The VVV Open Cluster Project II. Near-infrared sequences of 37 open clusters on eight-dimensional parameter space

K. Peña Ramírez,1⋆ L.C. Smith,2 S. Ramírez Alegría,1 A.-N. Chené,3 C. González-Fernández,2
P. W. Lucas,4 and D. Minniti5,6

1 Centro de Astronomía (CITEVA), Universidad de Antofagasta, Av. Angamos 601, Antofagasta, Chile.
2 Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK.
3 Gemini Observatory/NSF’s NOIRLab, 670 N. A’ohoku Place, Hilo, Hawai’i, 96720, USA.
4 Centre for Astrophysics Research, University of Hertfordshire, Hatfield AL10 9AB, UK.
5 Departamento de Ciencias Físicas, Universidad Andrés Bello, República 220, 8320000, Santiago, Chile.
6 Vatican Observatory, V00120 Vatican City State, Italy.

Accepted 2022 May 4. Received 2022 May 4; in original form 2022 March 23

ABSTRACT

Open clusters are key coeval structures that help us understand star formation, stellar evolution and trace the physical properties of our Galaxy. In the past years, the isolation of open clusters from the field has been heavily alleviated by the access to accurate large-scale stellar parallaxes and proper motions along a determined line of sight. Still, there are limitations regarding their completeness since large-scale studies rely on optical wavelengths. Here we extend the open clusters sequences towards fainter magnitudes complementing the Gaia photometric and astrometric information with near-infrared data from the VVV survey. We performed a homogeneous analysis on 37 open clusters implementing two coarse-to-fine characterization methods: extreme deconvolution Gaussian mixture models coupled with an unsupervised machine learning method on 8-dimensional parameter space. The process allowed us to separate the clusters from the field at near-infrared wavelengths. We report an increase of ∼47% new member candidates on average in our sample (considering only sources with high membership probability p≥0.9). This study is the second in a series intended to reveal open cluster near-infrared sequences homogeneously.

Key words: astronomical databases: miscellaneous – methods:data analysis – stars: evolution – open clusters and associations: individual: list.

1 INTRODUCTION

With the advancement of large-scale photometric and state-of-the-art astrometric instruments, the field of open clusters has been revisited in recent years with significant progress mainly in two aspects: accounting for large cluster samples while ensuring a uniform spatial distribution of the clusters and homogeneity in terms of observations and data processing. Building upon the first large-scale seminal works on open clusters (Dias et al. 2002; Kharchenko et al. 2013) and the Gaia satellite releases (Gaia DR1/DR2/EDR3, Gaia Collaboration et al. 2016, 2018, 2020), several authors have revisited the characterization of open clusters following a plethora of techniques leading to the most complete and accurate cluster membership determination scenario to date (Cantat-Gaudin et al. 2018; Castro-Ginard et al. 2018; Cantat-Gaudin et al. 2019; Castro-Ginard et al. 2019, 2020; Cantat-Gaudin & Anders 2020; Cantat-Gaudin et al. 2020; Olivares et al. 2019; Miret-Roig et al. 2019; Galli et al. 2020; Tarricq et al. 2022; Dias et al. 2022, among others). What is more, recently, Jackson et al. (2020, 2021) has built upon the kinematic space, including spectroscopic information from the GES internal data releases 5 and 6, to assign membership probabilities on more than 60 open clusters.

There is, however, a front in which the field is still lacking a proper analysis. Since most of the large-scale studies have been done at optical wavelengths, the near-infrared picture of the open clusters have been analyzed either on specific single cluster studies (e.g. Bonatto et al. 2006; Rangwal et al. 2019; Bisht et al. 2020a,b) or under the scrutiny of specific stellar populations (Mauzerah et al. 2011; Borissova et al. 2016, 2018, 2020; de la Fuente et al. 2021, among others). In the case of studies with a moderate number of clusters (Bica & Bonatto 2005; Santos-Silva & Gregorio-Hetem 2012; Kharchenko et al. 2016), the available near-infrared data was the Two Micron All-Sky Survey (2MASS, Skrutskie et al. 2006). In dense fields, such studies lead to source confusion issues.

The work presented here follows on from Peña Ramírez et al. (2021, hereafter PR21) where refined open cluster cen-
sus via Gaia DR2, 2MASS data, and high-quality multi-epoch near-infrared data from the VISTA Variables in the Vía Láctea Survey (VVV, Minniti et al. 2010) was combined to provide membership probabilities for six open clusters based on their three-dimensional kinematics and their near-infrared photometry. Here, we refine the methodology and extend the study to a new sample of 37 open clusters on the VVV footprint. We present the dataset used for the membership analysis in Section 2. The methodology implemented to select the cluster members and the derivation of the membership probabilities is covered in Section 3. Section 4 presents our main findings unveiling the open cluster near-infrared sequences of 37 open stellar clusters down to $K_s \sim 14.0$ mag, and a summary and conclusions are outlined in Section 5. Appendix A shows a set of plots per each of the clusters studied.

## 2 OBSERVATIONAL DATA

### 2.1 Open cluster sample

We selected a set of 37 clusters out of the list of clusters recovered from the literature and revisited by Cantat-Gaudin et al. (2018) and Cantat-Gaudin & Anders (2020) using an unsupervised membership assignment ensemble (see Section 3). These clusters fall within the VVV footprint, with a known considerable distance and age span, and were located at first instance through the coordinates and distances reported by Cantat-Gaudin & Anders (2020). The surveys involved are all-sky, and therefore our spatial coverage within the VVV footprint is homogeneous, and the photometric ones are only constrained by magnitude completeness. Table 1 presents the cluster names and their fundamental parameters recovered from the literature.

### 2.2 Multi-dimensional data sets

We kept our original purpose as outlined in PR21: to identify each open cluster sequence with the most extensive dynamical range possible in the near-infrared. We have access to the brightest sources using 2MASS photometry in the $J$, $H$ and $K_s$ bandpasses. For fainter targets we exploited the dynamic range of the VVV survey in the $J$, $H$ and $K_s$ bandpasses. The VVV survey, and its temporal and spatial extension the VVX, are ESO Public Surveys targeting the inner disc and the bulge of the Milky Way. The VVV and VVX utilise the VISTA telescope and its main imager, the VISTA Infrared Astrometric Catalogue version 2 (VIRAC2; Emerson et al. 2006), to obtain multi-epoch, infrared photometry. Observations taken with VISTA are reduced at the Cambridge Astronomy Survey Unit (CASU; http://casu.ast.cam.ac.uk/) as part of the VISTA Data Flow System (VDFS; Irwin et al. 2004). The latest data release uses v1.5 of their pipeline. At the bright end the VVV survey manages $\sim$[11.0, 11.0, 11.5, 12.0, 11.0] mag in $[Z, Y, J, H, K_s]$, at which point detector non-linearity brings photometric errors over 0.1 mag. At the faint end it reaches $\sim$[18.5, 18.0, 17.5, 16.5, 16.0] mag in those same bandpasses, at which point sky noise is the limiting factor for a typical observing sequence. More details about the performance and photometric properties of the data can be obtained in González-Fernández et al. (2018).

Regarding astrometric measurements, we also combined data from two surveys. For the brightest sources, the near-infrared photometric data sets were combined with astrometric information using a pre-computed nearest-neighbor, proper motion aware, cross-match between 2MASS and the latest release of Gaia\(^1\) to retrieve $J$, $H$, and $K_s$ magnitudes from 2MASS and the 5-parameter astrometric solution from Gaia EDR3 (right ascension, declination, proper motions in right ascension and declination, and parallaxes).\(^2\)

For the faintest sources, we relied on preliminary data from the VVV Infrared Astrometric Catalogue version 2 (VIRAC2; Smith et al. in prep; see Smith et al. 2018 for details of v1). VIRAC2 used PSF fitting to obtain astrometry and photometry of sources in VVV observations and VVVX observations of the original VVV area. The preliminary VIRAC2 data used in this work was calibrated astrometrically against Gaia DR2 (taking into account Gaia proper motions and parallaxes), and photometrically against the catalogues of Alonso-García et al. (2018). Individual detections were mapped to unique stars using a bespoke pipeline, which involved an iterative process of 5-parameter astrometric solution fitting and remapping until convergence.

As in PR21, our final working product is an astrophysical dataset in eight dimensions (five astrometric parameters and three photometric ones, $J$, $H$, and $K_s$) for each cluster covering an area of five times their published $r_{50}$ value\(^2\). In the 2MASS/VVV magnitude overlap, sources present in both catalogs are combined photometrically with an optimal inverse variance weighting. In the VVV/Gaia overlap the astrometric parameters of duplicated stars were also combined using an inverse variance weighted average. Since the magnitude range is broad, there is a significant difference in precision between the sources, from the brightest to the faintest. On the bright end, the nominal uncertainty reaches the Gaia precisions of 0.02–0.03 mas (yr\(^{-1}\)) in parallax and proper motions (Gaia Collaboration et al. 2020), while at $K_s \sim 15.0$ mag, the uncertainties are under $\sim$1.0 mas (yr\(^{-1}\)) in parallax and proper motions.

## 3 METHODOLOGY

The analysis used to determine membership probability here is similar to that described in PR21; however, the cluster tagging and their sequence determination have changed in some key aspects to tackle a broader mix of cluster ages, distances, and shapes. In our previous work, we applied the Gaussian Mixture Model (GMM; e.g. Everitt et al. 2011) technique to isolate the clusters from the field as a first instance. This technique is based on the assumption that the star’s distribution within an overdensity can be described by a superposition of multivariate Gaussian distributions (de Souza et al. 2017). Nevertheless, the technique does not account for the uncertainties on the parameter space at work. As noted by Jaehnig et al. (2021) it is ideal for a GMM to fit an intrinsic, underlying distribution rather than noisy discrete data. In that line, we implemented the method developed by Bovy

---

1. Using the Q3C software: Koposov & Bartonov (2006).

2. The $r_{50}$ value from Cantat-Gaudin & Anders (2020) is the radius (in degrees) from the cluster center that encompasses 50% of the members identified by the authors.
et al. (2011) in which the observed measurements are de-
convolved from their uncertainties and fitting the intrin-
sic distribution with a gaussian mixture model. The Ex-
tre Deconvolution Gaussian Mixture Model (XDGMM) has been
implemented successfully for the identification and char-
erization of specific open clusters (Oliveiras et al. 2019; Price-
Whelan et al. 2019; Jaehnig et al. 2021), globular clusters
(Vasiliev & Baumgardt 2021) and even supernova and host
galaxy populations (Holoien et al. 2017).

The clusters were initially identified purely from their pub-
lished coordinates, proper motions, and parallaxes as on
Cantat-Gaudin & Anders (2020). All the subsequent analy-
ces covered the entire area of study (5x 50) per cluster.
Once spatially identified each cluster (, δ parameter space),
the XDGMM was implemented in the 3-dimensional pa-
rameter space of proper motions and parallaxes (μ, δ, π).
In the pre-processing stage, we removed clear outlier sources
in the proper motion space by discarding sources with μ or
μ measurements more than 10σ from the cluster proper move-
ment. The central parallax values and their uncertainties
were recorded for each cluster. At this stage, we also discard
the sources with photometric uncertainties above the 0.5 mag-
itudes in J and K, to remove obvious spurious sources.
We used the ASTROM (Vanderplas et al. 2012) and Bovy et
al. (2011) implementations developed by (Holoien et al. 2017)3.
XDGMM takes into account the full covariance matrix repre-
sentation on μ, μ, and π which we considered ten Gaus-
sian components (and their respective covariance matrices)
to describe the above-mentioned parameter space. This
approach gives us results insensitive to the exact number of
Gaussians, as long as it is large enough to disentangle the
background from the members.

For selecting the gaussian component that traces the open
cluster, we followed the prescription given in Jaehnig et al.
(2021) to obtain the differential entropy information metric
(Ahmed & Gokhale 1989). It is related to the components co-
variance as in PR21 but extended to account for the dimen-
sion of the parameter space, therefore giving us a measure of
how compact a distribution is within a volume. The com-
ponent with the lowest differential entropy, i.e., the most
compact Gaussian component, was selected as the cluster com-
ponent. XDGMM resampling calculates the individual member-
ship probabilities via bootstrap. We resample component
assignments for each star for a total of 200 iterations.
Only sources with membership probability (p > 0.9) to the selected
cluster component were considered to our refinement stage.

3 https://github.com/holoien/XDGMM
At this point, the coherent motion of the conglomerates is assured since the input data was already clustered in the 3-dimensional proper motion-parallax parameter space ($\mu_\alpha$, $\mu_\delta$, $\pi$). Therefore, our refinement stage was focused on group stars according to their spatial and photometric distribution using the $J$, $H$, $K_s$ photometry, and the spatial positions. To do so, we employed the Unsupervised Photometric Membership Assignment in Stellar Clusters UPMASK (Krone-Martins, A. & Moitinho, A. 2014) on each cluster component. This approach has been successfully applied to identify and characterize open clusters (Cantat-Gaudin et al. 2018, 2019; Peña Ramírez et al. 2021) and also accounts for the uncertainties in the explored parameter space. We used the K-MEANS clustering (Forgy 1965) partition algorithm with a large $k$ to the dataset size guaranteeing at least 25 objects per group. For testing whether the distribution of stars within each group is more concentrated than what is expected for a random fluctuation in a uniform distribution, we use the total length of a minimum spanning tree (e.g. Graham & Hell 1985) and we iterate 100 times per cluster field. At each iteration, the photometry data were randomly sampled from the probability distribution function of each star’s positional parameters while taking into account uncertainties for each variable. The clustering score (i.e., membership probabilities) for a given source is derived directly from the number of iterations during which it was a member of a concentrated group and can also be interpreted as a membership probability.

Our final membership probabilities are then based on the 8-dimensional parameter space ($\alpha$, $\delta$, $\mu_\alpha^*$, $\mu_\delta$, $\pi$, $J$, $H$, $K_s$) information of each star and their associated nominal uncertainties. The figures in Appendix A show the near-infrared sequence for all the explored clusters, together with their spatial distributions and proper motions. The full membership list per cluster is available in electronic form (see Section 6).

4 DISCUSSION

4.1 Cluster parameters

As pointed out in PR21, data-driven and unbiased membership probabilities are required to derive a reliable list of cluster members and pursue further studies. High precision characterization of star cluster’s fundamental parameters, such as age, distance, reddening, and total mass, depends on the quality of the cluster membership determination. Therefore we need datasets that are as diverse as they are accurate. The decontamination of background interlopers must ensure that no additional biases are imposed and that whatever cluster members remain are robust and representative of the cluster itself.

Using only sources with a membership probability larger than or equal to 90%, we have re-derived the median spatial position, proper motion, and parallax for each cluster. We used the mean absolute deviation (from the median), or MAD values (Feigelson & Jogesh 2012), to estimate the dispersion of those parameters. The offset between our recalculated central positions and those from literature are minimal and range from -0.04 to 0.04 arcmin. We obtained average proper motion values that agree within the uncertainties with the values published in the literature. The same applies to the parallax values.

Figure 1. Number of high probability members (membership probabilities $p \geq 90\%$) per studied cluster in reverse alphabetical order. The black points represent the number density of members from the literature (Cantat-Gaudin & Anders 2020). The red points represent the number densities we are reporting here. Horizontal axis in logarithmic scale.

Due to the levity of interstellar extinction in the near-infrared, we can in this study map the cluster sequences better than the previous work based on optical data. From the 37 clusters analyzed, the literature reports 4435 high probability members. Here we surpass that number, reaching a total of 9357 high probability members. Therefore, we report an increase of $\sim 47\%$ new candidate members on average in our sample. This number is in total agreement with the increase of new member candidates reported in PR21. The improvement in quality of the near-infrared sequences of the clusters is evident compared with the available near-infrared passbands of the 2MASS survey. Figure 1 shows the difference in the number of high probability members on the studied clusters. As can be seen, most of the clusters show an increase in their populations except for two clusters (NGC 6134 and NGC 6192, where we identify less sources at intermediate magnitudes but increasing the number of highly probable members towards fainter magnitudes). It is important to note that the 90% threshold compared with literature members relies on fairly similar methodological procedures. We present in Figure 2 the color-magnitude diagrams of the clusters with the highest and lowest change in number density of high probability members. For the most impacted clusters, we report 348 new high probability members for UFGM 2, whereas for NGC 4349 at near-infrared wavelengths, we change from 11 highly probable members to 316. On the contrary, for ASCC 88, we report an addition of only 11 high probability members, and for Ruprecht 130, the addition of 36 new high probability members on a very diluted cluster sequence. All the cluster sequences are available on Appendix A.

For each cluster, we compare the absolute sequences in
Figure 2. $K_s$ vs. $J-K_s$ color-magnitude diagrams of the selected open stellar clusters. UFMG 2 and NGC 4349 are the clusters most highly impacted by this study (largest number of new highly probable members). The opposite for ASCC 88 and Ruprecht 130. Black dots correspond to the Cantat-Gaudin & Anders (2020) members with reported probability $p \geq 90\%$, whereas the high probability members presented here are represented with red circles. Text on the plots shows the cluster name.

the VISTA photometric system to the models of PARSEC-COLIBRI (Marigo et al. 2017). We selected the model corresponding to the age derived by Cantat-Gaudin et al. (2020). We applied the distance modulus and the reddening vector based on the distance and the median extinction $A_V$ from the same study. The reddening vector was transformed to $A_{K_s}$ using $A_{K_s}/A_V = 0.11802^{\pm 1}$, assuming the total-to-selective extinction ratios $R_J$ and $R_{K_s}$ of González-Fernández et al. (2018) for VISTA data. Similarly, we have placed the model tied to the cluster parameters reported by (Kounkel et al. 2020).

Our reference studies, Cantat-Gaudin et al. (2020) and Kounkel et al. (2020) used artificial neural networks to derive ages, distances and extinctions per cluster. This technique maps the input observables to the target output quantities (age, extinction, and distance modulus/log d) through a series of nodes. The nodes are arranged in layers that communicate among themselves through non-linear functions. The mentioned studies differ in the artificial neural network architecture, i.e., the way the nodes and the layers are designed refers to the neural network architecture. In the case of Cantat-Gaudin et al. (2020) they used a rectified linear unit (ReLU, see their Figure 2), whereas the work of (Kounkel et al. 2020, see their Appendix A) uses a convolutional artificial neural network based on the MNIST model (Hoffer et al. 2018). Their trained model (Auriga\(^5\)) is publicly available and considers inputs that were scattered by the errors as statistically comparable. Each realization from the neural network on the same dataset is independent (Olney et al. 2020), making it possible to measure scatter between them; therefore, it reports uncertainties on the output parameters.

The resulting isochrones (plotted in orange and blue in the figures of Section A) are not always well aligned with our observed sequences. Therefore, we run the Auriga artificial neural network on our highly probable members per cluster. One hundred iterations of each cluster were passed through the Auriga model to generate the parameter errors. We also present the isochrones using the updated Auriga distance and age determinations and adjusting the reddening. The result is presented as red isochrones in Figure 3 for NGC 4349 and ASCC 88 as an example, the complete figure set can be found in Section A. The derived age, distance, reddening and extinction values are presented in Table 2.

Adding the new high probability members helps the isochrone models to trace most of the cluster sequences better. As noted in PR21, for the vast majority of the clusters, even the fainter end of the sequences match the theoretical isochrones. Since we are still in the stellar regime at those magnitudes and for the range of cluster ages explored, we do not expect deviations due to the uncertainties in the models. That indicates a low rate of interlopers across the whole magnitude range covered in this study. In various cases (BH 202, Lynga 9, NGC 4349, NGC 5381, NGC 5715, NGC 5999, NGC 6568, UFMG 1, UFMG 2), it is reproduced the phenomenon found in PR21 for NGC 6259, where we identify a group of high probability members at $K \sim 11$ mag. Most sources are not identified as high probability members at optical wavelengths. On the other hand, the isochrone location for all Ruprecht clusters plus Teutsch 84 and Trumpler 26 is rather poor compared with the outlined cluster sequences. As discussed below, this is partly induced by the very dispersed sequences with large error bars on distance, extinction, and age. All the clusters mentioned earlier are also located in the IV Galactic quadrant (Sagittarius-Carina (Sag-Car) and Scutum arms). Recently, (Poggio et al. 2021) made a comparison between their map of density variations (built with upper

\(^4\) http://stev.oapd.inaf.it/cgi-bin/cmd
\(^5\) https://github.com/mkounkel/Auriga
main sequence stars, open clusters, and classical Cepheids) and the Galactic models of Taylor & Cordes (1993) and Reid et al. (2019). In quadrant IV, the authors find a possible inter-arm gap between Sag-Car and the inner Scutum arm. This inter-arm would be closer to the Sun than Taylor & Cordes (1993) and Reid et al. (2019) models suggest. Alternatively, the authors mentioned the possibility of an additional branch off the Sag-Car arm. Given this scenario, we avoid manually reallocating the cluster sequences on that specific region, and a dedicated study will address the phenomenon in future work.

Since our membership determination isolates a unique compact Gaussian distribution among the field, we rule out the inclusion of other known clusters and cluster candidates in the surveyed cluster area. Moreover, given that the cluster-selected Gaussian component was allowed full covariance, we did not restrict the study to a specific cluster shape. Recent studies have explored the open cluster shapes extensively in our Galaxy (Pang et al. 2021; Hu et al. 2021), leaving evident the enormous diversity of cluster forms.

The following presents the comparison of our results with those in the literature. For the broad comparison of our results with the literature, we have focused on the studies of Cantat-Gaudin et al. (2020) and Koukkel et al. (2020). We must first point out that our membership probabilities are skewed towards high values since we adopt only the highest members probabilities from the XDGMM procedure as input to the spatial/photometric membership refinement. The overall parameter comparison can be seen in Figure 4. The bulk of the Auriga ages ranges from 14 Myr to 870 Myr. Only NGC 5606 reports a young age of about 7 Myr where the Auriga model can overestimate the age. As can be seen, the most discrepant ages are linked to the most extincted clusters or to those where there is no agreement among published values (e.g., Ruprecht 130, Ruprecht 134, Lynga 6, Teutsch 84, ASCC 88). For sources below ~ 100 Myr we report younger ages with the inclusion of the newly identified members. In
terms of extinction values, the most distant clusters have the most distinct values, which is expected given the near-infrared nature of our dataset. The general trend is that we report larger extinction values than those available in the literature by about 0.5 mag. We cover cluster distances ranging from about 0.9 to 3.2 kpc. We found larger distance values for those clusters beyond ∼2.5 kpc. The overall distance values differ for the clusters with the literature’s most significant uncertainties (e.g., Ruprecht 130, Ruprecht 134, Pismis 19).

4.2 Cluster total mass

Following the methodology outlined on PR21, each cluster’s density profile was integrated to obtain the total number of stars contained in the cluster’s main sequence. We determined the completeness magnitude for each cluster as the faintest magnitude at which the number of sources per interval of magnitude does not deviate from an increasing distribution. Assuming the central values of Kroupa mass function (Kroupa 2001) and the PARSEC-COLIBRI isochrone on the rederived Auriga parameters, we extended the counting, adding all the masses of stars down to the stellar/substellar frontier (0.08 $M_\odot$). The evolved sources out from the main sequence were also linked to the best-fitting position on the corresponding isochrone. The membership probability of each source weighted the mass values.

The studied clusters have total masses in the range ∼120 – 2000 $M_\odot$, and the values are presented in Table 2. The adopted uncertainties consider the density profiles with and without weighting the mass values by each source’s membership probability.

5 CONCLUSIONS

This study unveils the near-infrared sequences of 37 open clusters located between ∼0.9 and ∼3.2 kpc. We employed a data-driven approach tailored toward disentangling the cluster population from field stars. The cluster members search comprises two main steps: a distribution-aware tagging of the central values and uncertainties in the astrometric space and a cluster membership assignment based on positional and photometric similarity. Our methodology has allowed us to disentangle the cluster populations from their dense backgrounds. Our study increased cluster membership by ∼47% on average, this increase in efficiency at recovering cluster members builds up our knowledge on the near-infrared cluster sequences. The measured physical parameters of the clusters generally agree with the literature values. However, we update their extinction values and total masses by unveiling a large portion of the low-to-intermediate mass population with a native near-infrared data-driven approach. In the near future a similar method can be applied to large samples of clusters to be optically discovered by the Vera Rubin telescope (LSST Science Collaboration et al. 2009).

The so-called Gaia revolution has changed the picture of our Galaxy in just a few years. Now we count in the thousands the number of over-densities in our Galaxy that diverse machine learning methods have tagged as open clusters. Here we show that the synergies with other instruments/data-sets, such as photometry and astrometry from VISTA, allow us to unveil the near-infrared counterpart of each studied cluster.

We have taken advantage of high quality parallaxes and deep near-infrared photometry, and machine learning methods to enable us to obtain clean near-infrared color-magnitude diagrams from which astrophysical parameters can be inferred.

ACKNOWLEDGEMENTS

We thank Marina Kounkel for her comments on the implementation of the Auriga model and Laurent Chemin for his feedback on the spiral Galactic structure. We also thank the referee for the received comments. This work was supported by ANID FONDECYT Iniciación 11201161, and ANID FONDECYT Regular 1201490. D.M. gratefully acknowledges support by the ANID BASAL projects ACE210002 and FB210003, and by Fondecyt Project No. 1220724. This paper made use of the Whole Sky Database (wsdb) created by Sergey Koposov and maintained at the Institute of Astronomy, Cambridge by Sergey Koposov, Vasily Belokurov and Wyn Evans with financial support from the Science & Technology Facilities Council (STFC) and the European Research Council (ERC). This work has made use of data from the European Space Agency (ESA) mission Gaia (https://www.cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

The preparation of this work has made extensive use of Topcat (Taylor 2005), of NASA’s Astrophysics Data System Bibliographic Services, of the open-source Python packages Astropy (Astropy Collaboration et al. 2013), NumPy (van der Walt et al. 2011), and scikit-learn (Pedregosa et al. 2012) as well as the R packages under the tidyverse suite (Wickham et al. 2019). The figures in this paper were produced with ggplot2 (Wickham 2016).

6 DATA AVAILABILITY

The data underlying this article are available in the Open Software Foundation, at https://dx.doi.org/10.17605/OSF.IO/8D2HJ.
Figure 4. Difference between cluster literature values (age, extinction, distance) and the calculated parameters presented here. Filled circles relate to (Cantat-Gaudin et al. 2020) values, filled diamonds for those from (Kounkel et al. 2020). The symbols are color coded based on the explored parameters. UFGM 1, UFGM 2, and UFGM 3, are not included.

REFERENCES
Ahmed N., Gokhale D., 1989, IEEE Transactions on Information Theory, 35, 688
Alonso-García J., et al., 2018, A&A, 619, A4
Astropy Collaboration et al., 2013, A&A, 558, A33
Bica E., Bonatto C., 2005, A&A, 443, 465
Bisht D., Elsanhoury W. H., Zhu Q., Sariya D. P., Yadav R. K. S., Rangwal G., Durgapal A., Jiang L.-G., 2020a, arXiv e-prints, p. arXiv:2006.13618
Bisht D., Zhu Q., Yadav R. K. S., Durgapal A., Rangwal G., 2020b, MNRAS, 494, 607
Bonatto C., Bica E., Ortolani S., Barbuy B., 2006, A&A, 453, 121
Borissova J., et al., 2016, AJ, 152, 74
Borissova J., et al., 2018, MNRAS, 481, 3902
Borissova J., et al., 2020, MNRAS, 499, 3522
Bovy J., Hogg D. W., Roweis S. T., 2011, Annals of Applied Statistics, 5, 1657
Cantat-Gaudin T., Anders F., 2020, A&A, 633, A99
Cantat-Gaudin T., et al., 2018, A&A, 618, A93
Cantat-Gaudin T., et al., 2019, A&A, 624, A126
Cantat-Gaudin T., et al., 2020, A&A, 640, A1
Castro-Ginard A., Jordi C., Luri X., Julbe F., Morvan M., Balaguer-Núñez L., Cantat-Gaudin T., 2018, A&A, 618, A59
Castro-Ginard A., Jordi C., Luri X., Cantat-Gaudin T., Balaguer-Núñez L., 2019, A&A, 627, A35
Cantat-Gaudin T., Anders F., 2020, A&A, 635, A45
Dias W. S., Alessi B. S., Moitinho A., Lépine J. R. D., 2002, A&A, 389, 871
Dias W. S., Monteiro H., Caetano T., Hickel G., Alves G., Sacchi M., Lépine J. R. D., Amaro N., 2022, MNRAS, 512, 4464
Emerson J., McPherson A., Sutherland W., 2006, The Messenger, 126, 41
Everitt B. S., Landau S., Leese M., Stahl D., 2011, Cluster Analysis, 5 edn. John Wiley & Sons
Feigelson E. D., Jogesh B. G., 2012, Modern Statistical Methods for Astronomy. Cambridge University Press
Forsy W., 1965, Biometrics, 21, 768
Gaia Collaboration et al., 2016, A&A, 595, A2
Gaia Collaboration et al., 2018, A&A, 616, A1
Gaia Collaboration Brown A. G. A., Vallenari A., Prusti T., de Bruijne J. H. J., Babusiaux C., Biermann M., 2020, arXiv e-prints, p. arXiv:2012.01533
Galli P. A. B., et al., 2020, A&A, 643, A148
González-Fernández C., et al., 2018, MNRAS, 474, 5459
Graham R. L., Hell P., 1985, Annals of the History of Computing, 7, 43
Hoffer E., Hubara I., Soudry D., 2018, in International Conference on Learning Representations. https://openreview.net/forum?id=S1hbStg0-
Holoien T. W. S., Marshall P. J., Wechsler R. H., 2017, AJ, 153,
APPENDIX A: DIAGRAMS OF STUDIED CLUSTERS

Figures produced using the supplementary online material (see Section 6).
Figure A.1. Left: Spatial distribution of the cluster members. Center-left: $K_s$ vs. $J-K_s$ color-magnitude diagram of the studied open stellar clusters. Black dots correspond to the Cantat-Gaudin & Anders (2020) members with reported probability $p \geq 90\%$, whereas the sources presented here are represented with red circles. Center-right: $K_s$ vs. $J-K_s$ color-magnitude diagrams along with the theoretical isochrones located based on the cluster parameters from Cantat-Gaudin & Anders (2020); Kounkel et al. (2020) and this work over our catalog members as black points. All the sequences preserved a common scale. Right: Members proper motion distribution color-coded based on their parallax value. Mean proper motions in right ascension and declination are plotted with dotted lines. Only sources with membership probabilities $p \geq 90\%$ are shown in all the diagrams. Text on the plots shows the cluster name.
Figure A.2. Same as Figure A.1
Figure A.3. Same as Figure A.1
Figure A.4. Same as Figure A.1
Figure A.5. Same as Figure A.1
Figure A.6. Same as Figure A.1
Figure A.7. Same as Figure A.1