members are kept, while the majority of background stars are rejected. These photometric selections might remove some evolved cluster members, but as aforementioned, we are not concerned with completeness at this stage.

3.3.3 Initial cluster parameters

In order to efficiently use the Levenberg-Marquardt fitting method, initial guesses for the fitted parameters must be supplied. In order to obtain reliable initial cluster

$E(B-V)$ and distance d estimates, the ZAMS is fit to the photometric observations of the astrometrically- and photometrically-selected candidate stars. $E(B-V)$ and d are inserted as parameters by adjusting the ZAMS photometry. For each member star, residuals are calculated for all available measured photometric bands and the corresponding ZAMS magnitudes for all possible masses. The residual is defined as difference between the photometric measurement and ZAMS model for a given photometric band, from which the χ^2 contribution for each star is derived. The mass, and subsequently the ZAMS photometry, yielding the minimum χ^2 contribution is then matched to the star. This method allows a direct comparison between measured and model magnitudes and ensures that the star has a mass consistent in all photometric bands.

We also consider the presence of unresolved binaries, which contribute to a spread in the cluster main sequence toward brighter magnitudes. For two stars of equal masses, the offset is 0.75 mag, but for stars of unequal masses, this offset is less (e.g., Fan et al. (1996) found binaries with mass ratios less than 0.5 are within 0.1 mag of the main sequence). The number and types of binaries in our clusters are unknown. After testing a range of values, we decided to use a mean offset of -0.1 mag, as given by Fan et al. (1996), for all passbands, ensuring that the residuals are minimized in such a way that the derived $E(B-V)$ and d fit between the sequences of single and multiple stars. In the future, we plan to determine the effect of binaries and this binary offset quantity in more detail by taking observed binary fractions and characteristics statistically into account, which will most likely lead to different offsets in the various photometric bands.

After the ZAMS fit, offsets of $\Delta J = +0.5$ mag and $\Delta(G-J) = +0.25$ mag are applied to the fitted ZAMS in the J vs. $G-J$ CMD to further remove nonmember stars. Again, stars above the ΔJ offset with $G-J < 0.5$ mag are kept. The ZAMS is refitted and stars are removed until the cluster membership no longer changes. However, if cluster membership is below 25 stars, the ZAMS fitting process is not iterated and no additional stars are removed; otherwise the cluster would contain too few members to properly fit. The final $E(B-V)$ and d values obtained from the ZAMS fit are then used as the starting values in the isochrone fitting.

3.3.4 Isochrone fitting

With highly probable cluster members determined and initial estimates for the cluster $E(B-V)$ and d, isochrones at a step of 0.1 dex are fitted to the cluster photom-

etry. Lower and upper limits for $E(B-V)$ are 0 and 0.5 mag. For d, the limits are $d\pm20$ pc. These limits are not too restrictive and allow the parameter space around the initial estimates to be explored, in case there is a bias from the ZAMS fitting.

As mentioned above, a given star can have up to six bands of photometric data; the bright Hipparcos stars, not in TGAS, can have up to five passbands, from ASCC-2.5 and 2MASS, available. The process of matching a stellar mass, and subsequently the isochrone model magnitudes, to the available photometry for a cluster member is the same as that described in Sect. 3.3.3 for the ZAMS fitting. By fitting the photometric measurements, instead of colors and magnitudes (in the CMD), the precision of each measurement is accounted for.

Here, the highly evolved members of the cluster are included and down-weighted (by increasing their error bars by a factor of 10). We down-weight the highly evolved stars for two reasons:

1. The isochrones are not very well-determined in the late phases of stellar evolution, that is, after the main sequence turn-off.

2. Since the RGB stars are very bright, they have very small photometric errors, which greatly affects the isochrone fitting.

Giving less weight to these stars allows them to still be considered when fitting for the cluster parameters, but prevents them from completely dominating the fitting.

The isochrone yielding the minimum reduced χ^2 is selected and the stars with the largest χ^2 contribution are removed. This process is repeated, starting with the ZAMS fit, until a minimum reduced $\chi^2 < 8$ is achieved. In some cases where this minimum could not be achieved - likely due to many bright Hipparcos stars with very small photometric errors in B and V - a reduced $\chi^2 < 14$ was adopted. After this reduced χ^2 criteria is met, all isochrones at $\Delta \log t = 0.01$ are then fitted to the photometry of the final cluster members, in order to fine-tune the cluster's age. The isochrone yielding the minimum reduced χ^2 gives the cluster's final age, $E(B-V)$, and d.

This reduced χ^2 threshold allows obvious nonmembers to be rejected, while keeping the evident cluster members, and was selected after experimenting with many different values for our cluster sample. Imposing the typical reduced $\chi^2 = 1$ would simply remove too many cluster members, but the best value to use cannot be determined because of multiple unquantifiable effects, including systematics in the theoretical models, photometric errors in the Gaia G band calibration, and

unresolved binaries. We do not expect a perfect reduced χ^2 agreement because the isochrone models do not perfectly match the data at all stellar masses and passbands, along small mass ranges of the isochrones. This seems to be evident in the low-mass end of the isochrones in the K_s vs. $J - K_s$ CMD. Additionally, the Gaia G isochrones used are based on the before launch G calibrations (Jordi et al. 2010), thus small deviations are expected, and in fact, do exist, when compared to G photometry (Carrasco et al. 2016).

Table 3.1: Derived parameters for 24 clusters

Name	MWSC	$\alpha_{\rm T}$ [h:m:s]	$\bar{\mu}_{\alpha^*,{\rm T}}$	$\bar{\mu}_{\delta,{\rm T}}$	$\bar{\varpi}_{\rm T}$	$\log t$ [yrs]	$E(B-V)$	d	$N_{\rm T}$	$N_{\rm H}$	$N_{\rm Hip}$
		$\delta_{\rm T}$ [d:m:s]	$\sigma_{\mu\alpha^*,{\rm T}}$	$\sigma_{\mu\delta,{\rm T}}$	$\sigma_{\varpi,{\rm T}}$		$\sigma_{E(B-V)}$	σ_d			
			[mas yr^{-1}]	[mas yr^{-1}]	[mas]	[dex]	[mag]	[pc]			
Blanco 1	7	00:04:13.47	18.65	2.63	4.11	8.16	0.007	251.6	48	237	5
		-29:55:40.26	0.08	0.08	0.10	$^{+0.59}_{-0.20}$	0.001	0.4			
Platais 2	109	01:11:41.91	14.99	-9.99	4.80	8.65	0.041	180.0	2	3	2
		32:03:34.40	0.91	0.66	0.71	$^{+0.14}_{-0.27}$	0.007	1.6			
α Per	274	03:25:49.74	22.80	-25.29	5.56	7.80	0.109	167.7	84	79	7
		49:08:21.03	0.05	0.04	0.06	$^{+0.05}_{-0.25}$	0.001	0.3			
Alessi 13	278	03:24:19.92	37.08	-4.30	9.62	8.75	0.027	97.8	9	9	0
		-35:49:26.86	0.11	0.12	0.16	$^{+0.05}_{-0.05}$	0.004	0.7			
Pleiades	305	03:46:16.73	19.92	-45.20	7.38	8.15	0.010	126.3	91	280	11
		24:13:27.06	0.05	0.04	0.06	$^{+0.08}_{-0.15}$	0.001	0.2			
Platais 3	395	04:39:37.44	3.83	-20.37	5.22	8.92	0.000	176.4	14	27	0
		71:16:03.03	0.15	0.17	0.20	$^{+0.18}_{-0.50}$	0.006	0.6			
Platais 4	467	05:06:55.86	1.99	-7.21	2.90	8.31	0.198	296.9	7	123	0
		22:36:15.54	0.21	0.20	0.28	$^{+0.09}_{-0.52}$	0.003	0.7			
Collinder 65	540	05:26:34.75	-0.17	-5.37	2.77	8.02	0.031	375.8	18	1222	2
		15:43:17.21	0.18	0.18	0.24	$^{+0.44}_{-0.23}$	0.001	0.3			
NGC 2232	871	06:27:50.50	-4.62	-1.80	3.56	8.02	0.000	356.6	8	218	4
		-04:47:30.02	0.19	0.19	0.24	$^{+0.13}_{-0.28}$	0.034	0.7			

Derived parameters for 24 clusters

Name	MWSC	$\alpha_{\rm T}$ [h:m:s]	$\bar{\mu}_{\alpha^*,{\rm T}}$	$\bar{\mu}_{\delta,{\rm T}}$	$\bar{\varpi}_{\rm T}$	$\log t$ [yrs]	$E(B-V)$	d	$N_{\rm T}$	$N_{\rm H}$	$N_{\rm Hip}$
		$\delta_{\rm T}$ [d:m:s]	$\sigma_{\mu\alpha^*,{\rm T}}$	$\sigma_{\mu\delta,{\rm T}}$	$\sigma_{\varpi,{\rm T}}$		$\sigma_{E(B-V)}$	σ_d			
			[mas yr^{-1}]	[mas yr^{-1}]	[mas]	[dex]	[mag]	[pc]			
Alessi 3	1157	07:16:08.68	-9.74	12.13	3.81	8.90	0.035	261.5	20	14	0
		-46:33:31.12	0.14	0.12	0.15	$^{+0.09}_{-0.10}$	0.004	1.1			
NGC2451A	1308	07:42:32.56	-21.21	15.42	5.27	8.17	0.014	196.6	24	77	7
		-38:19:44.51	0.13	0.13	0.15	$^{+0.22}_{-0.31}$	0.002	0.6			
Praesepe	1527	08:39:54.62	-36.03	-12.86	5.39	8.90	0.010	183.0	56	319	2
		19:36:05.43	0.07	0.06	0.08	$^{+0.12}_{-0.18}$	0.001	0.2			
IC 2391	1529	08:40:28.73	-24.51	23.28	6.74	7.91	0.057	158.5	24	18	4
		-53:10:02.84	0.10	0.09	0.12	$^{+0.39}_{-0.43}$	0.003	0.8			
Platais 8	1629	09:06:44.79	-15.83	14.73	7.45	7.90	0.024	143.3	12	25	1
		-58:59:11.66	0.12	0.12	0.16	$^{+0.26}_{-0.09}$	0.003	0.7			
Platais 9	1639	09:10:33.81	-24.62	12.91	5.89	8.09	0.005	190.7	9	63	4
		-43:53:06.96	0.21	0.20	0.24	$^{+0.45}_{-0.19}$	0.003	0.8			
IC 2602	1841	10:42:28.05	-17.63	10.57	6.79	8.00	0.004	149.0	32	99	8
		-64:14:37.49	0.08	0.07	0.10	$^{+0.05}_{-0.26}$	0.002	0.4			
Coma Ber	2020	12:24:23.72	-12.22	-9.01	11.55	8.75	0.053	85.6	33	10	7
		25:57:23.45	0.07	0.07	0.09	$^{+0.18}_{-0.15}$	0.001	0.1			
Platais 10	2150	13:41:50.68	-30.51	-10.52	4.02	8.29	0.093	231.0	8	43	0
		-59:07:48.14	0.14	0.14	0.19	$^{+0.12}_{-0.18}$	0.003	0.9			

Derived parameters for 24 clusters

Name	MWSC	$\alpha_{\rm T}$ [h:m:s]	$\bar{\mu}_{\alpha^*,{\rm T}}$	$\bar{\mu}_{\delta,{\rm T}}$	$\bar{\varpi}_{\rm T}$	$\log t$ [yrs]	$E(B-V)$	d	$N_{\rm T}$	$N_{\rm H}$	$N_{\rm Hip}$
		$\delta_{\rm T}$ [d:m:s]	$\sigma_{\mu\alpha^*,{\rm T}}$	$\sigma_{\mu\delta,{\rm T}}$	$\sigma_{\varpi,{\rm T}}$		$\sigma_{E(B-V)}$	σ_d			
			[mas yr^{-1}]	[mas yr^{-1}]	[mas]	[dex]	[mag]	[pc]			
Alessi 9	2670	17:44:59.78	9.71	-8.81	4.94	8.42	0.091	224.9	11	68	1
		-47:02:35.25	0.20	0.21	0.26	$^{+0.13}_{-0.32}$	0.002	0.6			
Collinder 350	2700	17:48:14.26	-5.28	-0.13	2.69	9.00	0.167	298.3	10	165	0
		01:20:25.42	0.19	0.19	0.24	$^{+0.14}_{-0.22}$	0.003	0.7			
NGC6475	2739	17:53:29.50	3.15	-5.51	3.45	8.29	0.156	300.8	49	1428	8
		-34:39:22.33	0.09	0.09	0.11	$^{+0.18}_{-0.31}$	0.001	0.3			
Ruprecht 147	3078	19:16:18.33	-1.04	-26.92	3.53	8.86	0.059	265.1	18	43	1
		-16:17:23.19	0.18	0.19	0.23	$^{+0.12}_{-0.65}$	0.003	0.9			
NGC7092	3521	21:32:05.91	-7.54	-20.13	3.33	8.70	0.010	290.6	26	24	0
		48:27:02.73	0.10	0.10	0.12	$^{+0.06}_{-0.23}$	0.002	0.9			
ASCC123	3654	22:41:36.00	11.67	-1.09	4.38	8.10	0.097	243.5	12	37	4
		54:09:56.43	0.15	0.13	0.17	$^{+0.40}_{-0.39}$	0.003	0.9			

3.4 Results

We analyzed 24 nearby open clusters with our automated pipeline. The cluster parameter results, $\log t$ [years], $\sigma_{\log t}$, $E(B-V)$, $\sigma_{E(B-V)}$, d, and σ_d, returned from the pipeline are listed for each cluster in Table 3.1. The errors in $E(B-V)$ and d are the formal 1σ errors from the χ^2 fit (Newville et al. 2016), which do not fully capture the real errors in these values. The real errors, including the systematics described in Sect. 3.4, are most likely larger. The error in $\log t$ is estimated from the plateau in the reduced χ^2 distribution of all possible ages at the cluster's $E(B-V)$ and d. The large range of errors for $\log t$ shows that the age is difficult to constrain. The smallest error in age is expected for clusters which harbor turn-off stars, but if turn-off stars are not present in any given cluster, then the age will be relatively uncertain. Cluster CMDs showing the final cluster membership and fitted isochrone are given in Appendix A.1. Three CMDs are provided for each cluster: V vs. $B-V$, J vs. $G-J$, and K_s vs. $J-K_s$. Tables with cluster membership and relevant stellar data (positions, proper motions, parallaxes, and photometry) are provided electronically. The tables also include stellar masses for each star, as determined by our isochrone fit. This data can be downloaded through the CDS online archive[2].

The contents of Table 3.1 also include mean cluster TGAS positions in RA and DEC, $\alpha_{\rm T}$ and $\delta_{\rm T}$ respectively, weighted mean cluster TGAS proper motions in RA and DEC, $\bar{\mu}_{\alpha^*,{\rm T}}$ and $\bar{\mu}_{\delta,{\rm T}}$ respectively, and parallaxes, $\bar{\varpi}_{\rm T}$, from the cluster's final TGAS members. The errors in cluster proper motion, $\sigma_{\mu\alpha^*,{\rm T}}$ and $\sigma_{\mu\delta,{\rm T}}$, and parallax, $\sigma_{\varpi,{\rm T}}$, provided are the formal errors of the weighted mean. The systematic error in parallax from Gaia is 0.3 mas (Lindegren et al. 2016). This systematic error does not average out and still needs to be considered in addition to the formal errors mentioned. The proper motion and parallax values quoted in the table may differ from those in the figures of Appendix B, in which the values are given for membership before photometric selection. Lastly, the final numbers of TGAS, HSOY, and Hipparcos stars, $N_{\rm T}$, $N_{\rm H}$, and $N_{\rm Hip}$ determined to be cluster members are provided in the last three columns of Table 3.1.

3.4.1 Comparison with MWSC

A comparison between our derived parameters to the MWSC values is shown in

[2]via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/A+A/615/A12

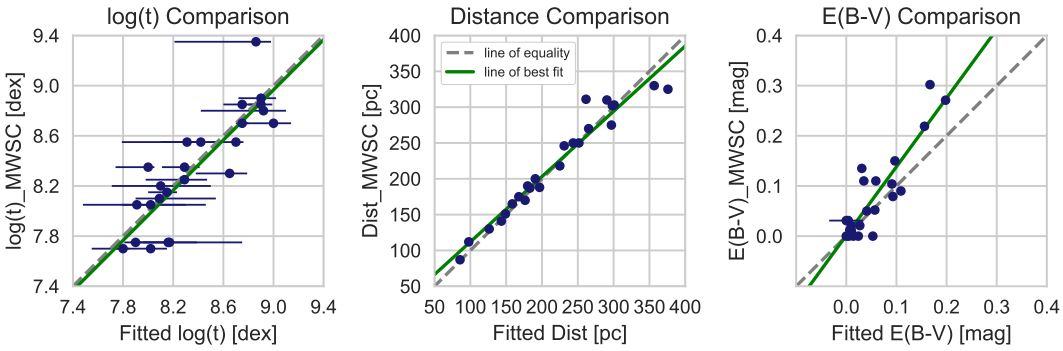

Figure 3.3: Comparison of 24 fitted cluster parameter results from this work to those determined in MWSC. Horizontal error bars, from this work, are plotted for all three comparisons, but for the fitted distance and fitted $E(B-V)$, the errors are generally smaller than the point marker used. The dashed gray line indicates the line of equality and the green line shows the best fit. Except for the $E(B-V)$, the agreement between MWSC and newly fitted pipeline values is rather good.

Fig. 3.3. As expected, there is a wide range in the determined ages. The median age difference is 0.13 dex and the dispersion in age is 0.22 dex. The ages of 15 clusters are within 1σ and 5 clusters are within 2σ. The clusters with the largest discrepancies in age are Ruprecht 147, NGC2451A, Blanco 1, and Platais 2 with differences of 0.49, 0.42, 0.41, and 0.35 dex respectively.

In the case of Platais 2, the difference could very well be due to the fact that the cluster does not really exist. There is no over density in the TGAS proper motion diagram and none of the few proper motion-selected members have a parallax consistent with the MWSC value. Furthermore, the small number of proper motion-selected members do not exhibit a very similar parallax. Another cluster in our sample, whose existence we also doubt, is Platais 4. The TGAS proper motion diagram of Platais 4 also does not show a strong over density and the peak of its TGAS parallax distribution is at 1 mas, which is mostly from many background and/or field stars, as the MWSC parallax is at 3.6 mas. Only four of its proper motion-selected members have TGAS parallaxes around 3.6 mas. If strong over densities are not observed in the TGAS proper motion and parallax domains, it is a possibility that the cluster is not real. The existence of these clusters will be clarified by the stellar parallaxes from Gaia DR2.

For Ruprecht 147, the age discrepancy clearly results from the addition of three early-type stars. These stars have TGAS proper motions and parallaxes consistent with the corresponding mean cluster values and thus, were considered cluster members by our pipeline. But for the age determined in the MWSC, these stars were

probably considered blue stragglers and excluded from the isochrone fitting. Manually rejecting these stars from our fitting procedure yields $\log t = 9.8$ dex, which is marginally more consistent with the MWSC age.

For the distance parameter, we find that our derived results are consistent with those listed in MWSC, as the middle panel of Fig. 3.3 shows most clusters accumulating along the line of equality. The median difference in distance is 6.7 pc, with a dispersion of 17.7 pc. The dispersion in distances is much larger than the median formal distance errors. As mentioned above, the formal distance errors are smaller than expected because they do not include systematic errors. The clusters with the largest differences in distance, of 50.8, 49.5, and 26.6 pc, are Collinder 65, Alessi 3, and NGC2232 respectively. Converting our distances to distance moduli, $(m_{K_s} - M_{K_s})$, we investigated whether a correlation exists between the difference in distance moduli and the difference in $E(B-V)$ of our determination and the MWSC. No correlation between the two was found.

Comparison of the derived and MWSC $E(B-V)$ values in the right panel of Fig. 3.3 illustrates that the derived values from this work are generally smaller than those found by the MWSC. The median difference and dispersion in $E(B-V)$ are 0.017 mag and 0.04 mag respectively. Due to the small formal errors derived, only 8 clusters are compatible within 3σ. The clusters with the largest discrepancies in $E(B-V)$ are Collinder 350, Collinder 65, and Alessi 3 with differences of 0.135, 0.101, and 0.075 mag respectively. Differences in $E(B-V)$ result in only small shifts of the isochrone, as seen from the two isochrones plotted in the cluster CMDs in Appendix A.

Overall, we find good agreement between MWSC values and our fitted cluster parameters, as the median deviations in $\log t$, d, and $E(B-V)$ are roughly 0.24%, 2%, and 18%. A linear best fit to the observed differences is almost indistinguishable from the line of equality (Fig. 3.3), especially for the distances.

3.4.2 Parallax comparison

Some discrepancies do exist between our mean parallaxes and those computed from our fitted distances, as illustrated in the left plot of Fig. 3.4. The median difference in parallaxes is 0.36 mas, with a dispersion of 0.42 mas. The largest discrepancies, with differences of 0.76 mas, are Platais 2 and NGC2232. Others include Collinder 350 and Platais 9, with parallax differences of 0.66 mas and 0.65 mas respectively. For most of these clusters, there is no clear peak in the TGAS parallax distributions

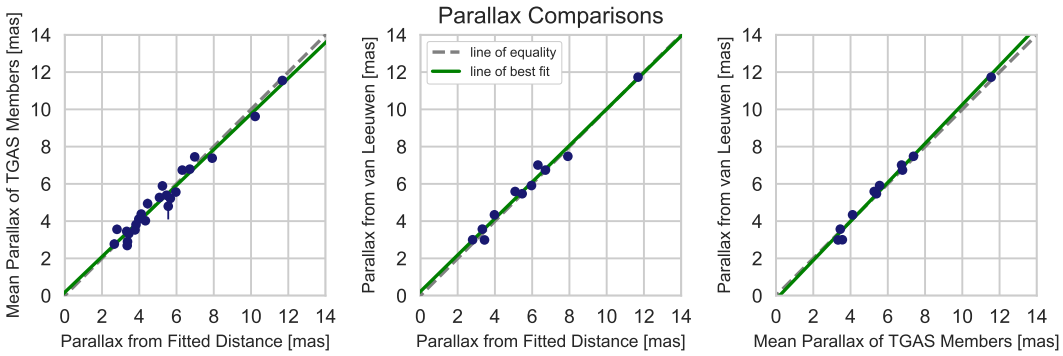

Figure 3.4: Left panel: Comparison of the weighted mean parallax of TGAS cluster members to parallaxes computed from the fitted distances for 24 clusters. The cluster with the largest error bar is Platais 2. Middle panel: Comparison of 11 cluster parallaxes from Gaia Collaboration, van Leeuwen et al. (2017b) to the parallaxes computed from the fitted distances. Right panel: Comparison of cluster parallaxes from Gaia Collaboration, van Leeuwen et al. (2017b) to the weighted mean parallax of TGAS cluster members for 11 clusters. In all plots, the error bars are generally smaller than the marker used. The dashed gray line indicates the line of equality and the green line represents the best fit.

at the MWSC parallax. Currently, TGAS parallaxes are not used to aid in the distance determination from isochrone fitting, so deviations are expected between the two values. We also checked to see if the discrepancy in parallaxes could have resulted from our assumption of solar metallicity for all clusters. For the 14 clusters with metallicity values, a plot of the differences between trigonometric parallaxes and photometric parallaxes as a function of metallicity did not reveal a strong correlation. Thus, the assumption of solar metallicity has little or no effect on the distances determined. Overall, the mean trigonometric TGAS parallaxes and parallaxes from the fitted distance are compatible. Comparison of these parallaxes show 21% agree within 1σ, 63% agree within 2σ, and 80% agree within 3σ.

Our result for the Pleiades is also somewhat inconsistent. Computing the weighted mean parallax from the final TGAS members yields a parallax of 7.38±0.06 mas, which is consistent with a parallax of 7.48±0.03 mas (Gaia Collaboration, van Leeuwen et al. 2017b). However, the parallax computed from our fitted distance is rather high, at about $7.92^{+0.01}_{-0.02}$ mas. The difference in parallaxes could be due to the small binary offset we used. Converse & Stahler (2008) found an usually large fraction of binaries in the Pleiades, which implies a larger binary offset would be required for our analysis of the Pleiades. As aforementioned, the effect of binaries in clusters will be determined more statistically in the future, and will also enforce

consistency between trigonometric parallax and fitted photometric distance once Gaia DR2 data is available.

The parallaxes for 11 of these 24 clusters were also recently investigated by Gaia Collaboration, van Leeuwen et al. (2017b). The clusters in common are listed in Table 3.2, as well as the number of TGAS members found by each study, denoted by $N_{\rm Y18}$ (this study) and $N_{\rm vL17}$, and the number of TGAS members in common, $N_{\rm com}$. For nearly all clusters, there is very good overlap in TGAS membership. On average, Gaia Collaboration, van Leeuwen et al. (2017b) has found more members per cluster. This is mainly due to the significantly larger sky areas they considered for the clusters. Additionally, their membership determination relied solely on astrometry, while our analysis further refines TGAS membership based on stellar photometry.

The parallaxes of these clusters are compared to the parallaxes derived from the fitted distance in the middle plot of Fig. 3.4. The median difference in parallaxes is 0.25 mas and the dispersion is 0.34 mas. Due to the small formal errors on the isochrone-fitted distances, only four clusters are compatible within 3σ.

Since the analysis of Gaia Collaboration, van Leeuwen et al. (2017b) focuses on TGAS astrometric data, we also compare those parallaxes with the weighted mean parallax of our TGAS cluster members. In this case, the median parallax difference is 0.23 mas with a dispersion of 0.27 mas. Overall, the TGAS parallaxes of these 11 clusters are in great agreement; 18% agree within 1σ, 82% within 2σ, and 91% within 3σ.

Table 3.2: Number of TGAS stars in common with Gaia Collaboration, van Leeuwen et al. 2017b for 11 clusters

Name	$N_{\rm Y18}$	$N_{\rm vL17}$	$N_{\rm com}$
Blanco 1	48	44	36
α Per	84	116	66
Pleiades	91	154	85
NGC2232	8	31	4
NGC2451A	24	37	19
Praesepe	56	79	46
IC2391	24	43	21
IC2602	32	66	32
Coma Ber	33	49	30
NGC6475	49	78	36
NGC7092	26	23	14

3.5 Summary and conclusions

In the Gaia era, it is paramount to develop new analysis techniques and to combine archived data with newly available data, in order to gain new insights about open clusters and subsequently, the Galactic disk. As aforementioned, efforts have already been made by several groups, including Monteiro et al. (2010), Dias et al. (2012), and Palmer et al. (2014). We have also taken a step toward this effort and developed an automatic isochrone-fitting procedure that determines both cluster membership and cluster parameters. Using precise proper motions and parallaxes predominately from TGAS and HSOY for initial membership determination and six-band photometry for parameter determination and membership refinement, we have returned cluster parameters and cluster parallaxes that are similar to those found by Kharchenko et al. (2013) and Gaia Collaboration, van Leeuwen et al. (2017b) respectively, for the 24 closest open clusters in MWSC.

Our pipeline was developed specifically to work with the quality and limitations of the data set used and currently only applies to nearby clusters. We are continuing development of the pipeline to make it more generalized and applicable to more distant clusters. We are also exploring the use of a maximum likelihood method, instead of least squares for our isochrone fitting procedure. We will continue to refine the astrometric and photometric selection criteria and adapt it for use with Gaia DR2. A major revision will be to couple the trigonometric parallaxes and parallaxes derived from photometric distances, as the precision of Gaia parallaxes will provide a stronger constraint on cluster distance than any other method. Nevertheless, our current technique, which determines cluster distance independent of TGAS parallaxes, will still be valuable for the analysis of very distant clusters, which will have poor parallaxes in Gaia DR2.

On the bright side, more precise proper motions and parallaxes from Gaia DR2 will allow for more straightforward membership selection. The present challenge involves developing a single algorithm, to obtain accurate estimates of cluster parameters, that can be applied to all clusters, near and far. A tool such as this is necessary to assemble a large, homogeneous catalog of open clusters and their parameters, which can then be used to investigate the structure, dynamics, and evolution of the Milky Way.

4

New insights into open clusters in the Milky Way Star Clusters catalog with Gaia DR2

Based on Yen et al. to be submitted to Astronomy & Astrophysics

4.1 Motivation

Gaia DR2 provides unprecedented high precision astrometric and photometric data for more than 1.3 billion stars. This rich data set is revolutionizing many aspects of astronomy and as discussed in Ch. 2, it is changing our knowledge about the open cluster population, as well as the way these objects are studied and characterized. Building upon the ground work from Ch. 3, we continued development of the cluster characterization pipeline, adapting it for use with DR2 data and generalizing it for analysis of clusters at all distances. This new tool will allow us to systematically and homogeneously reanalyze all MWSC open clusters, verifying their existence and providing improved parameters for true clusters.

4.2 Data and cluster sample

4.2.1 Open cluster list

The MWSC catalog surveyed a total of 3784 clusters from the literature and confirmed the existence of 3006 clusters. For about half of the confirmed clusters, cluster parameters were derived for the first time. The number of confirmed open clusters is 2808, of which 540 are sub-classified as moving groups, nebulosities, remnant clusters, and asterisms. This sample of open clusters was found to be nearly complete out to ∼1.8 kpc from the Sun.

This work primarily used kinematic data from PPMXL (Röser et al. 2010) and photometry from 2MASS (Skrutskie et al. 2006) to derive the fundamental cluster parameters. In our analysis, we make use of the following parameters from MWSC: cluster name, position, angular radius, distance, and reddening.

4.2.2 Gaia DR2 data

We begin with a cone search, in the DR2 data set, centered on the MWSC position for each cluster. The angular radius used is given by $r = r_{\rm cl} + r_{\rm ex}$, where $r_{\rm cl}$ corresponds to the angular radius of the cluster, *r2*, defined in the MWSC, and $r_{\rm ex}$ is an additional annulus added, which corresponds to an angle of 0.3°, to account for any systematics between the data sets.

To ensure we are using good quality photometric and astrometric sources from DR2, we use only sources with $G < 18$ mag and employ some of the selection suggestions from Lindegren et al. (2018). We require sources to have five parameter solutions, more than seven visibility periods, and good BP and RP photometry via the BP and RP photometric excess criterion (expression C.2 in Lindegren et al. (2018)). Furthermore, we adopt the re-normalized unit weight error (RUWE) criterion, RUWE $\leq$ 1.4, as suggested by Lindegren to obtain the good astrometric sources.

4.2.3 Isochrone models

The theoretical stellar isochrones used in this work are the PARSEC version 1.2S (Bressan et al. 2012) in the Johnson, 2MASS, and Gaia DR2 (Evans et al. 2018) photometric systems with $Z = Z_{\odot} = 0.0152$ (Caffau et al. 2009, 2011). We use solar metallicity because a measured metallicity is available for only a small fraction of open clusters; 13% of the ∼2000 open clusters in the updated catalog by Dias

et al. (2002). We note that for clusters with highly non-solar metallicities, this assumption will introduce a small bias in the derived cluster parameters. For our analysis with the future Gaia data releases, which include metallicity information for a large number of stars, we will compute cluster metallicities for member stars with this information and use the corresponding isochrone for analysis.

Our isochrone set spans the age range 5.7 dex $\leq \log t \leq$ 10.1 dex, at step sizes $\Delta \log t = 0.01$ dex. We also construct the corresponding zero-age main sequence for this isochrone set using the evolutionary tracks and "ptcri" file[1]. The age at which the ZAMS occurs for each mass in the isochrone set is determined by matching the MS_BEG point for a given mass in the ptcri file to the age at the MS_BEG point in the evolutionary track for that mass. The ZAMS B, V, J, H, K_s, G, G_{BP}, and G_{RP} magnitudes are then determined by finding the matching age and mass combination in the isochrone set.

4.3 Characterization pipeline methodology

Our cluster characterization pipeline is largely based on the one described in Yen et al. (2018). Several modifications and improvements have been made to take advantage of Gaia DR2. This includes a new procedure to re-identify MWSC clusters in the Gaia DR2 proper motion space, using the maximum likelihood method to fit stellar isochrones to the cluster photometry, and fitting the three fundamental parameters simultaneously.

The pipeline has three main components: (1) proper motion clustering detection, (2) membership selection from astrometry, and (3) isochrone fitting and membership refinement. In our analysis, we require an open cluster to be composed of at least 20 stars, so if at any point, the stars being considered falls below 20, the cluster exits the pipeline and is determined to be non-existent. The specific details of the pipeline are explained below and a schematic of the pipeline is summarized in Fig. 4.1.

4.3.1 Cluster signature in proper motion space

As shown in Yen et al. (2018), systematic offsets exist between the listed MWSC cluster proper motion, derived from PPMXL, and the updated cluster proper motion derived from DR1/TGAS data. Thus, to re-identify an MWSC cluster in DR2, its proper motion needs to be re-determined. Applying the iterative proper motion and

[1] http://people.sissa.it/∼sbressan/parsec.html

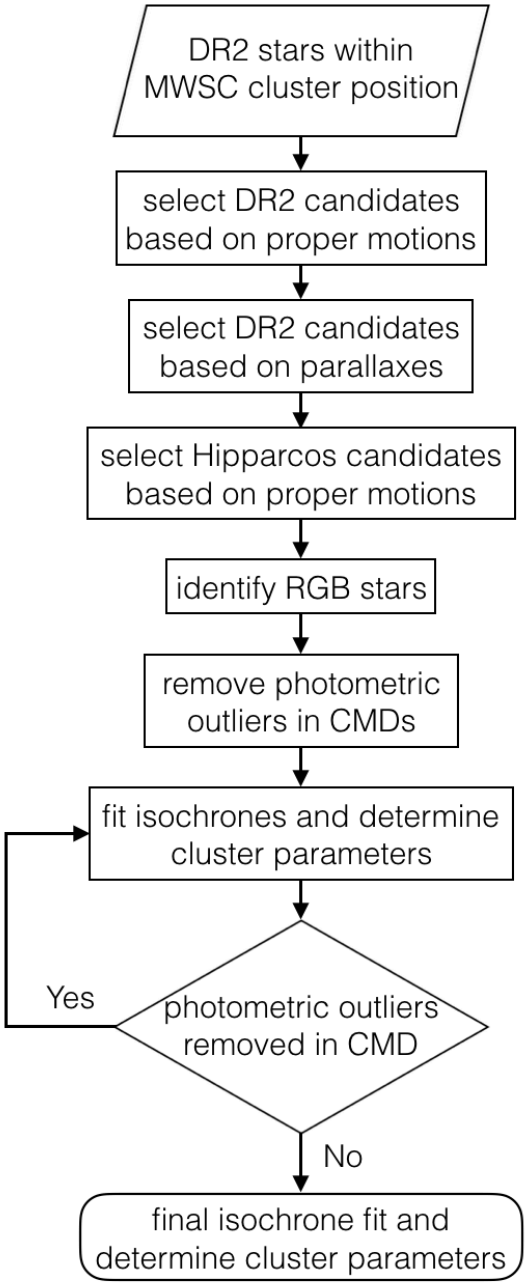

Figure 4.1: Flowchart of our cluster characterization pipeline. At each step preceeding the downward vertical arrow, there is a check on the number of member clusters. If at any point the cluster membership falls below 20 stars, the cluster is classified as not a cluster and exits the pipeline.

parallax selection procedure in Yen et al. (2018) on the DR2 data set led to misidentifications of the MWSC cluster, as the solution almost always arrived at the proper motion of the concentration of background stars. Formerly, this had worked well because the DR1/TGAS data set contained astrometry for only two million stars and the cluster sample was limited to nearby clusters, so there was little contamination from background stars and cluster stars could be easily distinguished from the background stars using parallaxes. The DR2 data set contains astrometric data for over one billion stars, so the number of stars within the cone search of the cluster position increases significantly, making it more difficult to distinguish the cluster from the background in the astrometric spaces, especially for distant clusters.

To identify possible signatures of the MWSC cluster in proper motion space, we limit our search to a square of $\mu_{\alpha,\,\mathrm{MWSC}} \pm 10$ mas yr^{-1} and $\mu_{\delta,\,\mathrm{MWSC}} \pm 10$ mas yr^{-1} around the MWSC proper motion in RA and DEC, $\mu_{\alpha,\,\mathrm{MWSC}}$ and $\mu_{\delta,\,\mathrm{MWSC}}$ respectively, where $\alpha^* = \alpha \cdot \cos\delta$. Open clusters are easily distinguished by their bright member stars, so a magnitude cut of $G < 16$ mag is applied in order to reduce the contamination from background stars and subsequently, more clearly see evidence of a cluster. Furthermore, we also limit the DR2 parallaxes of the stars to being no more than three times the MWSC cluster distance. For nearby clusters, this significantly reduces the background stellar population, so the cluster signature is apparent, while for distant clusters, this criterion does not drastically change the selection. These criteria ensure that we properly identify the same cluster recorded in the MWSC.

Next, we estimate the probability density function (PDF) of the selected proper motions by applying a gaussian kernel density estimation, where the gaussian kernels have a bandwidth of 0.08 mas yr^{-1}. This transforms a scatter plot of the selected proper motions into a smoothed image, where stars with similar proper motions form a peak. From the proper motion PDF, contours of 90%, 75%, 50%, and 25% levels are drawn to ascertain signatures of the cluster, if present.

In general, two peaks are expected in the proper motion PDF, one belonging to the background stars and the other to the cluster. However, sometimes the contamination from the background stars can be very strong, reducing the significance of the cluster or in extreme cases, even obscuring the cluster. In order to not miss the cluster, we select any strong and significant peaks, those with 90% closed contours, as well as, a single weaker peak, with either a 75%, 50%, or 25% closed contour, and keep them as possible evidence for the cluster identified in the MWSC. This

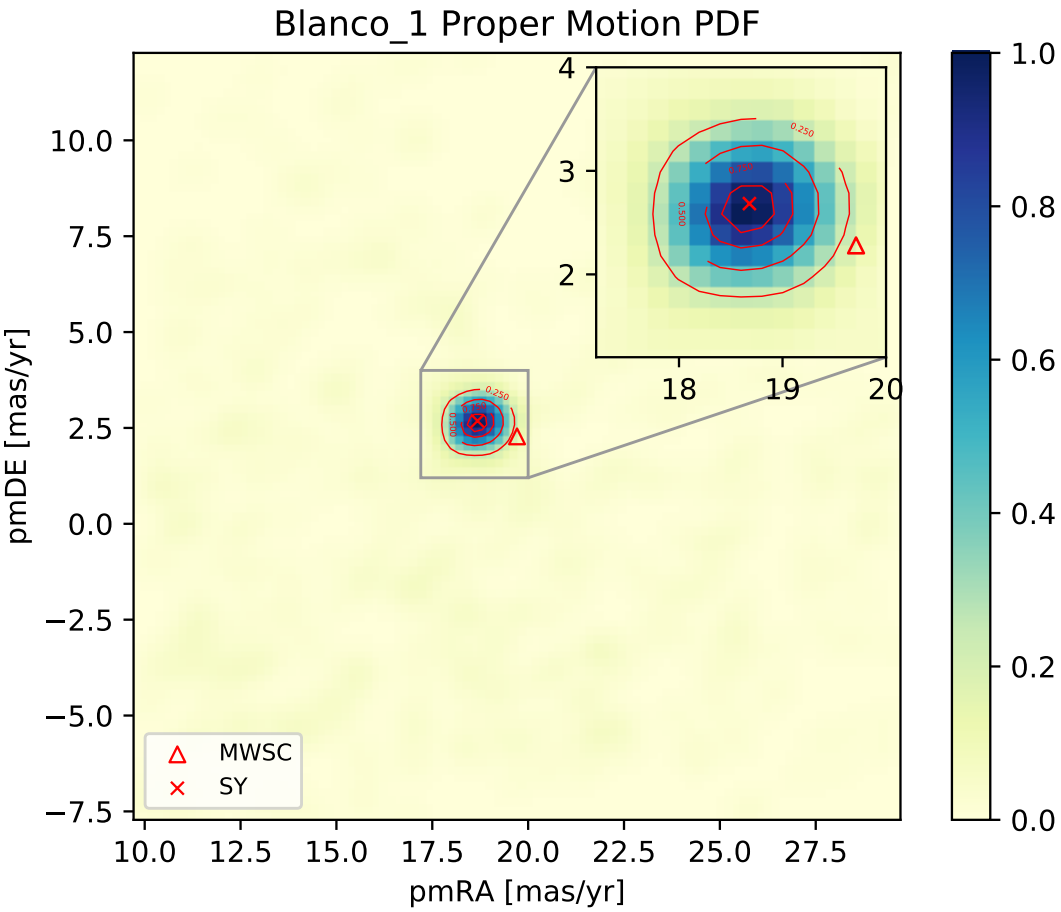

Figure 4.2: Blanco 1 proper motion PDF with contours. The stars of Blanco 1 have very high proper motions, so there is little confusion from background stars, and the cluster appears as a single peak in the KDE plot.The red cross denotes the initial cluster proper motion estimated by this work and the red triangle denotes the cluster proper motion as derived by the MWSC.

sometimes yields multiple initial proper motions to test per cluster, where each are used as inputs in the astrometric selection part of the characterization pipeline described in Sect. subsec:astrosel and successful completion through the pipeline will determine which is the true cluster proper motion.

In the exceptional cases, like Blanco 1, the background population is sparse and the cluster appears as the single peak and the only 90% closed contour in the proper motion PDF, as seen in the bottom left panel of Fig. 4.2. However, for clusters similar to Kharchenko 1, there is a strong background presence, the background peak will be represented by the 90% closed contour and the cluster will have a lower

level closed contour, as shown in the bottom right panel of Fig. 4.3.

For 38 clusters, we found no trace of a cluster in proper motion space, as there were less than 20 stars in our proper motion window. For the remaining 2770 clusters, we obtained a total of 4650 proper motion signatures, as 1380 clusters had more than one proper motion signature identified. Cantat-Gaudin et al. (2018a) determined DR2 proper motions for the MWSC open cluster sample also and recovered 1053 of them. For 95% of those clusters, we also derived cluster proper motions that coincide within 0.5 mas yr^{-1}.

We note that when comparing our results to Cantat-Gaudin et al. (2018a), for the 1053 clusters in common, five clusters either completely escape our detection or yield a drastically different proper motion than theirs because we limited our search to a 20 mas yr^{-1} × 20 mas yr^{-1} proper motion window centered on the MWSC value. These clusters are Collinder 271, NGC6005, Pismis 11, Ruprecht 107, and Trumpler 19. We performed a test using a larger search window, and while we obtained similar proper motions to Cantat-Gaudin et al. (2018a) for these five clusters, however, yielded mis-detections for the other clusters, due to the introduction of more background stars. Moreover, using a larger search window could lead to mis-identification of the MWSC cluster if there are two clusters with similar sky positions, thus we kept our original search window and note that these clusters will be missing from our analysis.

4.3.2 Astrometric membership selection

To determine the members of the cluster, we begin by selecting stars within our cluster proper motion estimate from the proper motion PDF. The median standard proper motion uncertainty of stars with $G = 18$ mag is about 0.2 mas yr^{-1} (Lindegren et al. 2018), but also recognizing that there exists some intrinsic scatter in the proper motions of cluster members, all stars within a radius of 1 mas yr^{-1} of the cluster proper motion estimate are selected as candidate members. Within a larger radius of 2.5 mas yr^{-1}, stars are also selected if their proper motion errors are within 3σ of the cluster proper motion estimate, to ensure we do not miss member stars of more dispersed clusters. Since the cluster proper motion estimates have been determined with only DR2 data, there is no need to iterate this step, unlike the membership selection procedure in Yen et al. (2018), which relied on the MWSC cluster proper motions.

The selection of candidate cluster members is further refined by their parallaxes.

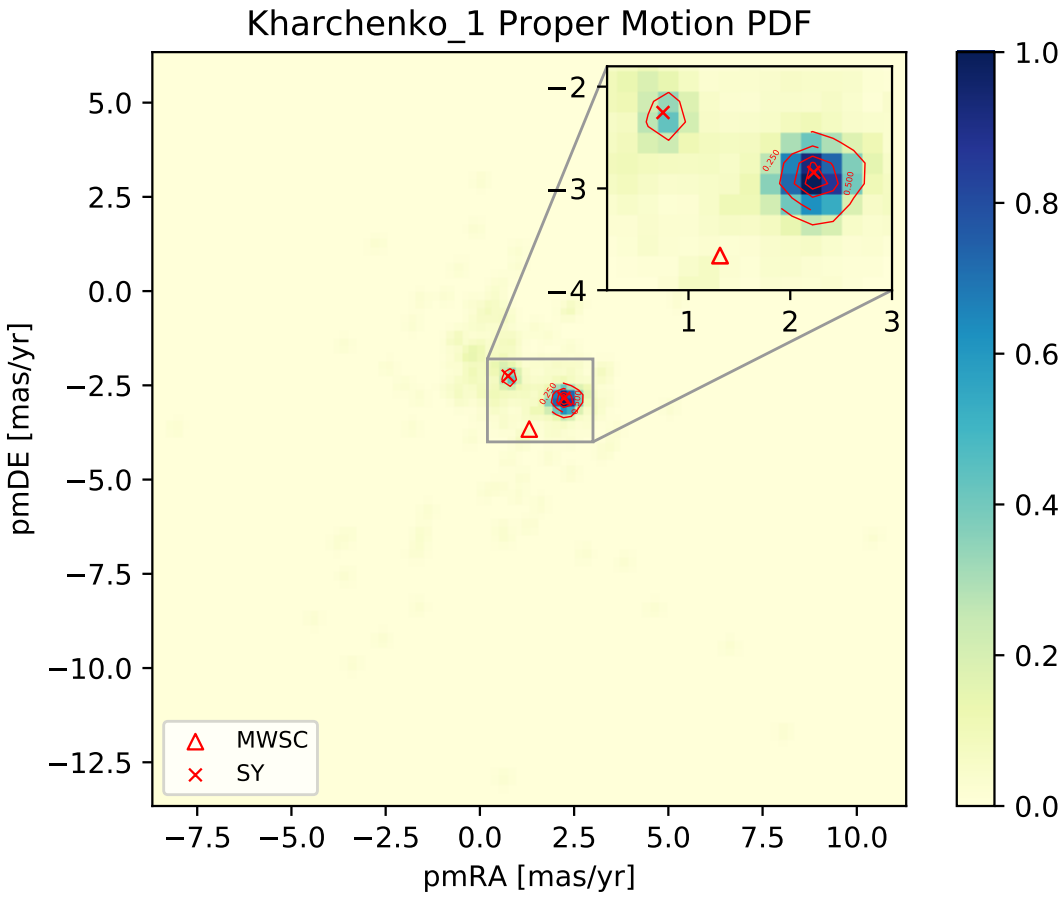

Figure 4.3: Kharchenko 1 proper motion PDF with contours. The stars of Kharchenko 1 have low proper motions, similar to the background, thus two peaks appear in the proper motion KDE. The red crosses denote the initial cluster proper motions estimated by this work and the red triangle denotes the cluster proper motion as derived by the MWSC.

Similar to the procedure in Yen et al. (2018), we first use the cluster's MWSC distance to compute a cluster parallax and select only the stars which have 3σ parallax errors consistent with this parallax value. The weighted mean cluster parallax, $\bar{\varpi}_{\mathrm{T}}$, of these stars is then computed and the 3σ parallax error of the stars are compared to this $\bar{\varpi}_{\mathrm{T}}$ to re-determine membership, in order to account for systematic differences between MWSC and DR2. This step is iterated until the cluster membership no longer changes and the astrometric members are selected.

If the mean cluster parallax falls below 75% of the MWSC parallax, that is, moving towards the parallax of the background stars, then the parallax iteration is canceled and candidate member stars are simply selected if their 3σ parallax error

is consistent with the MWSC parallax.

The cluster parallax determined at this step in the pipeline is an initial estimate. The final cluster parallax will be determined after the membership is confirmed from a photometric analysis.

4.3.3 Inclusion of Hipparcos stars

DR2 is mostly complete between $G = 12$ mag and $G = 17$ mag, but is still incomplete at the bright end, $G < 7$ mag (Gaia Collaboration et al. 2018b). Furthermore, the astrometric quality of sources brighter than $G \sim 6$ is much lower than that for the majority of the data set (Gaia Collaboration et al. 2018b). The cross-match between DR2 and Hipparcos yields 83,034 stars in common, which means 34,921 Hipparcos stars are missing from DR2. The brightest stars of a cluster are imperative to include because they set the age of the cluster. In an effort to recover these missing stars, we search for the Hipparcos stars within the cluster field and use the proper motion prescription described in Yen et al. (2018) to select the Hipparcos stellar candidates.

4.3.4 Photometric membership selection

Before fitting isochrones to the candidate cluster members, we first identify the RGB/highly-evolved stars and clean the CMDs of any obvious non-members. We seek to identify the highly-evolved stars because they will be treated slightly differently when fitting the isochrone, described in Sect. 4.3.5. In the G vs. $G - G_{\rm RP}$ diagram, the ZAMS is placed at the MWSC distance and reddening then shifted vertically by 5 magnitudes, so that the ZAMS is situated between the cluster main sequence and the highly-evolved stars. The stars to the right of the ZAMS are then selected as RGB/highly-evolved stars. Since highly-evolved stars are bright, this step is also performed in the V vs. $B - V$ CMD, as they might be Hipparcos-only stars and thus, missing Gaia photometry. For the next photometric selection, these highly-evolved stars are kept out of the sample.

To remove any obviously discrepant stars from the CMD, two cuts, below and above the cluster main sequence, are performed. In the $G_{\rm RP}$ vs. $G_{\rm BP} - G_{\rm RP}$ diagram, the ZAMS is placed using the DR2 parallax and MWSC reddening, then shifted vertically down by 0.5 magnitudes, to remove stars located below the cluster main sequence. We do not want to lose any bright members, so bright stars, with $G_{\rm RP} < 10$ mag, are kept. In the G vs. $G - J$ CMD, the ZAMS is placed using the DR2 parallax and MWSC reddening, then shifted up by 3 magnitudes,

so stars above this line are removed. Again, bright stars with $G < 12$ mag and $G - J < 1.0$ mag are kept. This step concludes the initial membership selection via astrometry and photometry for a given cluster.

4.3.5 Isochrone fitting

With the cluster membership determined, isochrones are fitted to the cluster photometry using the maximum likelihood technique. For isochrone fitting, all Gaia photometric errors are multiplied by a factor of 10 because the standard Gaia photometric errors are more precise than the isochrones. At the faint end, this discrepancy becomes quite large, about a magnitude. Also, the non-Gaia photometric errors of RGB/highly-evolved stars are increased by a factor of 10 because the isochrones in this evolutionary stage are not so accurate. The isochrone fitting process takes at minimum two iterations because cluster membership is refined using the result of each fit.

Likelihood

The likelihood of a cluster having certain parameters ($E(B-V)$, d, $\log t$) is the product of all likelihoods of the stars (i):

$$L_C = \prod_{i=0}^{n} L_i \tag{4.1}$$

For computational purposes, it can be better to compute in terms of log-likelihoods. This equation can be expressed as:

$$\log L_C = \sum_{i=0}^{n} \log L_i \tag{4.2}$$

The likelihood for a star having a certain mass in a model with certain cluster parameters is given by:

$$L = \exp\left(-0.5 * \left(\sum_{n} \left(\frac{x_n - X_n}{\sigma_{x_n}}\right)^2 + 6\left(\frac{\varpi - \overline{\varpi}_{\text{clus}}}{\sigma_{\varpi}}\right)^2\right)\right), \tag{4.3}$$

where the first term is the sum of the χ^2 residuals of the observed stellar apparent magnitudes, x_n, and the apparent magnitudes calculated from the absolute magnitudes from the isochrone, X_n, over all the available n photometric bands and the

second term is the χ^2 residual of the observed parallax of the star, ϖ, and the average DR2 parallax of the cluster members currently considered, $\overline{\varpi}_{\rm clus}$. The parallax is weighted by a factor of 6, the average number of photometric bands available, so that its influence on the final result is similar to the photometric bands combined. Without the increase weighting, the parallax information was essentially down-weighted by the photometry.

In this work, the photometric bands used include B, V, J, H, K_s, G, $G_{\rm BP}$, and $G_{\rm RP}$. To compare the observational photometry to the theoretical photometry, we convert the theoretical absolute magnitude X' into a theoretical apparent magnitude X. For a given photometric band this is expressed as:

$$X = X' + 5\log_{10} d - 5 + A_X, \tag{4.4}$$

where d is the distance in parsecs, as given in the parameter set tested, and A_X is the extinction measured in the given band corrected for the reddening being tested. Explicitly, for the G band, if $A_G/A_V = 0.86$, then $A_G = 0.86A_V$, where $A_V = 3.1E(B-V)$ for the Milky Way.

The precision of Gaia photometry allows us to clearly distinguish the equal-mass binary sequence of cluster members in the CMD as a 0.7 mag brighter sequence (Coronado et al. 2018) compared to the cluster main sequence. To take this into account, we compute the likelihoods for the star to be on the cluster main sequence and the equal-mass binary sequence, which introduces a -0.7 mag into Eq. 4.4. Furthermore, the expected fraction of equal mass binaries is 0.1 (Coronado et al. 2018), so we also downweight the likelihood for the star to be on the equal-mass binary sequence by 0.1. The larger of the two likelihoods is then taken as the likelihood for the star L_i for the considered parameter set.

In determining the likelihood of the star L_i, we also obtain a mass estimate for the star. For a given parameter set, the mass grid of the isochrones are interpolated to steps of 0.01 $M_\odot$, thus the mass at the maximum likelihood in the mass-likelihood distribution of the star gives the mass and likelihood for the star. These stellar likelihoods are then put into Eq. 4.2, where we then obtain an overall cluster log-likelihood, which corresponds to the likelihood for the cluster to have a given parameter set (age, distance, and reddening).

Given the likelihood definition in Eq. 4.3, stars with a large χ^2 (not in well agreement within errors to the model) will have a very small, or 0, likelihood and will not contribute to the overall likelihood of the cluster. On the other hand, stars

with a small χ^2 (good agreement within errors to the model), will have a positive, non-zero likelihood and will contribute to the likelihood of the cluster having the considered parameters.

Initial fitting

For the initial fit, large parameter spaces are explored, so that MWSC values are not strictly relied upon. In $\log t$[yr], a range of 6.6 to 10.2, at steps of 0.2, is used. For distance, the mean DR2 parallax of the cluster is used to generate the parameter space. The minimum distance value, $d_{\rm min}$ is computed by adding 4σ to the mean cluster parallax, that is 1000 / ($\bar{\varpi} + 4\sigma_{\bar{\varpi}}$). The difference between the distance given by the mean cluster parallax, $d_{\bar{\varpi}}$, and $d_{\rm min}$ is then added to $d_{\bar{\varpi}}$ to obtain the maximum distance value, $d_{\rm max}$. The distance grid is then $d_{\rm min}$ to $d_{\rm max}$ at ten evenly spaced increments. For $E(B-V)$, the minimum value is set at -0.1 and the maximum value is set to be $E(B-V)_{MWSC}$ + 0.6, with ten evenly spaced increments.

The overall cluster likelihood is computed for each combination of $\log t$, d, and $E(B-V)$ values. The maximum likelihood distribution for these three parameters are created and a spline-interpolation is used to determine the peak and the 1σ error bars. If the peak is at the edge of any of these parameter spaces, the parameter space is widened and the likelihoods are recomputed. Using the fitted parameters, the photometric selection described in Sect. 4.3.4, discrepant stars are removed and the intermediate fitting process begins. If there are no discrepant stars, the cluster proceeds to the final fitting, described in Sect. 4.3.5.

Intermediate fitting

For the intermediate isochrone fits, the parameter set is tuned to the preceding fitted results. The range of ages is $(\log t)_{\rm i} \pm 1.5$ dex, at steps of 0.1 dex, where $(\log t)_{\rm i}$ is the previously fitted age. The distance range is defined by $d_{\rm i} \pm 5\sigma_{d,\rm max}$ at ten evenly-spaced increments, where $d_{\rm i}$ is the previously fitted distance and $\sigma_{d,\rm max}$ is the maximum of the asymmetrical fitted distance error bars. Similarly, the reddening range is defined by $E(B-V)_{\rm i} \pm 5\sigma_{E(B-V),\rm max}$ at ten evenly-spaced increments, where $E(B-V)_{\rm i}$ is the previously fitted reddening and $\sigma_{E(B-V),\rm max}$ is the maximum of the asymmetrical reddening error bars.

As in Sect. 4.3.5, the likelihood for each combination of parameter sets is computed and the fitted parameter values and their errors are determined from their

respective likelihood distributions. Furthermore, if the peak in any of these distributions is at the edge of the parameter space, the corresponding parameter space is widened and the likelihoods recomputed. Discrepant stars are removed as described in Sect. 4.3.4, and this intermediate fitting process is repeated until there are no discrepant stars to remove.

Final fitting

When the cluster membership is finalized by the intermediate fitting process, the cluster goes through one final fitting at a finer grid of parameter values. The range of ages is $(\log t)_{\rm i} \pm 0.35$ dex, at steps of 0.02 dex, where $(\log t)_{\rm i}$ is the previously fitted age. The distance range is defined by $d_{\rm i} \pm 5\sigma_{d,\rm max}$ at 15 evenly-spaced increments, where $d_{\rm i}$ is the previously fitted distance and $\sigma_{d,\rm max}$ is the maximum of the asymmetrical fitted distance error bars. Similarly, the reddening range is defined by $E(B-V)_{\rm i} \pm 5\sigma_{E(B-V),\rm max}$ at 15 evenly-spaced increments, where $E(B-V)_{\rm i}$ is the previously fitted reddening and $\sigma_{E(B-V),\rm max}$ is the maximum of the asymmetrical reddening error bars.

Outputs of cluster characterization pipeline

The fundamental result of the characterization pipeline is to determine whether a given proper motion signature belongs to a true cluster. Our definition of an open cluster requires a concentration in proper motions and more than 20 member stars. If a cluster has less than 20 candidate members at any point in the pipeline analysis, it will exit the routine and classified as non-existent. If a cluster successfully completes the pipeline, it is classified as a true cluster. The outputs of the pipeline for true clusters is fully summarized in Tbl. 4.1, and includes mean astrometric parameters, the three fitted parameters, and quality flags. For each true cluster, various CMDs are provided, e.g. those in Fig. 4.5, as well as membership tables with all the stellar information used, the derived mass estimate, and the classification of being a single or binary. The specifics of the cluster membership tables are summarized in Tbl. 4.2. Publication of this work will allow all cluster tables and plots to be available electronically at the CDS[2].

[2]Until publication, requests for cluster information can be sent to me, email: syen@lsw.uni-heidelberg.de.

Table 4.1: Contents of main cluster table

Col	Label	Units	Explanations
1	MWSC	-	MWSC ID
2	Cluster	-	Cluster name
3	RA	deg	Right ascension (ICRS) at Ep 2015.5 from DR2
4	DE	deg	Declination (ICRS) at Ep 2015.5 from DR2
5	pmRA	mas/yr	Proper motion in RA (pmRA*cos(DE)) from DR2
6	e_pmRA	mas/yr	Error of proper motion in RA from DR2
7	pmDE	mas/yr	Proper motion in DE from DR2
8	e_pmDE	mas/yr	Error of proper motion in DE from DR2
9	Parallax	mas	Parallax from DR2
10	e_Parallax	mas	Error of parallax from DR2
11	Dist	pc	Fitted distance
12	e_Dist_min	pc	Lower 1-sigma error on fitted distance
13	e_Dist_max	pc	Upper 1-sigma error on fitted distance
14	E(B-V)	mag	Fitted E(B-V)
15	e_E(B-V)_min	mag	Lower 1-sigma error on fitted E(B-V)
16	e_E(B-V)_max	mag	Upper 1-sigma error on fitted E(B-V)
17	logt	dex	Fitted log(t)
18	e_logt_min	dex	Lower 1-sigma error on fitted log(t)
19	e_logt_max	dex	Upper 1-sigma error on fitted log(t)
20	num_MS	-	Number of main sequence stars
21	num_EMB	-	Number of equal-mass binary stars
22	flag_bin	-	Binary flag
23	flag_age	-	Age flag

4.4 Results

We have analyzed all 2808 MWSC clusters with our characterization routine. A summary of the statistics of our results is provided in Tbl. 4.3. The main goal of our work is to derive cluster parameters for MWSC clusters using the improved Gaia data and clean the MWSC cluster list, removing dubious clusters. In total, we obtain positive solutions for 1873 of the 2808 open clusters in the MWSC.

For the 1873 successfully re-confirmed clusters, we obtain a total of 2658 different solutions. Some clusters, which had more than one proper motion signature detected, yielded multiple solutions for its various proper motions. For several of these clusters, this is because the proper motions were rather similar, so there is significant overlap in membership between the final solutions. Duplicate solutions for clusters with more than 70% membership similarities are removed by keeping

Table 4.2: Contents of stellar data for a given cluster

Col	Label	Units	Explanations
1	ID	-	Gaia DR2 ID or Hipparcos ID
2	RA	deg	Right ascension J2000.0, Ep 2015.5 from DR2
3	DE	deg	Declination J2000.0, Ep 2015.5 from DR2
4	pmRA	mas/yr	Proper motion in RA from DR2
5	σ_{μ_α}	mas/yr	Error of proper motion in RA from DR2
6	pmDE	mas/yr	Proper motion in DE from DR2
7	σ_{μ_δ}	mas/yr	Error of proper motion in DE from DR2
8	parallax	mas	Parallax from DR2
9	σ_{ϖ}	mas	Error of parallax from DR2
10	B	mag	B magnitude
11	σ_B	mag	Error of B magnitude
12	V	mag	V magnitude
13	σ_V	mag	Error of V magnitude
14	J	mag	J magnitude 2MASS
15	σ_J	mag	Error of J magnitude 2MASS
16	H	mag	H magnitude 2MASS
17	σ_H	mag	Error of H magnitude 2MASS
18	K_s	mag	K_s magnitude 2MASS
19	σ_{K_s}	mag	Error of K_s magnitude 2MASS
20	G	mag	G magnitude
21	σ_G	mag	Error of G magnitude DR2
22	G_{BP}	mag	G_{BP} magnitude
23	$\sigma_{G_{BP}}$	mag	Error of G_{BP} magnitude DR2
24	G_{RP}	mag	G_{RP} magnitude
25	$\sigma_{G_{RP}}$	mag	Error of G_{RP} magnitude DR2
26	mass	$M_\odot$	Stellar mass estimated from isochrone fit
27	binary	-	Binary classification (0 = single, 1 = binary)

the solution with more member stars. The final cluster numbers and their number of solutions are listed in Tbl. 4.3. We obtain singular solutions for 1357 clusters and multiple solutions for 516 clusters. For the 516 clusters with multiple solutions, further analysis must be done to determine which solution matches to the cluster of interest, as explained in Sec. 4.4.3. For discussion of the results, we split our sample of true clusters into three categories: single-solution cluster with no flags, single-solution cluster with flags, and multi-solution clusters. The flags indicate a doubt in the quality of our fitted result and are further explained in Sec. 4.4.2.

Table 4.3: Summary of results

number of clusters	
MWSC OC	2808
with initial PMs	2770
with no initial PM	38
yes clusters	1873
not clusters	912
maybe clusters	23
number of solutions returned	
total initial PMs to test	4650
yes solutions	2658
no solutions	1992
yes cluster numbers	
with one *yes* solution	1357
with two *yes* solutions	420
with three *yes* solutions	73
with four *yes* solutions	20
with five *yes* solutions	3

Table 4.4: Summary of flags for single-solution clusters

Total	No Flags	Binary Flag	Age Flag	Red Flag
1357	913	239	292	15

4.4.1 Single-solution clusters with no flags

We consider our most reliable cluster results to be those with one solution and no flags, of which there are 913 clusters. An overview of the fits of some single-solution clusters with no flags is provided in Fig. 4.4. A full comparison of the results for these clusters to the literature is provided in Sect. 4.5.

An exceptional example is Blanco 1[3], a nearby cluster with a well-populated and thin main sequence and a few photometric binary stars. As seen from its CMDs in Fig. 4.5, the isochrones fit the brighter end of the cluster main sequence perfectly, but there is a small deviation in the fainter, low-mass end of the isochrone. This discrepancy in the Gaia bands can also be seen in the CMDs of other clusters and is due to the use of fixed Gaia extinction coefficients, k_X ($k_X = A_X/A_V$, where X could be G, G_{BP}, and G_{RP}), relations (Danielski et al. 2018). We used fixed k_X values provided by the Padova group, which were computed using Cardelli et al.

[3] Also, my personal favorite over the past four years.

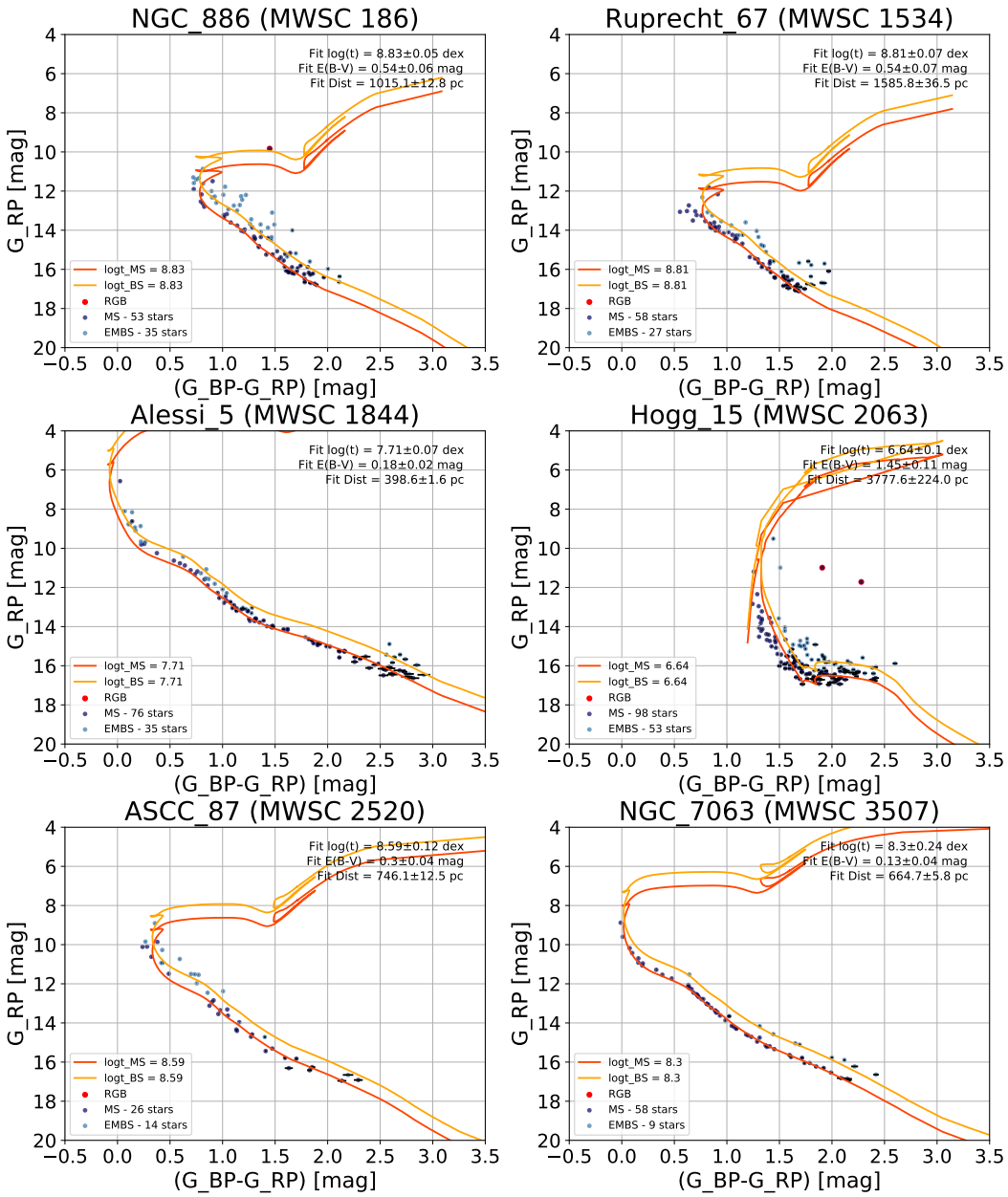

Figure 4.4: G_{RP} vs. $G_{BP} - G_{RP}$ CMDs for six clusters that returned a single solution without any flags. From *top* to *bottom* and *left* to *right*, these clusters are: NGC 886, Ruprecht 67, Alessi 5, Hogg 15, ASCC 87, and NGC 7063. Main sequence stars are shown as dark blue points with the corresponding main sequence isochrone as a solid red line. photometric binaries are shown as light blue points with the corresponding equal-mass binary isochrone as a solid orange line.

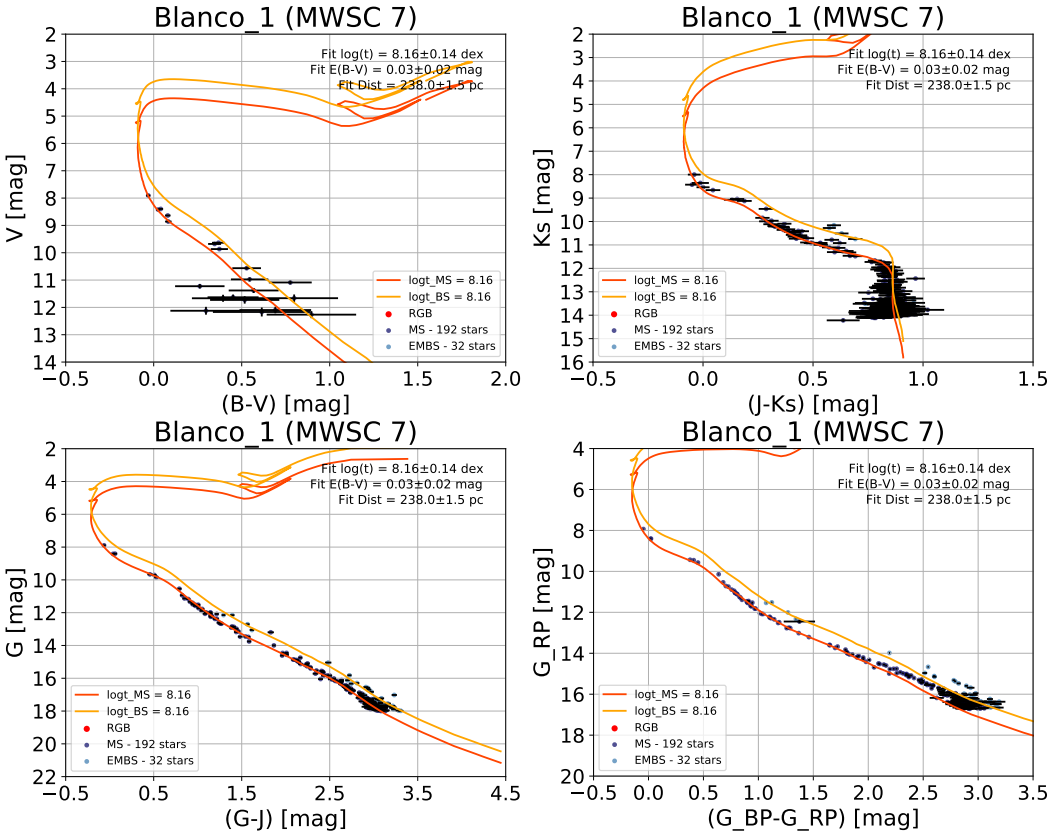

Figure 4.5: Various CMDs for Blanco 1, a cluster with a single solution returned without any flags. Main sequence stars are shown as dark blue points with the corresponding main sequence isochrone as a solid red line. photometric binaries are shown as light blue points with the corresponding equal-mass binary isochrone as a solid orange line.

(1989) and O'Donnell (1994) extinction curves with $R_V = 3.1$ (Schultz & Wiemer 1975). However, e.g. Jordi et al. (2010), has shown that k_X is actually dependent on the color (stellar effective temperature) and extinction in the large Gaia bands, thus a more sophisticated formalism for the extinction coefficients in the Gaia bands is required, see Danielski et al. (2018) and Gaia Collaboration et al. (2018a). This will be implemented in the future version of our pipeline analysis.

We have also found examples of clusters with differential reddening in our sample. In the CMDs of these clusters, the cluster sequence is found to be a couple magnitudes wide, extending beyond the equal-mass binary sequence, as shown in the top panel of Fig. 4.6 for NGC 2112 (MWSC 692). All these stars have roughly the same parallax, as shown in the bottom panel of Fig. 4.6, indicating that they

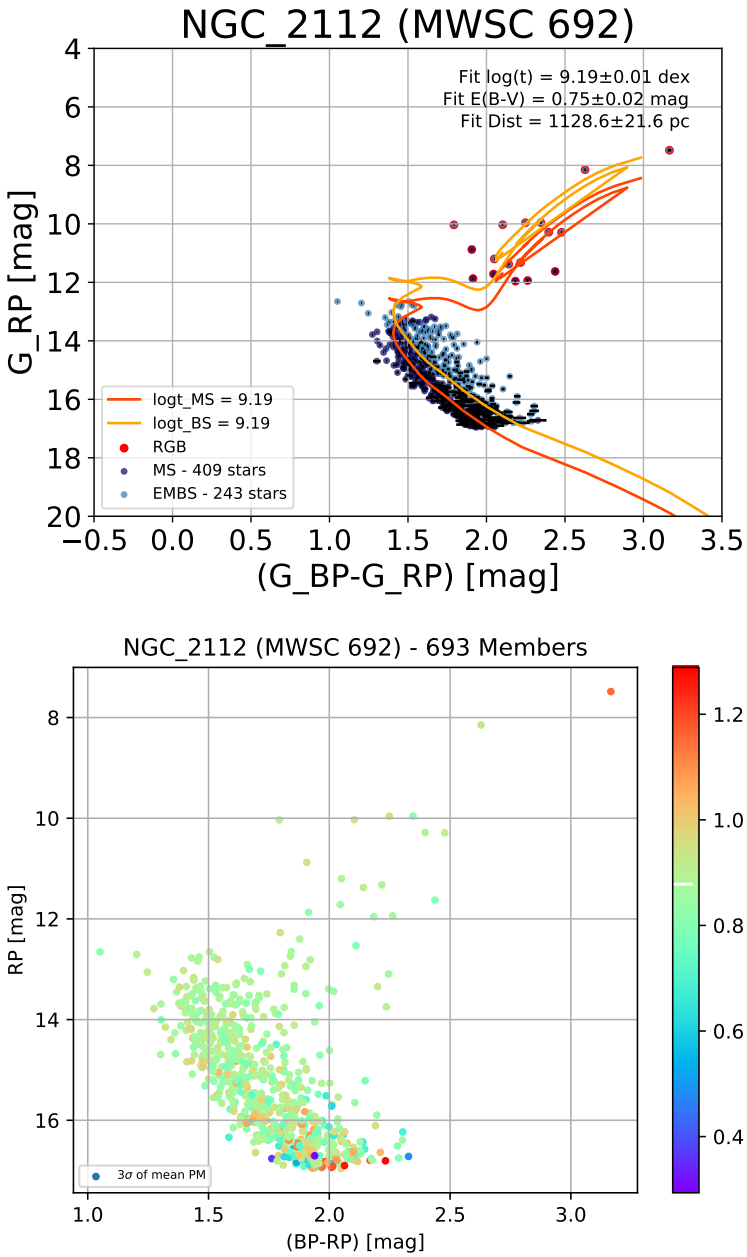

Figure 4.6: G_{RP} vs. $G_{BP} - G_{RP}$ CMDs for NGC 2112 (MWSC 692) showing variable extinction. *Top*: Final G_{RP} vs. $G_{BP} - G_{RP}$ diagram with fitted isochrone. Main sequence stars are shown as dark blue points with the corresponding main sequence isochrone as a solid red line. photometric binaries are shown as light blue points with the corresponding equal-mass binary isochrone as a solid orange line. RGB stars are indicated by a red circular outer edge. *Bottom*: Initial G_{RP} vs. $G_{BP} - G_{RP}$ diagram with stars color-coded by their parallaxes. The white marking on the color bar represents the mean cluster parallax.

are all in fact true members of the cluster and not field stars. Thus, the thickness of the cluster sequence is explained by the effects of variable extinction, caused by different layers of dust in between our line of sight to the cluster. Dutra & Bica (2000), for example, have studied this effect in NGC 2112, as well as 102 old open clusters.

4.4.2 Single-solution clusters with flags

In our determination of the cluster parameters, we identified three key features to judge the quality of our fitted results. The first is a binary flag, which indicates if the number of photometric binary members is more than 45% of the total cluster membership. As aforementioned, it is expected that equal-mass binaries account for ∼10% of the cluster population, however, the number of unequal-mass binaries, which lie between the main sequence and equal-mass binary sequence, is largely unknown. We expect the total binary fraction to be somewhere between 10% and 50%. The distribution of binary fractions in our single-solution clusters, in Fig. 4.7, shows a dip around 45%, thus we flag cluster solutions that have greater than 45% of binaries. This binary flag indicates that there is probably something wrong with the solution, as photometric binaries should not make up the majority of a cluster. For the 1357 clusters with single solutions, 239 clusters receive a flag on the fraction of binary stars.

Clusters end up with more stars identified as binary stars than single stars for a variety of reasons. One reason is due to a poorly matched theoretical isochrone. In the case of vdBergh-Hagen 164 (MWSC 2255), the final fit returns an age of $\log t = 8.91$, and visual inspection of the fit in the CMD shown in Fig. 4.8 shows a very poor fit. This age fits the brighter stars of the cluster to the main sequence and all the fainter stars, $G_{RP} > 14$ mag, to the binary sequence. The brighter end of the cluster sequence, between 10 mag and 14 mag in G_{RP}, shows the correct assignment of stars to their sequence, where the binary stars, are the few stars located 0.7 mag above the cluster main sequence. The initial age likelihood distribution of this cluster, Fig. 4.9, actually shows a bimodal age distribution. Plotting the younger, less-likely age visually shows a much better fit to the cluster, as seen in Fig. 4.10. This would yield proper main sequence or binary classification to the cluster members. Furthermore, the two brightest cluster members with $G_{RP} < 8$ mag

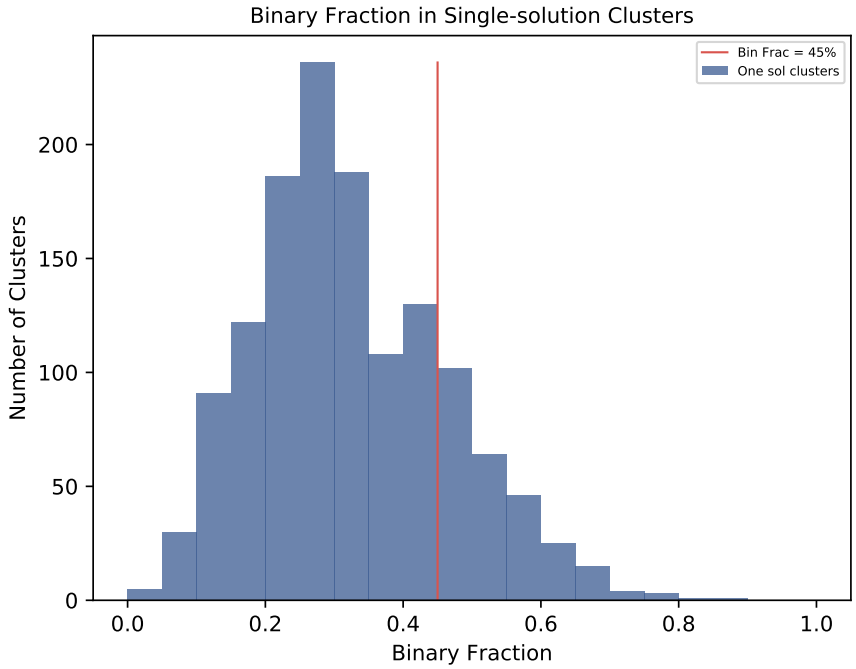

Figure 4.7: Fraction of photometric binary stars in single-solution clusters. Red line marks the 45% binary fraction; clusters to the right of this line are assigned a flag, as the fraction of photometric binaries should not make up the majority of cluster members.

are also included in the fit with the younger isochrone.

In other cases, the binary isochrone is fitted to the cluster main sequence instead of the cluster binary sequence, as in the case of Tombaugh 5 (MWSC 306), as shown in the top panel of Fig. 4.11. In this CMD, nearly all cluster members lie above the main sequence isochrone. However, it is important to remark that our fit considers photometry in all bands, and when the isochrone fit is viewed in another band, as shown in the bottom panel of Fig. 4.11, one sees that the main sequence isochrone does seem to fit the cluster main sequence better; in fact, in this CMD, a lower reddening value would give a better agreement but would create even more discrepancy in the G_{RP} vs. $G_{BP} - G_{RP}$ diagram. This might indicate an offset in some of the photometric bands with respect to each other. The cluster sequence of Tombaugh 5 is also quite wide, illustrating another possible cluster with differential reddening.

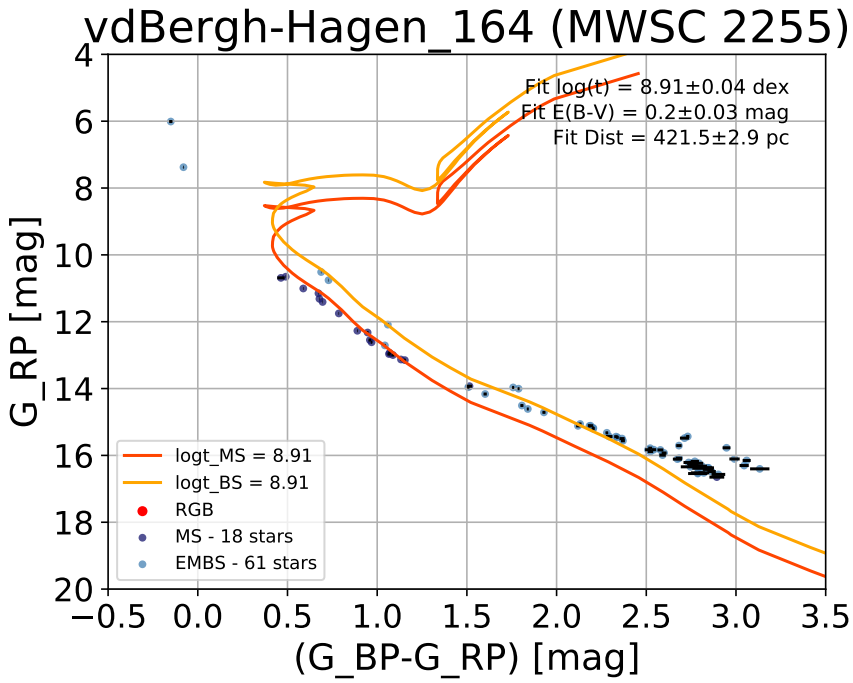

Figure 4.8: G_{RP} vs. $G_{BP} - G_{RP}$ diagram of vdBergh-Hagen 164, an example cluster with a single solution and binary flag. Dark blue points represent main sequence stars and light blue points are equal-mass binary stars. The two isochrones plotted are the main sequence isochrone in red and the equal-mass binary sequence in orange.

The second quality flag is based on the age determination. If the age-likelihood distribution for the cluster is singularly peaked, yielding very small errors on age, or yields an age with asymmetric error bars, we flag the result, as the viability of the age determination and overall cluster solution is doubtful. These specific criteria were imposed based on the results of an older version of the code where the age fitting was handled a bit differently. An investigation of flagged clusters shows that this criteria needs to be fine-tuned, as now, a singularly peaked age distribution does not always indicate a wrongly determined age, as in the case of NGC 7789 (MWSC3779), shown in Fig. 4.12, which is one of our most well-populated clusters. A number of these well-populated clusters actually received the age flag for having a singularly peaked age, but it is clear that a very limited number of isochrones would fit to the cluster, so for this thesis, we remove the age flag for clusters with more than 250 members. This cut off was determined because the age flag seemed to properly apply to clusters with less than 250 members. Of course, in the future publication of this work, a stronger criterium will be developed to properly flag the truly doubtful age determinations. For the clusters with a single solution, this

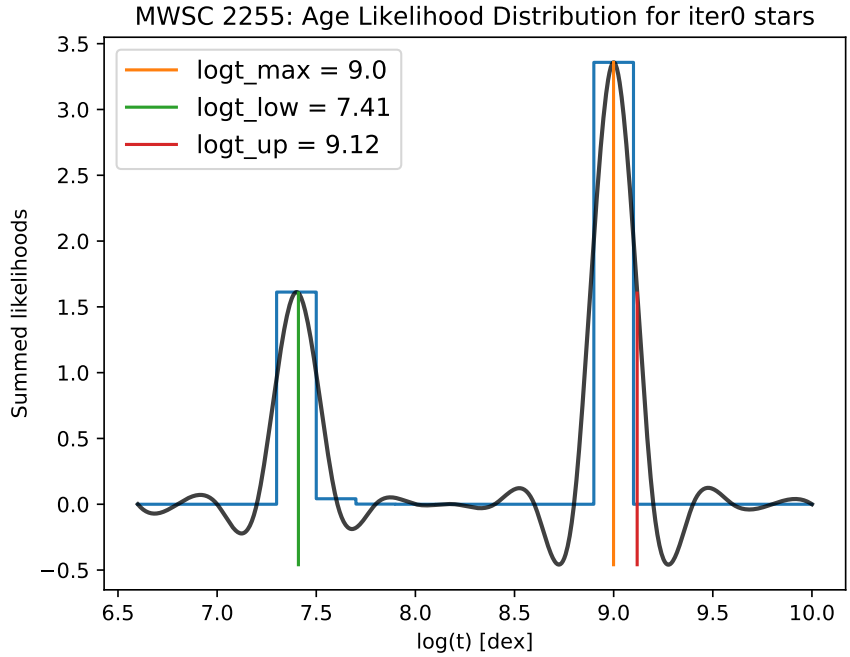

Figure 4.9: Initial age likelihood distribution of vdBergh-Hagen 164, an example cluster with a binary flag.

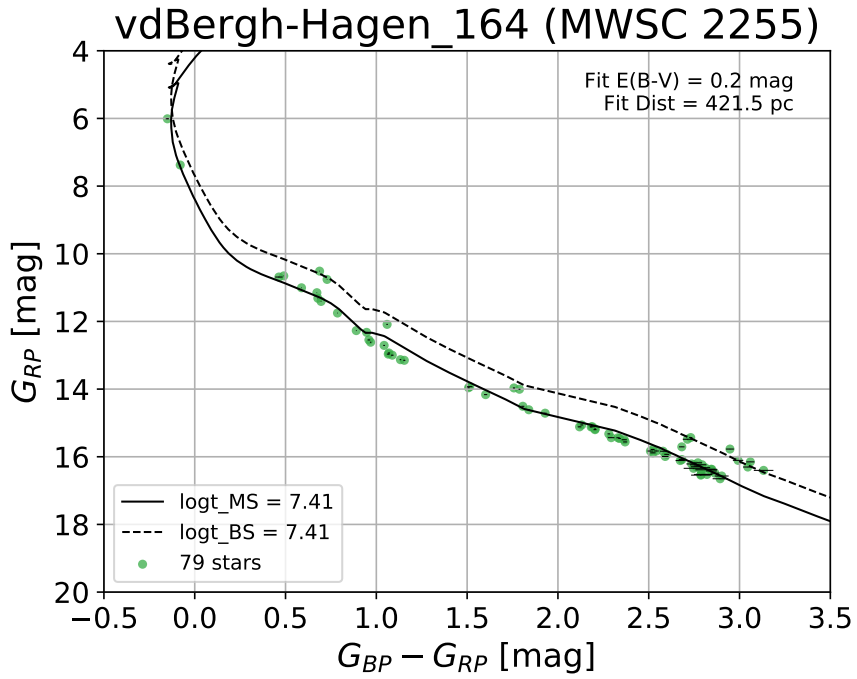

Figure 4.10: G_{RP} vs. $G_{BP} - G_{RP}$ diagram of vdBergh-Hagen 164 with an isochrone of the less likely age from Fig. 4.9. The solid black line represents the main sequence isochrone and the dashed black line represents the equal-mass binary sequence.

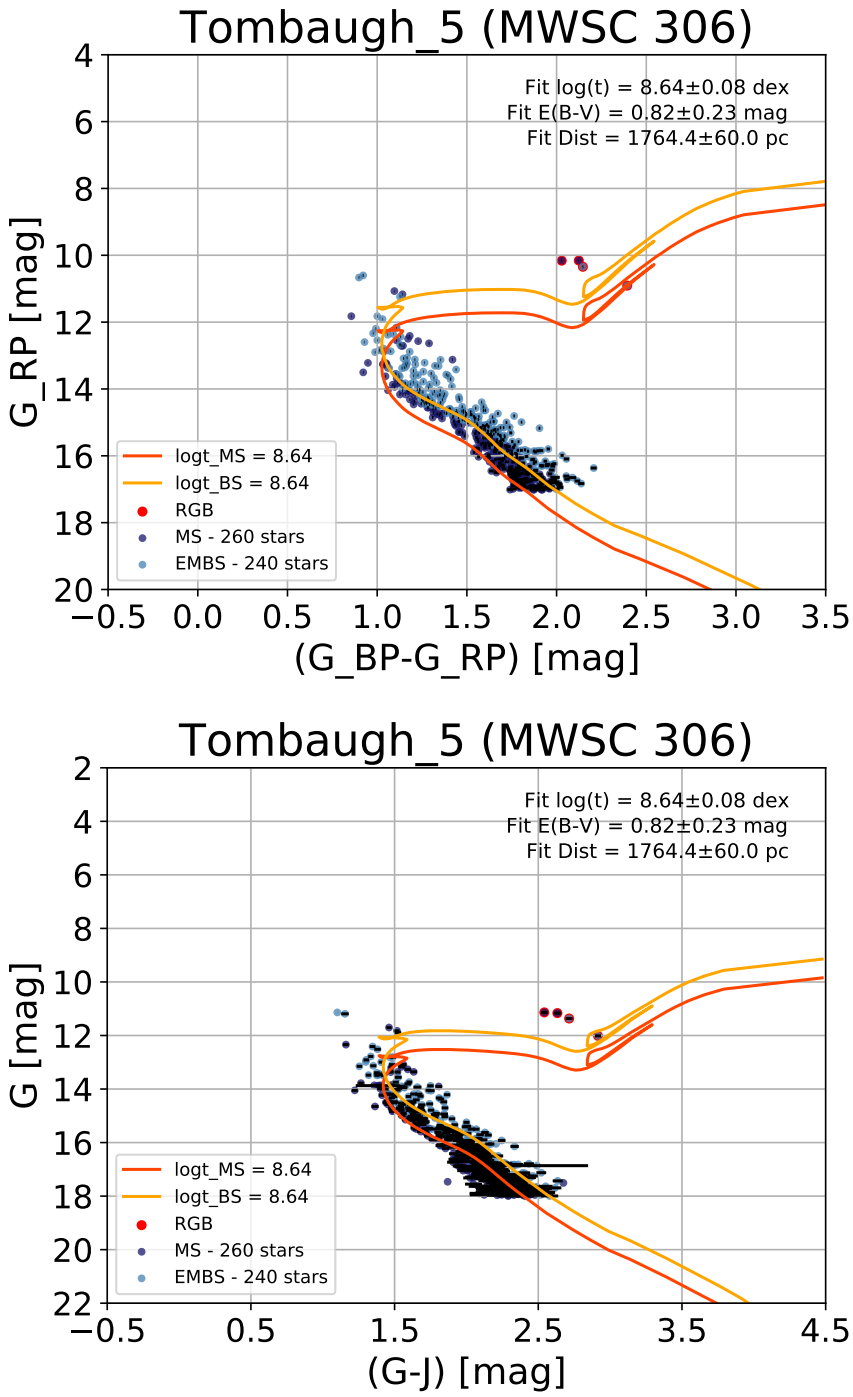

Figure 4.11: Two color magnitude diagrams of Tombaugh 5, G_{RP} vs. $G_{BP} - G_{RP}$ (top) and G vs. $G - J$ $(bottom)$. Main sequence stars are shown as dark blue points with the corresponding main sequence isochrone as a solid red line. photometric binaries are shown as light blue points with the corresponding equal-mass binary isochrone as a solid orange line. RGB stars are indicated by a red circular outer edge.

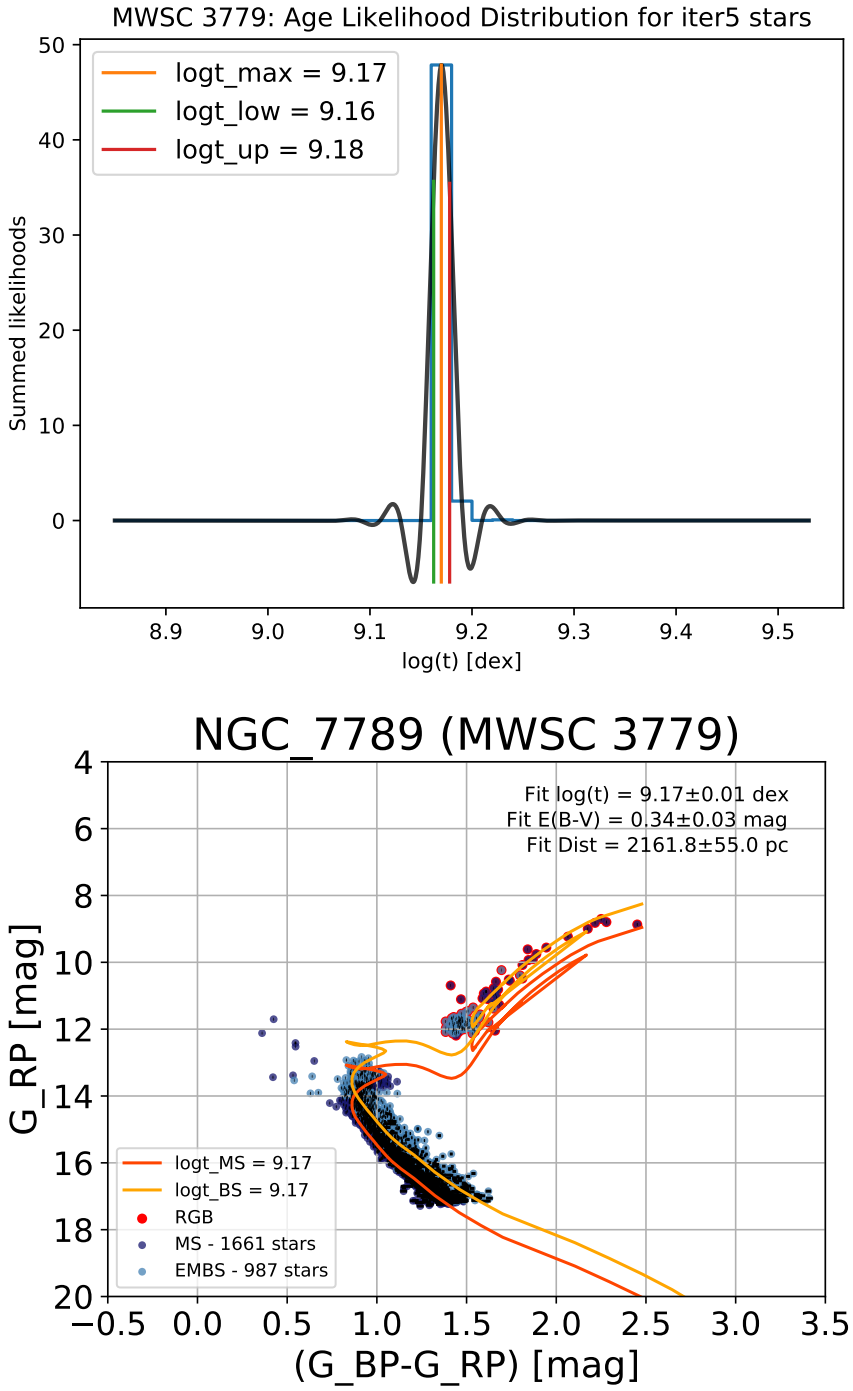

Figure 4.12: *Top:* Age likelihood distribution for NGC 7789 (MWSC 3779). The blue histogram shows the results from the likelihood fits and the black line is a spline interpolation to the likelihood distribution, from which the maximum age and 1σ error bars are derived. *Bottom:* G_{RP} vs. $G_{BP} - G_{RP}$ diagram for NGC 7789 with the fitted results. Main sequence stars are shown as dark blue points with the corresponding main sequence isochrone in red. photometric binaries are shown as light blue points with the corresponding equal-mass binary isochrone in orange. RGB stars are indicated by a red outer edge.

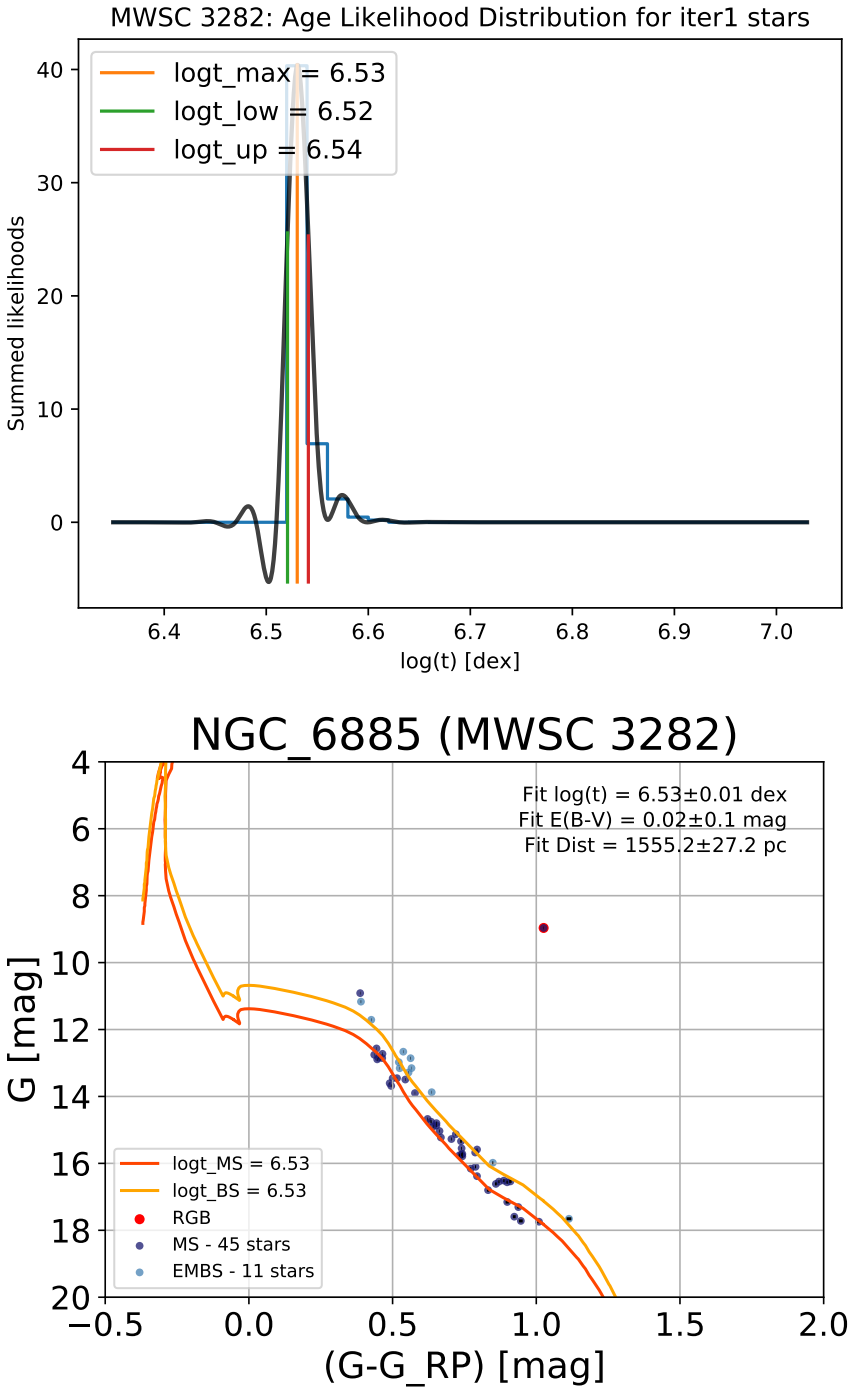

Figure 4.13: *Top:* Age likelihood distribution for NGC 6885 (MWSC 3282). The blue histogram shows the results from the likelihood fits and the black line is a spline interpolation to the likelihood distribution, from which the maximum age and 1σ error bars are derived. *Bottom:* G vs. $G - G_{RP}$ diagram for NGC 6885 with the fitted results. Main sequence stars are shown as dark blue points with the corresponding main sequence isochrone in red. photometric binaries are shown as light blue points with the corresponding equal-mass binary isochrone in orange. RGB stars are indicated by a red outer edge.

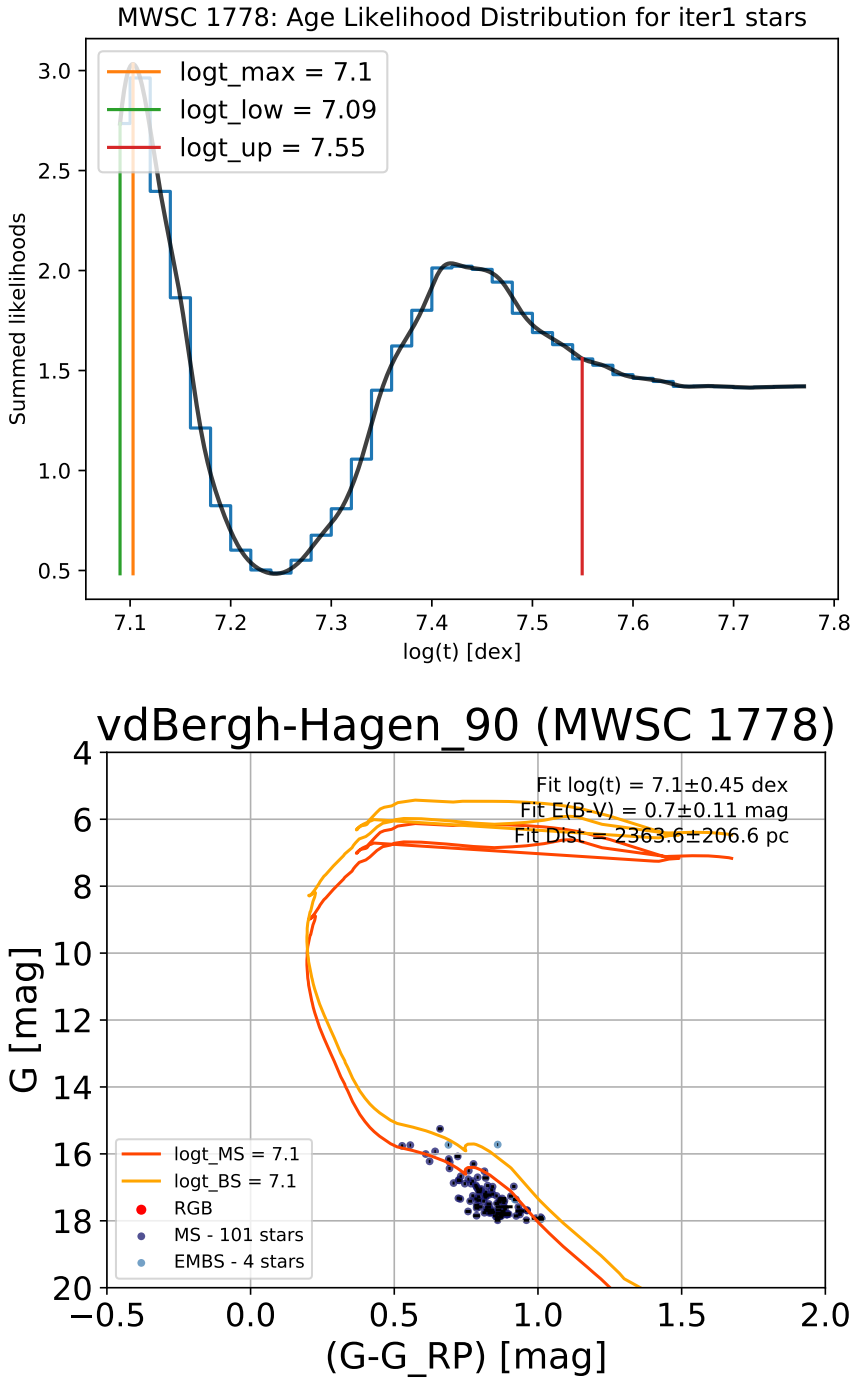

Figure 4.14: *Top:* Age likelihood distribution for vdBergh-Hagen 90 (MWSC 1778). The blue histogram shows the results from the likelihood fits and the black line is a spline interpolation to the likelihood distribution, from which the maximum age and 1σ error bars are derived. *Bottom:* G vs. $G-G_{RP}$ diagram for vdBergh-Hagen 90 with the fitted results. Main sequence stars are shown as dark blue points with the corresponding main sequence isochrone in red. photometric binaries are shown as light blue points with the corresponding equal-mass binary isochrone in orange. RGB stars are indicated by a red circular outer edge.

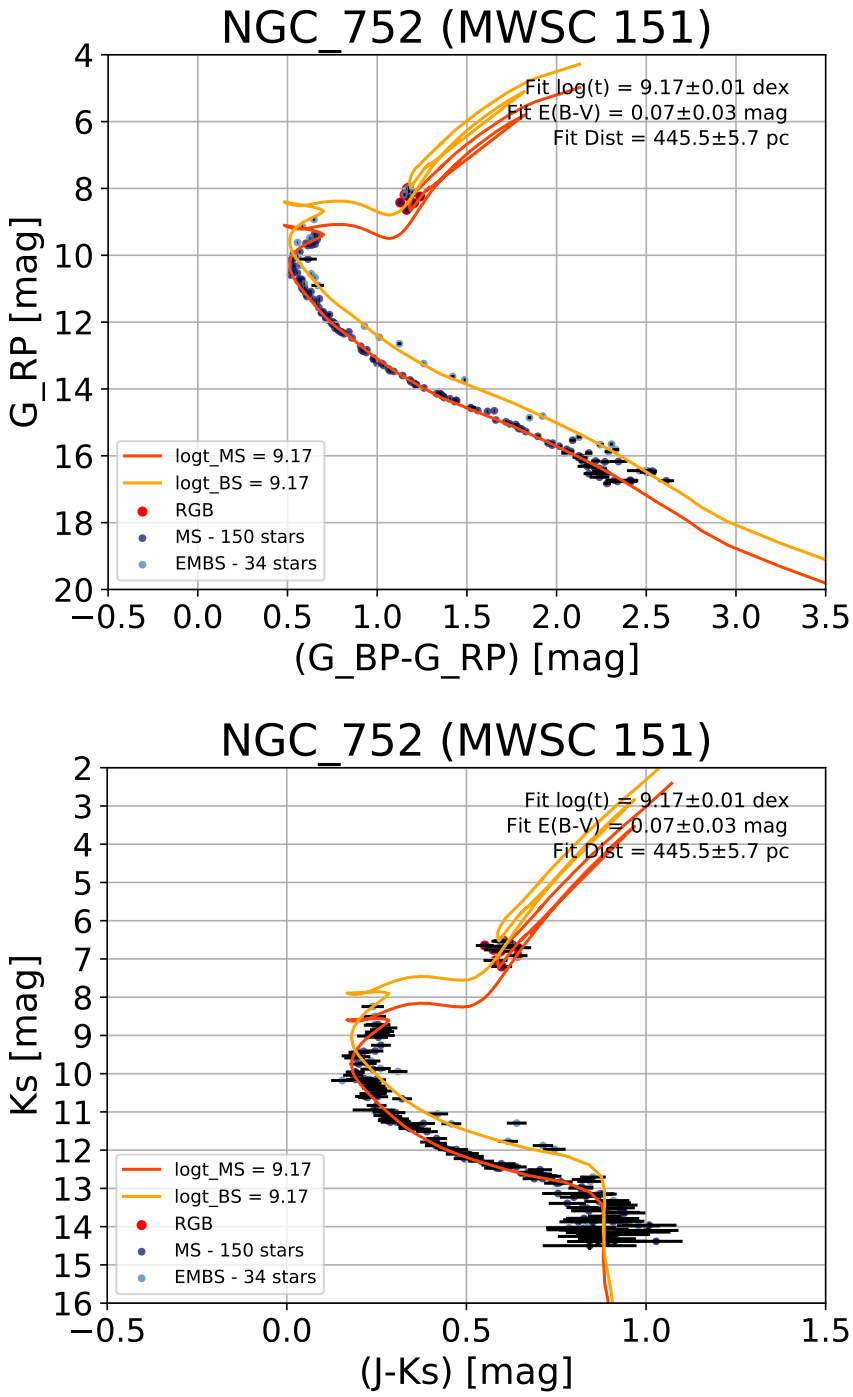

Figure 4.15: Two CMDs of NGC 752 (MWSC 151), G_{RP} vs. $G_{BP} - G_{RP}$ (top) and K_s vs. $J - K_s$ $(bottom)$. Main sequence stars are shown as dark blue points with the corresponding main sequence isochrone as a solid red line. photometric binaries are shown as light blue points with the corresponding equal-mass binary isochrone as a solid orange line. RGB stars are indicated by a red circular outer edge.

amounts to 292 clusters, of which 181 have only the age flag.

An example of a cluster with a properly flagged singularly peaked age is NGC 6885 (MWSC 3282). The singularly peaked age can be seen in the age likelihood distribution of the cluster, which is provided in the top panel of Fig. 4.13. From the CMD of NGC 6885 (MWSC 3282) in the bottom panel of Fig. 4.13, it is evident that this is not the true age of the cluster. The young age isochrone places all the stars on the pre-main sequence, which is highly unlikely. Also, there is one RGB star, which is not considered at all by the determined age isochrone. Further analysis into this cluster is required to determine why an older age is not considered.

A fitted age is also flagged if the one sigma age bounds are at either edge of age range being considered. Typically this indicates derived ages with asymmetric error bars. An example of a cluster flagged for this reason is vdBergh-Hagen 90 (MWSC 1778). From the age likelihood distribution of this cluster in the top panel of Fig. 4.14, the maximum age for this cluster is determined to be at the low end of the age range tested, yielding implausible error bars and a poorly constrained fit. In cases like this, the isochrone fitting should be redone with an extended age range so that the maximum age is not at the edge of the parameter range. Nevertheless, it is appropriate for this cluster to be flagged because the cluster distribution in the CMD, shown in the bottom panel of Fig. 4.14, is not very convincing of a true cluster. Instead it shows more or less a clump of stars spanning two very faint magnitudes. vdBergh-Hagen 90 should be classified as a dubious cluster.

A further investigation into the assumption on age flag made for this thesis, shows that while we recovered the very well-populated clusters with singularly peaked ages, several less-populated nearby clusters with distances less than 1000 pc yield fits with very small age errors and thus, are flagged for their age fit, but actually turn out to be well-determined clusters. An example of such a cluster is NGC 752 (MWSC 151), which shows, in Fig. 4.15, a beautifully thin cluster main sequence and a 0.7 mag brighter binary sequence with an equally as nice, well-determined isochrone fit in various photometric bands. This certainly showcases that a new criterium must be developed for flagging a cluster based on its age.

The last quality flag is based on the derived cluster reddening. For 15 clusters, the reddening is determined to be less than zero, which is implausible. In many of these cases, the negative reddening is linked with a very young age, causing the cluster main sequence to lie only on the pre-main sequence part of the isochrone.

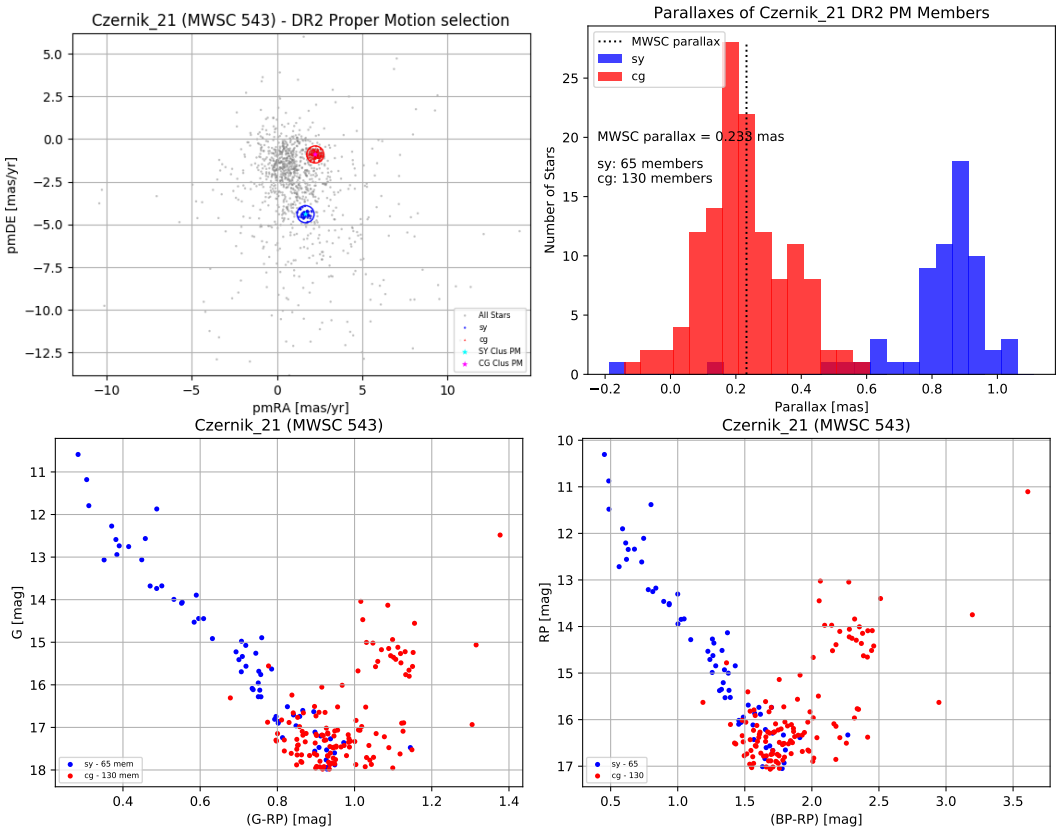

Figure 4.16: Example of two clusters in the same field of view: Czernik 21. Blue represents a solution of this work, red represents the other solution of this work, and that selected by Cantat-Gaudin et al. (2018a). Here it is clear the cluster intended as Czernik 21 by the MWSC is that given by the red selection, as the MWSC cluster parallax matches well with the DR2 parallax peak.

4.4.3 Multiple solution clusters

For 516 clusters, our analysis returns multiple possible solutions, each with a different cluster proper motion. The majority of these clusters have two solutions, which is likely due to an overlap of two clusters with the same sky position. An example of such a case is Czernik 21 (MWSC 543), where the initial proper motion selection yields two signatures, as shown in the top left panel of Fig. 4.16. In the parallax distributions, see top right panel of Fig. 4.16, one of these proper motion signatures shows a very strong concentration at the MWSC parallax, indicating that this is the true cluster labeled as Czernik 21. The other proper motion signature yields a peak at a slightly higher parallax, as well as a nice cluster sequence in the CMDs; this is probably another cluster listed in the MWSC or a potentially newly discovered

cluster. For now, our analysis retains both as solutions for Czernik 21. Further analysis is required to properly and accurately identify the correct cluster as meant by the MWSC and separate potentially newly discovered clusters.

In the other cases of multiple solutions, the determination of the true cluster is not as straightforward. The clusters with three or more successful solutions are likely to be dubious clusters, where the cluster signatures identified in the proper motion KDE come from random overdensities in the background distribution. They could also contain new cluster discoveries, but a more thorough investigation is required to confirm this. Nevertheless, for the remainder of this thesis, the solution yielding the most cluster members or the most peaked distribution in the proper motion KDE is selected as the temporarily "true" solution for the clusters with multiple solutions.

4.4.4 Parameter comparison with literature

To obtain an idea of the quality of our fitted parameters, we compare our derived cluster parameters to those listed in the MWSC. Cluster distance, $E(B-V)$, and log t comparison plots between the MWSC and our full sample of 1873 clusters, which consist of the three solution groups (single-solution clusters with no flags, single-solution clusters with flags, and multiple solution clusters), can be found in Appendix B.1.

Focusing on our best results, the 913 single-solution clusters with no flags, we notice very good agreement in cluster distance with the MWSC up about 1 kpc, as illustrated in Fig. 4.17. However, for clusters with distances greater than 1 kpc, our distance derivation tends to be higher than that in the MWSC. In total, for about half of these clusters, the distances match within 3σ.

The opposite trend seems to be true when comparing our isochrone-fitted distances to the mean DR2 cluster parallax.While we find excellent agreement between the two parallaxes, with a typical difference of less than 0.1 mas, as seen in Fig. 4.18, there is a noticeable offset, in which the mean DR2 cluster parallaxes are slightly larger than the parallaxes computed from our isochrone fitted distances. This observation is actually contradictory to the finding of Schönrich et al. (2019), who find the opposite offset when deriving Bayesian distances for the radial velocity stellar sample in DR2. They find that DR2 parallaxes need to be increased, by an average of 0.054 mas (Schönrich et al. 2019).

In general, large discrepancies exist between the parameter determinations from

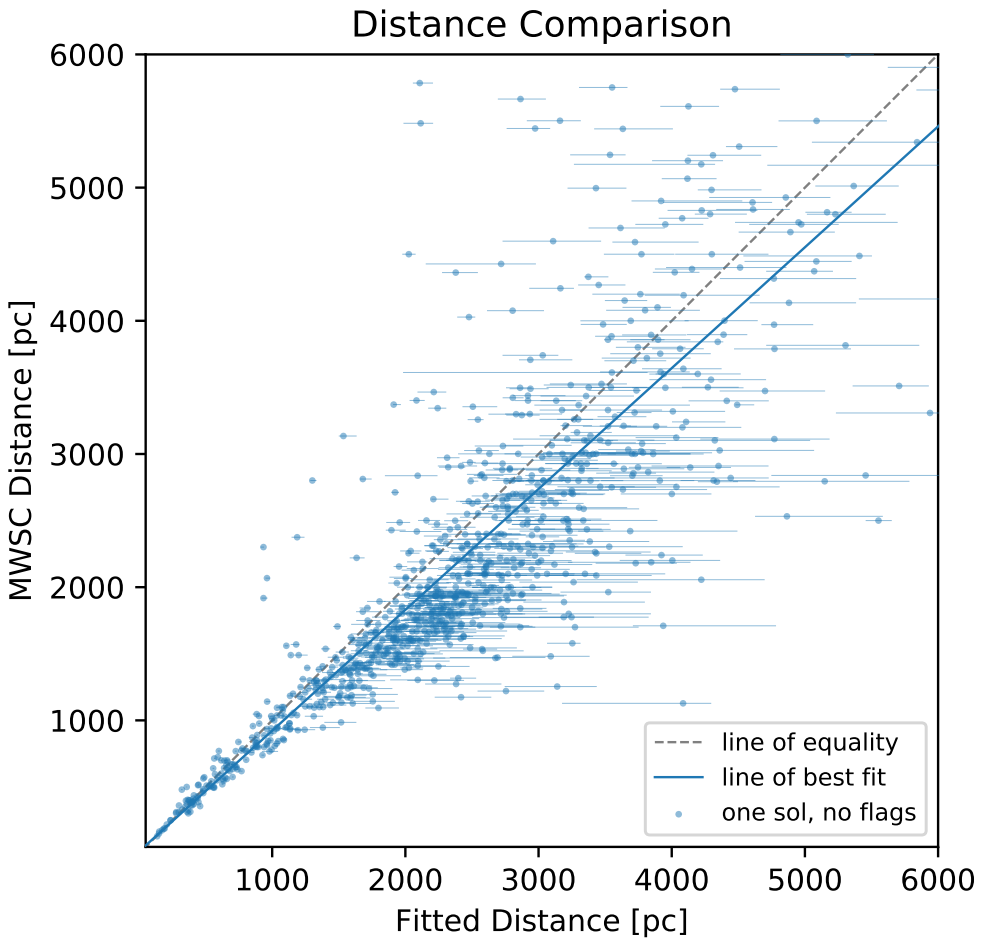

Figure 4.17: Zoom-in distance comparison of cluster distances derived from this work to those determined in MWSC. The dashed grey line indicates the line of equality and the blue line shows the best fit to the single solution clusters with no flags.

our study and the MWSC, especially for the cluster ages, where for many clusters, opposite ages are derived. For many young clusters in the MWSC, we derive older ages, and vice versa. We recognize that discrepancies are expected due to the difference in data quality and quantity between the two studies. Therefore, we look to more recent studies for a proper assessment of the derived results.

So far, the study by Bossini et al. (2019) consists of the largest published sample of parameter determinations for open clusters with DR2. They report new cluster parameters (distance modulus, extinction, and ages) for 242 MWSC clusters, of which 235 have parameters determined by our study. The parameter comparison plots for all 235 clusters in common can be found in Appendix B.2, but in this

discussion we focus again on our single-solution clusters with no flags, of which there are 171 in common.

Our distance moduli for these 171 clusters are in excellent agreement, as seen in Fig. 4.19. The two most discrepant clusters are Saurer 2 (MWSC 1489), a very poorly populated cluster at about 3 kpc, and Ruprecht 63 (MWSC 1504), a well populated cluster at roughly 4 kpc. Such large deviations in distance moduli suggests differences in the cluster members considered.

For cluster extinctions and ages, we observe larger deviations between the two samples, as seen in Figs. 4.20 and 4.21, respectively. Our study tends to report larger extinctions as compared to Bossini et al. (2019). There also seems to be a systematic offset, which could be due to the assumption of a fixed extinction relation in the Gaia bands, as discussed in Sec. 4.4.1. The comparison of ages yields a much better agreement with Bossini et al. (2019) than with the MWSC. A little more than half of the clusters agree within the 3σ age errors. We remark that many of the cluster parameters have zero errors as derived by Bossini et al. (2019).

4.5 Discussion

One of the main aims of this work is to provide updated statistics on the previously known open cluster population, as originally compiled by Kharchenko et al. (2013). Applying our automated cluster characterization tool on all MWSC open clusters, we confirm the existence of 1873 open clusters, roughly 67% of the initial sample. For 23 clusters, their parameters could not be adequately determined, though are expected to give results after some adjustments to the pipeline. We determine that the remaining 912 clusters (32%) are non-existent.

4.5.1 Removed open clusters

The MWSC identifier of the non-existent clusters is provided in Table 4.5. We recognize that our number of removed open clusters is lower than that determined by Cantat-Gaudin et al. (2018a), who began their analysis with a larger cluster sample, adding clusters from five other works (Dias et al. 2002; Froebrich et al. 2007; Schmeja et al. 2014; Scholz et al. 2015; Röser et al. 2016) to the MWSC list. They derive cluster astrometry for 1169 previously known open clusters, of which

1085 are in the MWSC. Furthermore, this implies that they do not find strong evidence for 61% of the MWSC clusters.

From visual checks of the final cluster CMDs for some of our clusters with solutions, we do find some questionable clusters. Thus, our list of non-existent clusters is likely the most obvious non-clusters. With more stringent definitions on what constitutes an open cluster, we may end up discarding more MWSC clusters.

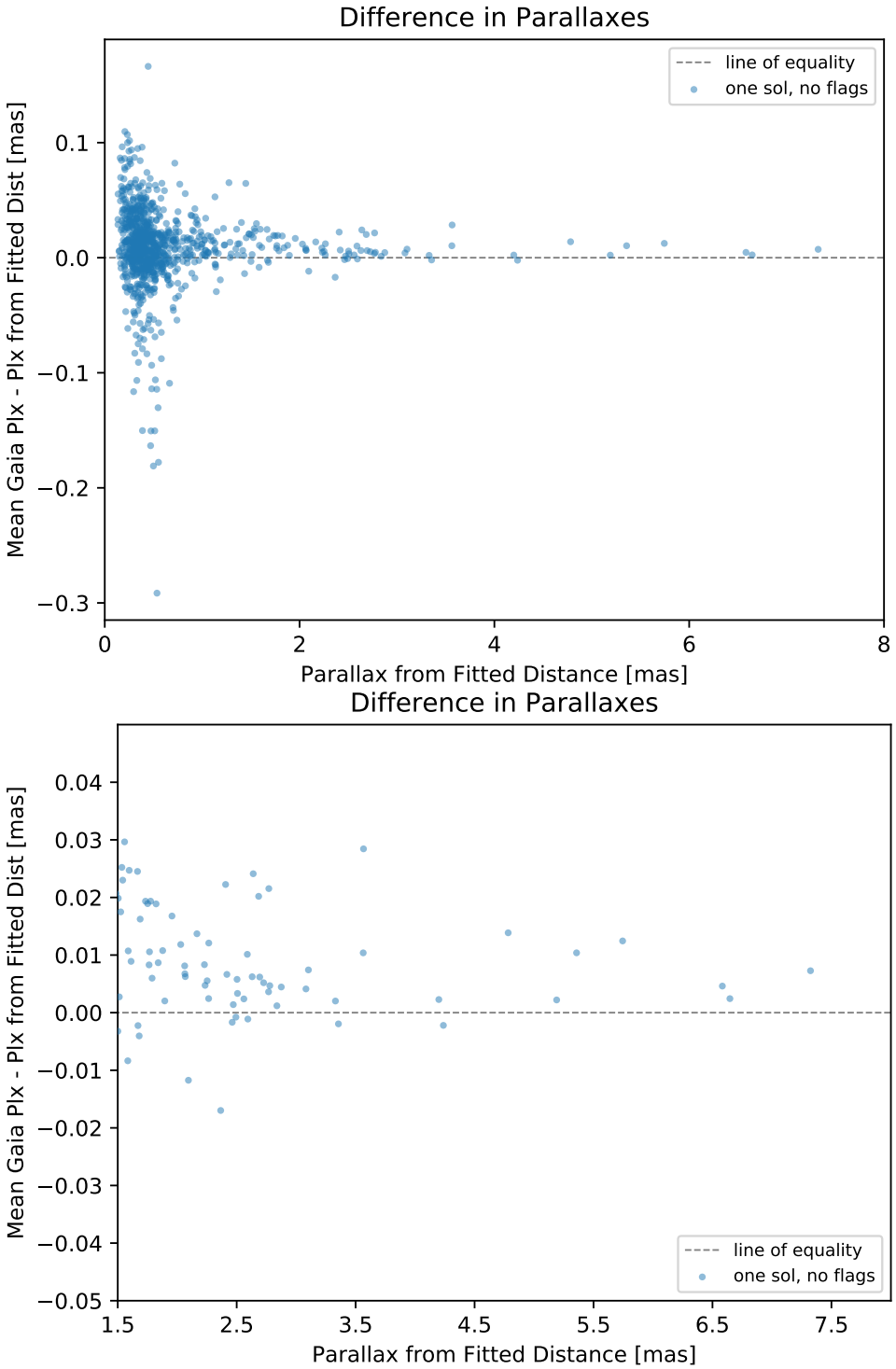

Figure 4.18: *Top*: Parallax comparison of mean DR2 cluster parallaxes and fitted cluster distances for single-solution clusters with no flags. *Bottom*: A zoom-in on the plot for closer distances. In both figures, the dashed grey line indicates the line of equality. The single solution clusters with no flags are shown in blue.

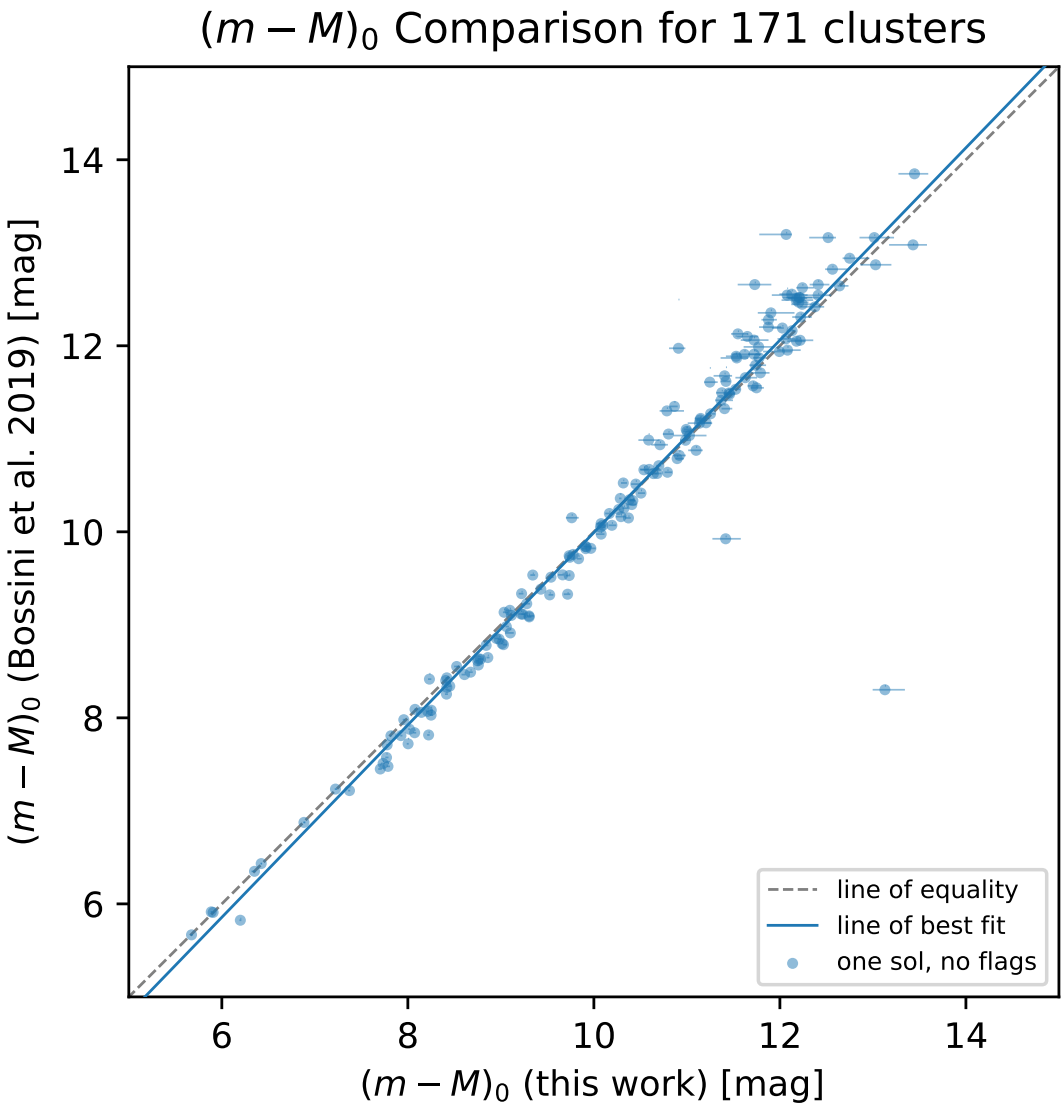

Figure 4.19: Distance modulus comparison of 171 clusters in common with Bossini et al. (2019). The dashed grey line is the line of equality and the solid blue line indicates the line of best fit.

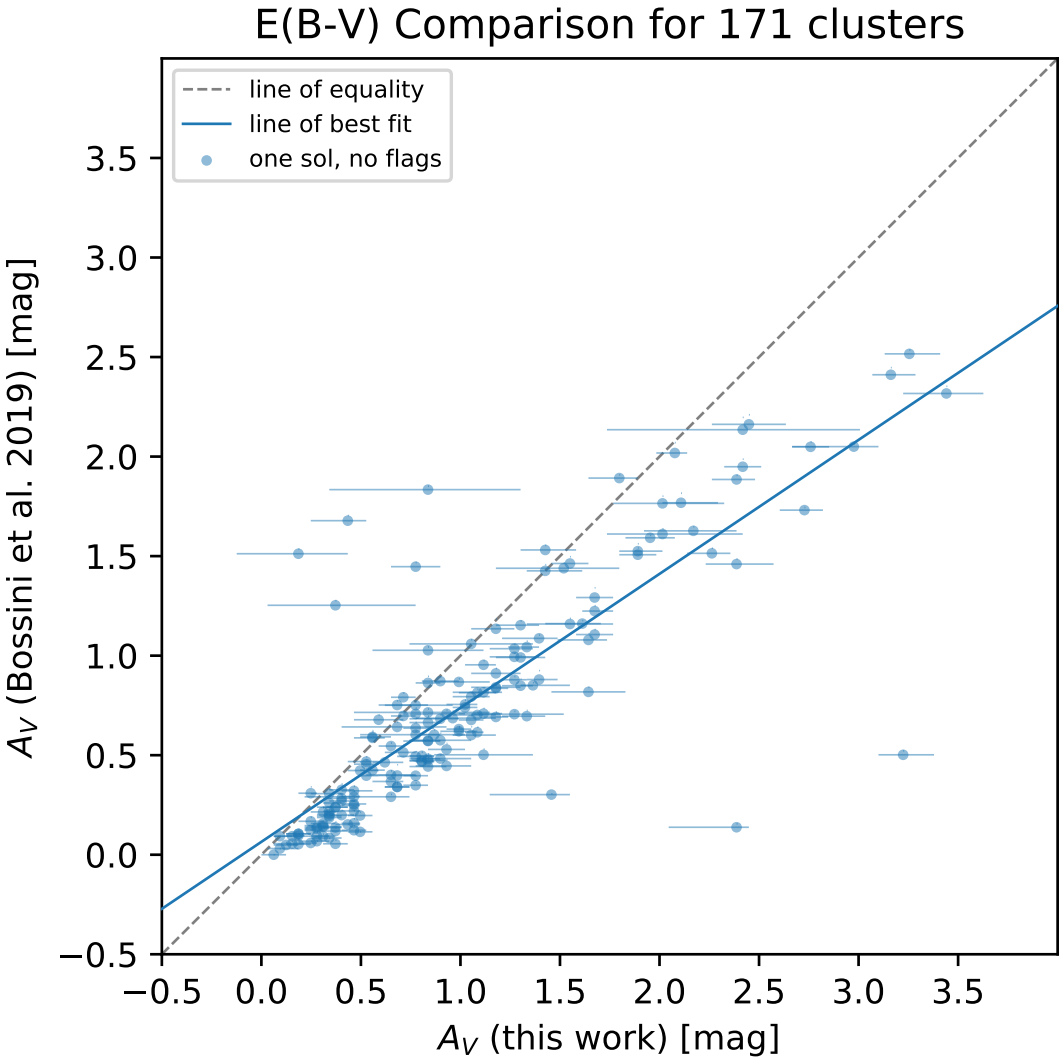

Figure 4.20: Extinction comparison of 171 clusters in common with Bossini et al. (2019). The dashed grey line is the line of equality and the solid blue line indicates the line of best fit. **need to change plot title**

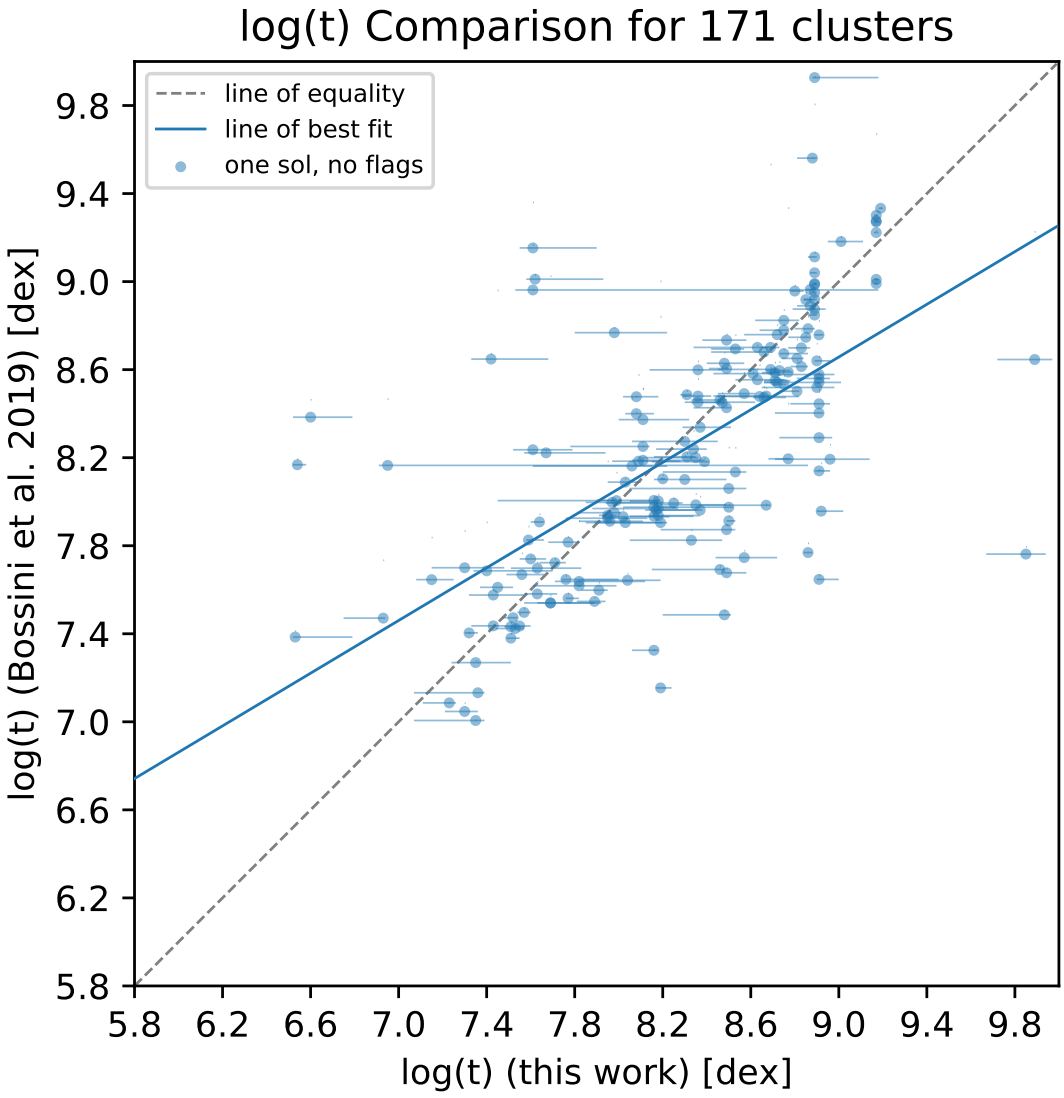

Figure 4.21: Age comparison of 171 clusters in common with Bossini et al. (2019). The dashed grey line is the line of equality and the solid blue line is the line of best fit.

Table 4.5: List of 912 removed open clusters

MWSC	MWSC	MWSC	MWSC	MWSC	MWSC	MWSC	MWSC
2	491	1147	1594	1940	2323	2971	3401
4	493	1148	1599	1943	2324	2979	3402
9	497	1150	1600	1946	2325	2983	3403
14	501	1151	1601	1952	2328	2991	3407
16	503	1153	1604	1955	2329	2992	3408
18	506	1163	1608	1956	2333	2997	3410
19	513	1169	1609	1958	2336	2999	3413
24	521	1171	1612	1959	2337	3000	3416
27	524	1174	1615	1965	2339	3008	3418
32	525	1183	1619	1969	2349	3015	3422
33	535	1191	1620	1972	2354	3016	3423
34	536	1194	1622	1975	2358	3022	3428
36	544	1196	1623	1977	2378	3028	3430
39	545	1200	1624	1979	2379	3031	3439
45	548	1201	1625	1981	2388	3035	3442
48	551	1202	1627	1983	2390	3038	3443
77	560	1205	1628	1986	2393	3043	3448
78	567	1207	1629	1988	2395	3044	3450
80	568	1208	1631	1990	2397	3060	3455
83	569	1227	1634	1994	2398	3063	3464
87	603	1228	1635	1996	2401	3071	3469
89	607	1232	1637	1998	2402	3072	3470
93	614	1235	1638	2000	2413	3076	3472
98	629	1239	1640	2005	2421	3080	3479
104	631	1240	1641	2009	2424	3085	3482
106	635	1242	1642	2011	2429	3087	3484
108	639	1251	1643	2012	2437	3090	3487
109	643	1254	1645	2014	2442	3093	3488
112	650	1262	1646	2015	2453	3114	3491
121	655	1264	1649	2025	2454	3116	3492
130	656	1267	1653	2034	2466	3119	3494
131	657	1276	1655	2036	2467	3123	3499

List of 912 removed open clusters continued

MWSC	MWSC	MWSC	MWSC	MWSC	MWSC	MWSC	MWSC
134	664	1286	1656	2037	2489	3125	3503
135	665	1298	1659	2041	2501	3131	3504
143	667	1300	1661	2043	2504	3135	3509
147	671	1301	1663	2044	2505	3136	3511
148	678	1302	1664	2046	2510	3141	3513
152	680	1307	1673	2048	2521	3142	3516
159	681	1316	1680	2049	2522	3145	3524
160	684	1317	1683	2053	2524	3164	3527
164	685	1318	1684	2058	2527	3174	3531
166	690	1329	1685	2064	2529	3178	3532
172	694	1330	1687	2065	2540	3179	3535
174	700	1332	1689	2066	2548	3183	3537
178	705	1334	1696	2067	2558	3185	3540
184	709	1337	1703	2068	2575	3190	3542
192	716	1339	1713	2073	2591	3202	3546
193	721	1341	1715	2074	2597	3204	3549
207	724	1342	1719	2079	2598	3205	3552
208	732	1349	1720	2083	2606	3212	3555
215	737	1352	1724	2084	2609	3215	3556
220	739	1361	1726	2089	2610	3217	3567
235	742	1365	1727	2091	2611	3220	3568
237	748	1366	1728	2092	2617	3224	3570
243	759	1367	1731	2099	2623	3231	3573
248	763	1374	1735	2100	2646	3233	3576
249	769	1376	1736	2107	2650	3234	3579
252	770	1389	1740	2109	2652	3236	3582
258	784	1390	1743	2111	2659	3247	3587
260	807	1395	1744	2112	2679	3249	3594
266	814	1396	1746	2114	2682	3264	3596
270	821	1397	1753	2115	2695	3265	3597
272	827	1399	1758	2117	2702	3267	3598
278	840	1401	1760	2126	2706	3269	3601
279	841	1402	1761	2135	2709	3278	3607

List of 912 removed open clusters continued

MWSC	MWSC	MWSC	MWSC	MWSC	MWSC	MWSC	MWSC
285	852	1403	1763	2137	2724	3279	3614
288	859	1405	1766	2140	2727	3283	3615
291	860	1407	1768	2141	2733	3285	3623
293	875	1409	1770	2145	2734	3288	3634
304	876	1414	1772	2146	2735	3290	3638
314	884	1418	1775	2150	2738	3292	3639
315	887	1423	1776	2157	2746	3299	3640
317	900	1427	1777	2169	2748	3306	3641
319	905	1431	1779	2170	2750	3307	3643
320	909	1440	1780	2171	2756	3310	3656
329	910	1441	1786	2178	2759	3313	3657
335	911	1445	1787	2179	2767	3314	3660
338	915	1452	1791	2180	2769	3315	3665
339	926	1453	1793	2181	2776	3321	3667
341	927	1457	1796	2184	2781	3323	3670
345	944	1458	1798	2191	2784	3328	3671
347	949	1462	1799	2193	2787	3329	3676
351	950	1463	1802	2195	2799	3330	3682
355	952	1470	1807	2197	2805	3335	3687
358	956	1471	1808	2199	2806	3336	3692
360	962	1474	1812	2200	2808	3337	3696
371	972	1477	1814	2208	2809	3339	3699
372	973	1479	1815	2209	2811	3340	3702
374	995	1486	1817	2212	2818	3341	3707
375	998	1495	1823	2213	2822	3346	3710
380	1017	1496	1825	2228	2830	3347	3711
383	1027	1497	1827	2230	2832	3348	3713
385	1028	1503	1829	2232	2839	3349	3718
393	1030	1505	1837	2233	2840	3352	3720
395	1036	1506	1856	2236	2842	3353	3723
396	1042	1508	1858	2240	2843	3356	3724
399	1051	1512	1862	2260	2846	3361	3729
400	1062	1516	1867	2262	2848	3362	3731

List of 912 removed open clusters continued

MWSC	MWSC	MWSC	MWSC	MWSC	MWSC	MWSC	MWSC
407	1066	1521	1872	2264	2854	3366	3734
415	1073	1522	1874	2266	2855	3367	3736
426	1075	1531	1880	2268	2856	3368	3738
427	1093	1535	1881	2274	2863	3372	3740
429	1095	1536	1891	2279	2864	3376	3742
432	1096	1540	1894	2280	2869	3378	3745
437	1097	1541	1896	2283	2870	3379	3746
446	1103	1545	1897	2291	2881	3380	3747
451	1113	1552	1902	2294	2889	3382	3748
452	1115	1559	1904	2297	2891	3385	3751
458	1123	1564	1905	2299	2894	3391	3755
465	1127	1565	1907	2311	2895	3392	3767
467	1134	1568	1915	2312	2930	3394	3769
476	1139	1571	1922	2318	2934	3395	3770
488	1141	1575	1925	2319	2952	3396	3772
490	1145	1589	1939	2320	2966	3400	3775

4.5.2 Current census of open clusters

We have created a new and updated sample of the Milky Way open cluster population. Using the distances derived by our fitting procedure, we can create a plot of the positions of the clusters on the Galactic plane, as shown in Fig. 4.22. Furthermore, using the ages derived in our work, we observe that younger clusters are found near the plane of the galaxy, while older clusters are distributed across various Galactic altitudes, see Fig. 4.23, which is consistent with previous studies (e.g. Lynga (1982); Bonatto et al. (2006); Bossini et al. (2019)).

Of interest is also the completeness of this new cluster sample. In Fig. 4.24, the surface densities of our sample of 1873 clusters is shown as a function of their distance d_{XY} from the Sun in the Galactic plane. We notice a sharp peak in the distribution around $d_{XY}\sim$400 pc, which is related to the young clusters of the Orion star formation complex (Kharchenko et al. 2013). However, unlike the original sample of MWSC clusters, which yielded a census completeness to 1.8 kpc, we find that there is no clear evidence to suggest any completeness of this new cluster sample. The total distribution of surface density varies widely, illustrating that a dedicated search for new clusters in the Gaia data is needed in order to derive conclusions about the Milky Way open cluster population. In fact, new clusters have already been discovered with Gaia, including several closer than 500 pc (Castro-Ginard et al. 2018), proving the census of open clusters at nearby distances is also incomplete.

4.5.3 Conclusions

We further developed the cluster characterization tool we first created for analysis of nearby clusters with DR1 (Yen et al. 2018). The newest version of our pipeline contains several improvements over the first implementation. The most notable change is that it now determines cluster signatures in proper motion space, no longer relying on the past cluster proper motion determinations. It also treats photometric binaries in a more robust manner. Furthermore, it is more versatile, as it can be applied to clusters of all distances from the Sun.

With our new cluster characterization tool, we have reanalyzed all open clusters in the MWSC catalog and present a new homogeneous catalog of the true open clusters. We confirm the existence of 1873 (67%) MWSC open clusters, deriving updated cluster parameters and cleaner membership lists. For cluster member stars, we also

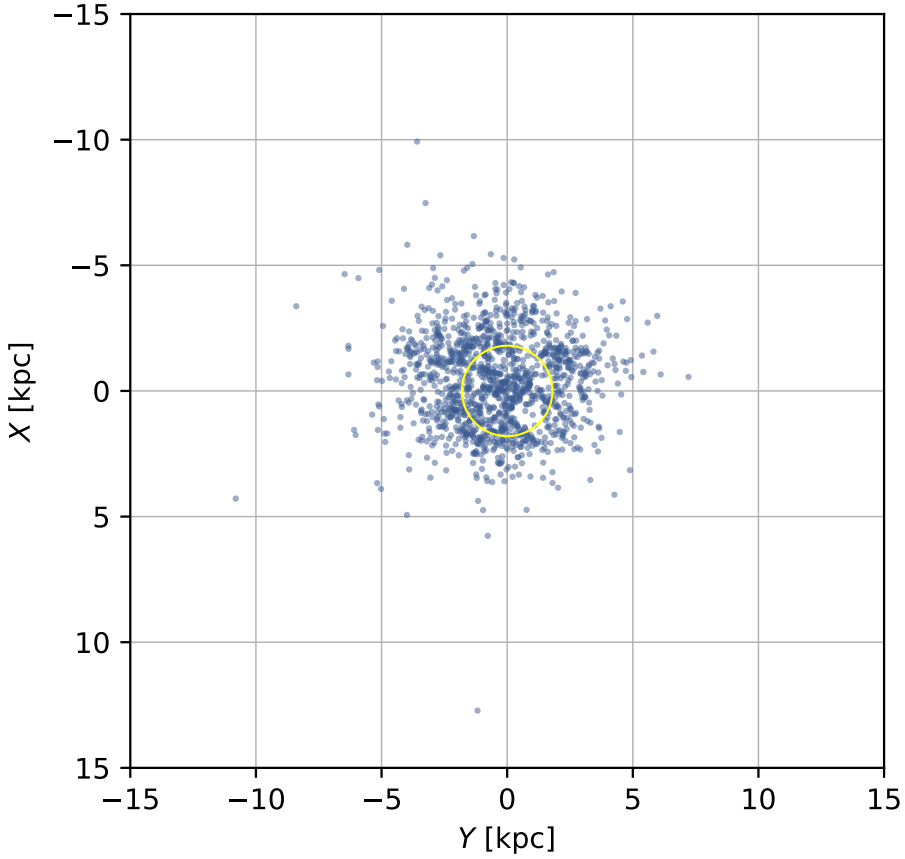

Figure 4.22: Distribution of 1873 open clusters in the Galactic plane. The yellow circle marks the 1.8 kpc completeness limit from the Sun, as derived in the MWSC.

provide mass estimates and. Furthermore, we also classified 912 (32%) MWSC clusters as non-existent clusters, changing our previous picture of the Galactic open cluster population. Our main finding is that the open cluster census is not complete to 1.8 kpc, as previously thought. We conclude that this new sample of open clusters represents a very incomplete sample of the open cluster population in the Galactic disk, as no trend is clearly observed in the distributions of cluster surface density in the Galactic plane.

Data from Gaia DR2 is truly revolutionizing our understanding of the Galaxy's open clusters. With the unprecedented volume of highly precise data, we are obtaining a clearer view on these clusters. The next big step with this work is to perform a systematic search for new clusters in the DR2 data set, looking at near and far distances. A new and more complete census of the open cluster population will shed new light into the properties of the Galactic disk.

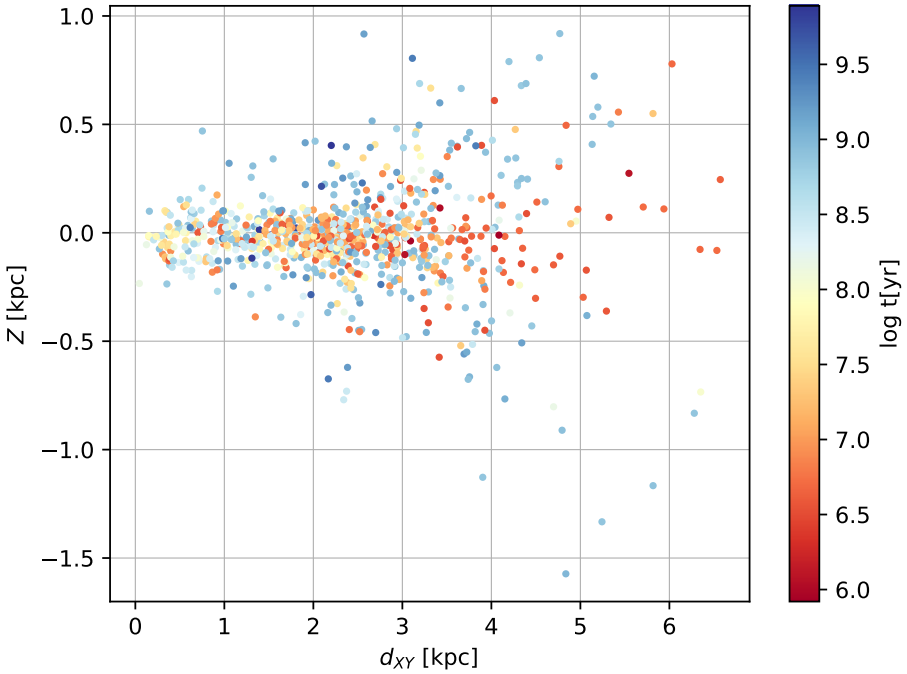

Figure 4.23: Galactic scale height (distance to the Galactic plane) vs. distance from the Sun, color-coded by the derived cluster age.

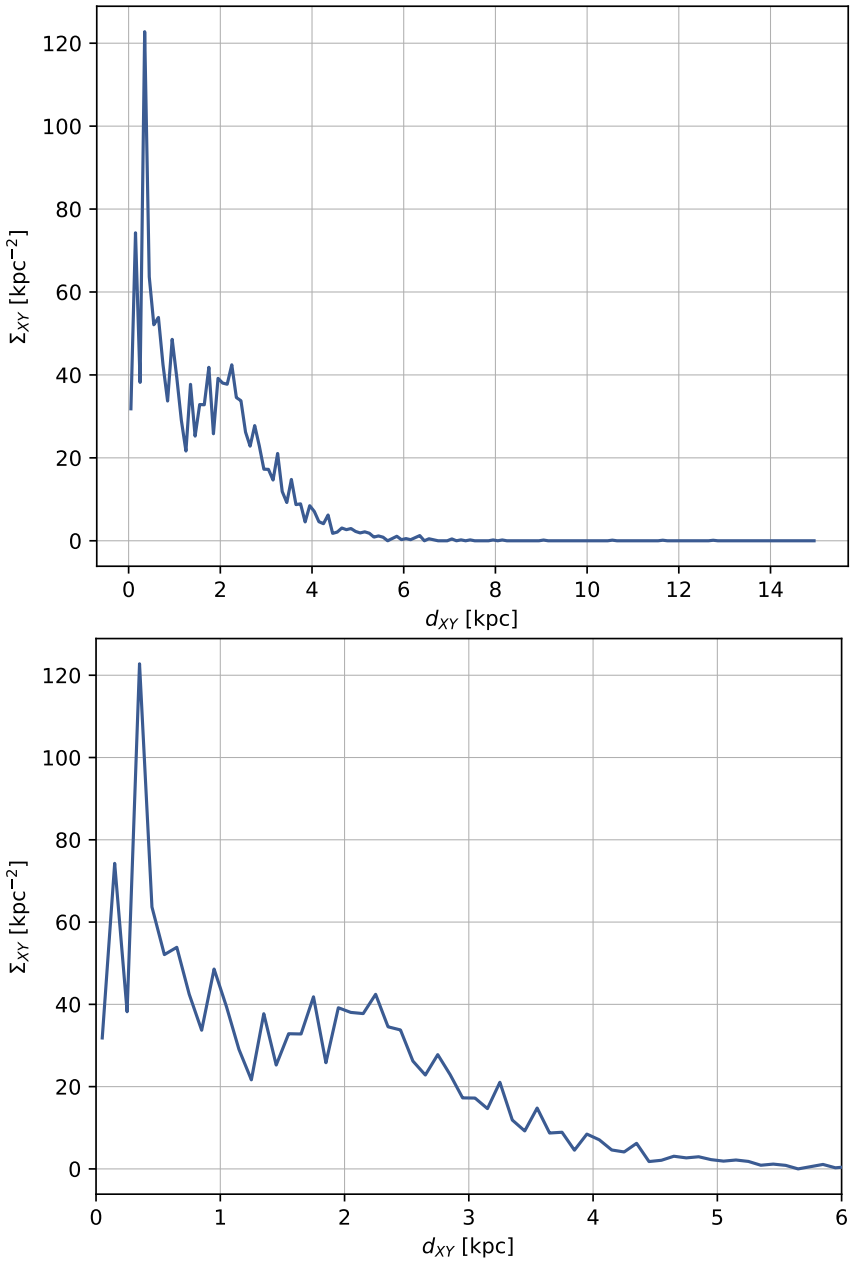

Figure 4.24: Distribution of the surface density Σ_{XY} of open clusters vs. distance d_{XY} from the Sun projected on the Galactic plane. The density distribution of all 1873 clusters is shown in blue. Top: Surface density distribution for all 1873 clusters. $Bottom$: A zoom-in on the surface density distribution for clusters within 6 kpc.

5

Summary and Outlook

5.1 Summary

In the past, studies of open clusters and their parameters were largely plagued by poor data quality, confusion between true cluster members and background stars, and subjective determinations of cluster parameters using a fit-by-eye approach. However, we are now in the age of Gaia, where we have high quality stellar data for more than one billion Milky Way stars at our fingertips. The precision of Gaia data allows us to easily distinguish between true cluster members and non-member field stars, leading to cleaner cluster membership lists and improved cluster parameters, and resulting in a more complete and better characterization of the Milky Way open cluster population. To take full advantage of this data, new tools must be developed for cluster analysis. Thus, the aim of this thesis is two-fold: first, to develop a cluster characterization pipeline to derive homogeneous cluster parameters in an unsupervised way, and second, to use this pipeline to compile an updated catalog of MWSC open clusters and their parameters with data from the Gaia mission.

In Chapter 3, I described an initial version of the cluster characterization pipeline I developed. I used the pipeline to analyze a subset of 24 nearby ($d_{MWSC} < 333$ pc) with data primarily Gaia DR1/TGAS and HSOY. From this analysis, I obtained

cleaner cluster main sequences and improved cluster parameters. I also noticed that some of the nearby clusters might not be true clusters after all, highlighting an important outcome of the quality of Gaia data. The pipeline worked well for the small sample of clusters, showing good agreement to literature values, but was developed specifically for the quality and limitations of the data sets used and is applicable only to nearby clusters.

In Chapter 4, I described the steps of the most recent version of the cluster characterization pipeline, which is largely an extension of that in Ch. 3, but includes significant changes to handle the large amount of DR2 stellar data, less reliance on the previously determined MWSC parameters, more robust treatment of photometric binaries, and a consistent procedure to derive cluster parameters and membership regardless of cluster distance. I used the pipeline to reanalyze the full MWSC sample of 2808 open clusters, using astrometric data from Gaia DR2 and Hipparcos and photometric data from ASCC-2.5 (B and V bands), 2MASS (JHK_s bands), and Gaia DR2 (G, G_{BP}, G_{RP} bands). Parameter results were returned for 1873 clusters, with 912 clusters determined to be non-existent, and 23 clusters requiring further analysis.

This is the first study, to date, which uses a single pipeline to homogeneously reanalyze cluster parameters and membership for a large sample of open clusters. The study by Cantat-Gaudin et al. (2018a) also analyzed the full list of MWSC open clusters (and more), but confirms their existence using only DR2 astrometry and reporting only mean cluster proper motions and parallaxes for 1229 clusters. These new results give us a much better and cleaner view of the previously known Milky Way open clusters. Several MWSC cluster areas also appear to boast more than one open cluster in Gaia proper motion space, illustrating the great potential of the data set. Removing 32% of the MWSC clusters has also shown that the previously determined cluster completeness limit at 1.8 kpc from the Sun is no longer valid. I find a large number of clusters within 1 kpc from the Sun and argue that a dedicated search for new clusters with Gaia data is necessary in order to draw conclusions about the Milky Way open cluster population.

5.2 Outlook

The work presented in this thesis is a great step towards obtaining a better understanding of open cluster and Galactic disk formation and evolution. I homogeneously determined cluster parameters for a large collection of open clusters, with largely

uncontaminated cluster membership, and also provided a revision to the census of previously known open clusters. While the cluster parameters derived are generally reliable and consistent with other works (e.g. Bossini et al. (2019)), there are a few shortcomings in the pipeline methodology.

First and foremost, a big caveat of our method is the use of solar metallicity isochrones, as studies (e.g. Netopil et al. (2015)) have shown that open clusters exhibit a range of metallicities. Inclusion of a metallicity parameter would require much more computation time and while parameter estimates would improve, this does not justify the exponential increase in computation time. While it would be beneficial to use complimentary spectroscopic data, e.g. from the Galactic Archaeology with HERMES (GALAH) survey (Buder et al. 2018), to derive mean cluster metallicities from member stars with measured metallicities and use the respective isochrone in our parameter determination for a given cluster, it would cause a slight inhomogeneity in our analysis as this information is only available for a small sample of stars (about 350,000 in GALAH DR2) and thus, a limited fraction of clusters. Second, variable extinction coefficients for the Gaia bands should be used to obtain better agreement between the observed Gaia photometry and the theoretical models, as explained in Sec. 4.4.1 and references therein. Third, slight modifications need to be made to the determination of mean cluster parameters and their error bars. This would eliminate some of the quality flags mentioned in Sec. 4.4.2. Furthermore, a more stringent definition of what constitutes an open cluster is necessary. We impose a limit of at least 20 stars to be classified as an open cluster, which greatly improves upon the membership numbers for clusters as compared to those in the MWSC. However, in some of our results, in particular clusters with multiple solutions, one or more of the solutions showcase a clump of more than 20 stars in the cluster CMDs, which is clearly not indicative of a true cluster. Some of these changes will be made for the publication of our final results, however, it should not significantly affect our overall conclusions.

As illustrated in the results of thesis, the census of open clusters in the Milky Way is far from complete. The next focus of this work will be to identify new clusters in Gaia DR2 and, of course, future data releases. This thesis has shown that while the highly precise DR2 data does provide cleaner cluster membership lists, the plethora of data does create some complications with identification of clusters simply from overdensities in proper motion space. In this vein, we have explored the possibility of using various analytical methods. One of the most promising methods is to adapt AMICO (Adaptive Matched Identifier of Clustered Objects)

(Bellagamba et al. 2018), which was developed to detect galaxy clusters in noisy fields, to search for new open clusters in the Gaia data set. In addition to using cluster positions and astrometry, it would be able to include colors and magnitudes as inputs, taking advantage of the fact that stars in a cluster occupy a specific region of the HR diagram. This tool would simultaneously use all available data, 5D phase space parameters, as well as photometry, to identify a cluster. I started exploring the possibility of this venture with Maturi, one of the developers of AMICO, and believe this should be further pursued.

Future Gaia data releases will continue to shed new light on our galaxy's open clusters. EDR3, the early version of DR3, will become available in mid-2020, providing improved astrometry and photometry for all stars, which will allow the identification of more distant clusters. The full DR3 catalog will become available a year later in 2021, providing, for the first time, spectral data and radial velocities for well-behaved stars, allowing consideration of metallicities in cluster parameter determinations, which would yield more accurate parameters for a full sample of open clusters.

Finally, a substantially enhanced cluster catalog with homogeneous and accurate cluster parameters will enable deep investigations into the formation and evolution of stars and open clusters, e.g. derivations of luminosity and mass functions of stars in open clusters, as well as insights into the formation history of the Galactic disk. Such a catalog will also improve the models of stellar evolution and simulations of stellar clusters. Just as open clusters are defined by their brightest stars, the future of all open cluster and Milky Way studies, aided by the forthcoming data releases from Gaia, is bright.

A

Supplementary material for Chapter 3

A.1 Cluster color-magnitude diagrams

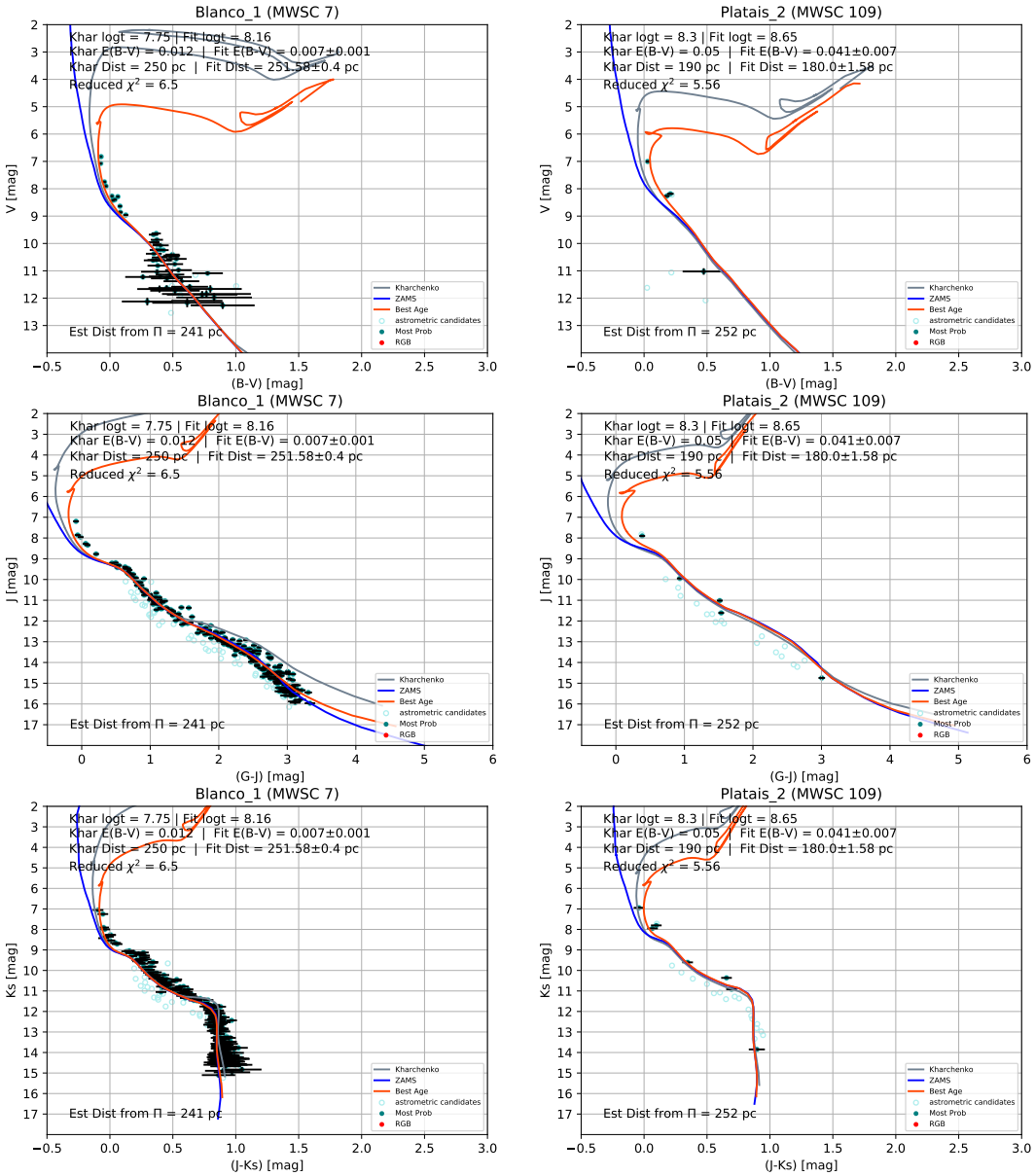

Figure A.1: Color-magnitude diagrams for Blanco 1 (*left*) and Platais 2 (*right*). From *top* to *bottom*: their respective V vs. $(B-V)$, J vs. $(G-J)$, and K_s vs. $(J-K_s)$ CMDs. The cluster members determined from the pipeline are given by teal circles with their corresponding magnitude and color error bars. The cluster astrometric candidates that were later rejected as cluster members are shown by light blue open circles. RGB stars, if any, are indicated by red circles. The red isochrone is the pipeline selected age, plotted with the fitted cluster $E(B-V)$ and d. This isochrone is plotted without the binary offset. The gray isochrone shows the age, $E(B-V)$, and d as determined by Kharchenko et al. (2013). The blue line is the ZAMS plotted with the fitted cluster $E(B-V)$ and d.

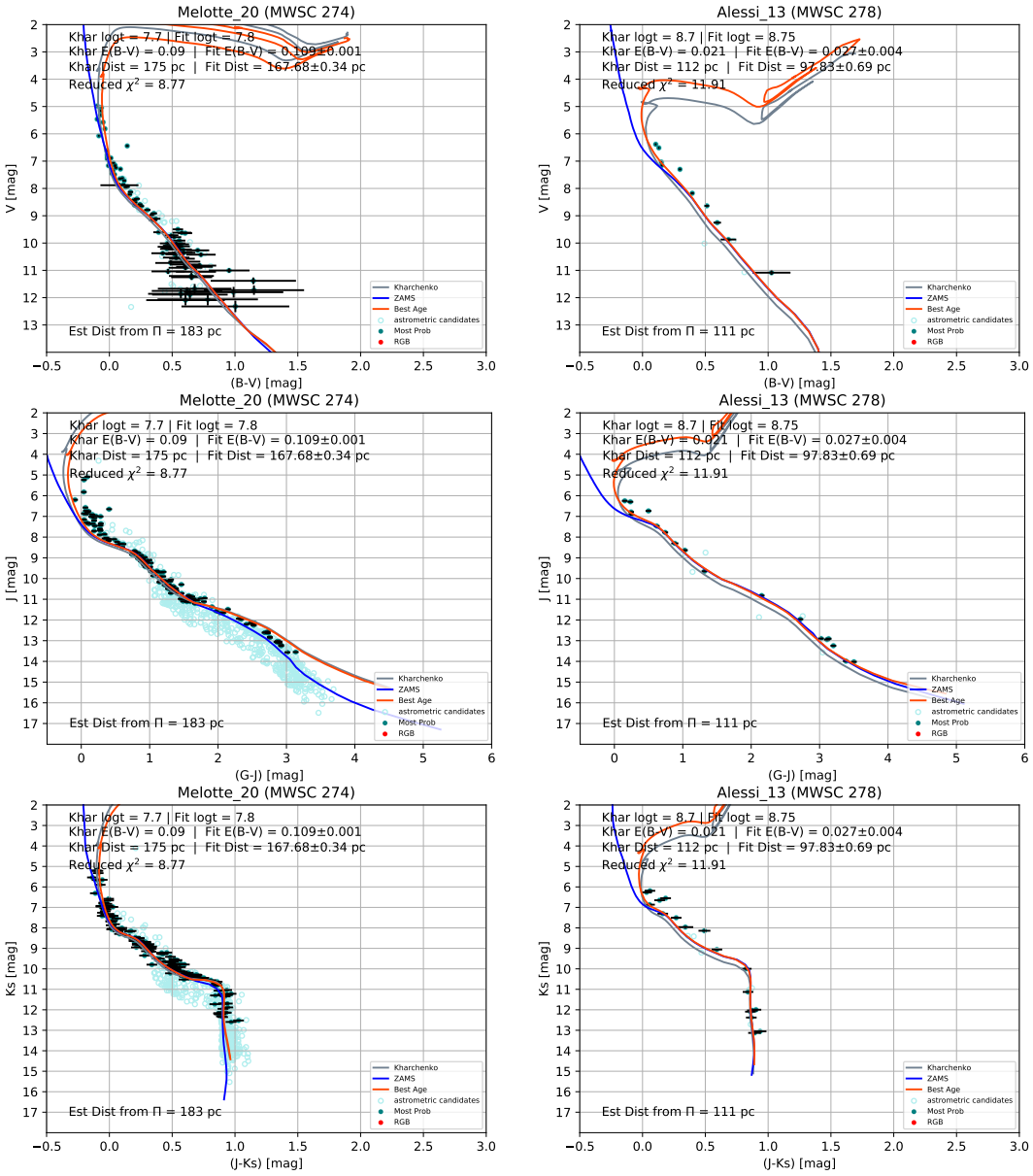

Figure A.2: Color-magnitude diagrams for α Per (Melotte 20) (*left*) and Alessi 13 (*right*). From *top* to *bottom*: their respective V vs. $(B-V)$, J vs. $(G-J)$, and K_s vs. $(J-K_s)$ CMDs. The cluster members determined from the pipeline are given by teal circles with their corresponding magnitude and color error bars. The cluster astrometric candidates that were later rejected as cluster members are shown by light blue open circles. RGB stars, if any, are indicated by red circles. The red isochrone is the pipeline selected age, plotted with the fitted cluster $E(B-V)$ and d. This isochrone is plotted without the binary offset. The gray isochrone shows the age, $E(B-V)$, and d as determined by Kharchenko et al. (2013). The blue line is the ZAMS plotted with the fitted cluster $E(B-V)$ and d.

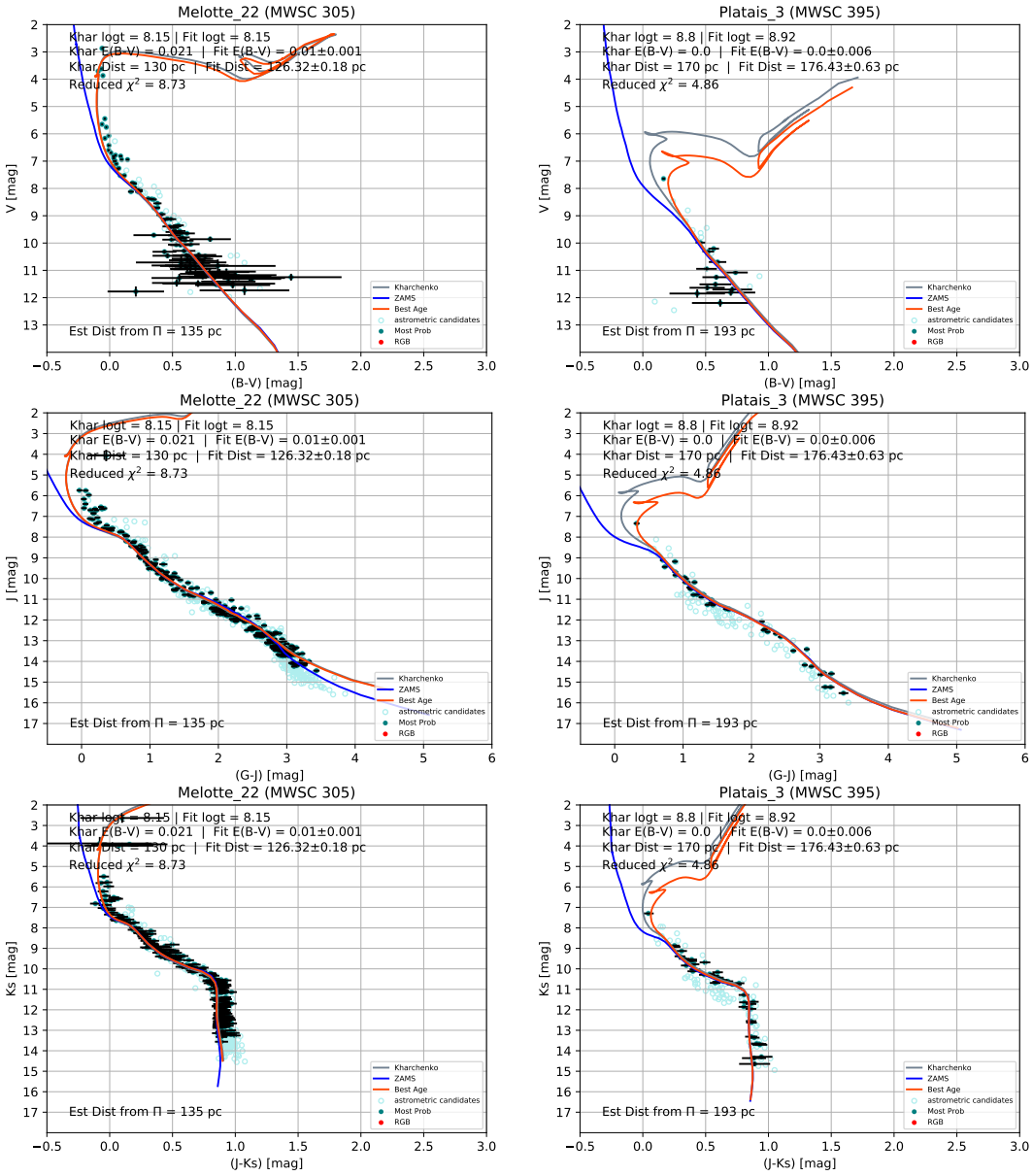

Figure A.3: Color-magnitude diagrams for the Pleiades (Melotte 22) (*left*) and Platais 3 (*right*). From *top* to *bottom*: their respective V vs. $(B-V)$, J vs. $(G-J)$, and K_s vs. $(J-K_s)$ CMDs. The cluster members determined from the pipeline are given by teal circles with their corresponding magnitude and color error bars. The cluster astrometric candidates that were later rejected as cluster members are shown by light blue open circles. RGB stars, if any, are indicated by red circles. The red isochrone is the pipeline selected age, plotted with the fitted cluster $E(B-V)$ and d. This isochrone is plotted without the binary offset. The gray isochrone shows the age, $E(B-V)$, and d as determined by Kharchenko et al. (2013). The blue line is the ZAMS plotted with the fitted cluster $E(B-V)$ and d.

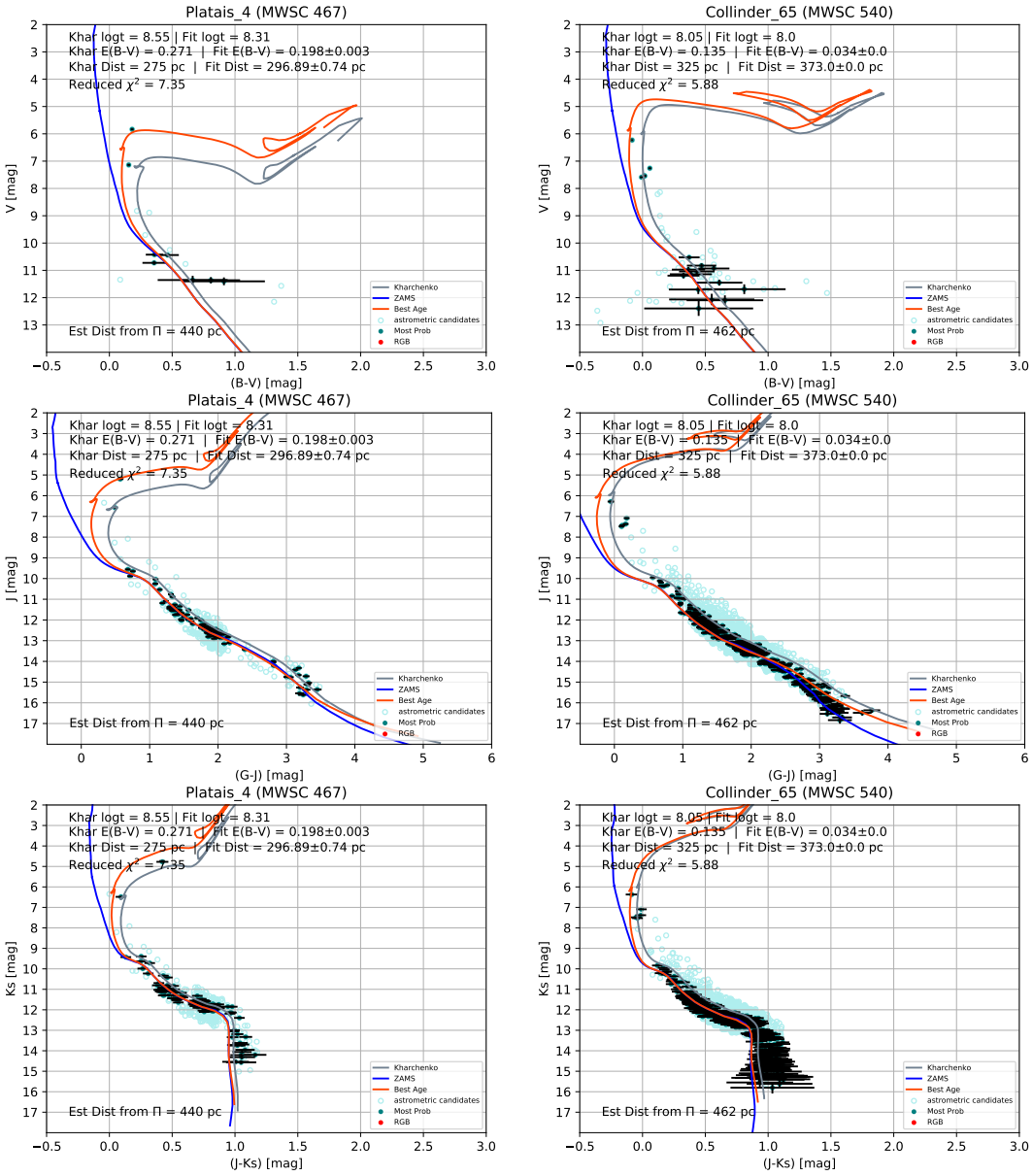

Figure A.4: Color-magnitude diagrams for Platais 4 (*left*) and Collinder 65 (*right*). From *top* to *bottom*: their respective V vs. $(B-V)$, J vs. $(G-J)$, and K_s vs. $(J-K_s)$ CMDs. The cluster members determined from the pipeline are given by teal circles with their corresponding magnitude and color error bars. The cluster astrometric candidates that were later rejected as cluster members are shown by light blue open circles. RGB stars, if any, are indicated by red circles. The red isochrone is the pipeline selected age, plotted with the fitted cluster $E(B-V)$ and d. This isochrone is plotted without the binary offset. The gray isochrone shows the age, $E(B-V)$, and d as determined by Kharchenko et al. (2013). The blue line is the ZAMS plotted with the fitted cluster $E(B-V)$ and d.

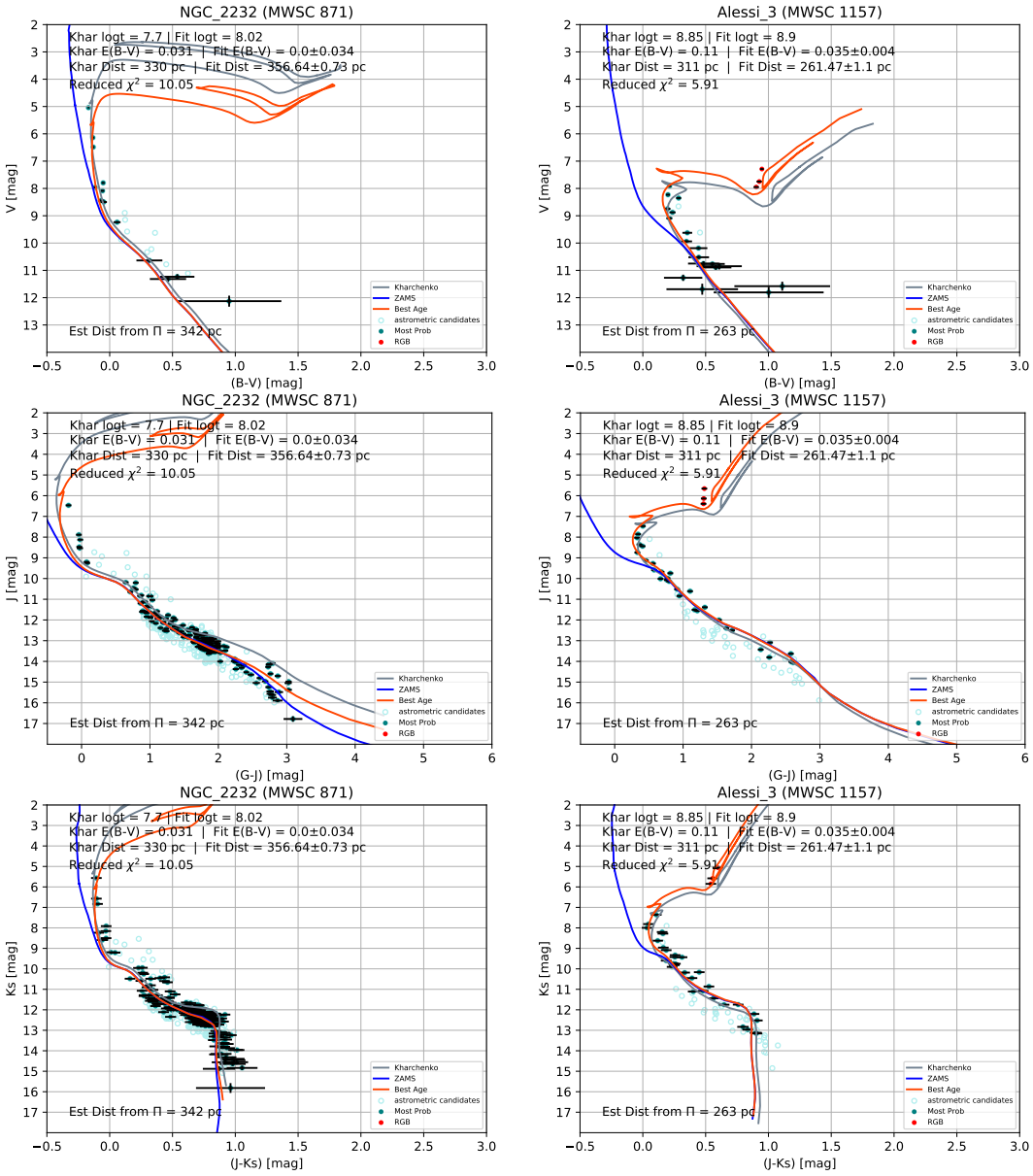

Figure A.5: Color-magnitude diagrams for NGC2232 (*left*) and Alessi 3 (*right*). From *top* to *bottom*: their respective V vs. $(B-V)$, J vs. $(G-J)$, and K_s vs. $(J-K_s)$ CMDs. The cluster members determined from the pipeline are given by teal circles with their corresponding magnitude and color error bars. The cluster astrometric candidates that were later rejected as cluster members are shown by light blue open circles. RGB stars, if any, are indicated by red circles. The red isochrone is the pipeline selected age, plotted with the fitted cluster $E(B-V)$ and d. This isochrone is plotted without the binary offset. The gray isochrone shows the age, $E(B-V)$, and d as determined by Kharchenko et al. (2013). The blue line is the ZAMS plotted with the fitted cluster $E(B-V)$ and d.

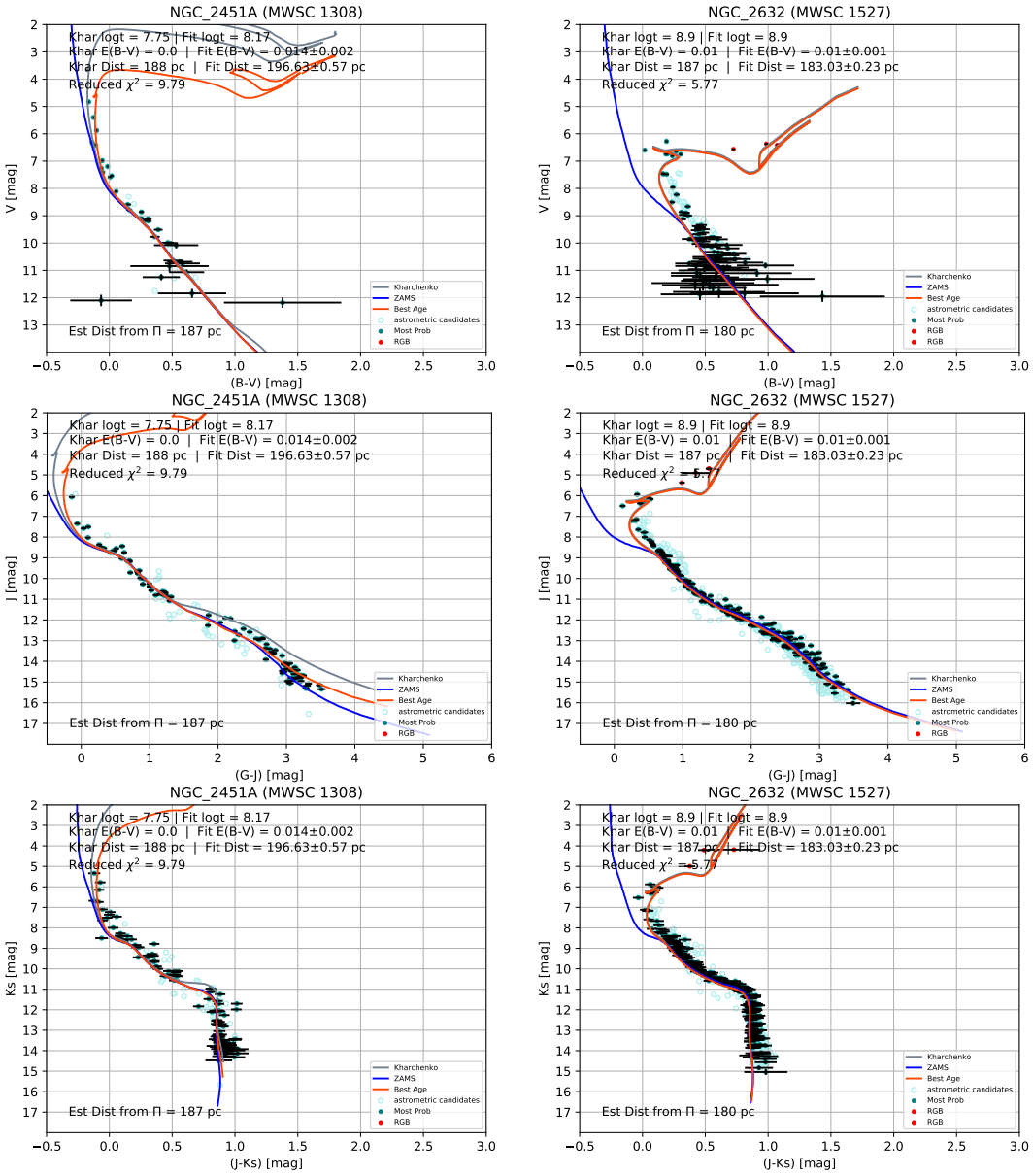

Figure A.6: Color-magnitude diagrams for NGC2451A (*left*) and Praesepe (NGC2632) (*right*). From *top* to *bottom*: their respective V vs. $(B-V)$, J vs. $(G-J)$, and K_s vs. $(J-K_s)$ CMDs. The cluster members determined from the pipeline are given by teal circles with their corresponding magnitude and color error bars. The cluster astrometric candidates that were later rejected as cluster members are shown by light blue open circles. RGB stars, if any, are indicated by red circles. The red isochrone is the pipeline selected age, plotted with the fitted cluster $E(B-V)$ and d. This isochrone is plotted without the binary offset. The gray isochrone shows the age, $E(B-V)$, and d as determined by Kharchenko et al. (2013). The blue line is the ZAMS plotted with the fitted cluster $E(B-V)$ and d.

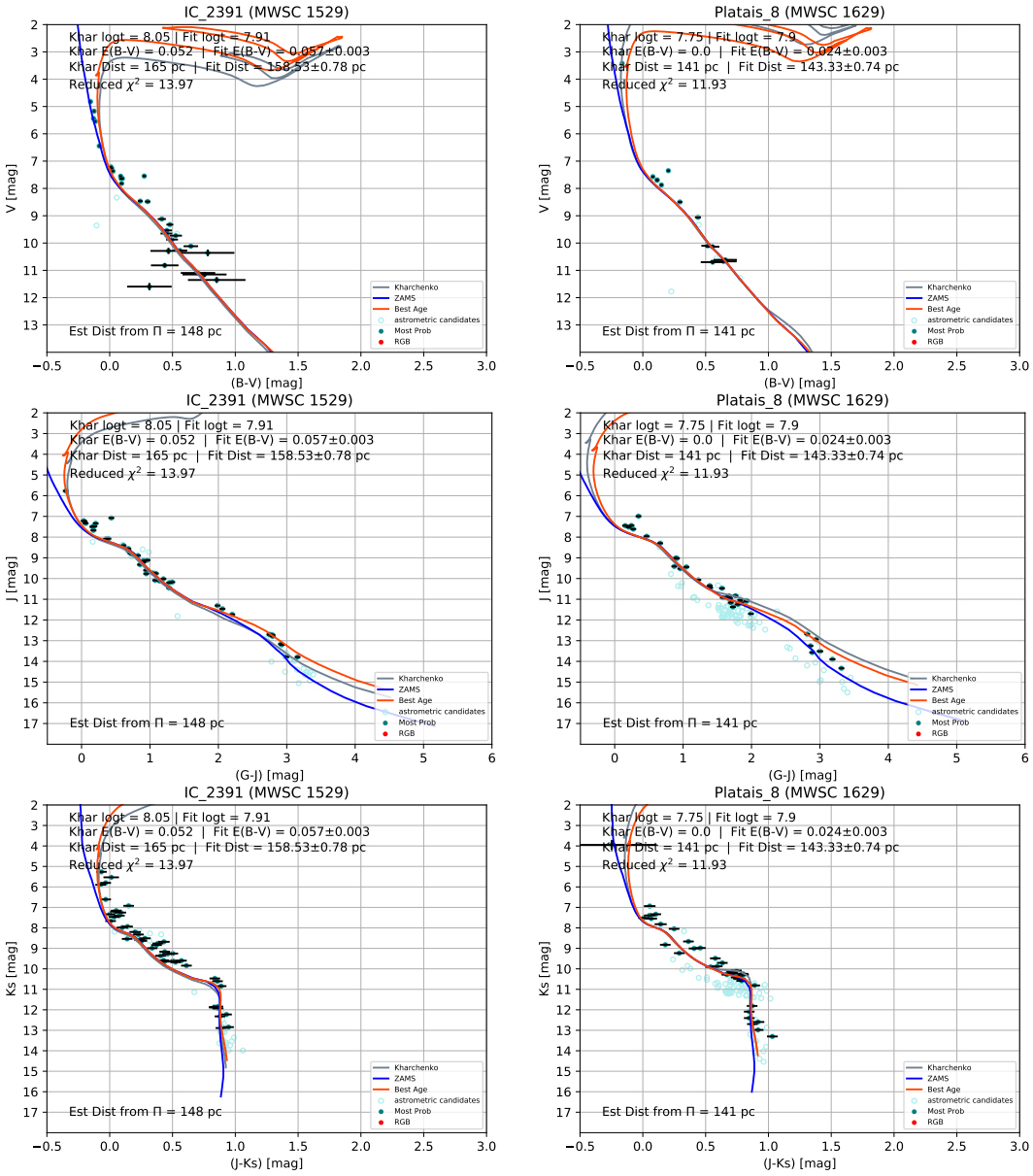

Figure A.7: Color-magnitude diagrams for IC2391 (*left*) and Platais 8 (*right*). From *top* to *bottom*: their respective V vs. $(B-V)$, J vs. $(G-J)$, and K_s vs. $(J-K_s)$ CMDs. The cluster members determined from the pipeline are given by teal circles with their corresponding magnitude and color error bars. The cluster astrometric candidates that were later rejected as cluster members are shown by light blue open circles. RGB stars, if any, are indicated by red circles. The red isochrone is the pipeline selected age, plotted with the fitted cluster $E(B-V)$ and d. This isochrone is plotted without the binary offset. The gray isochrone shows the age, $E(B-V)$, and d as determined by Kharchenko et al. (2013). The blue line is the ZAMS plotted with the fitted cluster $E(B-V)$ and d.

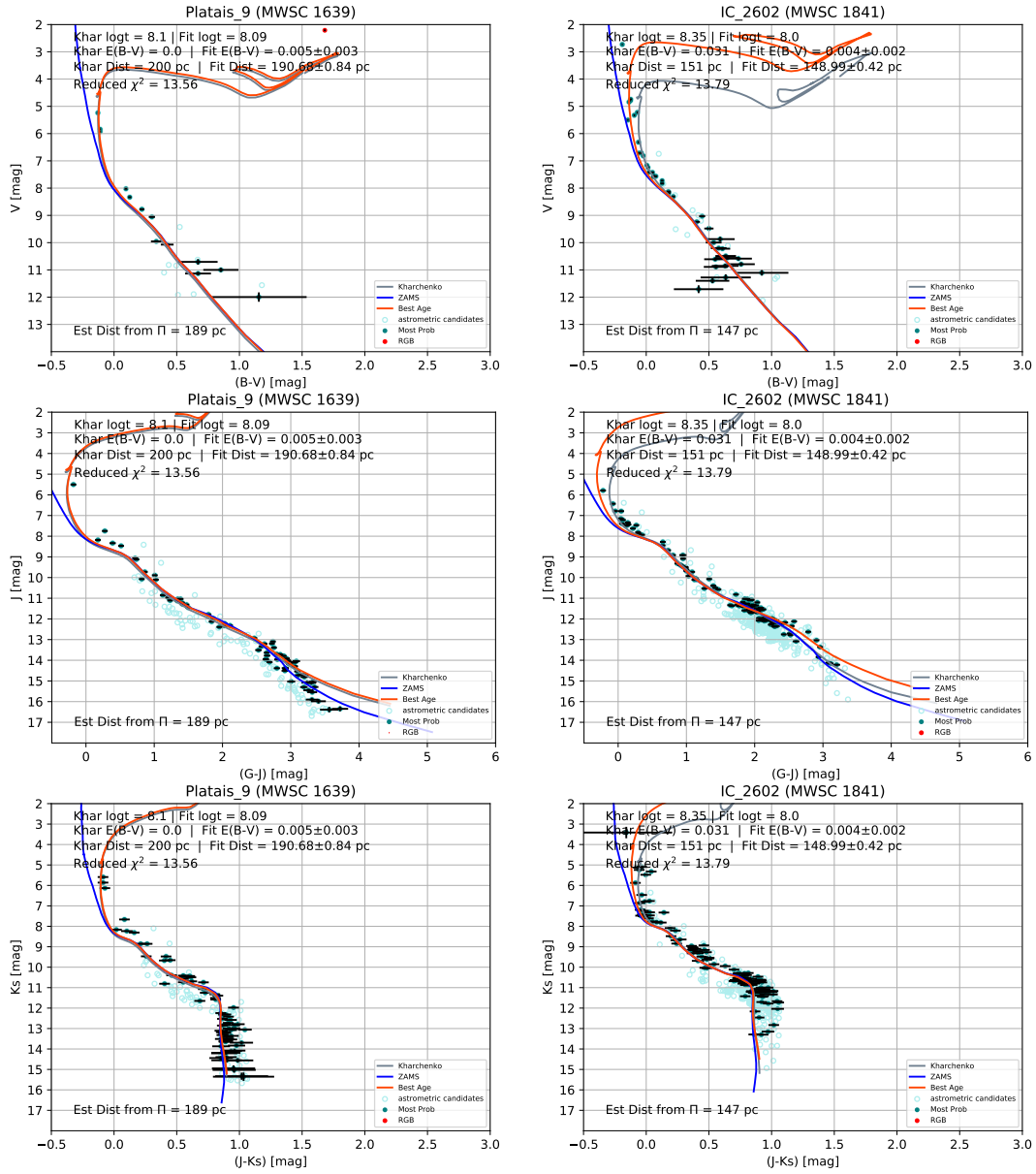

Figure A.8: Color-magnitude diagrams for Platais 9 (*left*) and IC2602 (*right*). From *top* to *bottom*: their respective V vs. $(B-V)$, J vs. $(G-J)$, and K_s vs. $(J-K_s)$ CMDs. The cluster members determined from the pipeline are given by teal circles with their corresponding magnitude and color error bars. The cluster astrometric candidates that were later rejected as cluster members are shown by light blue open circles. RGB stars, if any, are indicated by red circles. The red isochrone is the pipeline selected age, plotted with the fitted cluster $E(B-V)$ and d. This isochrone is plotted without the binary offset. The gray isochrone shows the age, $E(B-V)$, and d as determined by Kharchenko et al. (2013). The blue line is the ZAMS plotted with the fitted cluster $E(B-V)$ and d.

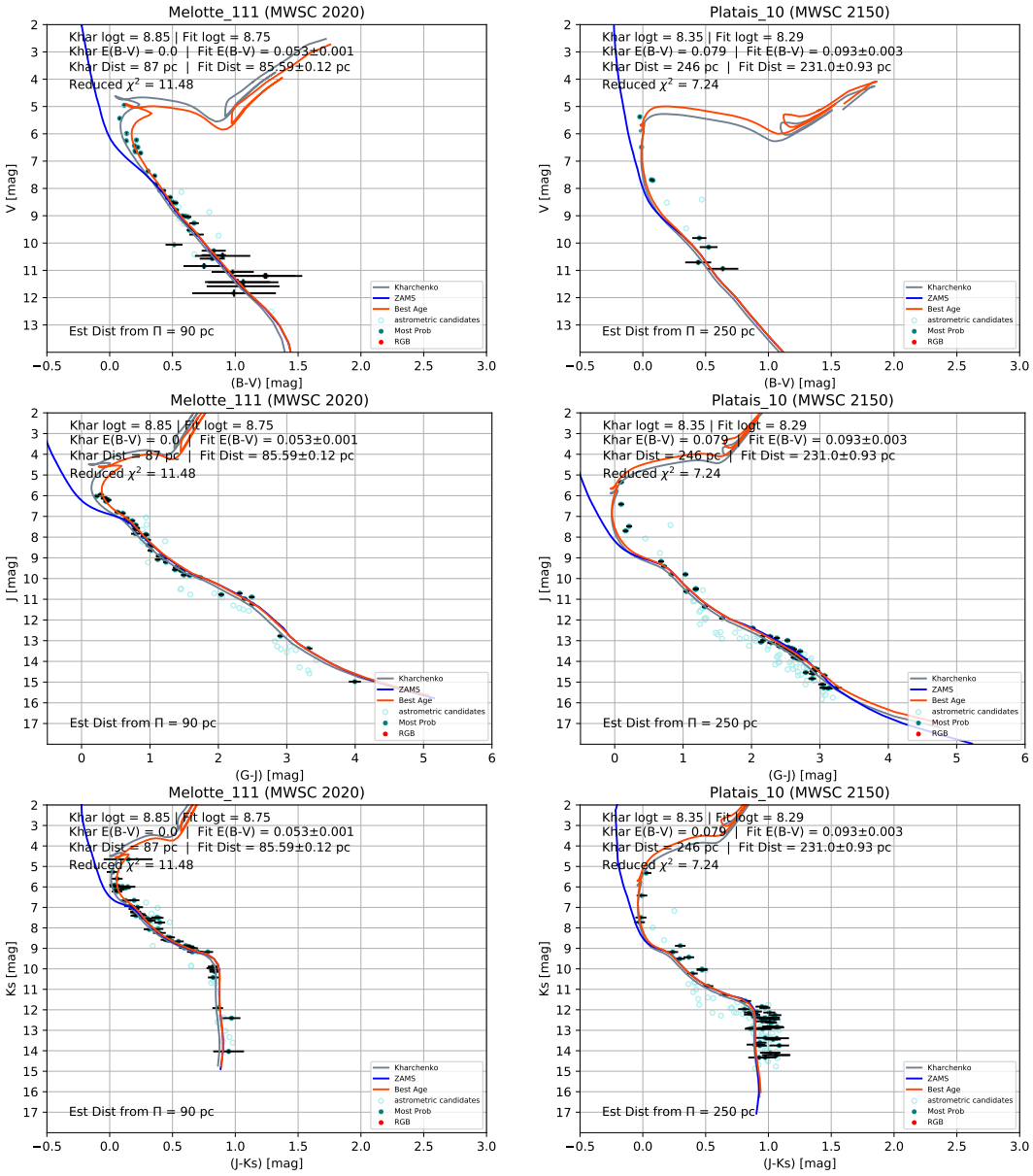

Figure A.9: Color-magnitude diagrams for Coma Ber (Melotte 111) (*left*) and Platais 10 (*right*). From *top* to *bottom*: their respective V vs. $(B-V)$, J vs. $(G-J)$, and K_s vs. $(J-K_s)$ CMDs. The cluster members determined from the pipeline are given by teal circles with their corresponding magnitude and color error bars. The cluster astrometric candidates that were later rejected as cluster members are shown by light blue open circles. RGB stars, if any, are indicated by red circles. The red isochrone is the pipeline selected age, plotted with the fitted cluster $E(B-V)$ and d. This isochrone is plotted without the binary offset. The gray isochrone shows the age, $E(B-V)$, and d as determined by Kharchenko et al. (2013). The blue line is the ZAMS plotted with the fitted cluster $E(B-V)$ and d.

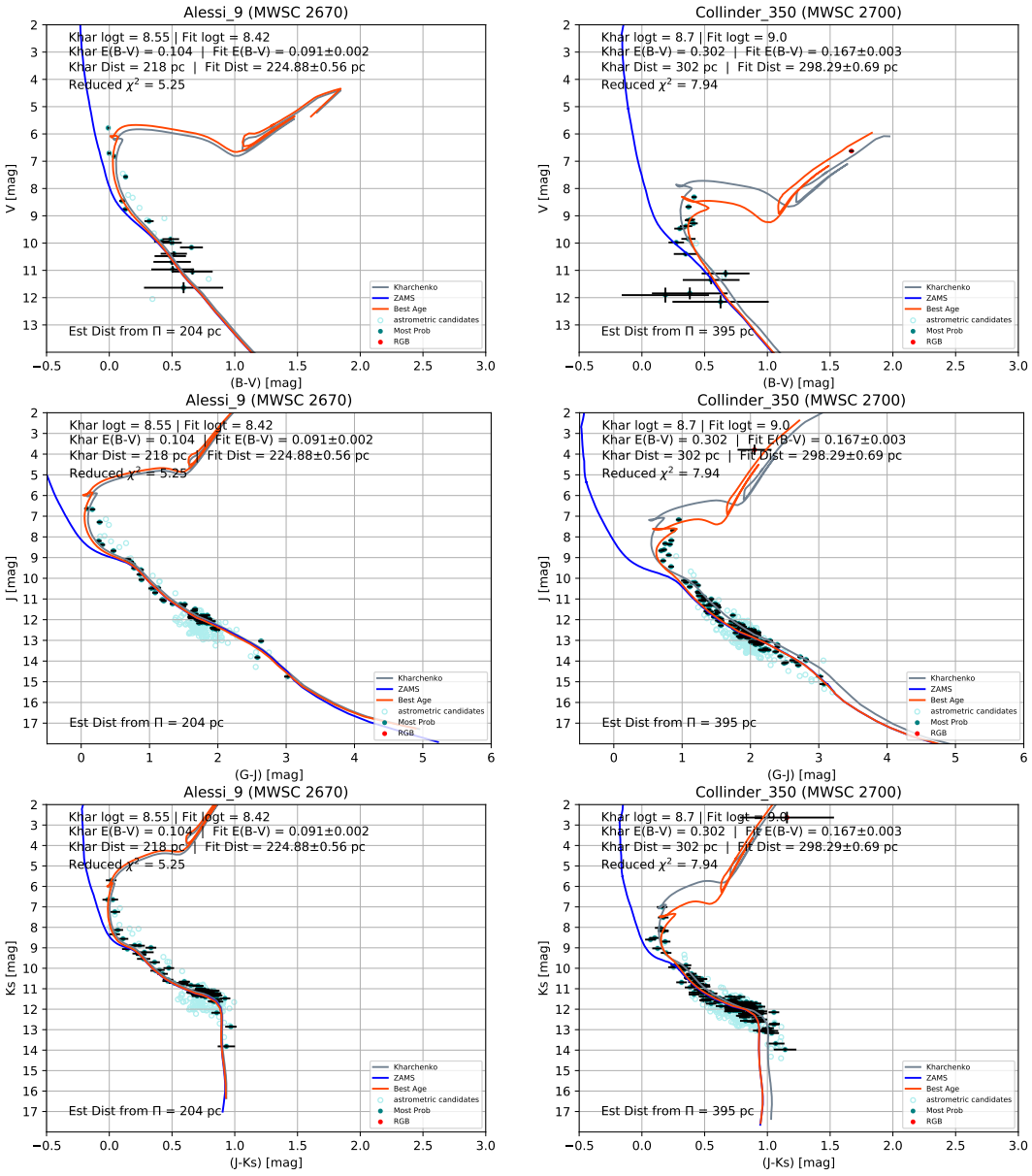

Figure A.10: Color-magnitude diagrams for Alessi 9 (*left*) and Collinder 350 (*right*). From *top* to *bottom*: their respective V vs. $(B-V)$, J vs. $(G-J)$, and K_s vs. $(J-K_s)$ CMDs. The cluster members determined from the pipeline are given by teal circles with their corresponding magnitude and color error bars. The cluster astrometric candidates that were later rejected as cluster members are shown by light blue open circles. RGB stars, if any, are indicated by red circles. The red isochrone is the pipeline selected age, plotted with the fitted cluster $E(B-V)$ and d. This isochrone is plotted without the binary offset. The gray isochrone shows the age, $E(B-V)$, and d as determined by Kharchenko et al. (2013). The blue line is the ZAMS plotted with the fitted cluster $E(B-V)$ and d.

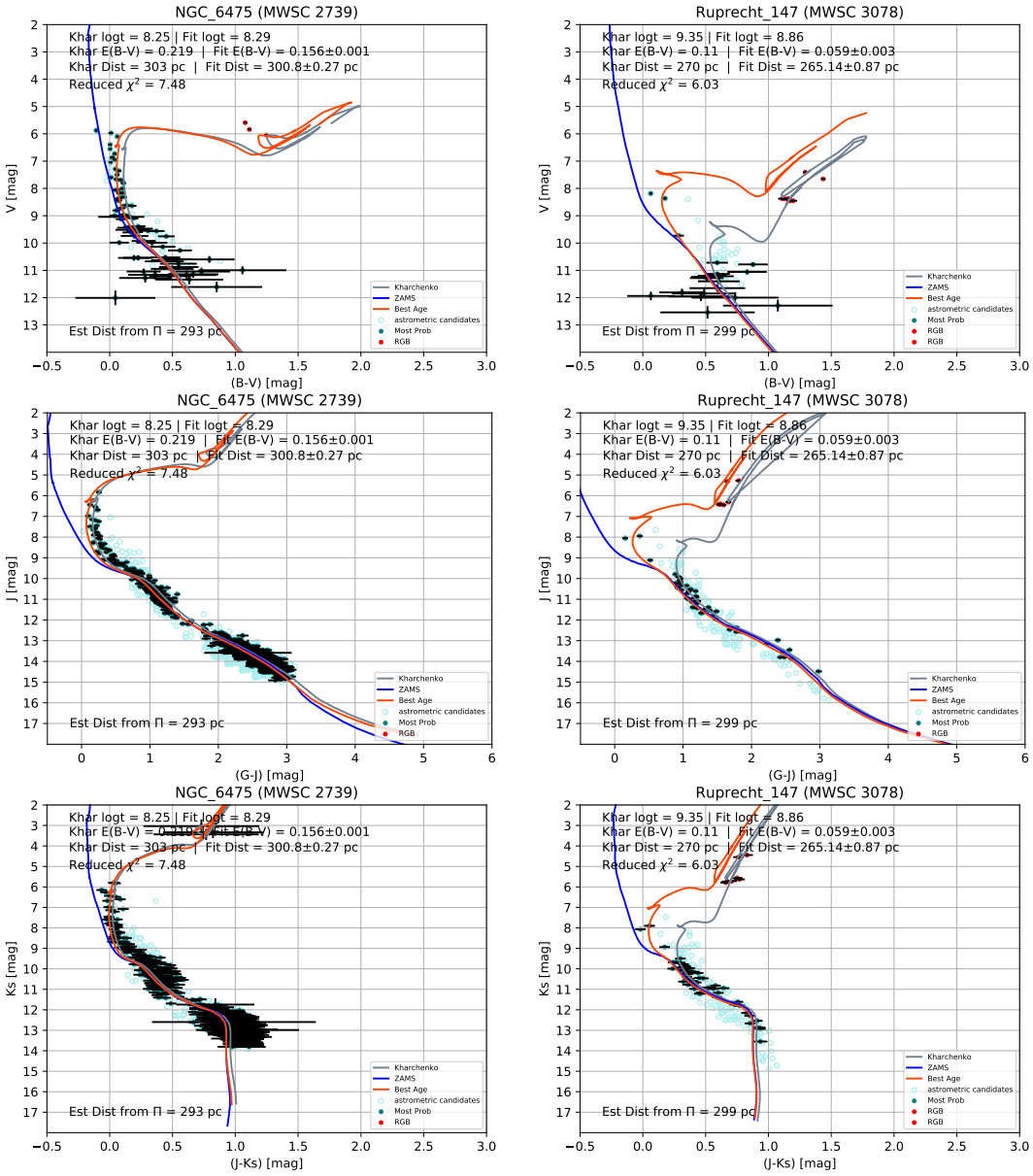

Figure A.11: Color-magnitude diagrams for NGC6475 (*left*) and Ruprecht 147 (*right*). From *top* to *bottom*: their respective V vs. $(B-V)$, J vs. $(G-J)$, and K_s vs. $(J-K_s)$ CMDs. The cluster members determined from the pipeline are given by teal circles with their corresponding magnitude and color error bars. The cluster astrometric candidates that were later rejected as cluster members are shown by light blue open circles. RGB stars, if any, are indicated by red circles. The red isochrone is the pipeline selected age, plotted with the fitted cluster $E(B-V)$ and d. This isochrone is plotted without the binary offset. The gray isochrone shows the age, $E(B-V)$, and d as determined by Kharchenko et al. (2013). The blue line is the ZAMS plotted with the fitted cluster $E(B-V)$ and d.

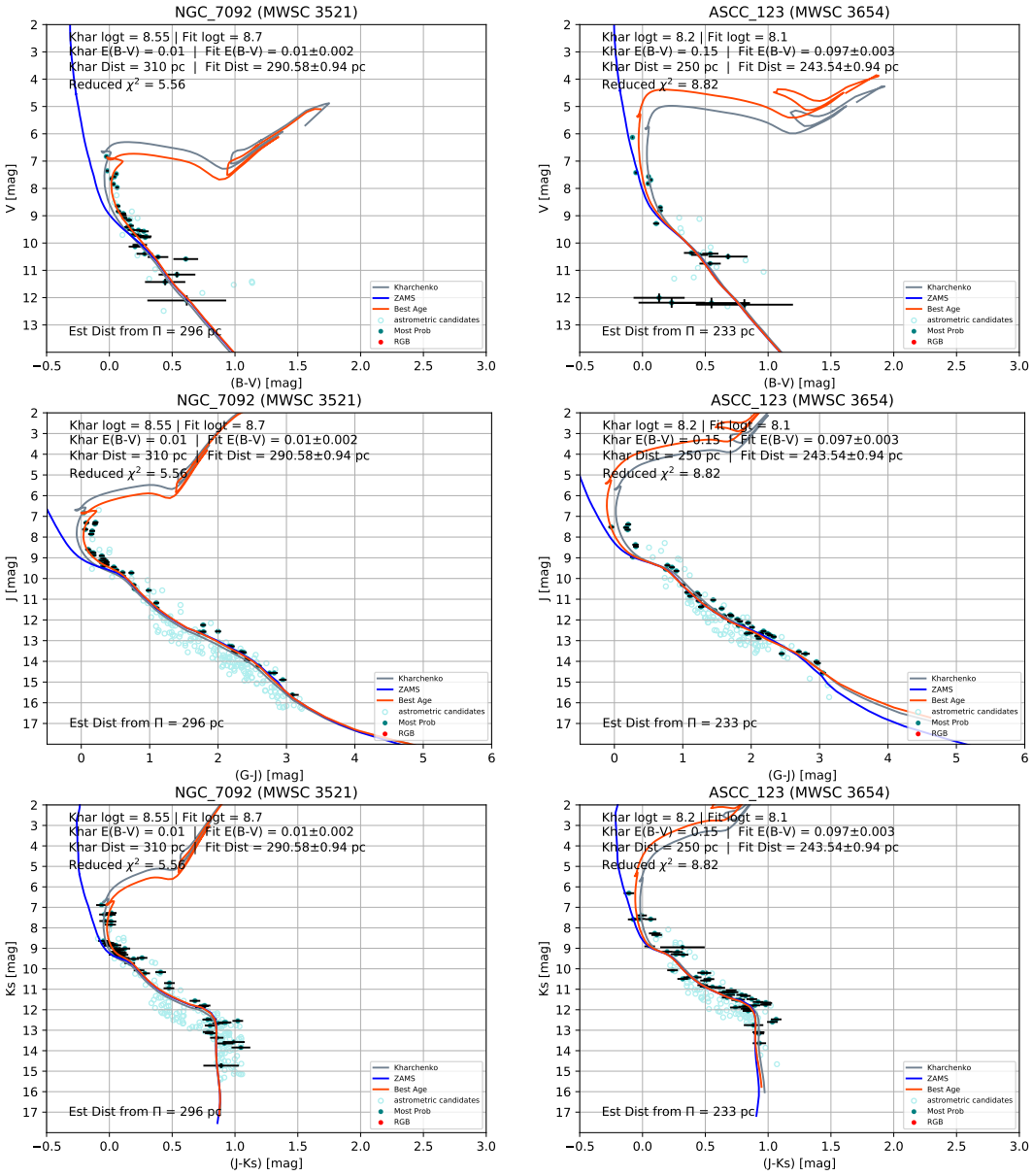

Figure A.12: Color-magnitude diagrams for NGC7092 (*left*) and ASCC 123 (*right*). From *top* to *bottom*: their respective V vs. $(B-V)$, J vs. $(G-J)$, and K_s vs. $(J-K_s)$ CMDs. The cluster members determined from the pipeline are given by teal circles with their corresponding magnitude and color error bars. The cluster astrometric candidates that were later rejected as cluster members are shown by light blue open circles. RGB stars, if any, are indicated by red circles. The red isochrone is the pipeline selected age, plotted with the fitted cluster $E(B-V)$ and d. This isochrone is plotted without the binary offset. The gray isochrone shows the age, $E(B-V)$, and d as determined by Kharchenko et al. (2013). The blue line is the ZAMS plotted with the fitted cluster $E(B-V)$ and d.

A.2 Cluster TGAS proper motion diagrams & parallax histograms

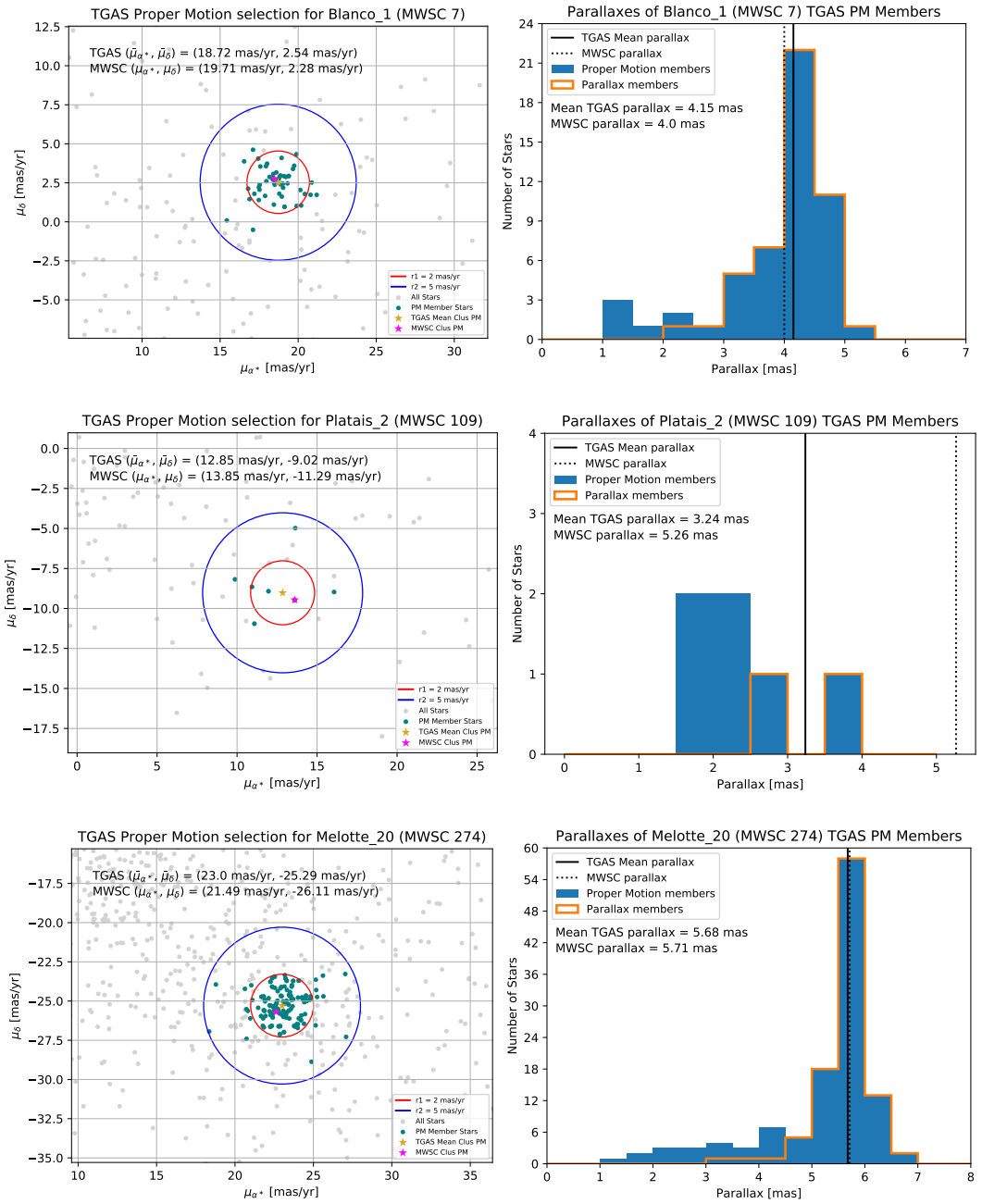

Figure A.13: TGAS proper motion (*left*) and parallax (*right*) selection diagrams for Blanco 1, Platais 2, and α Per (Melotte 20), from *top* to *bottom*. The proper motion and parallax criteria for membership selection is described in Sect. 3.1; the values for cluster proper motion and parallaxes in the diagrams are based on initial cluster membership. The final cluster proper motion and parallaxes are provided in Table 3.1. *Left panels*: The teal points represent the proper motion members, where all stars within the 2 mas yr^{-1} radius (red circle) of the mean cluster proper motion are selected and the stars within 5 mas yr^{-1} (blue circle) are only selected if their 3σ errors are consistent with the mean cluster proper motion. *Right panels*: The orange outline illustrates the stars with 3σ errors consistent with the mean cluster parallax; these stars are the TGAS astrometrically-selected candidates of the cluster.

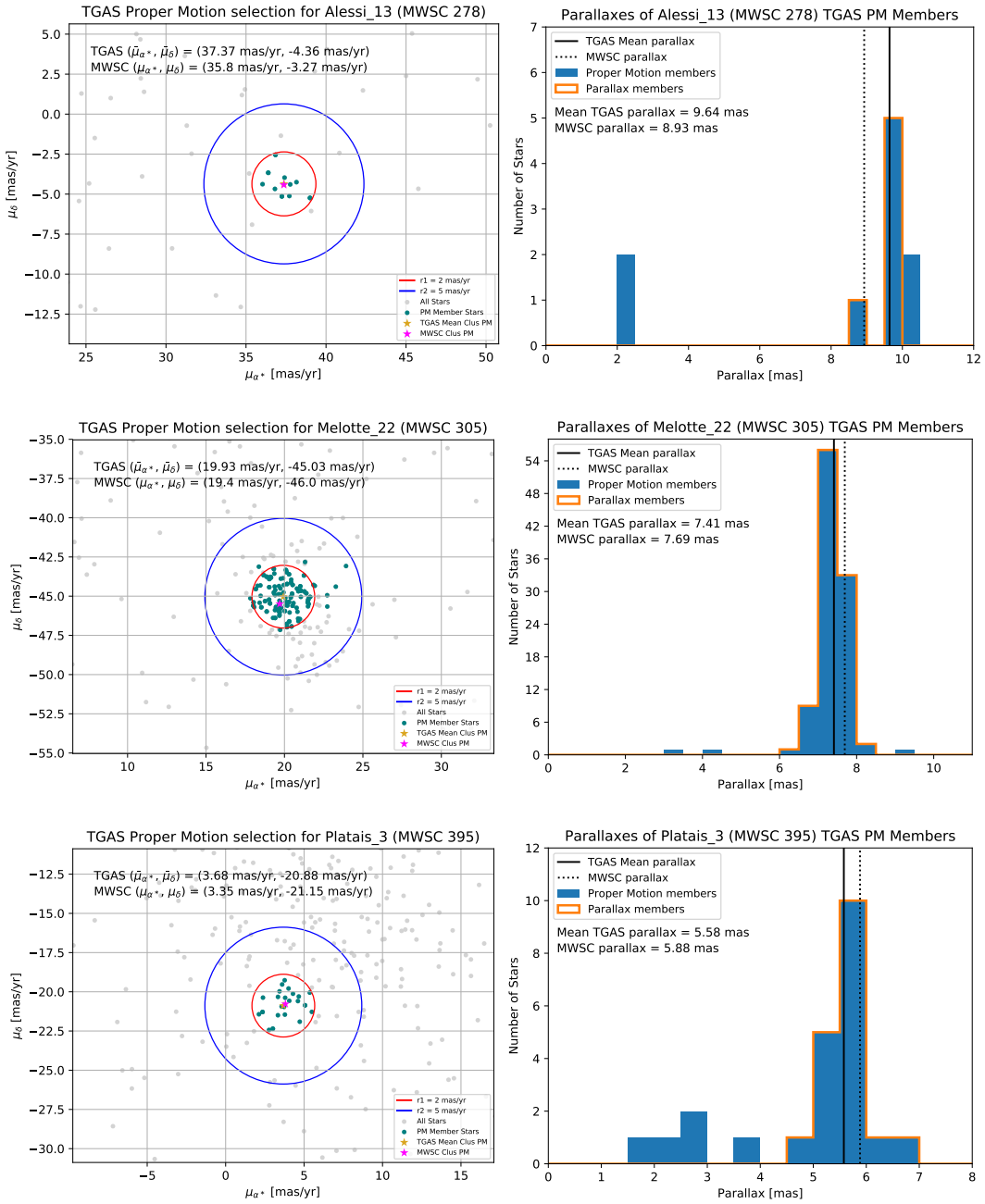

Figure A.14: TGAS proper motion (*left*) and parallax (*right*) selection diagrams for Alessi 13, the Pleiades (Melotte 22), and Platais 3, from *top* to *bottom*. The proper motion and parallax criteria for membership selection is described in Sect. 3.1; the values for cluster proper motion and parallaxes in the diagrams are based on initial cluster membership. The final cluster proper motion and parallaxes are provided in Table 3.1. *Left panels*: The teal points represent the proper motion members, where all stars within the 2 mas yr^{-1} radius (red circle) of the mean cluster proper motion are selected and the stars within 5 mas yr^{-1} (blue circle) are only selected if their 3σ errors are consistent with the mean cluster proper motion. *Right panels*: The orange outline illustrates the stars with 3σ errors consistent with the mean cluster parallax; these stars are the TGAS astrometrically-selected candidates of the cluster.

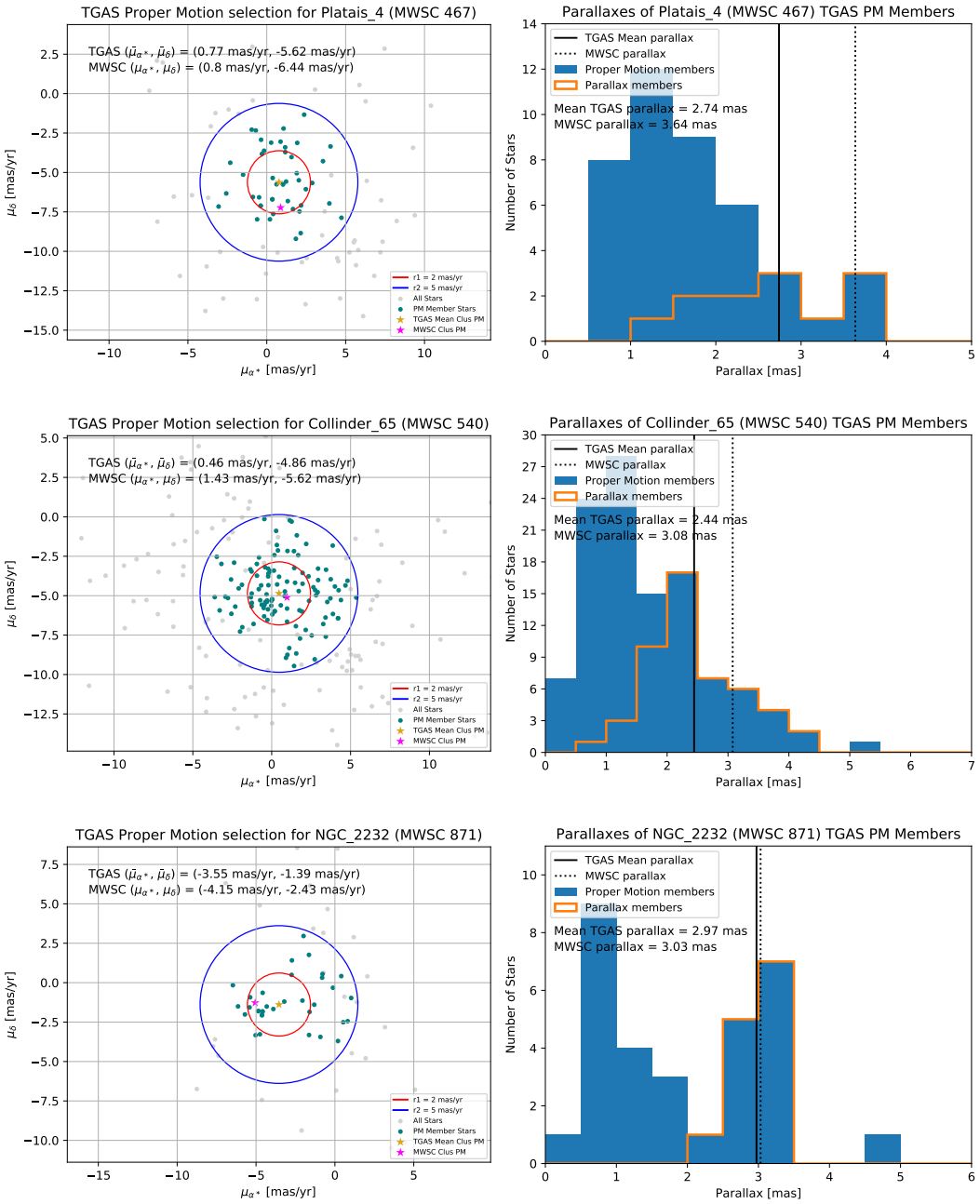

Figure A.15: TGAS proper motion (*left*) and parallax (*right*) selection diagrams for Platais 4, Collinder 65, and NGC2232, from *top* to *bottom*. The proper motion and parallax criteria for membership selection is described in Sect. 3.1; the values for cluster proper motion and parallaxes in the diagrams are based on initial cluster membership. The final cluster proper motion and parallaxes are provided in Table 3.1. *Left panels*: The teal points represent the proper motion members, where all stars within the 2 mas yr^{-1} radius (red circle) of the mean cluster proper motion are selected and the stars within 5 mas yr^{-1} (blue circle) are only selected if their 3σ errors are consistent with the mean cluster proper motion. *Right panels*: The orange outline illustrates the stars with 3σ errors consistent with the mean cluster parallax; these stars are the TGAS astrometrically-selected candidates of the cluster.

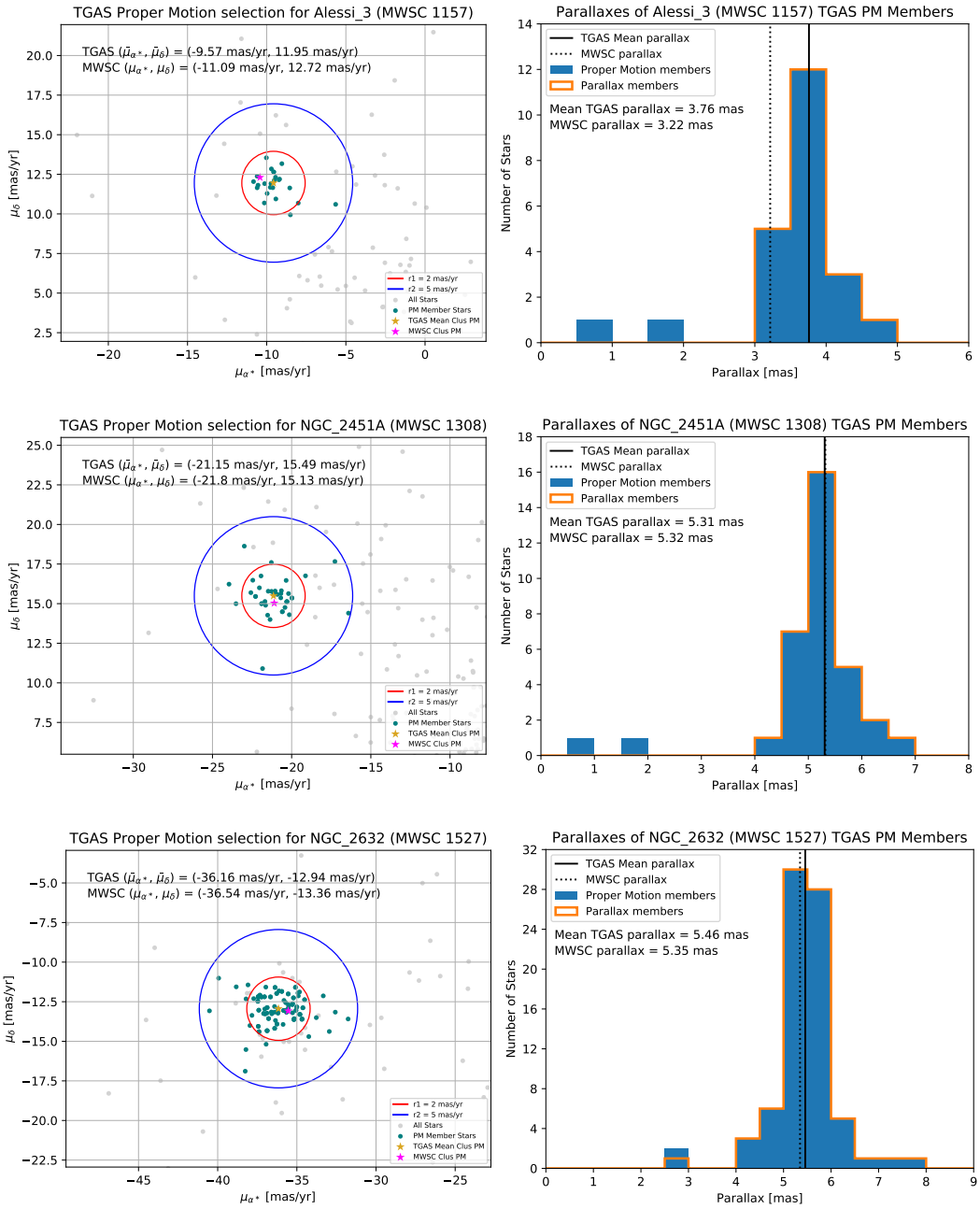

Figure A.16: TGAS proper motion (*left*) and parallax (*right*) selection diagrams for Alessi 3, NGC2451A, and Praesepe (NGC2632), from *top* to *bottom*. The proper motion and parallax criteria for membership selection is described in Sect. 3.1; the values for cluster proper motion and parallaxes in the diagrams are based on initial cluster membership. The final cluster proper motion and parallaxes are provided in Table 3.1. *Left panels*: The teal points represent the proper motion members, where all stars within the 2 mas yr^{-1} radius (red circle) of the mean cluster proper motion are selected and the stars within 5 mas yr^{-1} (blue circle) are only selected if their 3σ errors are consistent with the mean cluster proper motion. *Right panels*: The orange outline illustrates the stars with 3σ errors consistent with the mean cluster parallax; these stars are the TGAS astrometrically-selected candidates of the cluster.

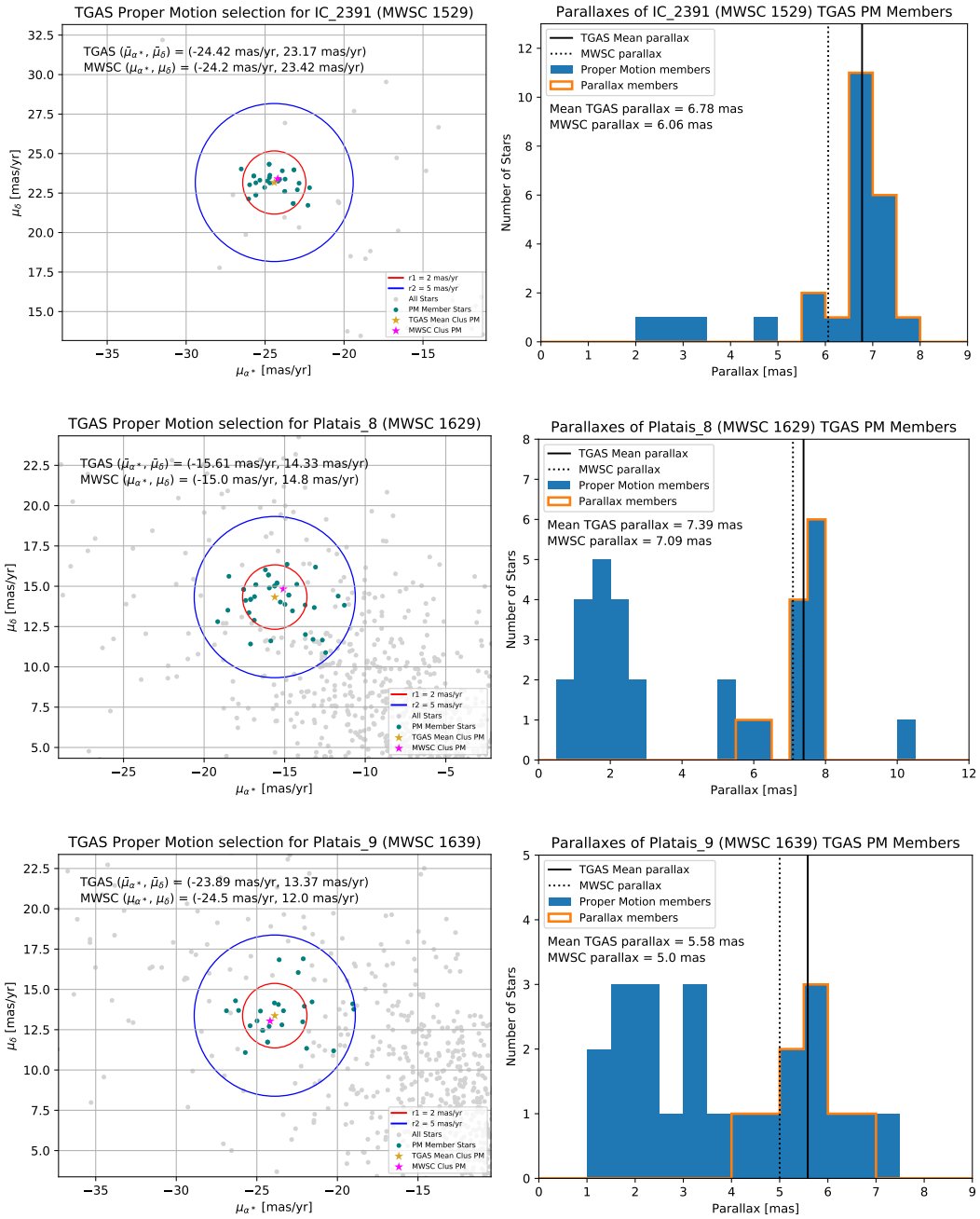

Figure A.17: TGAS proper motion (*left*) and parallax (*right*) selection diagrams for IC2391, Platais 8, and Platais 9, from *top* to *bottom*. The proper motion and parallax criteria for membership selection is described in Sect. 3.1; the values for cluster proper motion and parallaxes in the diagrams are based on initial cluster membership. The final cluster proper motion and parallaxes are provided in Table 3.1. *Left panels*: The teal points represent the proper motion members, where all stars within the 2 mas yr^{-1} radius (red circle) of the mean cluster proper motion are selected and the stars within 5 mas yr^{-1} (blue circle) are only selected if their 3σ errors are consistent with the mean cluster proper motion. *Right panels*: The orange outline illustrates the stars with 3σ errors consistent with the mean cluster parallax; these stars are the TGAS astrometrically-selected candidates of the cluster.

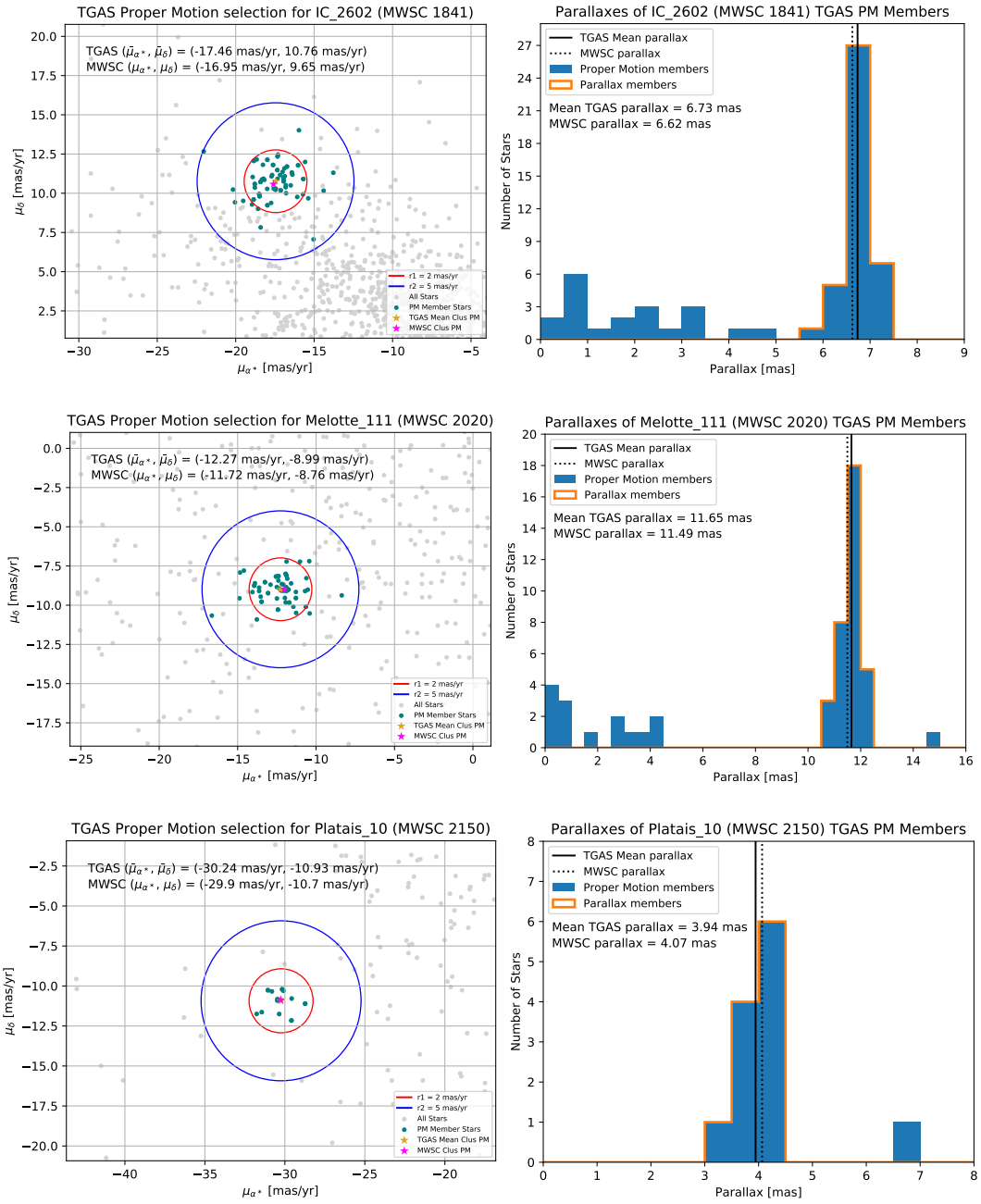

Figure A.18: TGAS proper motion (*left*) and parallax (*right*) selection diagrams for IC2602, Coma Ber (Melotte 111), and Platais 10, from *top* to *bottom*. The proper motion and parallax criteria for membership selection is described in Sect. 3.1; the values for cluster proper motion and parallaxes in the diagrams are based on initial cluster membership. The final cluster proper motion and parallaxes are provided in Table 3.1. *Left panels*: The teal points represent the proper motion members, where all stars within the 2 mas yr^{-1} radius (red circle) of the mean cluster proper motion are selected and the stars within 5 mas yr^{-1} (blue circle) are only selected if their 3σ errors are consistent with the mean cluster proper motion. *Right panels*: The orange outline illustrates the stars with 3σ errors consistent with the mean cluster parallax; these stars are the TGAS astrometrically-selected candidates of the cluster.

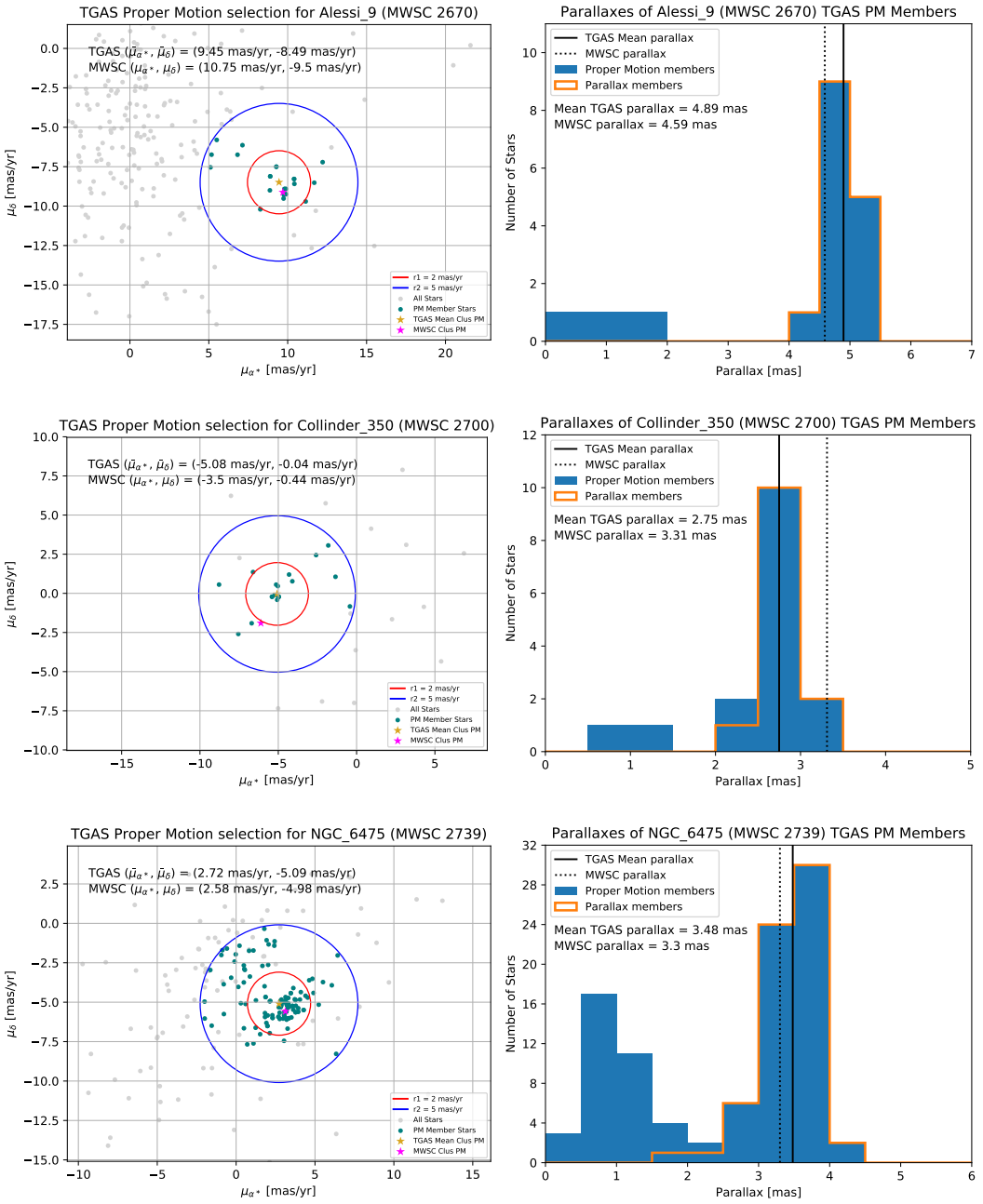

Figure A.19: TGAS proper motion (*left*) and parallax (*right*) selection diagrams for Alessi 9, Collinder 350, and NGC6475, from *top* to *bottom*. The proper motion and parallax criteria for membership selection is described in Sect. 3.1; the values for cluster proper motion and parallaxes in the diagrams are based on initial cluster membership. The final cluster proper motion and parallaxes are provided in Table 3.1. *Left panels*: The teal points represent the proper motion members, where all stars within the 2 mas yr^{-1} radius (red circle) of the mean cluster proper motion are selected and the stars within 5 mas yr^{-1} (blue circle) are only selected if their 3σ errors are consistent with the mean cluster proper motion. *Right panels*: The orange outline illustrates the stars with 3σ errors consistent with the mean cluster parallax; these stars are the TGAS astrometrically-selected candidates of the cluster.

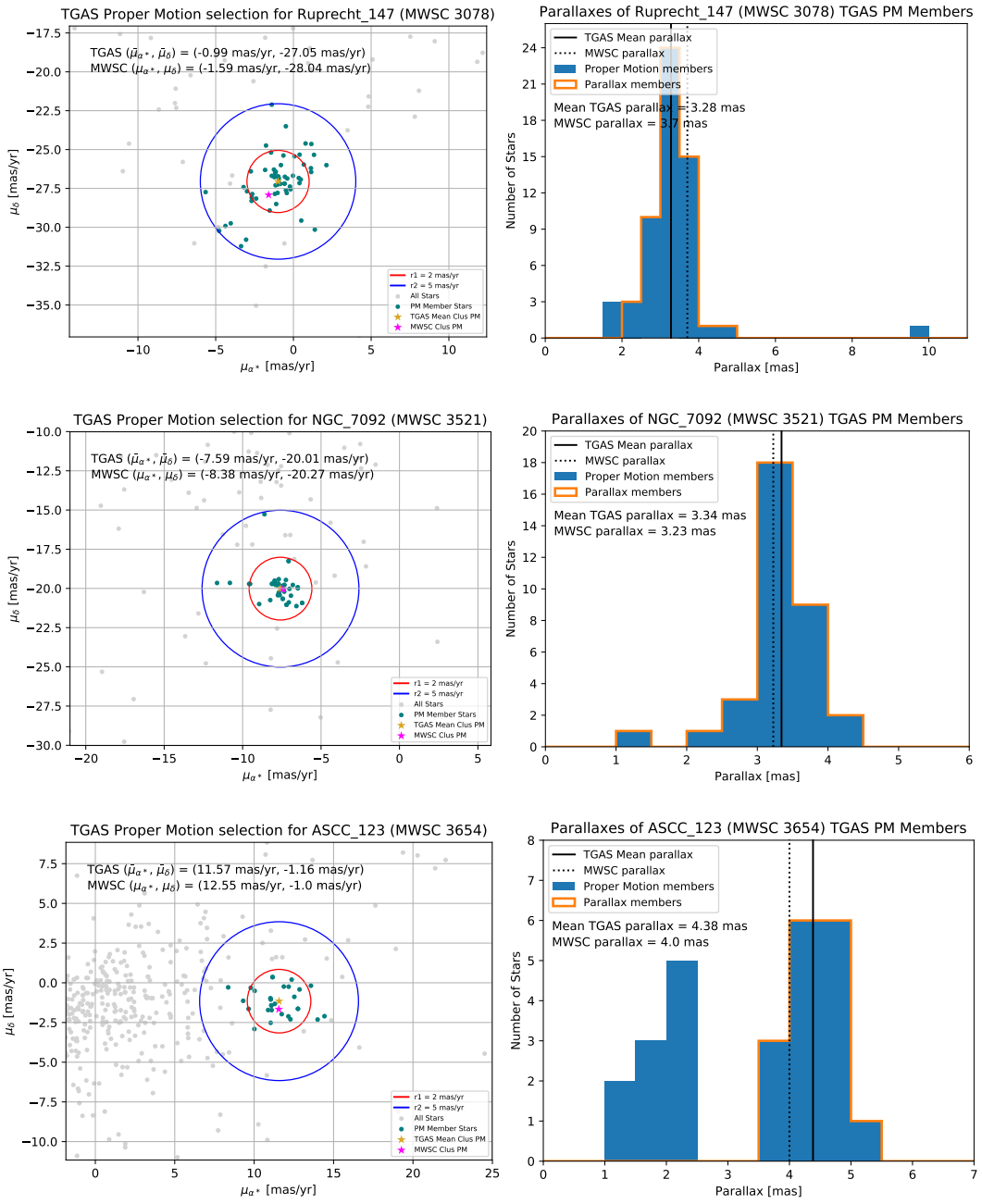

Figure A.20: TGAS proper motion (*left*) and parallax (*right*) selection diagrams for Ruprecht 147, NGC7092, and ASCC123, from *top* to *bottom*. The proper motion and parallax criteria for membership selection is described in Sect. 3.1; the values for cluster proper motion and parallaxes in the diagrams are based on initial cluster membership. The final cluster proper motion and parallaxes are provided in Table 3.1. *Left panels*: The teal points represent the proper motion members, where all stars within the 2 mas yr^{-1} radius (red circle) of the mean cluster proper motion are selected and the stars within 5 mas yr^{-1} (blue circle) are only selected if their 3σ errors are consistent with the mean cluster proper motion. *Right panels*: The orange outline illustrates the stars with 3σ errors consistent with the mean cluster parallax; these stars are the TGAS astrometrically-selected candidates of the cluster.

B

Supplementary material for Chapter 4

B.1 Parameter comparisons with MWSC

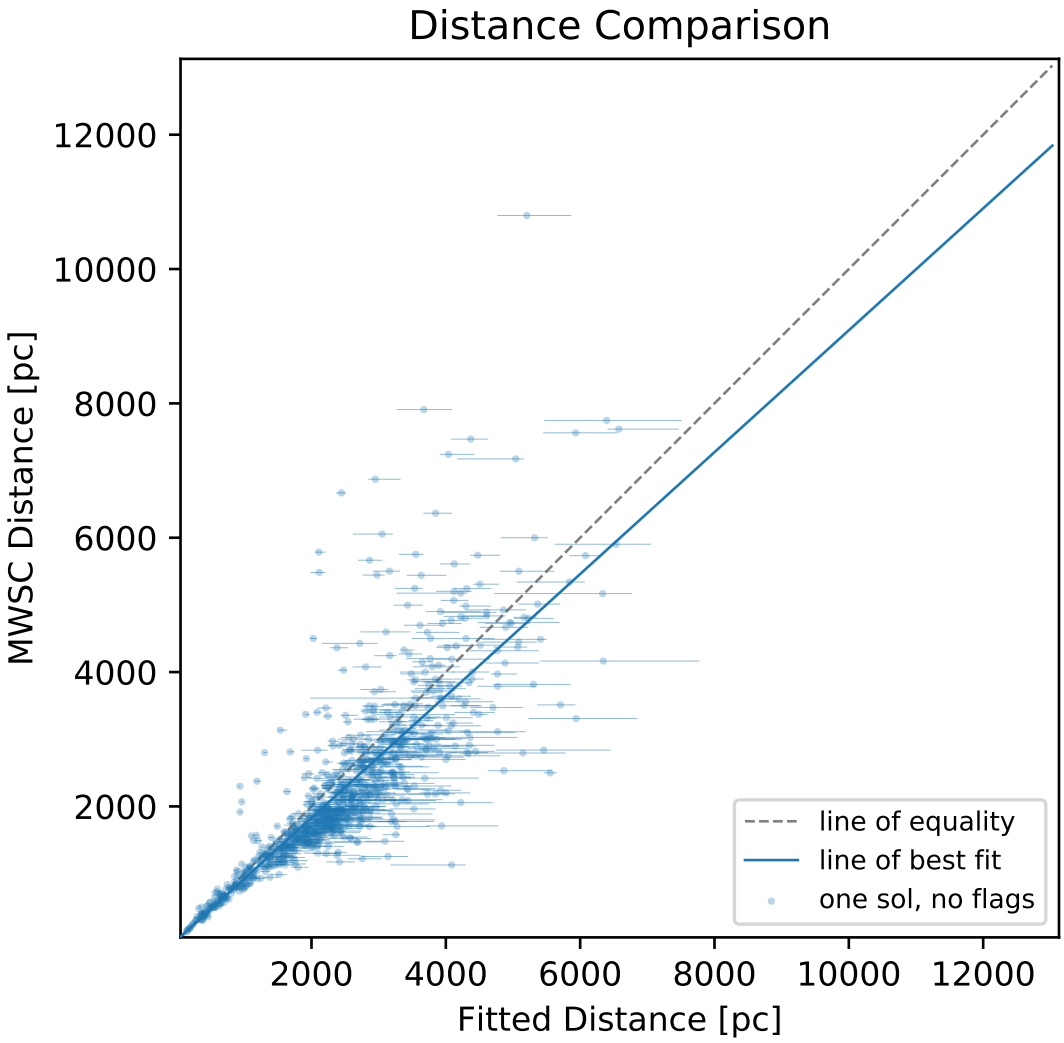

Figure B.1: Comparison of cluster distances derived from this work to those determined in MWSC. The dashed grey line indicates the line of equality and the solid blue line shows the best fit to the single solution results with no flags. The single solution clusters with no flags are shown in blue.

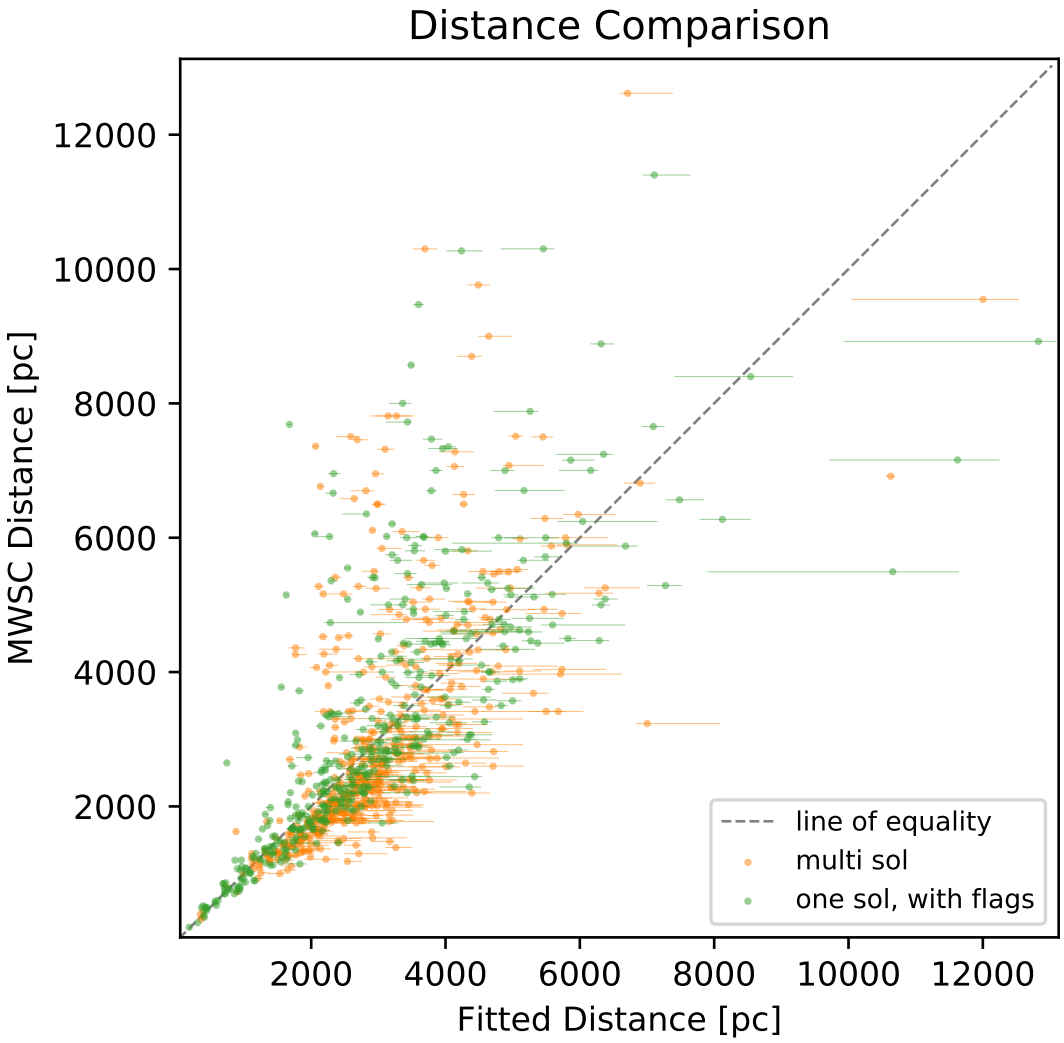

Figure B.2: Comparison of cluster distances derived from this work to those determined in MWSC. The dashed grey line indicates the line of equality. The single solution clusters with one or more flags are shown in green and one result from each of the multiple solutions clusters is shown in orange.

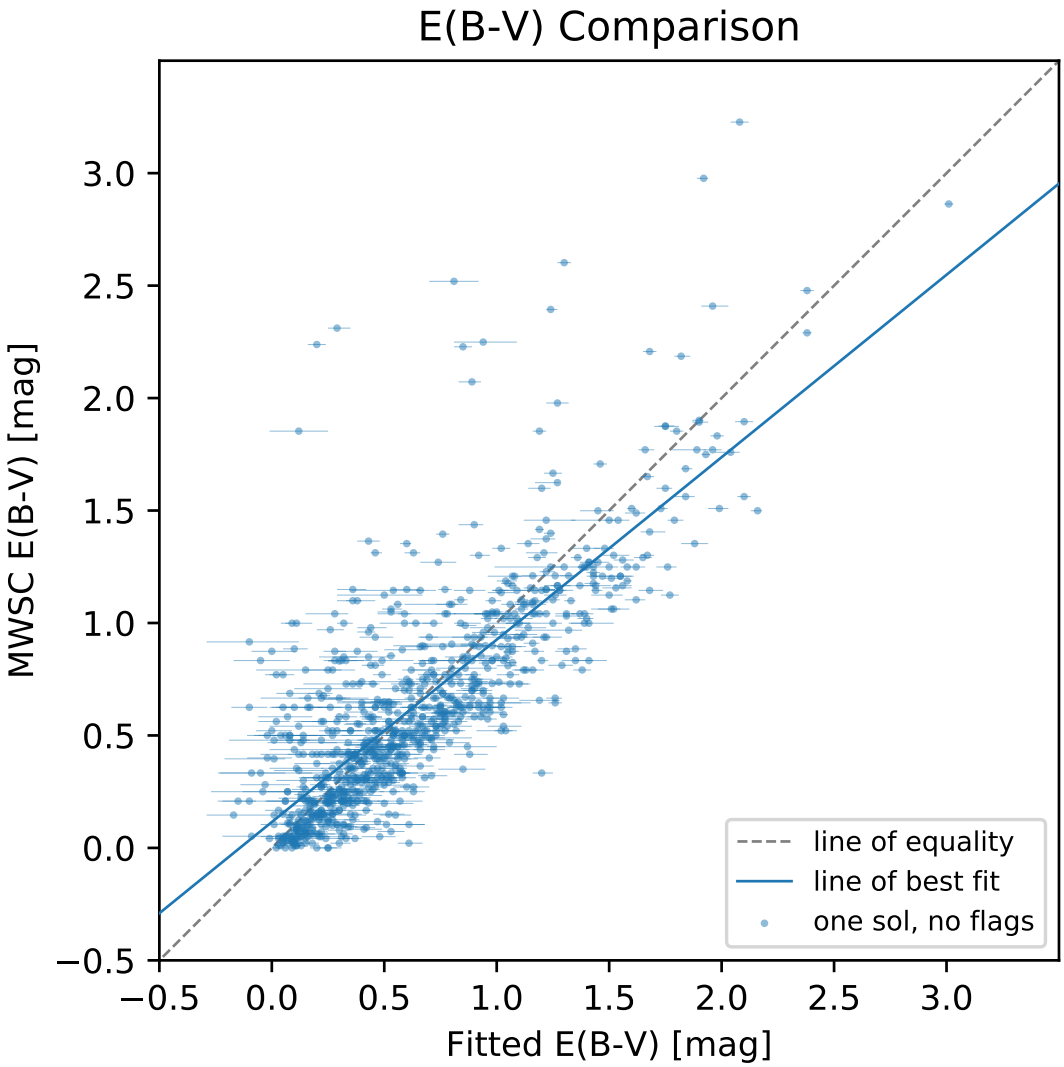

Figure B.3: Comparison of cluster reddenings derived from this work to those determined in MWSC. The dashed grey line indicates the line of equality and the blue line shows the best fit to the single solution results with no flags. The single solution clusters with no flags are shown in blue.

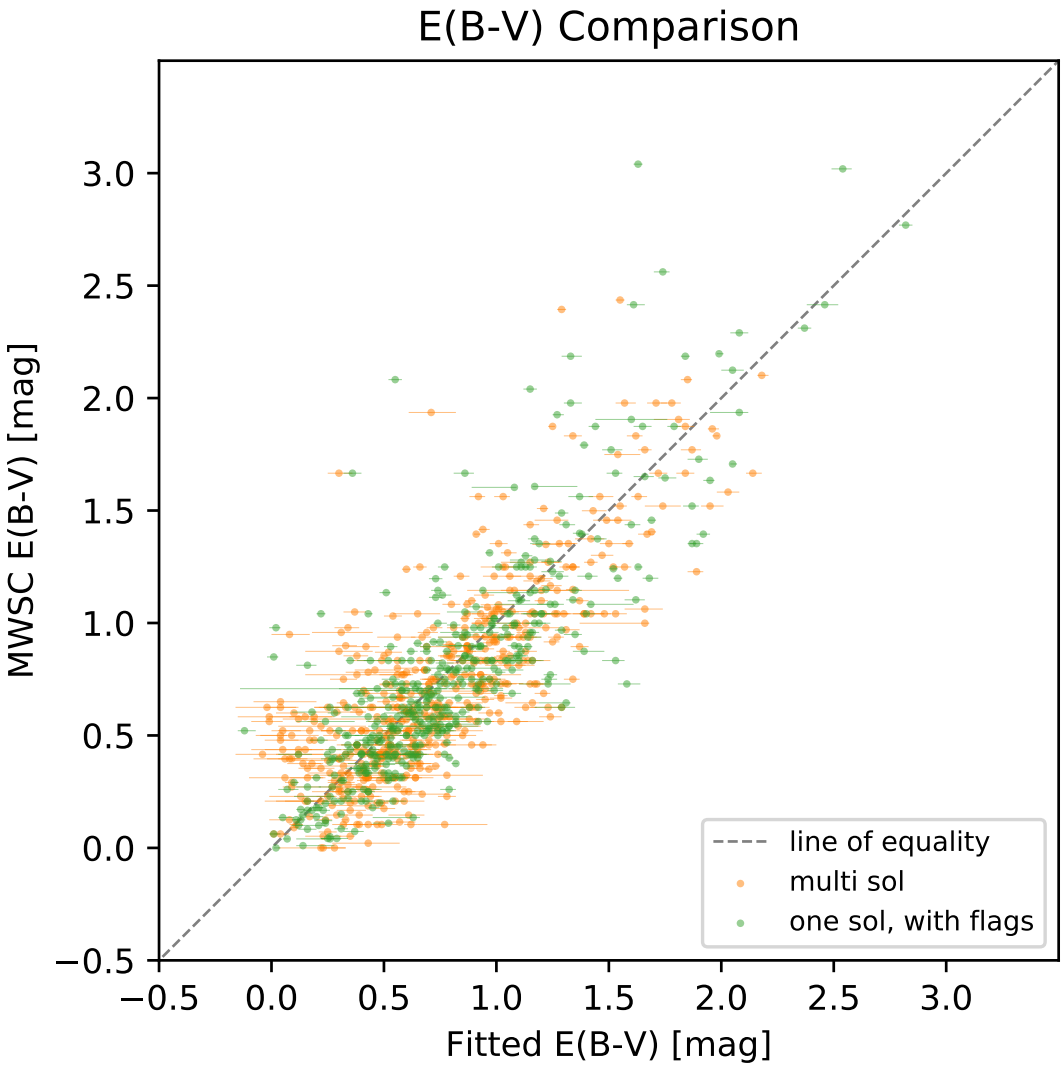

Figure B.4: Comparison of cluster reddenings derived from this work to those determined in MWSC. The dashed grey line indicates the line of equality. The single solution clusters with one or more flags are shown in green and one result from each of the multiple solutions clusters is shown in orange.

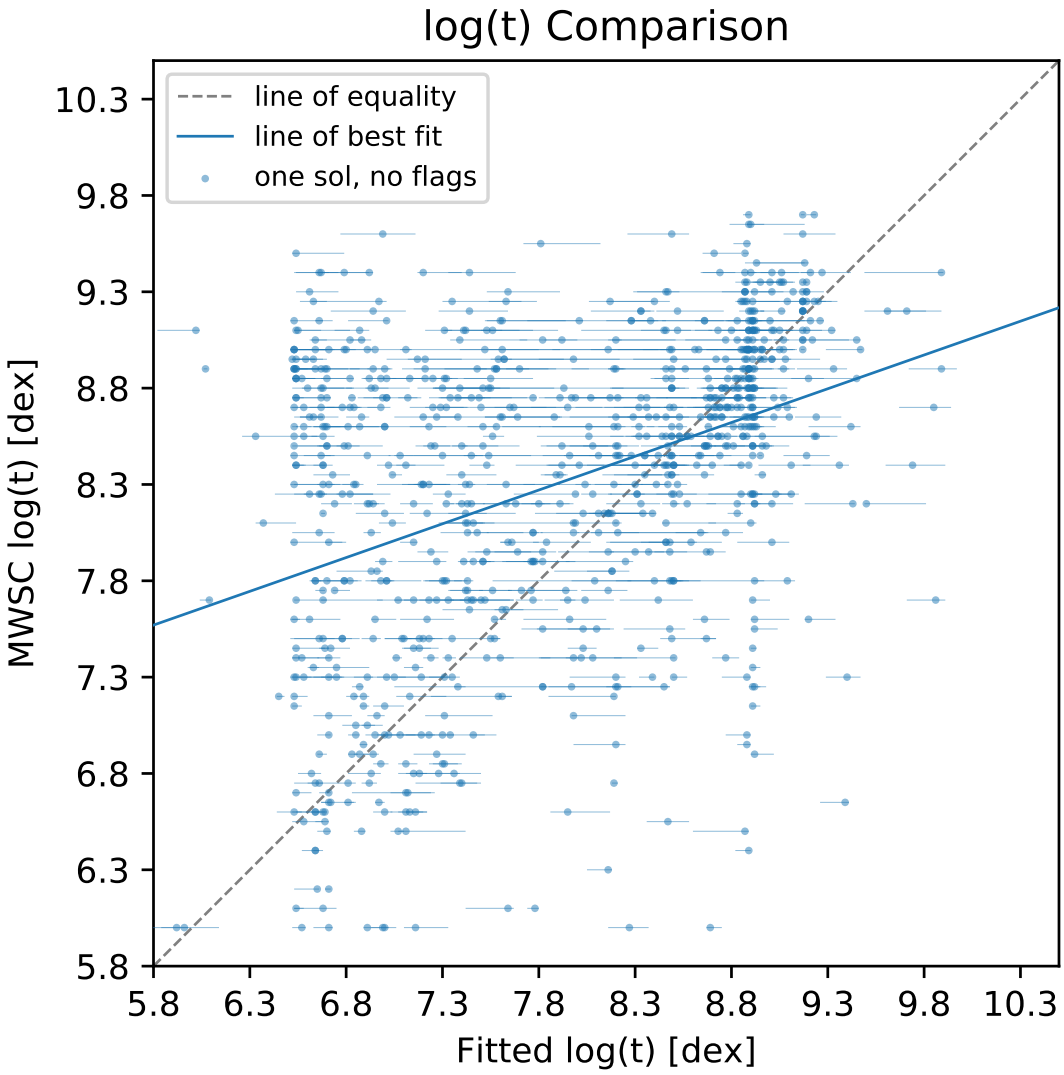

Figure B.5: Comparison of cluster ages from the MWSC to those derived in this work. The dashed grey line indicates the line of equality and the solid blue line shows the best fit to the single solution results with no flags. The single solution clusters with no flags are shown in blue.

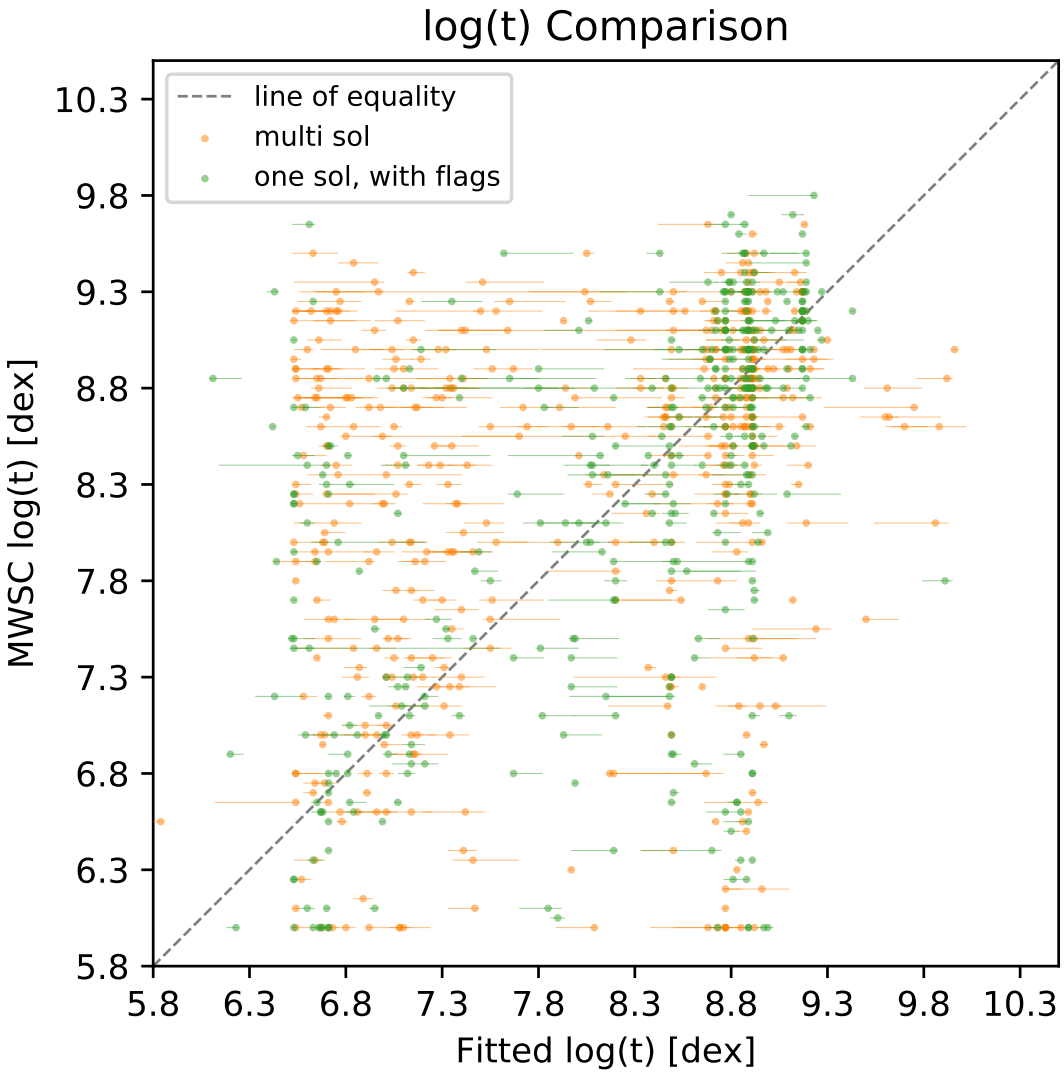

Figure B.6: Comparison of cluster ages from the MWSC to those derived in this work. The dashed grey line indicates the line of equality. The single solution clusters with one or more flags are shown in green and one result from each of the multiple-solution clusters is shown in orange.

B.2 Parameter comparisons with Bossini et al. (2019)

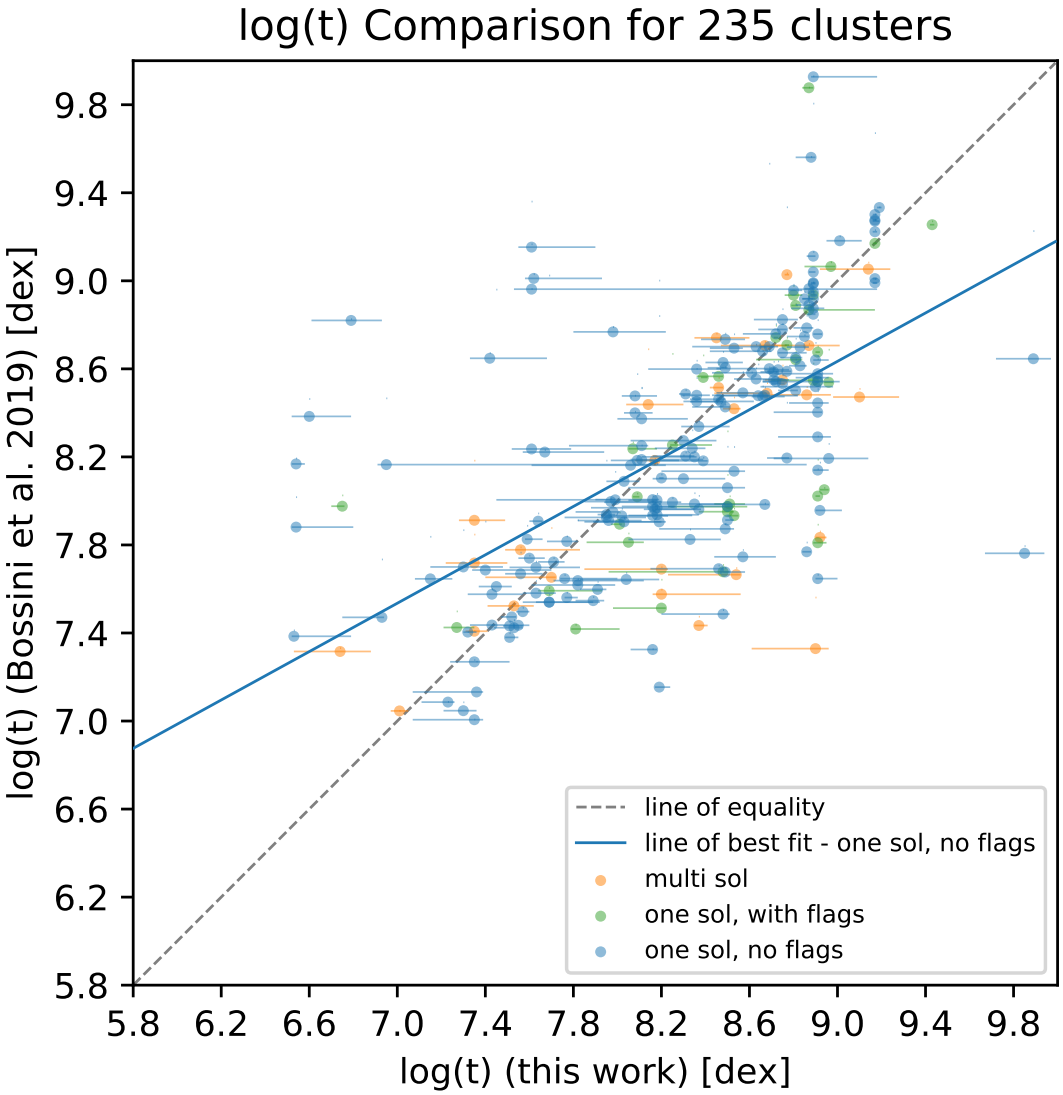

Figure B.7: Age comparison of 235 clusters in common with Bossini et al. (2019). The dashed grey line is the line of equality and the solid blue line is the line of best fit for the clusters in this work which have a single solution and no flags. The single solution clusters with no flags is represented by blue circles, the single solution clusters with flags is represented by green circles, and one result from each of the multiple solution clusters is represented by orange circles.

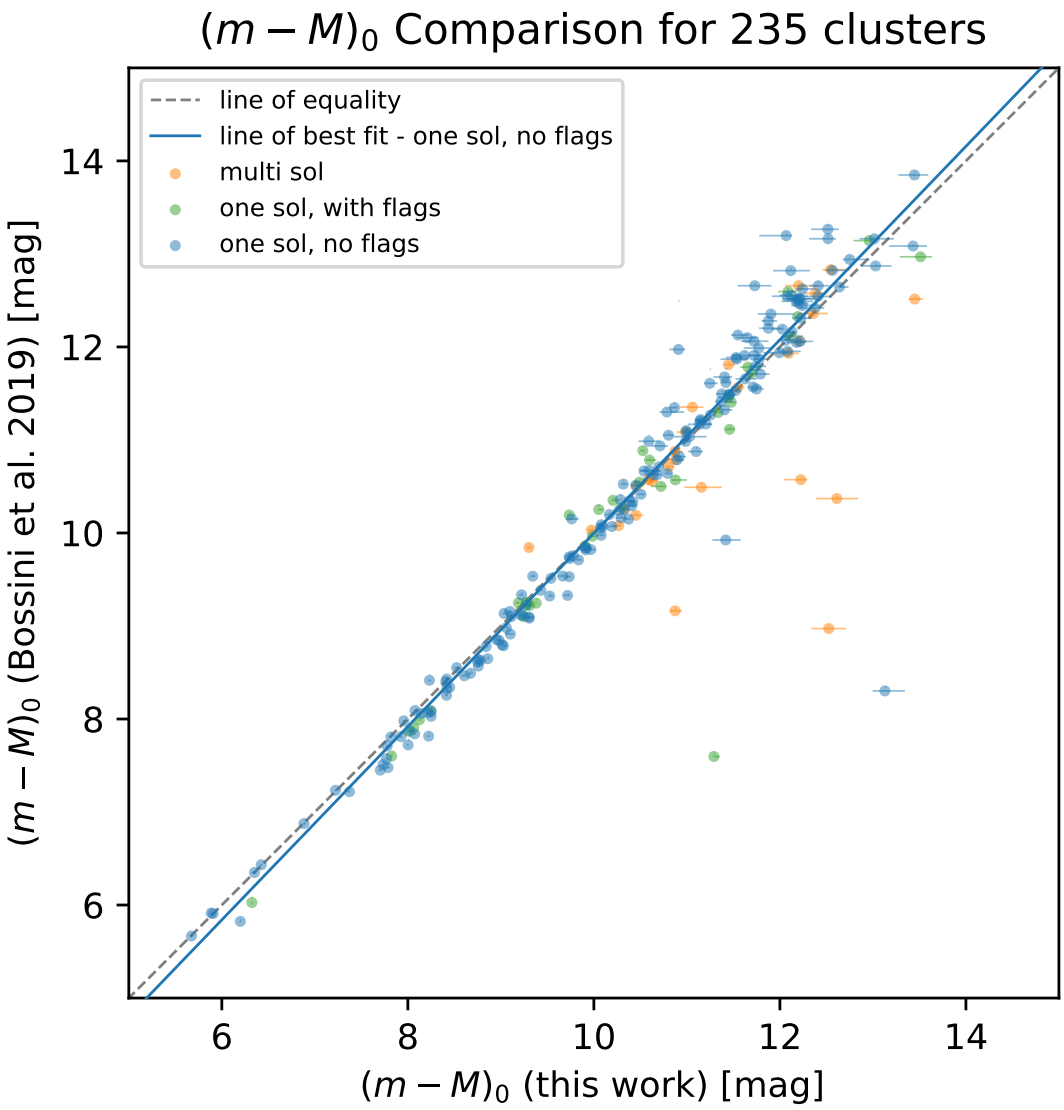

Figure B.8: Distance modulus comparison of 235 clusters in common with Bossini et al. (2019). The dashed grey line is the line of equality and the solid blue line is the line of best fit for the clusters in this work which have a single solution and no flags. The single solution clusters with no flags is represented by blue circles, the single solution clusters with flags is represented by green circles, and one result from each of the multiple solution clusters is represented by orange circles.

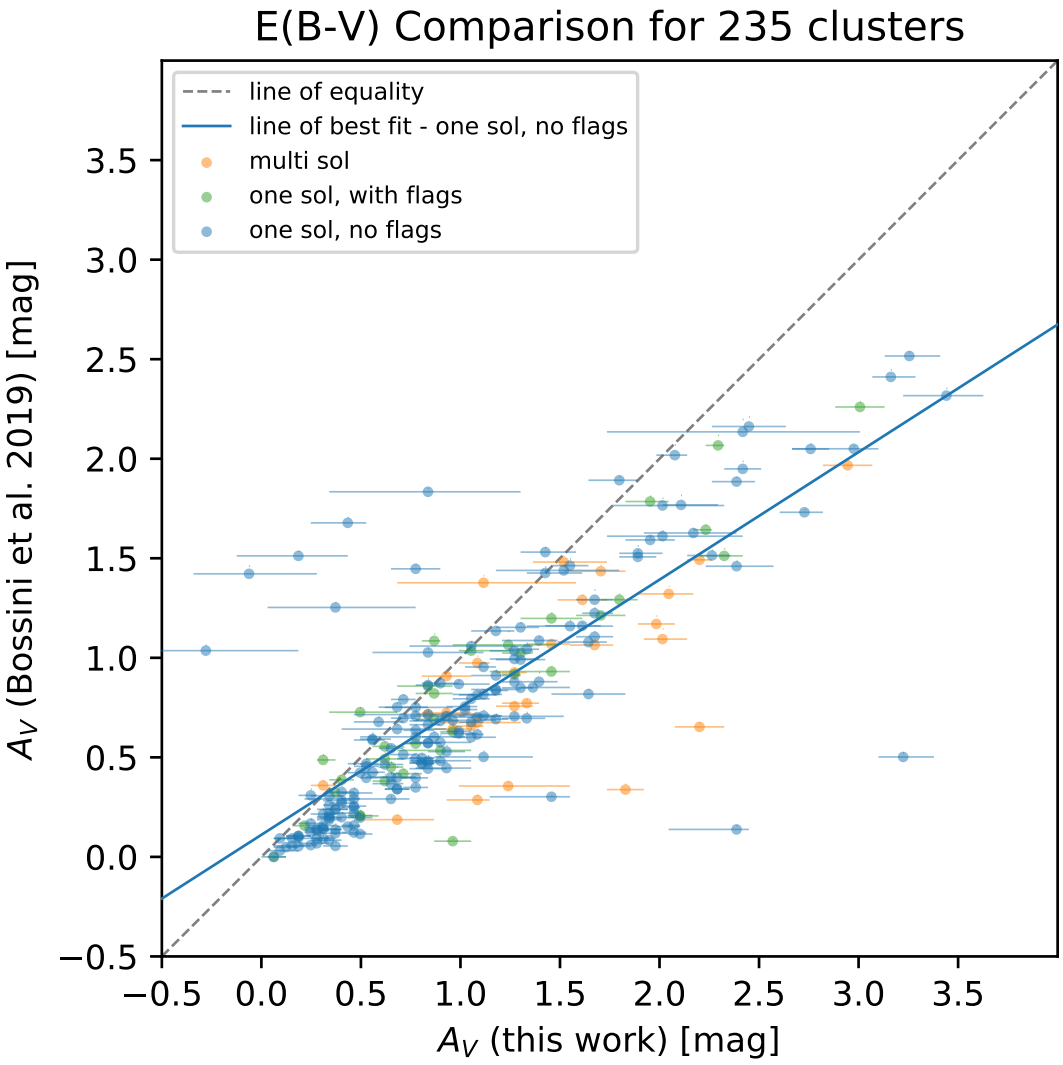

Figure B.9: E(B-V) comparison of 235 clusters in common with Bossini et al. (2019). The dashed grey line is the line of equality and the solid blue line is the line of best fit for the clusters in this work which have a single solution and no flags. The single solution clusters with no flags is represented by blue circles, the single solution clusters with flags is represented by green circles, and one result from each of the multiple solution clusters is represented by orange circles.

Publications

The following publications were used in this thesis:

- **Yen, S. X.**, Reffert, S., Schilbach, E., and Röser, S. (2019). 'Rediscovering open clusters in the Milky Way Star Clusters catalog with *Gaia* DR2.' To be submitted to A&A.
- **Yen, S. X.**, Reffert, S., Schilbach, E., Röser, S., Kharchenko, N. V. , and Piskunov, A. E. (2018). 'Reanalysis of nearby open clusters using *Gaia* DR1/TGAS and HSOY.' In A&A 615, A12.

I am also a co-author on the following publications, which are not included in this thesis:

- Piskunov, A. E., Just, A., Kharchenko, N. V., Berczik, P., Scholz, R.-D., Reffert, S., **Yen, S. X.** (2018). 'Global survey of star clusters in the Milky Way. VI. Age distribution and cluster formation history.' In A&A 614, A22.
- Kovaleva, D. A., Piskunov, A. E., Kharchenko, N. V., Röser, S., Schilbach, E., Scholz, R.-D., Reffert, S., **Yen, S. X.** (2017). 'Open star clusters in the Milky Way. Comparison of photometric and trigonometric distance scales based on Gaia TGAS data.' In A&A 606, L8.

References

Altmann, M., Roeser, S., Demleitner, M., Bastian, U., & Schilbach, E. 2017, A&A, 600, L4

Baade, D. 1983, A&AS, 51, 235

Becker, W. & Fenkart, R. 1971, A&AS, 4, 241

Bellagamba, F., Roncarelli, M., Maturi, M., & Moscardini, L. 2018, MNRAS, 473, 5221

Bica, E., Dutra, C. M., & Barbuy, B. 2003, A&A, 397, 177

Bonatto, C., Kerber, L. O., Bica, E., & Santiago, B. X. 2006, A&A, 446, 121

Bossini, D., Vallenari, A., Bragaglia, A., et al. 2019, A&A, 623, A108

Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127

Buder, S., Asplund, M., Duong, L., et al. 2018, MNRAS, 478, 4513

Bukowiecki, Ł., Maciejewski, G., Konorski, P., & Strobel, A. 2011, Acta Astron., 61, 231

Cabrera-Caño, J. & Alfaro, E. J. 1985, A&A, 150, 298

Caffau, E., Ludwig, H.-G., Steffen, M., Freytag, B., & Bonifacio, P. 2011, Sol. Phys., 268, 255

Caffau, E., Maiorca, E., Bonifacio, P., et al. 2009, A&A, 498, 877

Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018a, A&A, 618, A93

Cantat-Gaudin, T., Vallenari, A., Sordo, R., et al. 2018b, A&A, 615, A49

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Carraro, G. 2006, Bulletin of the Astronomical Society of India, 34, 153

Carraro, G., Baume, G., Seleznev, A. F., & Costa, E. 2017, Ap&SS, 362, 128

Carrasco, J. M., Evans, D. W., Montegriffo, P., et al. 2016, A&A, 595, A7

Castro-Ginard, A., Jordi, C., Luri, X., Cantat-Gaudin, T., & Balaguer-Núñez, L. 2019, A&A, 627, A35

Castro-Ginard, A., Jordi, C., Luri, X., et al. 2018, A&A, 618, A59

Chaboyer, B., Demarque, P., Kernan, P. J., Krauss, L. M., & Sarajedini, A. 1996, MNRAS, 283, 683

Chereul, E., Crézé, M., & Bienaymé, O. 1999, A&AS, 135, 5

Converse, J. M. & Stahler, S. W. 2008, ApJ, 678, 431

Coronado, J., Rix, H.-W., & Trick, W. H. 2018, MNRAS, 481, 2970

Danielski, C., Babusiaux, C., Ruiz-Dern, L., Sartoretti, P., & Arenou, F. 2018, A&A, 614, A19

Dias, W. S., Alessi, B. S., Moitinho, A., & Lépine, J. R. D. 2002, A&A, 389, 871

Dias, W. S., Monteiro, H., Caetano, T. C., & Oliveira, A. F. 2012, A&A, 539, A125

Dutra, C. M. & Bica, E. 2000, A&A, 359, 347

Dutra, C. M. & Bica, E. 2001, A&A, 376, 434

Dutra, C. M., Bica, E., Soares, J., & Barbuy, B. 2003, A&A, 400, 533

Evans, D. W., Riello, M., De Angeli, F., et al. 2018, A&A, 616, A4

Fan, X., Burstein, D., Chen, J.-S., et al. 1996, AJ, 112, 628

Froebrich, D., Scholz, A., & Raftery, C. L. 2007, MNRAS, 374, 399

Fürnkranz, V., Meingast, S., & Alves, J. 2019, A&A, 624, L11

Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018a, A&A, 616, A10

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018b, A&A, 616, A1

Gaia Collaboration, Brown, A. G. A., Vallenari, A., Prusti, T., et al. 2016, A&A, 595, A2

Gaia Collaboration, Evans, D. W., Riello, M., De Angeli, F., et al. 2017, A&A, 600, A51

Gaia Collaboration, van Leeuwen, F., Evans, D. W., De Angeli, F., et al. 2017a, A&A, 599, A32

Gaia Collaboration, van Leeuwen, F., Vallenari, A., Jordi, C., et al. 2017b, A&A, 601, A19

Glushkova, E. V., Koposov, S. E., Zolotukhin, I. Y., et al. 2010, Astronomy Letters, 36, 75

Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27

Janes, K. & Adler, D. 1982, ApJS, 49, 425

Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48

Kharchenko, N. V. 2001, Kinematika i Fizika Nebesnykh Tel, 17, 409

Kharchenko, N. V., Piskunov, A. E., Röser, S., Schilbach, E., & Scholz, R.-D. 2005a, A&A, 440, 403

Kharchenko, N. V., Piskunov, A. E., Röser, S., Schilbach, E., & Scholz, R.-D. 2005b, A&A, 438, 1163

Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R.-D. 2012, A&A, 543, A156

Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R.-D. 2013, A&A, 558, A53

Koposov, S. E., Glushkova, E. V., & Zolotukhin, I. Y. 2008, A&A, 486, 771

Kos, J., de Silva, G., Buder, S., et al. 2018, MNRAS, 480, 5242

Kovaleva, D. A., Piskunov, A. E., Kharchenko, N. V., et al. 2017, A&A, 606, L8

Krone-Martins, A. & Moitinho, A. 2014, A&A, 561, A57

Lada, C. J. & Lada, E. A. 2003, ARA&A, 41, 57

Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2

Lindegren, L., Lammers, U., Bastian, U., et al. 2016, A&A, 595, A4

Lynga, G. 1982, A&A, 109, 213

Malo, L., Doyon, R., Lafrenière, D., et al. 2013, ApJ, 762, 88

Meingast, S. & Alves, J. 2019, A&A, 621, L3

Mermilliod, J.-C. 1993, in Databases for Galactic Structure, ed. A. G. D. Philip, B. Hauck, & A. R. Upgren, 27–34

Moffat, A. F. J. 1972, A&AS, 7, 355

Monteiro, H., Dias, W. S., & Caetano, T. C. 2010, A&A, 516, A2

Moraux, E. 2016, in EAS Publications Series, Vol. 80-81, EAS Publications Series, 73–114

Netopil, M., Paunzen, E., & Carraro, G. 2015, A&A, 582, A19

Newville, M., Stensitzki, T., Allen, D. B., et al. 2016, Lmfit: Non-Linear Least-Square Minimization and Curve-Fitting for Python, Astrophysics Source Code Library

O'Donnell, J. E. 1994, 422, 158

Palmer, M., Arenou, F., Luri, X., & Masana, E. 2014, A&A, 564, A49

Perren, G. I., Vázquez, R. A., & Piatti, A. E. 2015, A&A, 576, A6

Platais, I., Kozhurina-Platais, V., & van Leeuwen, F. 1998, AJ, 116, 2423

Röser, S., Demleitner, M., & Schilbach, E. 2010, AJ, 139, 2440

Röser, S. & Schilbach, E. 2019, A&A, 627, A4

Röser, S., Schilbach, E., & Goldman, B. 2016, A&A, 595, A22

Röser, S., Schilbach, E., & Goldman, B. 2019, A&A, 621, L2

Ruprecht, J., Balazs, B. A., & White, R. E. 1981, Catalogue of star clusters and associations (Akadémiai Kiadó, Budapest)

Sanders, W. L. 1971, A&A, 14, 226

Sarro, L. M., Bouy, H., Berihuete, A., et al. 2014, A&A, 563, A45

Schmeja, S., Kharchenko, N. V., Piskunov, A. E., et al. 2014, A&A, 568, A51

Scholz, R. D., Kharchenko, N. V., Piskunov, A. E., Röser, S., & Schilbach, E. 2015, A&A, 581, A39

Schönrich, R., McMillan, P., & Eyer, L. 2019, MNRAS, 487, 3568

Schultz, G. V. & Wiemer, W. 1975, A&A, 43, 133

Seleznev, A. F. 2016, MNRAS, 456, 3757

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Soubiran, C., Cantat-Gaudin, T., Romero-Gómez, M., et al. 2018, A&A, 619, A155

Tang, S.-Y., Pang, X., Yuan, Z., et al. 2019, ApJ, 877, 12

van Leeuwen, F. 2007, A&A, 474, 653

Vasilevskis, S., Klemola, A., & Preston, G. 1958, AJ, 63, 387

Vogt, N. 1971, A&A, 11, 359

von Hippel, T., Jefferys, W. H., Scott, J., et al. 2006, ApJ, 645, 1436

Yen, S. X., Reffert, S., Schilbach, E., et al. 2018, Astronomy and Astrophysics, 615, A12

Acknowledgments

These last four and a bit years have been wonderfully eye-opening and truly challenging. In these last pages and with these last words, I would like to thank everyone who has helped and supported me throughout my PhD.

First and foremost, I would like to thank my supervisor, Sabine Reffert, for all her help, patience, and support over the last four and a half years and for the opportunity to pursue a PhD. I could not have asked for a more supportive and caring supervisor. Thank you for always being available to answer questions and for your dedication to our project. I had a great time learning about open clusters with you.

I would also like to thank my unofficial, secondary supervisors Siegfried Röser and Elena Schilbach for lending their expertise over the years. Their quick knowledge on parameters for specific open clusters never ceases to amaze me. Thank you so much for your time, dedication, and patience. I appreciate all of your insightful ideas.

Another special thanks also goes to Nina Kharchenko and Anatoly Piskunov for all their help and advice during the beginning of my PhD. Perhaps most importantly, thanks for putting together the MWSC – without it, my PhD would have been a lot tougher!

I would also like to extend gratitude towards my IMPRS thesis committee members, Anna Pasquali and Andreas Quirrenbach, for making sure I was successfully on track these last four years. To Eva Grebel, thank you so much for agreeing to be the second referee of my thesis and for leading the SFB881: The Milky Way System. To Geneviève Parmentier and Björn Malte Schäfer, thank you for agreeing to be my additional examiners.

A big thank you goes to Stefan for translating the abstract into German and to Holly and Vera for proofreading parts of this thesis. Thanks to ESA and the whole Gaia team for all the wonderful data to work with.

Next, I would like to thank my LSW family. Words cannot describe how lucky I feel to have ended up pursuing my PhD here and with all of you. The past four and a half years seem to have gone by in a blink of an eye, but they are all so full of wonderful memories with each of you. First, my officemate and fellow Stef, Stefan – I seriously could not have asked for a better officemate (long live the Stef^2 office) to endure the PhD with. Thank you for all of your help and support over the years and especially in the last few weeks. Und danke, dass du mich dazu gebracht

haben, mehr Deutsch zu sprechen und zu schreiben. The rest of my LSW family: Marcelo (I remember the day we first met in my office. I thought, "Who is this guy with these crazy glasses?" I am so glad to have you and Stefan as my PhD buddies. I'll never forget the Steffi songs.), Rob (My older brother from another mother. Thanks for always lending an ear, for all of the advice, and for always fixing my bike.), Paul (My coolest buddy! Thanks for always being around and up for anything! Thanks for keeping my head warm ;)), Vera (My anime buddy and fellow dog lover – thank you for everything!!!), Mauricio (Thank you for all your wisdom and help, especially during the beginning of the PhD. I couldn't have done it without you!), Sepideh (Thanks for all the support, encouragement, and hugs over the years.), Katja (Thanks for translating my German mail and for the flight in the glider!), Stephan (Thanks for the game night invites!), Theo (Perpendicular, perpendicular!!! ;) Thanks for the blue Pikachu!), Hotti (OMG, thank you for teaching me the ssh shortcut and for managing lx33!), Dane (You are the absolute sweetest. Thanks for being on my side when the others were not!), Kai (Thanks for all the laughs and puzzling!), and the rest of LSW – thanks for being you and for making LSW such a wonderful working environment. Otmar, thank you for being so quick to solve any computer problems!

To my MPIA officemates, Mayte, Matthias, Aida, Neven, Manuel, Miriam (and Micah) – thanks for adopting me and all of the coffee. I have so many happy memories with each of you. Mayte – Thanks for always listening and for being such a great travel buddy. Thanks for also motivating me to be the best version of myself! Matthias – Thanks for answering all my Python questions and showing me the wonderful world of anime beyond Pokèmon and Studio Ghibli. Aida – Thanks for being such a wonderful friend and for all the support, especially over these last few weeks. I couldn't have asked for a better thesis writing buddy. Neven – You are truly one of a kind! The last few months in Heidelberg have not been as colorful without you. Manuel – My Tesla, dino, and skiing buddy, thanks for all the wonderfully long chats over coffee. Miriam – You are the sweetest! Don't work too hard! Micah – Thanks for being one of my North American buddies and for all the laughs on our Black Forest trip! I hope you and Aida will have my room in Leiden ready!

To my dear Heidelberg friends, Rosalie, Sara, Gabriele, Hans, Victor, Reza, Diane, Qian, and others I've missed – thanks for your friendships and all the wonderful memories and adventures. I cannot wait to see where life will take us!

To my IMPRS dinner crew (Alex, Christos, Matthias, Neven, Sabina, and Yulong), thanks for all the laughs and crazy conversations over our weekly Thursday dinners. To my IMPRS generation, thanks for the super fun retreats. To Christian and Huong, thank you so much for all you do and creating such a wonderful student program here in Heidelberg!

To my former supervisors, Neal, Regina, and Adam, you all are an inspiration and I would not be where I am today without you. Thanks for all the scientific lessons and all the encouragement. To my IfA REU & LEAPS buddies, thanks for

amazing summers and the lifelong friendships. Philip, aka the brightest person I know, thanks for all your help and for pushing me to do better.

To my friends back home, Autumn, Andrew, Kaela, Holly, Timmy, Yannie, and Ryan – you guys are the best friends I could have ever asked for! <3 I cannot tell you how much it means to me that we are still so close, despite all these years and the distance. Thank you for the long FaceTime, Skype, Google Hangout calls about all of life's ups and downs, for your advice with science and work, and for the warmest welcome home whenever I visited. I cannot wait to see where life takes us. To my Aunt Nancy and Uncle Don – thanks for all the life lessons (and ski lessons!), love, and support.

Last, but certainly not least, my family. There are not enough words to describe how I feel and how much you all mean to me. Thank you for the constant love and support. It has not been easy spending so much time away from home, but thank goodness for FaceTime! Mommy and Daddy, thank you for all the sacrifices you have made for Brandon and I. We appreciate everything you have done for us. Brandon, my favorite and only little brother, you amaze me. I am so proud of all you have accomplished and cannot wait to see what more the future has for you. I love you all and I cannot wait to see you all very soon!